A Hierarchical Data-Driven Method for Event-Based Load Shedding Against Fault-Induced Delayed Voltage Recovery in Power Systems

Qiaoqiao Li, *Student Member, IEEE*, Yan Xu, *Senior Member, IEEE*, and Chao Ren, *Student Member, IEEE*

***Abstract*—Load shedding (LS) is an effective control strategy against voltage instability in power systems. With increasing uncertainties and complexity in modern power grids, there is a pressing need for faster and more accurate control decisions. In this article, a hierarchical data-driven method is proposed for the online prediction of event-based load shedding (ELS) against fault-induced delayed voltage recovery. The ELS problem is hierarchically modeled as a multi-output classification subproblem for identifying the best shedding location and a regression subproblem to predict the minimum shedding amount. To solve the two subproblems, the weighted kernel extreme learning machine is adopted to construct a direct mapping between the system pre-fault operating conditions and the corresponding control variables. The method is tested on the ELS database, which is analytically generated via a novel adaptive sensitivity-based process on the New England 39-bus system. Compared with other methods, the proposed method is very accurate in prediction with excellent control performance, which maintains superior prediction ability under an imbalanced data distribution.**

***Index Terms*—Data-driven, event-based load shedding (ELS), fault-induced delayed voltage recovery (FIDVR), imbalanced learning.**

Nomenclature

LS	Load shedding.
ELS	Event-based load shedding.
FIDVR	Fault-induced delayed voltage recovery.
UVLS	Undervoltage load shedding.
ELM	Extreme learning machine.
KELM	Kernel extreme learning machine.
WKELM	Weighted kernel extreme learning machine.
SLFN	Single hidden layer feedforward network.
DT	Decision tree.
TSAT	Transient stability assessment tool.
OPF	Optimal power flow.
TDS	Time-domain simulation.
TVSI	Transient voltage severity index.
TVDI	Transient voltage deviation index.
PDF	Probability density function.
MAD	Mean absolute deviation.
MAPE	Mean average percentage error.
MOSR	Mean overshedding rate.
MOSA	Mean overshedding amount.
ER	Effectiveness rate.

Manuscript received February 13, 2020; revised April 16, 2020; accepted April 25, 2020. Date of publication May 11, 2020; date of current version October 23, 2020. This work was supported in part by the Ministry of Education, Republic of Singapore, under Grant AcRF TIER 1 2019-T1-001-069 (RG75/19). The work of Yan Xu was supported by Nanyang Assistant Professorship from Nanyang Technological University, Singapore. Paper no. TII-20-0718. *(Corresponding author: Yan Xu.)*

Qiaoqiao Li and Yan Xu are with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798, Singapore (e-mail: qiaoqiao001@e.ntu.edu.sg; eeyanxu@gmail.com).

Chao Ren is with the Interdisciplinary Graduate School, Nanyang Technological University, Singapore 639798, Singapore (e-mail: renc0003@e.ntu.edu.sg).

Color versions of one or more of the figures in this article are available online at https://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TII.2020.2993807

I. Introduction

Fault-induced delayed voltage recovery (FIDVR) is the phenomenon that the power system experiences a significant long-time delayed voltage recovery following a severe fault [1]. The driving force of the delayed voltage recovery and further system degradation (e.g., voltage collapse) is the unfavorable response of dynamic loads, such as induction motors [2]. In response to the fault-induced voltage sag, the induction motor loads may stall to sustain the reduced voltage and draw excessive reactive power from the grid [3]. Meanwhile, for systems with high-level inverter-interfaced renewable power generation, the wind and the photovoltaic power may fail to meet the low-voltage ride-through requirement and disconnect from the main grid, causing further power supply loss in the system [4]. The recent South Australia blackout has clearly illustrated how a severe FIDVR event leads to a blackout [5].

Load shedding (LS) control is an effective countermeasure against voltage instability [6], [7]. According to different intervention timings and logics, the strategy of LS can be classified into under-voltage load shedding (UVLS) and event-driven load shedding (ELS). UVLS is response-based, which is triggered when the voltage decreases below pre-defined thresholds after certain time delays [8], [9]. The ELS, by contrast, does not need the real-time measurement of the post-fault voltage trajectory. It is determined preventively and initiated immediately after a contingency is detected [10]–[12]. Due to the open-loop control logic, ELS may not be as precise and robust as UVLS [9], but it is much faster and effective to mitigate voltage drop and restore voltage magnitude [12]. In practice, both of the two LS schemes are necessary for the power system voltage stability enhancement and should be appropriately coordinated [13]. This article mainly focuses on the ELS method, which is the earlier control action to take.

Traditionally, the ELS strategy is offline calculated and stored in a decision table. According to the decision table, given pre-fault system operation conditions and contingency information, the ELS location and amount can be immediately indexed. The traditional scheme to generate the ELS decision table needs to screen all possible combinations of expected contingencies and typical system operating conditions [14]. As power systems operate in ever-changing power flow patterns, the system operating condition may differ significantly from what is expected. Thus, this scheme generally suffers from the mismatching problem during the online application. With the development of advanced real-time system monitoring and parallel computing techniques, the “on-line pre-decision” scheme is widely adopted [15]. Based on the online measured system condition, this method preventively calculates ELS solutions and periodically refresh the decision table. In this way, it can be more adaptive to the change in system operating conditions. However, under the circumstances of uncertain and highly unpredictable power grids with deregulated utility and penetration of renewable energy, it is usually required to refresh the table in a short time. The real-time computational load can be extensive for large systems, as optimization and time-consuming simulations are required for the calculation of ELS control strategy. Therefore, there is a pressingly need for both adaptive and faster online ELS approaches.

In recent years, the data-driven method has been identified as an efficient and effective tool to improve real-time power system situational awareness and secure operation. The main principle is to learn the highly non-linear relationships between the system security status and the operating conditions from massive offline simulated cases. During the online application, the trained data-driven model could make a decision immediately and accommodate both seen and unseen scenarios. Currently, most related studies are concentrated on power system stability monitoring and assessment, and much less work focused on the stability control problems.

In general, two control approaches based on the data-driven methods have been investigated in the literature. The first kind of approach utilizes interpretable machine learning methods, e.g., decision tree (DT) [16]–[18], pattern discovery [19], to construct system security regions by the inverse reading of the supervised learning results, which can also be further embedded as constraints in a cost-efficient (e.g., [16], [17], and [19]) or risk-averse (e.g., [18]) optimization model to derive the required control actions. Although such rule-based approaches are effective, there remain some challenges. First, as the uncertainty and complexity of modern power system increases, it will be difficult for these rules to accurately identify the high-dimensional security boundary and generalize to unseen scenarios. Second, for the online ELS control process with limited response time after the contingency occurs, the additional computational burden will be introduced by the optimization part.

The other type of control method extracts the knowledge of decision-making from the offline optimized control cases, which can be directly used to predict the control strategy online. In [20], to mitigate the FIVDR problem, a deep reinforcement learning-based platform is developed to optimize the generator dynamic braking and UVLS in discrete action space. Based on the significant features extracted by the principal component analysis, an artificial neural network is used to predict the optimal coordinated control actions for long-term voltage stability in [21]. In [22], a graph convolutional network is embedded in the LS scheme, where the network topology information is well mined to predict an optimal LS ratio. In [23], to counteract frequency instability, an extreme learning machine (ELM) based method is proposed to predict the LS amount of each load bus in terms of both seen and unseen contingency. Although these methods achieve good prediction accuracy, there remain some limitations in terms of practical control performance. First, the ELS control problem is essentially a complex optimization problem with discrete and continuous decision variables (e.g., LS location and amount). A fully discrete control without specific shedding amount (e.g., [20] and [21]) or fully continuous control without identifying the best shedding location (e.g., [22] and [23]) is not cost-efficient for ELS strategy design. Furthermore, in a realistic power system, severe post-fault conditions usually happen less frequently, which will result in a sparse and imbalanced ELS database. In this case, the trained model tends to overlook the minority yet more severe cases and the accuracy may be affected. However, none of the aforementioned methods have considered the influence of the imbalanced data distribution.

To overcome the inadequacies in the existing approaches, this article proposes a new hierarchical data-driven ELS method. It consists of a classification model to predict the best ELS location and a regression model to predict the minimum ELS amount on the identified locations. Considering the challenge of the imbalanced database, the two sub-models are sequentially trained using a cost-sensitive learning algorithm, weighted kernel extreme learning machine (WKELM). For the online application, given the pre-fault system operating condition and contingency information, ELS decisions can be made in real-time. The novel contributions of this work over existing works are summarized as follows.

1) A hierarchical data-driven control structure is proposed to achieve a more efficient and effective online ELS control. It could identify the critical load buses that are mainly responsible for the delayed voltage recovery and effectively reduce the number of over-shedding buses. Meanwhile, under the hierarchical structure, the prediction accuracy of LS amount can be significantly improved.
2) The proposed method takes the merits of the WKELM algorithm to extract the complex relationship of system

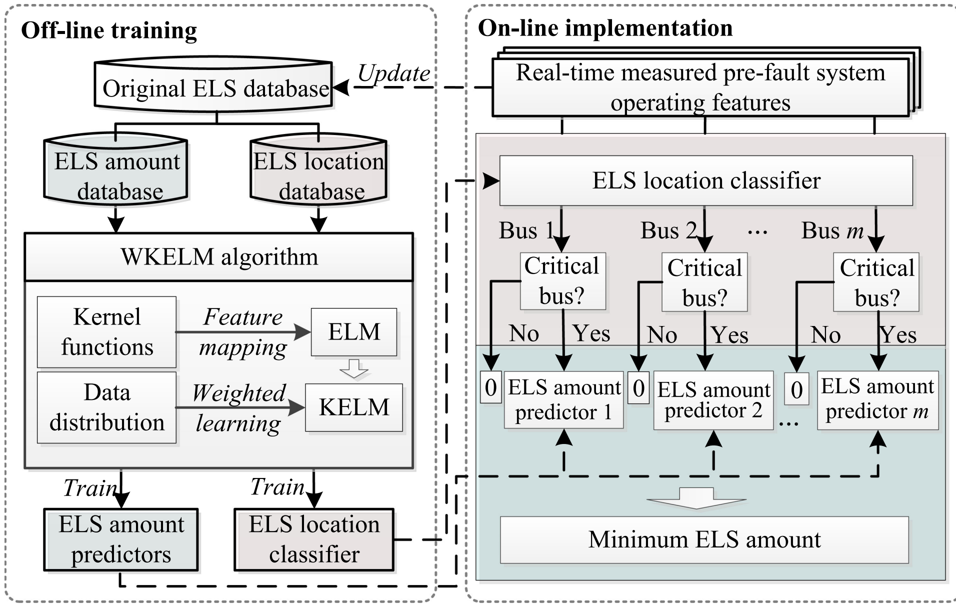

Fig. 1. Framework of the proposed hierarchical data-driven method.

operating conditions and control actions from an imbalanced ELS database. Meanwhile, a data-distribution-based weighted learning scheme is established, which can provide additional emphasis on the rare yet critical cases. Therefore, the risk of insufficient ELS prediction is reduced and better ELS effectiveness can be achieved.

The proposed method is numerically tested on the New England 39-bus system. The simulation result has verified its high prediction accuracy and excellent control performance over the other methods.

II. Proposed Method

A. Framework

Generally, the proposed data-driven control scheme could be divided into the offline and online stages, as illustrated in Fig. 1. At the offline stage, a database containing a large number of ELS control cases is constructed based on a sensitivity-based optimization method. Then, the original database is converted into an ELS location database with binary multi-dimensional outputs and meanwhile divided to m separated ELS amount database of each bus with numerical output. The classifier and the predictors are trained by an imbalanced learning algorithm, WKELM. Based on the kernel ELM (KELM), which could maintain fast speed, high accuracy, and much more stable performance than the original ELM, WKELM could further provide additional emphasis on the rare yet critical cases to improve the online control effectiveness. At the online implementation stage, once the contingency is detected, the input prefault system measurement (including generator outputs, bus voltage magnitude, bus voltage angle, load power, power flow on the branches, etc.) will be fed to the well-trained model, and the predicted ELS location and amount could be immediately provided to the control device. Due to the superiority of learning speed, the proposed WKELM-based prediction model can also be online updated to retrain the model and enhance its prediction ability.

B. ELS Database Generation

The whole procedure of the offline database generation of the proposed method is illustrated in Fig. 2. At first, based on the forecasted load demands, a steady-state operating point is obtained by running optimal power flow (OPF) calculation. Then, by performing the time-domain simulation (TDS) using the transient stability assessment tool (TSAT), the dynamic voltage performance of the post-fault system can be quantified by a FIDVR severity index, which is called the transient voltage severity index (TVSI) [24]. Based on the quantified security index, trajectory sensitivities with respect to LS location and amount can be accurately calculated, which can be used to search the optimal ELS iteratively. During the iteration, if the security requirement is fulfilled, the process will stop, and the pre-fault system operating features and the corresponding ELS strategy will be recorded in the database; otherwise, the iteration will continue. The details of the quantified FIDVR severity index and iterative optimization method will be introduced in Section III.

C. Hierarchical Model Design

There are two main concerns on the prediction of ELS. First, the optimal ELS location, which consists of the most critical buses, will vary with system conditions. The traditional methods do not distinguish the critical buses under different conditions; thus, uncritical buses may also be shed due to the unreasonable model design and inevitable regression error. Moreover, considering the variation of optimal ELS location, on certain buses (e.g., buses that are not close to the fault location), ELS may only be performed when very severe system conditions occur. For such a sparse and imbalanced database, the trained ELS

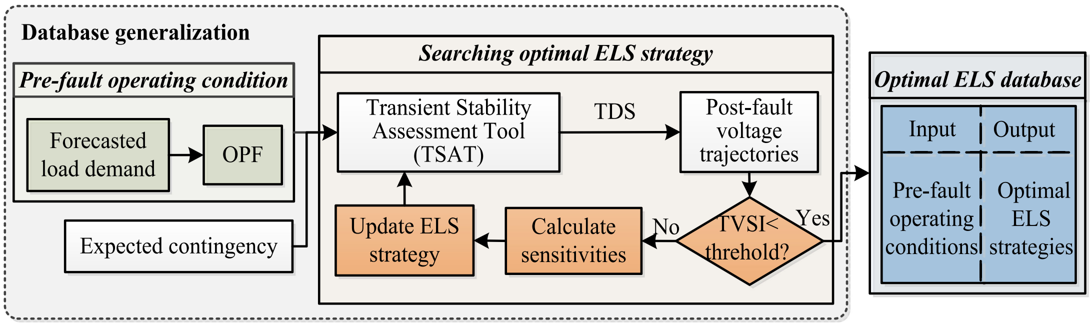

Fig. 2. Framework of database generation.

model will tend to be biased to the majority cases with rare scenarios overlooked and mistaken. In this case, the whole ELS shedding accuracy may be affected.

Considering the above concerns, this article proposes a hierarchical data-driven method for higher control efficiency and effectiveness. In the first hierarchy, an ELS location classifier is trained to detect all the critical buses that are most affected in case of an FIDVR event. Then, in the second hierarchy, the ELS amount predictor of each load bus is separately trained to evaluate the minimum amount of LS. It is worth noticing that, in each ELS amount database, cases with zero ELS value should be eliminated to ensure the efficient training of ELS amount predictors. At the online stage, once a fault is detected, the pre-fault system measurements are first provided to the classifier. If a load bus is identified as critical, the measurements are further provided to the corresponding ELS amount predictor to estimate the minimum load to shed. Otherwise, the bus is considered to be uncritical for voltage recovery; thus, ELS will not be executed on that bus. Finally, the predictions of the two hierarchy will be aggregated to provide a comprehensive ELS strategy to the control device. In this way, the over-shedding amount and the number of over-shedding buses can be reduced. Meanwhile, the data sparsity is eliminated by removing zero cases; thus, training efficiency and regression accuracy can also be improved.

D. Imbalanced Learning Scheme

As mentioned above, ELS is performed on certain buses only when severe system conditions occur. Therefore, the number of observations is quite different for both classes in the database. Such imbalanced data distribution can make it difficult to train regular learning algorithms on these buses. As illustrated in the literature [25], the classification boundary can be significantly pushed toward the majority cases with more minority cases misclassified. The problem becomes more unacceptable for high-reliability applications, such as the LS, in this article. Due to the fact that mistaking a critical bus as non-critical will lead to insufficient ELS, it has a high possibility of causing further voltage oscillation and system instability.

Given the above concerns, this article designs a data-distribution-based cost-sensitive learning algorithm based on the WKELM to address imbalanced data distribution problems existing in the procedure of ELS prediction. The general idea of this algorithm is to assign larger weights to the training error of more critical cases. Based on the information of data distribution, the weights can be determined automatically by the inverse of the class frequency or estimated probability density. By this means, this method could raise more attention to the rarely occurred yet more critical cases. Moreover, it can accommodate both balanced and imbalanced situations, which is much convenient in the application, with no extensive offline analysis.

III. ELS Database Generation

A. Quantitative FIDVR Severity Index

After a system fault is cleared, the unacceptable FIDVR events appear as a long-time low-voltage phenomenon. In this article, based on the average accumulated relative voltage deviation of all buses, TVSI, which has been widely adopted in the short-term voltage stability analysis [24], [26], [27], is applied to quantify the FIDVR severity as follows:

$$\mathrm{TVSI} = \frac{\sum_{i=1}^{m}\sum_{t=t_c}^{T}\mathrm{TVDI}_{i,t}}{N_b \times (T - t_c)} \tag{1}$$

where m is the number of buses in the system; T is the considered voltage recovery time frame, t_c is the fault clearing time, and the transient voltage deviation index (TVDI) can be formulated as follows:

$$\mathrm{TVDI}_{i,t} = \begin{cases} \frac{|V_{i,t}-V_{i,0}|}{V_{i,0}}, & \text{if } \frac{|V_{i,t}-V_{i,0}|}{V_{i,0}} \geq \alpha \\ 0, & \text{otherwise} \end{cases} \quad \forall t \in [t_c, T] \tag{2}$$

where $V_{i,t}$ denotes the voltage magnitude of bus i at time t; $V_{i,0}$ is the pre-fault voltage magnitude of bus i; and α is the threshold of unacceptable voltage deviation magnitude. Commonly, 10% is adopted for α in the industrial criteria. Also, the value of α could be adjusted depending on the actual stability requirement. The above indexes would be calculated via performing TDS to obtain the voltage trajectories. TVSI is the average of TVDIs over all buses, indicating the whole system voltage performance. A smaller value means the voltage performance is better. Depending on the size and parameters of the actual power system, a threshold of TVSI can be determined to distinguish the acceptable and unacceptable FIDVR events.

B. Objective of ELS Optimization

The calculation of ELS can be formulated as an optimization problem with dynamic security constraints [10], [23]

$$\min_{L} \ f\ (L) \ = \sum_{i=1}^{m} \ l_i \tag{3-a}$$

$$\text{s.t.} \begin{cases} \text{TVSI} = \varphi(L), \quad L = [l_1,\ l_2\ , \ldots,\ l_m]^T \\ \text{TVSI} - \text{TVSI}_{\text{thre}} \ \geq \ \varepsilon \\ 0 \ \leq \ l_i \ \leq \ \overline{l_i}, \quad i = 1,\ 2,\ \ldots,\ m \end{cases} \tag{3-b-c-d}$$

where the objective is to minimize the total LS amount of all the *m* buses, assuming that the selected load buses are of the same power supply priority and electricity cost. L is the LS vector, where l_i is the LS value on the *i*th load bus, $i = 1, 2, \ldots, m$. The value of TVSI is a function of L, which could be calculated by performing TDS. $\text{TVSI}_{\text{thre}}$ denotes the TVSI threshold to distinguish the acceptable and unacceptable FIDVR events. ε is a small positive value denoting stability margin threshold, considering the unavoidable errors of system model and control decision. The constraint in (3-d) denotes the availability of load to be shed, and $\overline{l_i}$ is the maximum load that can be shed at the *i*th load bus.

As indicated in (3-b), the constraints of TVSI is a nonlinear nonconvex function with respect to the LS variables and have no exhibit expression. To solve this problem, an adaptive trajectory-sensitivity-based method is proposed in this article.

C. Trajectory Sensitivity Analysis

The concept of trajectory sensitivity indicates the influence of small changes in one of the parameters (e.g., LS variable) to the system dynamic behavior (e.g., FIDVR severity). Based on the sensitivity information, the control variables could be analytically adjusted to fulfill the objective and constraints. In this way, an optimal or suboptimal control solution could be derived [28]. In this article, the trajectory sensitivity is defined as follows:

$$s_i = \frac{\varphi(L,\ \Delta l_i) - \varphi(L)}{|\Delta l_i|} \tag{4}$$

where *i* is the number of load bus, $i = 1, \ldots, m$. $\varphi(L)$ is the value of TVSI under LS vector L and Δl_i is a small perturbation on the $i\text{th}$ load bus.

It is worth noting that Δl_i is a user-defined step size of LS, and it should not be too large to avoid unnecessary over shedding or too small to be computational massive. Thus, in the proposed method, TVSI index and sensitivity analysis with an adaptive step size are designed for ELS database generation. When the stability margin comes closer to the security threshold as the iteration number increases, the shedding step of the control variable will be adjusted smaller in order to avoid unnecessary over shedding

$$\Delta {l_i}^k = \Delta {l_i}^0 . \ (\mu^k / \ \mu^0)^\gamma . \tag{5}$$

where k denotes the iteration number of sensitivity analysis; $\Delta {l_i}^k$ is the LS increment step in the *k*th iteration; μ^k is the stability margin of TVSI, $\mu^k = \varphi(L^k) - \text{TVSI}_{\text{thre}}$, and L^k is the LS amount vector in the *k*th iteration. $\Delta {l_i}^0$ is a user-defined basement step size and μ^0 is the basement margin. γ is a user-defined parameter, which is set as 0.25 in this article. Thus, as the iteration number increases, the step size will reduce slowly at first and then shrink fast when the TVSI margin comes close to zero. The algorithm for ELS optimization is as follows:

ELS optimization based on the adaptive sensitivity analysis

Input: System pre-fault operating point and expected contingency.
Output: The optimized ELS vector.
Initialize: $\Delta {l_i}^0$, μ^0, and ELS vector $L^1 = [0,\ \ldots, 0]^T$;
for $k = 1$ **to** K **do**
  Provide the input to TSAT, and then run TDS to obtain the TVSI stability margin, $\mu^k = \ \text{TVSI}_{\text{thre}} - \varphi(L^k)$, **if** $\mu^k \ < \ \varepsilon$, which means the stability constraints are not fulfilled, the iteration continues; **else**, stop;
  Update step size, $\Delta {l_i}^k = \Delta {l_i}^0 (\mu^k / \ \mu^0)^\gamma$;
  For bus $i = 1, 2, \ldots, m$, calculate its trajectory sensitivity s_i with respect to the step size $\Delta {l_i}^k$ using (4);
  Select the bus with the largest sensitivity, $h = \max_i s_i$, increase its LS value by Δl_h^k, ${l_h}^k = {l_h}^k \ + \ \Delta {l_h}^k$ or $L^k = \ L^k + [0,\ \ldots,\ \Delta {l_h}^k,\ \ldots,\ 0]^T$;
end for

IV. Imbalanced Learning

The proposed method takes the merits of the WKELM algorithm to extract the highly non-linear non-convex relationship of system operating conditions and ELS strategy from an imbalanced control database. Besides, a data-distribution-based cost-sensitive learning scheme is established, which can provide additional emphasis on the extremely distributed yet critical cases. The proposed method also generalizes the imbalanced learning scheme to solve the regression problem by adopting the kernel density estimate method, which can address both imbalanced classification and regression tasks.

A. ELM With Kernel Feature Mapping

ELM is a single hidden layer feedforward network (SLFN) proposed by Huang *et al.* [29]. The structure of SLFN is illustrated in Fig. 3. Considering a prediction problem with the *n*-dimensional input vector x_i and *m*-dimensional output vector y_i, (6) demonstrates the output function of ELM with *P* hidden nodes and activation function *G*

$$f_P(\mathbf{x}_i) = \sum_{j=1}^{P} \boldsymbol{\beta}_j \ \cdot \ G(\mathbf{w}_j,\ \mathbf{x}_i,\ b_j) = y_i, \quad i = 1,\ 2,\ \ldots,\ N \tag{6}$$

where $\boldsymbol{w}_j$ is the vector of weight linking input nodes to the $j\text{th}$ hidden node; b_j is the threshold of the $j\text{th}$ hidden node; and

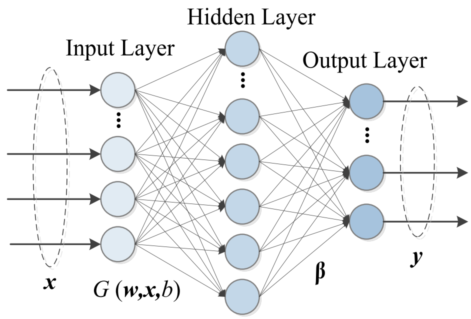

Fig. 3. Structure of SLFN.

$\boldsymbol{\beta}_j$ is the weight vector from the $j\text{th}$ hidden nodes to the output nodes. N is the number of total instances.

Then, the output vector of P hidden nodes could be presented as $\mathbf{h}(\mathbf{x}) = [h_1(\mathbf{x}), \ldots, h_P(\mathbf{x})]$ and $h_j(\mathbf{x}) = G(\mathbf{w}_j, \mathbf{x}, b_j)$. It is also called feature mapping in ELM, which maps the input samples from its original feature space to hidden layer space by an activation function. Then, a compact format of (6) is given as follows:

$$\mathbf{H}\,\boldsymbol{\beta} = \mathbf{Y} \tag{7}$$

where $\mathbf{H} = [\mathbf{h}(\mathbf{x}_1), \ldots, \mathbf{h}(\mathbf{x}_N)]^T$ is called the hidden layer output matrix, and $\boldsymbol{\beta}$ is the output weight matrix. $\mathbf{Y}$ is the target matrix.

If the feature mapping $\mathbf{h}(\mathbf{x})$ is unknown, a kernel matrix could be defined for ELM under Mercer's conditions and calculated by input vectors from the training set.

$$\Omega_{\text{ELM}\ i,j} = \mathbf{h}(\mathbf{x}_i)\mathbf{h}(\mathbf{x}_j) = K(\mathbf{x}_i, \mathbf{x}_j). \tag{8}$$

Then, by the method of the least square error with the minimum norm, the output of ELM could be kernelized as

$$\begin{aligned}\boldsymbol{f}(\mathbf{x}) &= \mathbf{h}(\mathbf{x})\boldsymbol{\beta}\\ &= [K(\mathbf{x}, \mathbf{x}_1), \ldots, K(\mathbf{x}, \mathbf{x}_N)]\left(\frac{\mathbf{I}}{C} + \boldsymbol{\Omega}_{\mathbf{ELM}}\right)^{-1}\mathbf{Y}\end{aligned} \tag{9}$$

where $\mathbf{x}_i,\ \ i = 1, \ldots, N$ are the input vectors from the training data, and $\mathbf{Y}$ is the training set target matrix. By this means, neither the specific feature mapping need to be known to users nor is the number of hidden nodes to be tuned. Instead, its corresponding kernel $K(\mathbf{x}_i, \mathbf{x}_j)$ should be provided. Then, the network of ELM could be fully determined by the method of the least square error with the minimum norm, and the detailed solution procedure can be found in [30].

B. Weighted KELM

The conventional learning algorithms mainly aim to improve prediction accuracy without considering complex data distribution. However, if trained by an imbalanced database, models will tend to be biased to the most frequent cases and neglect the others. In previous studies, sampling techniques have been proposed to rebalance the data under the risk of information loss or redundancy [31]. Another popular approach is to assign the misclassification cost to weight the sample space. Although it is effective and easy to apply, this method would introduce extra difficulty for the users to set appropriate weights for different cases.

Based on the unweighted ELM, WELM has been designed to cope with the class imbalance problem, as a sort of cost-sensitive learning algorithm [25]. Given a training set $[\mathbf{x}_i, y_i],\ \ i = 1, \ldots, N$, an $N \times N$ diagonal matrix $\mathbf{W}$ will be utilized to regulate the penalty factor C for each input sample. In general, the training error of samples in the minority class will be assigned larger C, while the majority class will have smaller C. The mathematical formulation of the least square error with the minimum norm with weight matrix $\mathbf{W}$ could be written as follows:

$$\min_{\boldsymbol{\beta}} \mathbf{L_{ELM}} = \frac{1}{2}\|\boldsymbol{\beta}\|^2 + C\mathbf{W}\frac{1}{2}\sum_{i=1}^{N}\|\xi_i\|^2 \tag{10-a}$$

$$\text{s.t.}\ \ \mathbf{h}(\mathbf{x}_i)\,\boldsymbol{\beta} = y_i^T - \xi_i^T,\quad i = 1, \ldots, N \tag{10-b}$$

where ξ_i is the training error of input vector $\mathbf{x}_i$. C is the penalty factor that represents the tradeoff between the total training errors and the norm of the output parameter $\boldsymbol{\beta}$.

In terms of kernel feature mapping, the solution could be derived as

$$\begin{aligned}\mathbf{f}(\mathbf{x}) &= \mathbf{h}(\mathbf{x})\,\boldsymbol{\beta} = [K(\mathbf{x}, \mathbf{x}_1), \ldots, K(\mathbf{x}, \mathbf{x}_N)]\\ &\quad \left(\frac{\mathbf{I}}{\mathrm{C}} + \mathbf{W}\boldsymbol{\Omega}_{\mathbf{ELM}}\right)^{-1}\mathbf{W}\mathbf{Y}.\end{aligned} \tag{11}$$

The weight matrix $\mathbf{W} = \operatorname{diag}\{w_{ii}\}$, $i = 1, \ldots, N$ could be defined by the users according to the required rebalance degree.

C. Imbalanced Learning Scheme

In this article, for the classification problem, an automatic weighting scheme is employed following the class distribution:

$$w_{ii} = \frac{1}{c(y_i)} \tag{12}$$

where $c(y_i)$ is the number of samples belonging to the class y_i; and y_i is the $i\text{th}$ sample.

However, most works focused on imbalanced classification, this article generalizes the weighted learning scheme to solve the imbalanced regression problem. The value of weight is designed as a function that is inversely proportional to the probability density function (PDF) of the output values. Based on the available infinite samples, the histogram is one of the viable methods for PDF estimation. Then kernel density estimate is used to obtain a smooth and continuous PDF [32]

$$w_{ii} = \frac{1}{\hat{f}(y_i)} \tag{13}$$

$$\hat{f}(y) = \frac{1}{N\,b}\sum_{i=1}^{N} K\left(\frac{y - y_i}{b}\right) \tag{14}$$

where $\hat{f}(y)$ is the estimated PDF of target y; K is the kernel function; b is a smoothing parameter called the bandwidth, which is a tradeoff between the estimation bias and variance.

In general, based on the imbalanced learning algorithm, the calculation steps of the learning and validation procedure of the proposed method can be summarized as follows.

Step 1. Database preparation and division: Prepare the ELS location database and m ELS amount database, and separately divide into a training set, a validation set, and a testing set. Remove the cases with zero output in the training set and the validation set of ELS amount databases.

Step 2. Calculation of data distribution: Calculate the class distribution of ELS location variables and the PDF of numerical ELS amount variables by the training data, and formulate the weight matrix $\mathbf{W}$.

Step 3. Initialization: Select the parameter of kernel function and penalty factor C from the candidate set.

Step 4. Calculate the kernel matrix $\boldsymbol{\Omega_{ELM}}$ by the input vectors of the training set.

Step 5. Calculate the output of validation cases by (11); save the validation error and go to Step 3 until all the candidates of hyperparameters have been tested.

Step 6. Select the parameters with the lowest validation error for the classifier and predictors, respectively.

Step 7. Calculate the output of the classifier on the testing cases. For each load bus on each case, if the identification result is positive, calculate the output of the corresponding ELS amount predictor; otherwise, the predicted ELS amount is set as zero.

V. Simulation Results

The proposed method is verified on the New England 10-machine 39-bus system. Considering that the FIDVR problem is mainly caused by the fast load dynamics, the industry-standard composite load model CLOD [33] defined in PSS/E is employed to simulate the typical induction-motor-dominant load components. The dynamic load parameter is set as follows:

1) large motor—20%;
2) small motor—30%;
3) discharge lighting—5%;
4) transformer exciting current—5%;
5) constant power load—20%;
6) induction motor load—20%.

A. Database Generation

1) Pre-fault System Operating Conditions: In this test, 3795 power system operating points are simulated based on a variety of load demand levels sampling from 80% to 120% of the base power. Then, the other system operating variables are obtained by OPF calculations. The input vector of the ELS control database consists of the following:

1) generator P, Q generation, voltage magnitude, and angle;
2) load bus P, Q load, voltage magnitude, and angle;
3) power flow on the branches.

TABLE I
Class Distribution of the ELS Database

Number	Fault #1			Fault #2		
	Bus8	Bus9	Bus27	Bus9	Bus24	Bus25
Positive class	1429	2470	520	909	3263	2548
Negative class	2366	1325	3275	2886	532	1247

In total, there are 383 input features. In a realistic power system, the operating conditions could be online updated and enriched by the measured data to maintain the accuracy of the database.

2) Contingency Setting: In this article, two typical three-phase short-circuit faults yielding two respective databases are considered to test the prediction performance of the proposed method. A three-phase short-circuit fault, denoted as fault #1, is applied to bus 17. The fault duration time is 0.1 s. After fault #1, the line between bus 17 and bus 27 is tripped. The extension test fault #2 occurs on bus 21. After the fault, the line between bus 16 and bus 21 is tripped. Other severe contingencies can also be considered in need.

3) Simulation Setting: Provided system operating conditions and contingency information, TDS is then conducted using the TSAT software, where post-fault voltage trajectories could be obtained to calculate the TVSI value. In TSAT, the simulation step size is set to 0.01 s. The maximum simulation time is 5 s. A total of 10% is adopted for α, the threshold of voltage deviation magnitude, as illustrated in (2). According to the size and parameter setting of an adopted power system model, the unacceptable FIDVR threshold, defined as $\mathrm{TVSI_{thre}}$, is set as 4.24 for this New England 10-machine 39-bus system. Then, the corresponding ELS for each case is iteratively searched based on the sensitivity-based optimization method, as introduced in Section III.

4) Data Distribution: In this 10-machine 39-bus system, there are 19 candidate load buses in total. In the ELS database, 16 load buses are identified not covered in the potential LS area of all scenarios; thus, they are excluded from prediction performance assessment. For the remaining three buses, the class distribution is illustrated in Table I. In the table, the negative class means that the bus does not need to perform ELS in this case, and vice versa. Then, the samples that belong to the positive class will further be used for constructing the amount prediction model of each bus. It is found that for the 3795 cases, the corresponding ELS amount range from 0 to 300 MW.

B. Model Tuning

In this article, the ELS database is randomly divided into a training set, a validation set, and a testing set by the ratio of 3:1:1. According to the KELM theory, the number of hidden neurons is decided by the number of training samples, which is 2277 in this model. The Gaussian kernel is selected as the kernel function for feature mapping. Then, there are two hyperparameters need to be tuned, the regulation factor C and σ for the Gaussian kernel, $K\ (\mathbf{x},\ \mathbf{y}) = \exp(-||\mathbf{x}-\mathbf{y}||^2/\sigma)$. In the validation stage, grid search is conducted, where C and σ are selected in a wide range $\{2^{-18}, 2^{-16}, \ldots, 2^{16}, 2^{18}\}$. Then, the hyperparameters of the lowest validation error are adopted for testing.

C. Performance Indices

To evaluate the prediction performance of the proposed method, three commonly used indices are employed in this article. First, the classification accuracy is used to quantify the success rate of critical bus identification in the first hierarchy. Then, the regression error of ELS amount prediction is evaluated by mean absolute deviation (MAD) as

$$\mathrm{MAD} = \frac{1}{N_t}\sum_{i=1}^{N_t} |y_i^{\mathrm{real}} - y_i^{\mathrm{predicted}}| \,. \tag{15}$$

And mean average percentage error (MAPE) is as follows:

$$\mathrm{MAPE} = \frac{1}{N_t}\sum_{i=1}^{N_t} |y_i^{\mathrm{real}} - y_i^{\mathrm{predicted}}|/|y_i^{\mathrm{real}}|. \tag{16}$$

In addition, to further evaluate the control performance of the proposed method, the other three indices are proposed as the mean over-shedding rate (MOSR), the mean overshedding amount (MOSA), and effectiveness rate (ER). The MOSR is the average percentage of over-predicted cases to the total tested cases. The MOSA is the average amount of over-predicted cases

$$\mathrm{MOSA} = \frac{1}{N_{\mathrm{os}}}\sum_{i,\mathrm{os}=1}^{N_{\mathrm{os}}} \left(y_{i,\mathrm{os}}^{\mathrm{predicted}} - y_{i,\mathrm{os}}^{\mathrm{real}} \right) \tag{17}$$

where i,os is the ith over-shedding cases, N_{os} is the the total number of over-predicted samples in the testing set.

The ER is designed to test the actual success rate of the predicted ELS strategy

$$\mathrm{ER} = \sum_{i=1}^{N_t} e_i/N_t \tag{18}$$

where e_i is the TDS result under the predicted ELS control strategy. N_t is the total number of testing cases. As illustrated in (3-c), if the security constraint is fulfilled, e_i is set to 1; otherwise, e_i will be set to 0.

D. Testing Results

Based on the tuning results of the validation set, the trained model is then applied to the testing samples to verify its performance. The comparison results of model-based, nonhierarchical, and unweighted methods are illustrated in the following sections to demonstrate the advantages of the proposed method. Besides, other standard and popular machine learning techniques, including basic ELM, support vector machine (SVM), and random forest (RF) are also tested in the same data set. In the test, the number of hidden layer nodes of basic ELM is set as the same number of KELM and WKLEM. Grid search is also conducted to find the hyperparameters of the other algorithms with the minimum validation loss. Especially, 100 single DTs are ensembled in the RF method, where the minimum leaf size is finely tuned as 2 and the maximum number of splits is 5. The adaptive boosting method is used as the ensemble aggregation method for classification RF and the least squares boosting is adopted for regression RF.

TABLE II
CONTROL PERFORMANCE OF THE PROPOSED METHOD

	Fault #1			**Fault #2**		
	Bus8	Bus9	Bus27	Bus9	Bus24	Bus25
Accuracy	99.18%	99.79%	98.15%	97.71%	99.69%	97.52%
MAPE	2.26%	2.61%	0.68%	3.42%	1.67%	3.58%
MAD	1.37	2.39	0.39	3.51	2.18	4.78
MOSA	1.08	2.33	0.30	3.48	2.10	4.82
ER		99.60%			98.51%	

TABLE III
SUMMARY OF COMPUTATIONAL TIME

Prediction time	Proposed method	Model-based method
t	0.137 ms	28.47 min

1) Comparison With Model-Based Method: The proposed data-driven method is comprehensively compared with the trajectory-sensitivity-based conventional method in terms of both control effectivity and computational efficiency. The results of six performance indices are listed in Tables II and III. First, the first hierarchy classification accuracy shows that the critical buses of more than 98% of testing cases can be successfully identified. In terms of ELS amount prediction, the mean deviation to the true values (that are offline calculated by the conventional mathematical method) is less than 5 MW with MAPE less than 4%. Moreover, TDS under the predicted ELS actions is performed for the whole testing cases one by one, where the case with TVSI less than the pre-defined threshold will be regarded as an effective control. In this circumstance, the ER is higher than 98% of both faults with around 2 MW load over-shed on average, which shows that the control performance is quite satisfactory.

Considering the calculation speed of the online control, the average computational time of the testing cases is listed in Table III. It costs less than 1.0 ms for the proposed method to make an online ELS decision, while the model-based method needs almost half an hour to find a feasible solution. Thus, in terms of calculation speed, the proposed method shows great superiority over the conventional model-based method. The main reason is that the computational time of the data-driven method is fully determined by the neuron network structure and dimension of the input feature, which will not be affected by the TDS speed or optimization procedure.

To intuitively show the system behavior under different control methods, post-fault voltage trajectories are collected through TDS on a new system operating point. The ELS actions calculated by each method are individually performed 0.15 s after fault 1# occurred. From the average bus voltage curves given in Fig. 4(a), it can be seen that the difference of ELS performance of the two methods is negligible. As shown in Fig. 4(b), ELS action predicted by the proposed method is performed, the

TABLE IV
PERFORMANCE COMPARISON IN TERMS OF HIERARCHICAL AND NONHIERARCHICAL METHODS

Model		Fault #1				Fault #2			
		MOSR	MOSA	MAPE / MAD	ER	MOSR	MOSA	MAPE / MAD	ER
WK-ELM	**Hierartical**	**23.20%**	**1.24**	**1.85% / 1.38**	**99.60%**	**37.45%**	**3.47**	**2.89% / 3.49**	**98.51%**
	Nonhierartical	51.34%	2.16	3.93% / 2.46	99.60%	62.58%	4.33	3.19% / 4.76	98.02%
K-ELM	Hierartical	22.79%	1.21	1.82% / 1.43	98.99%	37.02%	3.54	3.40% / 3.99	98.16%
	Nonhierartical	51.15%	2.05	3.98% / 2.37	98.99%	63.11%	4.12	4.12% / 4.91	98.02%
ELM	Hierartical	24.29%	3.41	4.63% / 3.62	96.24%	39.92%	6.59	6.53% / 7.11	93.65%
	Nonhierartical	50.34%	6.95	9.90% / 9.69	91.30%	51.50%	7.23	11.16% / 9.57	89.86%
RF	Hierartical	24.73%	1.81	2.16% / 1.74	99.14%	35.37%	4.97	3.68% / 5.02	97.49%
	Nonhierartical	51.54%	3.32	5.34% / 3.51	98.81%	51.90%	6.11	7.53% / 7.23	95.53%
SVM	Hierartical	23.53%	2.12	3.09% / 2.29	97.23%	38.08%	5.12	4.76% / 5.59	97.10%
	Nonhierartical	58.12%	4.37	7.61% / 5.15	95.65%	54.28%	6.41	8.67% / 7.98	94.49%

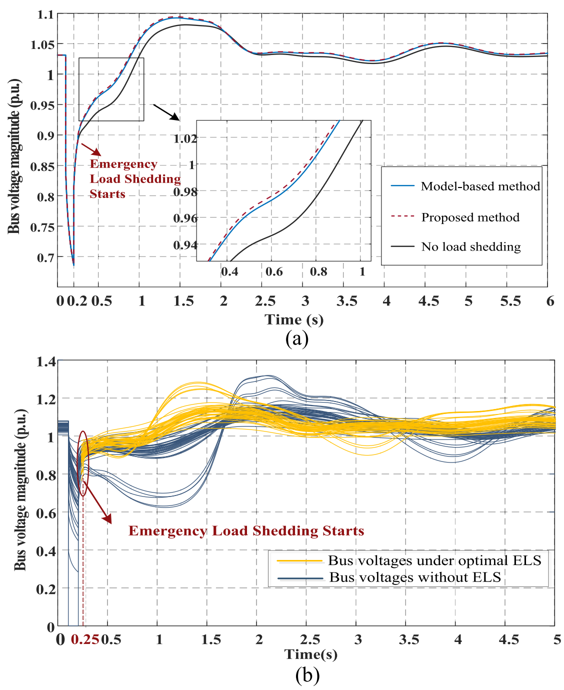

Fig. 4. (a) Average bus voltage curve and (b) all the 39-bus voltage curves of a new operating point after fault#1.

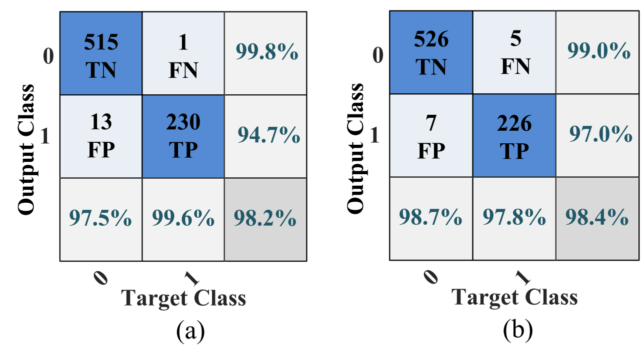

Fig. 5. Comparison of the weighted and unweighted algorithms. (a) Confusion matrix of WKELM. (b) Confusion matrix of KELM.

time of system voltage delay is significantly reduced, and the unacceptable FIDVR is effectively mitigated.

2) Comparison With the Nonhierarchical Method: The proposed method hierarchically predicts ELS control strategy to reduce the over-shedding amount and meanwhile increase the prediction accuracy. For the nonhierarchical method, the ELS prediction is directly conducted on each load bus without location identification. In this article, the two prediction structures are compared on five machine learning algorithms, the WKELM-based method, the KELM-based method, the ELM-based method, the SVM-based method, and the RF-based method. As observed from Table IV, more than half of the test samples are over-predicted by the nonhierarchical method. The reason is that under the nonhierarchical structure, the uncritical buses are also shed due to the regression error, resulting in near twice the over-shedding area of the hierarchical method. Moreover, for the prediction of ELS amount, the hierarchical method can achieve almost half of the regression error of direct regression as well as the over-shedding amount. Since the nonhierarchical database is highly sparse and imbalanced, the trained ELS model will tend to be biased to the majority of cases that need not conduct ELS. In this case, the severe and critical cases are overlooked and the regression accuracy will be decreased. Besides, for ELM-based and SVM-based methods, the ER is significantly improved due to the reduction in prediction error. In summary, through identifying the best ELS location in prior, the proposed hierarchical method can predict the ELS amount more accurately and improve the control performance in terms of both efficiency and effectivity.

3) Comparison With the Other Learning Algorithms: In the proposed method, the WKELM algorithm is adopted to learn from the imbalanced data and incorporate cost in the training procedure. To show its advantage, the performance of the WKELM-based method, KELM-based, and the other popular learning algorithms is compared in this section.

To demonstrate the superiority of the weighted learning scheme, in Fig. 5(a) and (b), the precision evaluation is carried out for the critical bus identification results by confusion matrix, where TP, TN, FP, and FN stand for true positive, true negative, false positive, and false negative, respectively. On the confusion matrix of the plot, the rows correspond to the predicted class

TABLE V
SUMMARY OF PERFORMANCE WITH OTHER STATE-OF-THE-ART ALGORITHMS

Model	Fault #1			Fault #2		
	1st Hierarchy Average classification accuracy	2nd Hierarchy Average prediction MAPE / MAE	ER	1st Hierarchy Average classification accuracy	2nd Hierarchy Average prediction MAPE / MAE	ER
Proposed method	**99.04%**	**1.85% / 1.38**	**99.60%**	**98.31%**	**2.89% / 3.49**	**98.51%**
KELM	99.09%	1.82% / 1.43	98.99%	98.50%	3.40% / 3.99	98.16%
ELM	96.64%	4.63% / 3.62	96.24%	94.57%	6.53% / 7.11	93.65%
RF	99.29%	2.16% / 1.74	99.14%	97.99%	3.68% / 5.02	97.49%
SVM	97.49%	3.09% / 2.29	97.23%	97.76%	4.76% / 5.59	97.10%

(output class) and the columns correspond to the true class (target class). The columns on the far right of the plot are negative predictive rate and positive predicted rate. The rows at the bottom of the plot are true negative rate and true positive rate. The detailed description of these statistical rates can be found in [34]. The block in the bottom right shows the overall accuracy.

In the power system real-time control, one of the most significant indices should be the number of FN. A case of FN means a critical bus (the target class is 1) that need to be shed is classified as uncritical, (the output class 0). Mistaking such cases would lead to insufficient ELS, which has a high possibility of causing further voltage oscillation and system instability. The results of classification based on both weighted and unweighted algorithm on bus 27 are compared in Fig. 5 with a class imbalance ratio 0.159:1 (label 0: label 1). Under this situation, the weighted scheme effectively decreases the FN cases to 1 with a little sacrifice on increasing FP cases, and the sacrifice can further be compensated in the second hierarchy. Finally, as summarized in Table V, due to good prediction capability on scarce cases, the WKELM can improve the control performance of unweighted KELM by rising the ER from 98.99 to 99.6%.

Besides, the proposed method shows superior prediction ability against the other algorithms. It significantly increases the prediction performance in both classification and regression compared with the basic ELM and SVM. Compared with the RF algorithm that ensembles 100 DTs with accuracy boosting, the WKELM algorithm also behaves quite satisfactory.

VI. CONCLUSION

In this article, a hierarchical data-driven method was proposed to mitigate the FIDVR events. The adaptive trajectory sensitivity analysis was proposed to optimize the ELS solution based on the quantitative TVSI index. Meanwhile, the WKELM-based hierarchical data-driven method was developed to learn the complex relationship between the pre-fault system operating condition and the online ELS actions. The proposed hierarchical structure and the weighted learning scheme could extract the control knowledge from the highly sparse and imbalanced data set. The case studies demonstrated the high prediction accuracy and excellent control performance of the proposed method. In terms of the tested two databases, more than 98% of critical buses were identified and the mean percentage error of ELS amount prediction was less than 4%. Moreover, the TDS results showed that 98% of testing cases were effectively controlled with around 2 MW load overshed on average.

In practical implementation, the proposed method could effectively accelerate the decision process and facilitate a more adaptive and faster voltage stability control. As a generalized data-driven method, it could also be applied to the online UVLS and other stability control problems.

REFERENCES

[1] "Fault-induced delayed voltage recovery," in *Proc. NERC Transmiss. Issues Subcommittee Syst. Protect. Control Subcommittee*, Princeton, NJ, USA, Report, Jun. 2009.

[2] E. G. Potamianakis and C. D. Vournas, "Short-term voltage instability: Effects on synchronous and induction machines," *IEEE Trans. Power Syst.*, vol. 21, no. 2, pp. 791–798, May 2006.

[3] M. Glavic *et al.*, "See it fast to keep calm: Real-time voltage control under stressed conditions," *IEEE Power Energy Mag.*, vol. 10, no. 4, pp. 43–55, Jul./Aug. 2012.

[4] Y. Zhang, Y. Xu, Z. Y. Dong, and P. Zhang, "Real-time assessment of fault-induced delayed voltage recovery: A probabilistic self-adaptive data-driven method," *IEEE Trans. Smart Grid*, vol. 10, no. 3, pp. 2485–2494, May 2019.

[5] Australian Energy Market Operator, "Black system South Australia 28 Sep 2016," 3rd Preliminary Rep., Dec. 2016. [Online]. Available: https://www.aemo.com.au/

[6] C. W. Taylor, "Concepts of undervoltage load shedding for voltage stability," *IEEE Trans. Power Del.*, vol. 7, no. 2, pp. 480–488, Apr. 1992.

[7] S. Arnborg, G. Andersson, D. J. Hill, and I. A. Hiskens, "On undervoltage load shedding in power systems," *Int. J. Elect. Power Energy Syst.*, vol. 19, no. 2, pp. 141–149, Feb. 1997.

[8] R. M. Larik, M. W. Mustafa, and M. N. Aman, "A critical review of the state-of-art schemes for under voltage load shedding," *Int. Trans. Elect. Energy Syst.*, vol. 29, no. 5, 2019, Art. no. e2828.

[9] D. Wang, Y. Xue, Q. H. Wu, and T. Xu, "Optimization and coordination of fault-driven load shedding and trajectory-driven load shedding," *J. Autom. Elect. Power Syst. (Chin.)*, vol. 33, no. 13, pp. 1–6, 2009.

[10] Y. Wang, I. R. Pordanjani, W. Li, W. Xu, and E. Vaahedi, "Strategy to minimize the load shedding amount for voltage collapse prevention," *IET Gener., Transmiss. Distrib.*, vol. 5, no. 3, pp. 307–313, Mar. 2011.

[11] X. Xu, H. Zhang, C. Li, Y. Liu, W. Li, and V. Terzija, "Optimization of the event-driven emergency load-shedding considering transient security and stability constraints," *IEEE Trans. Power Syst.*, vol. 32, no. 4, pp. 2581–2592, Jul. 2017.

[12] V. C. Nikolaidis and C. D. Vournas, "Design strategies for load-shedding schemes against voltage collapse in the Hellenic system," *IEEE Trans. Power Syst.*, vol. 23, no. 2, pp. 582–591, May 2008.

[13] Y. Xu, Y. Dai, Z. Y. Dong, Y. Xue, and K. P. Wong, "Load shedding and its strategies against frequency instability in power systems," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, San Diego, CA USA, Jul. 2012, pp. 1–7.

[14] P. Kundur and G. K. Morison, "Techniques for emergency control of power systems and their implementation," *IFAC Proc.*, vol. 30, no. 17, pp. 639–644, Aug. 1997.

[15] Y. Fang and Y. Xue, "An on-line pre-decision based transient stability control system for the Ertan power system," in *Proc. Int. Conf. Power Syst. Technol.*, Perth, WA, Australia, 2000, vol. 1, pp. 287–292.
[16] I. Genc, R. Diao, V. Vittal, S. Kolluri, and S. Mandal, "Decision trees-based preventive and corrective control applications for dynamic security enhancement in power systems," *IEEE Trans. Power Syst.*, vol. 25, no. 3, pp. 1611–1619, Aug. 2010.
[17] C. Liu *et al.*, "A systematic approach for dynamic security assessment and the corresponding preventive control scheme based on decision trees," *IEEE Trans. Power Syst.*, vol. 29, no. 2, pp. 717–730, Mar. 2014.
[18] J. L. Cremer, I. Konstantelos, S. H. Tindemans, and G. Strbac, "Data-driven power system operation: Exploring the balance between cost and risk," *IEEE Trans. Power Syst.*, vol. 34, no. 1, pp. 791–801, Jan. 2019.
[19] Y. Xu, Z. Y. Dong, L. Guan, R. Zhang, K. P. Wong, and F. Luo, "Preventive dynamic security control of power systems based on pattern discovery technique," *IEEE Trans. Power Syst.*, vol. 27, no. 3, pp. 1236–1244, Aug. 2012.
[20] Q. Huang, R. Huang, W. Hao, J. Tan, R. Fan, and Z. Huang, "Adaptive power system emergency control using deep reinforcement learning," *IEEE Trans. Smart Grid*, vol. 11, no. 2, pp. 1171–1182, Mar. 2020.
[21] H. Cai, H. Ma, and D. Hill, "A data-based learning and control method for long-term voltage stability," *IEEE Trans. Power Syst.*, to be published.
[22] C. Kim, P. Balaprakash, and M. Anitescu, "Graph convolutional neural networks for optimal load shedding under line contingency," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, Atlanta, GA, USA, 2019, pp. 1–5.
[23] Y. Dai, Y. Xu, Z. Y. Dong, K. P. Wong, and L. Zhuang, "Real-time prediction of event-driven load shedding for frequency stability enhancement of power systems," *IET Gener., Transmiss. Distrib.*, vol. 6, no. 9, pp. 914–921, Sep. 2012.
[24] Y. Xu, Z. Y. Dong, K. Meng, W. F. Yao, R. Zhang, and K. P. Wong, "Multi-objective dynamic VAR planning against short-term voltage instability using a decomposition-based evolutionary algorithm," *IEEE Trans. Power Syst.*, vol. 29, no. 6, pp. 2813–2822, Nov. 2014.
[25] W. Zong, G.-B. Huang, and Y. Chen, "Weighted extreme learning machine for imbalance learning," *Neurocomputing*, vol. 101, pp. 229–242, 2013.
[26] Y. Zhang, Y. Xu, Z. Y. Dong, and R. Zhang, "A hierarchical self-adaptive data-analytics method for real-time power system short-term voltage stability assessment," *IEEE Trans. Ind. Inform.*, vol. 15, no. 1, pp. 74–84, Jan. 2019.
[27] C. Ren, Y. Xu, Y. Zhang, and R. Zhang, "A hybrid randomized learning system for temporal-adaptive voltage stability assessment of power systems," *IEEE Trans. Ind. Inform.*, vol. 16, no. 6, pp. 3672–3684, Jun. 2020.
[28] I. A. Hiskens and M. A. Pai, "Power system applications of trajectory sensitivities," in *Proc. IEEE Power Eng . Soc . Winter Meeting*, 2002, vol. 2, pp. 1200–1205.
[29] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, pp. 489–501, Dec. 2006.
[30] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Trans. Syst., Man, Cybern. B*, vol. 42, no. 2, pp. 513–529, Apr. 2012.
[31] P. Branco, L. Torgo, and R. P. Ribeiro, "A survey of predictive modeling on imbalanced domains," *ACM Comput. Surv.*, vol. 49, no. 2, 2016, Art. no. 31.
[32] L. Torgo and R. Ribeiro, "Utility-based regression," in *Proc. 11th Eur. Conf. Princ. Pract. Knowl. Discov. Databases*, Warsaw, Poland, 2007, pp. 597–604.
[33] PSSE 33.0 Program Application Guide: Volume-II, Siemens Power Technol. Int., Munich, Germany, Mar. 2013.
[34] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," *BMC Genomics*, vol. 21, no. 1, Dec. 2020, Art. no. 6.

Qiaoqiao Li (Student Member, IEEE) received the B.S. degree in electrical engineering from North China Electric Power University, Beijing, China, in 2018. She is currently working toward the Ph.D. degree in power engineering with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore.

Her current research interests include power system stability and control, and intelligent system applications to power engineering.

Yan Xu (Senior Member, IEEE) received the B.E. and M.E degrees both in electrical engineering from the South China University of Technology, Guangzhou, China, in 2008 and 2011, respectively, and the Ph.D. degree from the University of Newcastle, Callaghan, Australia, in 2013.

He is currently an Assistant Professor with the School of Electrical and Electronic Engineering, Nanyang Technological University (NTU), Singapore, and a Cluster Director with Energy Research Institute, NTU, Singapore. He held the University of Sydney Postdoctoral Fellowship in Australia. His current research interests include power system stability and control, microgrid, and data analytics for smart grid applications.

Dr. Xu is an Editor for the IEEE TRANSACTIONS ON SMART GRID, the IEEE TRANSACTIONS ON POWER SYSTEMS, the *IEEE Power Engineering Letters*, and *CSEE Journal of Power and Energy Systems*, and an Associate Editor for the *IET Generation, Transmission, and Distribution*.

Chao Ren (Student Member, IEEE) received the B.E. degree in computer science from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2017. He is currently working toward the Ph.D. degree in cross-disciplinary of computer science and electrical engineering from the Interdisciplinary Graduate School, Nanyang Technological University, Singapore.

His current research interests include machine learning, data analytics, optimization, security assessment, and its applications to power engineering.

Mr. Ren was a recipient of several programming contest awards, including the Champion of Chinese Software Cup, etc.