energies

MDPI

Article

Online Prediction and Correction of Static Voltage Stability Index Based on Extreme Gradient Boosting Algorithm

Huiling Qin [1], Shuang Li [2,*], Juncheng Zhang [1], Zhi Rao [2], Chengyu He [1], Zhijun Chen [1] and Bo Li [3]

[1] Guangxi Power Grid Corporation, Nanning 530004, China; qin_hl@gx.csg.cn (H.Q.); zhangjc28131@gx.csg.cn (J.Z.); hechengyu1003@163.com (C.H.); chen_zj.xt@gx.csg.cn (Z.C.)
[2] China Southern Power Grid Co., Ltd., Guangzhou 510530, China; raozhi@csg.cn
[3] School of Electrical Engineering, Guangxi University, Nanning 530004, China; boli@gxu.edu.cn
* Correspondence: li_shuang2024@126.com

Abstract: With the increasing integration of renewable energy sources into the power grid and the continuous expansion of grid infrastructure, real-time preventive control becomes crucial. This article proposes a real-time prediction and correction method based on the extreme gradient boosting (XGBoost) algorithm. The XGBoost algorithm is utilized to evaluate the real-time stability of grid static voltage, with the voltage stability L-index as the prediction target. A correction model is established with the objective of minimizing correction costs while considering the operational constraints of the grid. When the L-index exceeds the warning value, the XGBoost algorithm can obtain the importance of each feature of the system and calculate the sensitivity approximation of highly important characteristics. The model corrects these characteristics to maintain the system's operation within a reasonably secure range. The methodology is demonstrated using the IEEE-14 and IEEE-118 systems. The results show that the XGBoost algorithm has higher prediction accuracy and computational efficiency in assessing the static voltage stability of the power grid. It is also shown that the proposed approach has the potential to greatly improve the operational dependability of the power grid.

Keywords: static voltage stability; preventive control; real-time prediction; extreme gradient boosting (XGBoost) algorithm; sensitivity approximation

Citation: Qin, H.; Li, S.; Zhang, J.; Rao, Z.; He, C.; Chen, Z.; Li, B. Online Prediction and Correction of Static Voltage Stability Index Based on Extreme Gradient Boosting Algorithm. *Energies* **2024**, *17*, 5710. https://doi.org/10.3390/en17225710

Academic Editor: Miguel Castilla

Received: 23 May 2024
Revised: 24 October 2024
Accepted: 29 October 2024
Published: 15 November 2024

1. Introduction

In recent years, many countries have experienced power grid accidents, leading to voltage collapse and significant shutdown events [1–3]. As new energy technologies continue to evolve and industrial electricity consumption rapidly escalates, the voltage stability margin in power grid operations is nearing a precarious tipping point. When the voltage stability margin surpasses its threshold, voltage collapse will manifest within the power grid, significantly impacting the country's economy and society. It is crucial to promptly and precisely evaluate the power grid's voltage stability margin to ensure the safe and stable operation of the power grid. So far, there are many indicators for evaluating power system voltage stability [1], such as singular value, eigenvalue index, voltage stability L index, etc. However, these methods typically require the construction of complex mathematical models and involve significant computational resources. It is challenging to achieve real-time monitoring of voltage stability states for large-scale power grid data [4,5].

To address the issue, machine learning-based methods are widely used to evaluate the voltage stability of the power grid [2,3]. This type of research is currently divided into classification problems and regression problems. The first type of research focuses on voltage stability as a classification problem. The objective of this problem is to determine whether the current voltage of the power grid is in a stable state [4]. This treatment method

is equivalent to simplifying the problem. The second type of research focuses on voltage stability as a regression problem: the voltage stability index under various operating conditions is used as the target for fitting [5]. However, it is significant to know the value of the system voltage stability index, as it enables the rational adjustment of the operating mode based on the prevailing conditions of the power grid. Therefore, the research of this paper addresses the topic of voltage stability by framing it as a regression problem.

The accuracy of predictions is equally crucial for the online evaluation model of static voltage stability. To enhance the precision of online voltage stability margin prediction, a model for assessing voltage stability within the transmission system through online methods was introduced in [6] utilizing an active machine learning algorithm. In [7], an improved online random forest model that incorporates drift detection and online bagging methodologies was proposed. A radial basis function (RBF) network to evaluate power grid voltage stability was proposed in [8]. The feature selection technique based on mutual information is used to reduce the dimension of input features, which reduces the training time of the model. A method for monitoring voltage stability in power systems through ANN was proposed in [9]. By calculating the sensitivity value of each input feature relative to the voltage stability margin and then selecting the feature with higher value as the input characteristic of the model, the prediction accuracy and training speed of the network are improved. These studies aim to improve the prediction accuracy by simplifying the input features of the model [6–9]. In [10], a continuous online prediction approach for anticipating power system instability was proposed utilizing a Convolutional Neural Network (CNN) model, which utilizes heatmap representations of the measurements as input for predicting instability. In order to monitor long-term voltage instability online, a support vector machine (SVM) based on a genetic algorithm was proposed in [11]. The genetic algorithm was used to calculate the optimal values of SVM parameters to improve the accuracy and speed of the algorithm. The authors of reference [12] used various machine learning methods to predict photovoltaic power, and the XGBoost algorithm had higher prediction accuracy. In order to predict hospital energy consumption, a prediction model that combines XGBoost and Random Forest (RF) was proposed in [13]. In [14], a method for predicting wind power at a regional level was introduced, which utilized XGBoost and multi-stage feature selection for short-term forecasting. In [15], a rapid and precise short-term voltage stability assessment technique employing Joint Mutual Information Maximization (JMIM) and XGBoost was introduced. JMIM is used to select key input features from high-dimensional original features, thus reducing the complexity of the model. The literature combines the strengths of two algorithms, thereby enhancing the efficiency of the model [13–15]. A strategy utilizing a deep learning model based on Multi-Layer Perceptron to enhance short-term voltage stability prediction accuracy was proposed in [16]. A data-driven method utilizing enhanced Gradient Boost Decision Tree (eGBDT) for the assessment of long-term voltage stability was introduced in [17]. In [18], the authors presented a hybrid convolutional long short-term memory (ConvLSTM) approach aimed at forecasting voltage stability. Additionally, in [19], a cascaded neural network (CNN) framework was proposed for the real-time online monitoring of voltage stability, utilizing load ability margin (LM) estimation. The predictive efficacy of the integrated solar and wind power generation system demonstrated in this study is noteworthy.

While the aforementioned techniques improve the precision of voltage stability index predictions, they have certain limitations: (1) the models involve numerous hyperparameters, resulting in extended training times; (2) during training, there is a risk of underfitting and susceptibility to local optimization; and (3) these prediction models are only functional when the system's L-index reaches a critical value. It is highly important to investigate the utilization of machine learning techniques for system adjustment in order to ensure the system remains operational within a stable range. The primary contributions outlined in this study are as follows:

(1) The L-index is a metric employed to evaluate the level of voltage stability within the power grid, and it is predicted online using the XGBoost algorithm.

(2) XGBoost is used to obtain the importance of system characteristics, calculate the approximate sensitivity value of highly important characteristics, and correct the system according to the sensitivity of these characteristics in order to guarantee that the system voltage can operate safely and steadily for an extended period.

The subsequent sections of the document are organized as follows: Section 2 provides a concise overview of the voltage stability L-index concept. Section 3 elaborates on the prediction and correction control model. Section 4 outlines the dataset generation method and scrutinizes the model's results. The paper concludes with Section 5.

2. Voltage Stability Evaluation Based on Machine Learning

2.1. Power System Voltage Stability Index

In this paper, the L index is used as the static voltage stability index of the power grid. The L index is chosen due to its well-defined physical definition and clear upper and lower bounds, and the index value can be normalized for any different system. Furthermore, the L index has previously been applied in an actual power grid [20].

The local voltage stability L index is used to estimate the operation state of the power system [20]. The electric network nodes can be divided into two categories: generator nodes (α_G) and load nodes (α_L). The node voltage can be calculated based on Kirchhoff's current law, formulated as following:

$$\begin{bmatrix} I_G \\ I_L \end{bmatrix} = \begin{bmatrix} Y_{GG} & Y_{GL} \\ Y_{LG} & Y_{LL} \end{bmatrix} \begin{bmatrix} V_G \\ V_L \end{bmatrix} \tag{1}$$

where I_G and I_L are the currents associated with the generator and load buses, respectively; V_G represents the terminal voltage of the generator; and VL represents the voltage at the load node. Y_{GG}, Y_{GL}, Y_{LG} and Y_{LL} represent subsets of the node admittance matrix. When $Z_{LL} = {Y_{LL}}^{-1}$, Equation (1) can be written as

$$\begin{bmatrix} I_G \\ V_L \end{bmatrix} = \begin{bmatrix} Y_{GG} - Y_{GL}Z_{LL}Y_{LG} & Y_{GL}Z_{LL} \\ -Z_{LL}Y_{LG} & Z_{LL} \end{bmatrix} \begin{bmatrix} V_G \\ I_L \end{bmatrix} \tag{2}$$

when $F_{LG} = -Z_{LL}Y_{LG}$, the L_j of load node j is defined as:

$$L_j = \left| 1 - \sum_{k \in \alpha_G} F_{jk} \frac{V_k}{V_j} \angle(\theta_{jk} + \delta_k - \delta_j) \right| \tag{3}$$

where V_k represents the voltage magnitude of the k-th generator; δ_k denotes the phase angle of V_j; and V_j, δ_j represent the voltage magnitude and phase of j-th node, respectively. F_{jk} is the j-th row and k-th column element of matrix F_{LG}, the angle of θ_{jk} with respect to F_{jk}.

The set of local voltage stability indexes for load nodes, denoted as $L = [L_1, L_2, \cdots L_m]$ $m \in \alpha_L$, is composed of the individual local voltage stability indexes of all load nodes within the system.

The node with the highest L value is considered the most vulnerable node within the system. The voltage stability index of the entire system is referring to:

$$L = \|L\|_\infty \tag{4}$$

In the context of local voltage stability, the L-index is indicative of the power grid's voltage stability level. Specifically, it has been noted that in order to maintain a stable voltage within a system, the L-index should be below 1. At the critical point of voltage stability, the L-index equals 1, signifying a delicate balance. Conversely, in instances of system voltage instability, the L-index exceeds 1.

2.2. Extreme Gradient Boosting (XGBoost)

The XGBoost algorithm is a highly effective algorithm in machine learning [21]. The XGBoost algorithm is a unified tree model that is formed by amalgamating various CART decision tree functions. The basic idea is to perform the second-order Taylor expansion of the objective function while assuming that a training dataset comprises n samples, denoted as $D = \{(x_i, y_i), i = 1, \ldots, n\}$, $x_i = (x_{i1}, x_{i2}, \ldots, x_{im})$ represents the m-dimensional eigenvector, and y_i is the label associated with x_i.

$$y_i^* = \varphi(x_i) = \sum_{k=1}^{K} f_k(x_i); \;\; f_k \in F \tag{5}$$

where K represents the number of decision trees in the ensemble, $F = \left\{ f(x) = \omega_{q(x)} \right\}$ $\left(q: R^M \rightarrow T, \omega \in R^T \right)$ denotes the collection of decision trees, q signifies the configuration of individual decision trees, T represents the number of terminal nodes within a tree, and each f_k is associated with a distinct tree configuration q along with the weight ω assigned to its terminal nodes.

This study examines the voltage stability index within the context of a regression problem and employs the regression model of the XGBoost algorithm. The loss function associated with the regression class is incorporated into the error link during the training phase. The loss function of the XGBoost model consists of the training error and a regularization term used to measure the complexity of the model:

$$L(\varphi) = \sum_{i=1}^{N} (y_i - y_i^*)^2 + \sum_{k=1}^{K} \Omega(f_k) \tag{6}$$

where N is the total number of samples under consideration. The initial component is the training error, which measures the difference between the actual label value y_i and the predicted value y_i^* in the predictive modeling. The larger the disparity between the projected and observed values of the L-index, the higher the training error value. The second term is a regular term, which is defined as:

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \|\omega\|^2 \tag{7}$$

In the tree structure, the initial parameter regulates the quantity of terminal nodes, whereas the subsequent parameter governs the significance or influence of these terminal nodes. γ and λ are super parameters used to adjust the proportion of the two terms.

The smaller the training error, the smaller the loss function. Considering the issue of overfitting the objective function, the regularization term value needs to be reduced. Then, the objective function is expected to exhibit increased precision and an accelerated convergence rate.

2.3. Steady State Feature Set of Power System

The voltage stability L index corresponding to the power system under any operating condition is uniquely determined. Therefore, establishing the relationship between the system characteristics under the current operating mode and the L index is essential. The characteristics of each node in the static voltage stable state of the system are selected as the input characteristics of the model. The steady-state characteristics of the system are detailed in Table 1.

Table 1. Characteristics of an electrical power system operating in a stable condition.

Feature Symbol	Feature Meaning
V_i, θ_i	The magnitude of voltage and the angle of phase at node i
P_{Gi}, Q_{Gi}	The power output in terms of active and reactive components form the i-th generator bus
P_{Lj}, Q_{Lj}	The power demand in terms of active and reactive components from the j-th load bus

Among these features, P_{Gi}, Q_{Gi} represent the generator node output and P_{Lj}, Q_{Lj} represent load demand and supply demand, respectively. These features reflect the supply–demand relationship, which changes according to the operational mode changes in the system. The output and load demand of each node in the system are determined by the power grid structure and parameters after the supply and demand are determined. These characteristics collectively represent the L-index of the system's voltage stability during a specific operational state.

3. Correction Adjustment Model

3.1. Correction Control Model

The voltage stability L index undergoes dynamic changes during power grid operation, potentially leading to voltage collapse if it surpasses a critical threshold. To maintain system voltage stability, a corrective and adjustment model based on the XGBoost algorithm is proposed. Firstly, the XGBoost model is employed to assess the impact of each system node's characteristic variable on the L index and identify the variable with the most significant influence. The XGBoost algorithm is used to determine the sensitivity of these features to the L index. Then, linear optimization techniques are used to devise an optimal adjustment plan to ensure that the L index remains below the safety limit. The correction control process of the model is illustrated in Figure 1.

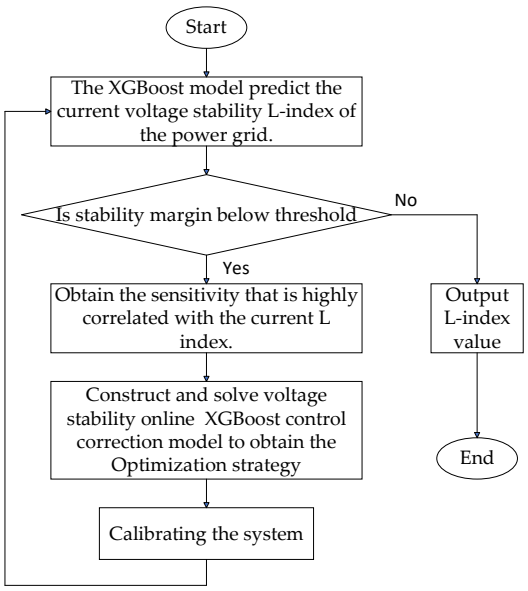


Figure 1. Correction control optimization process.

The adjusted data are fed into the trained XGBoost prediction model for validation to ascertain whether the L index falls within a secure operational range. Despite potential nonlinear model errors, the corrected L index prediction may still exceed the safety threshold. Therefore, further optimization is conducted on the refined solution, resulting in the final calibration plan when the L index value falls within the safe range.

When the operational state of the power system violates the stability threshold, the system is fine-tuned to reduce the regulatory changes. The linear optimization technique aims to minimize the regulatory changes [22,23]:

$$\min F = a \cdot \sum_{i \in S_G} \Delta V_i^2 + b \cdot \sum_{i \in S_G} \Delta P_i^2 + c \cdot \sum_{j \in S_L} \Delta P_j^2 + d \cdot \sum_{j \in S_L} \Delta Q_j^2 \tag{8}$$

Stability constraint of system operation:

$$\begin{cases} P_{Gi} - P_{Li} - V_i \sum\limits_{j \in S_n} V_j Y_{ij} \cos(\delta_i - \delta_j - \alpha_{ij}) = 0 \\ Q_{Gi} - Q_{Li} - V_i \sum\limits_{j \in S_n} V_j Y_{ij} \sin(\delta_i - \delta_j - \alpha_{ij}) = 0 \end{cases} \quad i \in S_n \tag{9}$$

where S_n represents the set of all nodes of the system; Y_{ij} and α_{ij} are the amplitude and phase of the transfer impedance between nodes i and j; and δ_i and δ_j represent the voltage phases of nodes i and j.

In order to attain an optimal solution for a given system, it is essential to impose constraints on the active power output of generators, the reactive power output of reactive power sources, and the voltage magnitudes at the system nodes, ensuring they remain within designated limits. These limitations are articulated through inequality constraints in the model:

$$\begin{cases} V_{i,\min} \le V_i \le V_{i,\max} \\ P_{Gi,\min} \le P_{Gi} \le P_{Gi,\max}, i \in S_G \\ P_{Lj,\min} \le P_{Lj} \le P_{Lj,\max} \\ Q_{Lj,\min} \le Q_{Lj} \le Q_{Lj,\max}, j \in S_L \end{cases} \tag{10}$$

where S_G represents a set of generator nodes and S_L represents a set of load nodes. $V_{i,\max}$, $V_{i,\min}$ are the upper/lower limit of voltage of unit i, respectively; $P_{Gi,\max}$, $P_{Gi,\min}$ are the upper/lower limit of real power output of unit i; $P_{Lj,\max}$,$P_{Lj,\min}$ are the upper/lower limit of real power output of node j; and $Q_{Lj,\max}$, $Q_{Lj,\min}$ are the upper/lower limit of reactive power of node j.

To improve the voltage stability of a system, there is a voltage stability operation constraint:

$$0 \le L + \left(\sum_{i \in S_G} \frac{\partial L}{\partial V_i} \cdot \Delta V_i + \sum_{i \in S_G} \frac{\partial L}{\partial P_i} \cdot \Delta P_i + \sum_{j \in S_L} \frac{\partial L}{\partial P_j} \cdot \Delta P_j + \sum_{j \in S_L} \frac{\partial L}{\partial Q_j} \cdot \Delta Q_j \right) \le t \tag{11}$$

In the process of correcting the node variables of the power system, it is essential to impose limits on the changes in the active output of generators, the changes in the reactive output, and the changes in the voltage magnitude at the system nodes, ensuring that they remain within specified limits. These constraints are articulated through inequality constraints in the model:

$$\begin{cases} \Delta V_{i,\min} \le \Delta V_i \le \Delta V_{i,\max} \\ \Delta P_{Gi,\min} \le \Delta P_{Gi} \le \Delta P_{Gi,\max}, i \in S_G \\ \Delta P_{Lj,\min} \le \Delta P_{Lj} \le \Delta P_{Lj,\max} \\ \Delta Q_{Lj,\min} \le \Delta Q_{Lj} \le \Delta Q_{Lj,\max}, j \in S_L \end{cases} \tag{12}$$

The objective function represents the total adjustment in features during the calibration control optimization process. ΔV_i is the voltage fluctuation of generator i within the chosen generator set S_G; ΔP_i is the output of generator i within the selected generator set S_G; ΔP_j and ΔQ_j are the variations in active and reactive power of node j, respectively; $\Delta V_{i,\max}$, $\Delta V_{i,\min}$ represent the upper and lower limits of voltage variation for unit i, respectively; $\Delta P_{Gi,\max}$, $\Delta P_{Gi,\min}$ represent the upper and lower limits of the real power output variation of unit i; $\Delta P_{Lj,\max}$,$\Delta P_{Lj,\ \min}$ represent the upper and lower limits of the real power output variation of node j; $\Delta Q_{Lj,\max}$, $\Delta Q_{Lj,\min}$ represent the upper and lower limits of the reactive

power variation of node j; $\partial L/\partial V_i$, $\partial L/\partial P_i$, $\partial L/\partial P_j$, and $\partial L/\partial Q_j$ represent L-V sensitivity, L-P sensitivity, L-P sensitivity, and L-Q sensitivity, respectively; and t is the value of the threshold.

3.2. Assessment Indicators

To assess the model rigorously and thoroughly, this study employs two evaluation metrics to gauge its efficacy: Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). The calculation procedures for these metrics are outlined below:

$$\begin{cases} MAPE = \dfrac{100\%}{n}\displaystyle\sum_{k=1}^{n}\left|\dfrac{y_k - \hat{y_k}}{y_k}\right| \\ RMSE = \sqrt{\dfrac{1}{n}\displaystyle\sum_{k=1}^{n}\left(y_k - \hat{y_k}\right)^2} \end{cases} \tag{13}$$

where n represents the test sample; y_k denotes the true value of the kth sample; and $\hat{y_k}$ is the estimated value of the kth sample. MAPE represents the sum of each absolute error divided by the actual value. These are common indicators used to evaluate prediction accuracy. The RMSE, or Root Mean Square Error, serves as a prevalent metric employed for assessing the predictive performance of a model.

4. Experiments and Discussion

4.1. Case Introduction

The IEEE-14 system is utilized to verify the accuracy of the proposed method, as illustrated in Figure 2. The fluctuation range is shown in Table 2, leading to various power flow samples within the tested dataset. A total of 5000 groups of training samples and 500 groups of test samples are sampled. Each specimen comprises the voltage amplitude, voltage phase angle, active power and reactive power of nodes, and the system voltage stability index L.

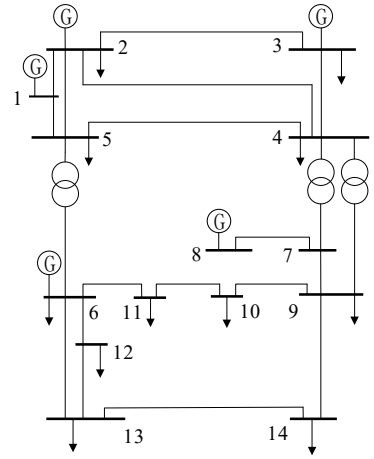


Figure 2. Illustration of the IEEE-14 system diagram.

Table 2. The range of parameters for each individual node varies.

Type of Data	Fluctuation Range
The active power demand	The data amplitude experiences ±20% variation
The active power output of generators	The data amplitude experiences ±20% variation
The voltage magnitude of node connected to generator	The data amplitude experiences ±3% variation

Upon receiving the samples, the data undergo normalization using the following calculation method:

$$x' = \frac{x - A_{\min}}{A_{\max} - A_{\min}} \tag{14}$$

where x represents the initial value; x' represents the normalized value; and $A_{\max}$ and $A_{\min}$ represent the highest and lowest values within the dataset, respectively.

4.2. Predictive Analysis of XGBoost Model

4.2.1. Comparative Analysis of Different Models in IEEE-14 System

The IEEE-14 system is utilized to verify the accuracy of the XGBoost model. The model is trained on a training dataset and then validated using a test set to assess its performance. Typical algorithms, including SVR, RF, and GBRT, are compared with the XGBoost algorithm. The comparison results are presented in Table 3. The results indicate that the XGBoost algorithm outperforms other algorithms in the context of voltage stability margin prediction. Both the RMSE and the MAPE are lower than those of the competing algorithms, suggesting that the XGBoost is capable of effectively learning from the dataset and achieving superior regression prediction performance.

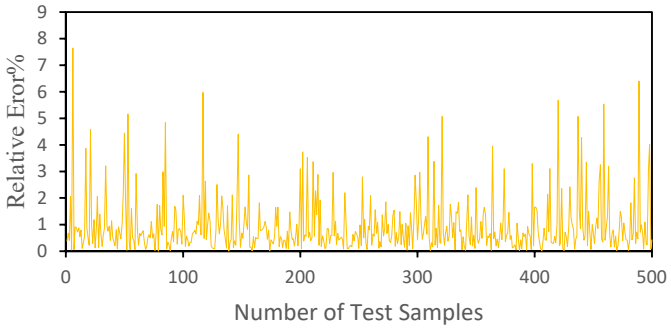
Table 3. Comparison of prediction performance of IEEE-14 dataset with different machine learning models.

Model	MAPE	RMSE
SVR	1.024	0.00301
RF	1.373	0.00624
GBRT	0.711	0.00280
XGBoost	0.615	0.00251

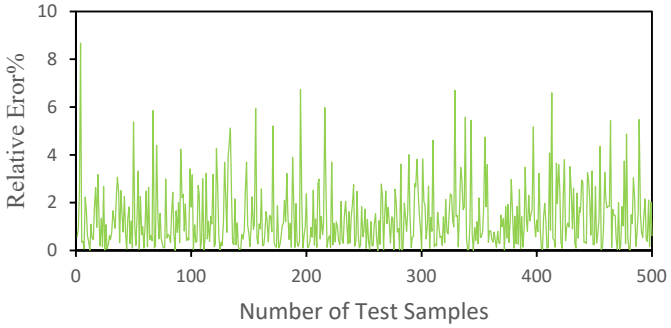
Figure 3 illustrates the fitting performance of each model on the test samples, providing a visual representation of the discrepancies between the predicted values and the actual values. The y-axis depicted in the chart denotes the discrepancy between the model's predicted value and the actual value, and the x-axis represents the number of test samples. A lower relative error indicates a greater prediction accuracy of the model. As illustrated in Figure 3, the XGBoost model demonstrates a lower overall relative error during the prediction process compared to the other three models evaluated. Consequently, the test fit of the XGBoost model is more favorable.

In the above results, for the IEEE-14 node system, the XGBoost algorithm outperforms the SVR, RF, and GBRT algorithms. For the three tree-based algorithms, RF, GBRT, and XGBoost, RF is an algorithm based on bagging, while GBRT and XGBoost are algorithms based on boosting. For the GBRT and XGBoost algorithms, XGBoost expands the loss function by utilizing the second derivative of the Taylor formula. It also incorporates a regularization term into the objective function to mitigate the issue of overfitting. Hence, the objective function demonstrates superior accuracy and quicker convergence rates compared to GBRT.

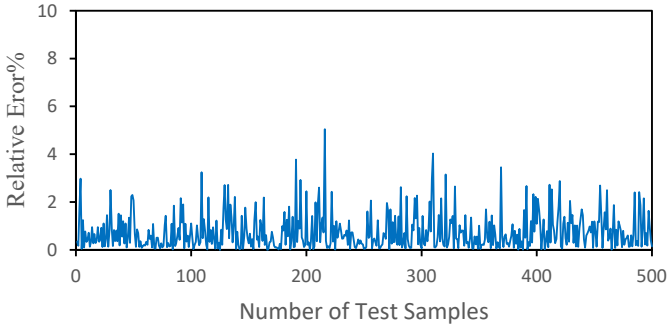
The training time T_1 and prediction time T_2 of the four machine learning algorithms in the IEEE-14 system are shown in Table 4. Due to the low model complexity of SVR, the training time consumption is less than that of the other three models. However, the training speed of XGBoost is better than that of other three models. During the prediction process, all four machine learning models have very short processing times, with the prediction time for 5000 sets of test samples being less than 0.1 s. The XGBoost model demonstrates a high level of predictive accuracy coupled with rapid prediction capabilities, and its prediction time can be as low as 0.004 s. The results show that the prediction time of the XGBoost algorithm fully meets the requirements of online applications.



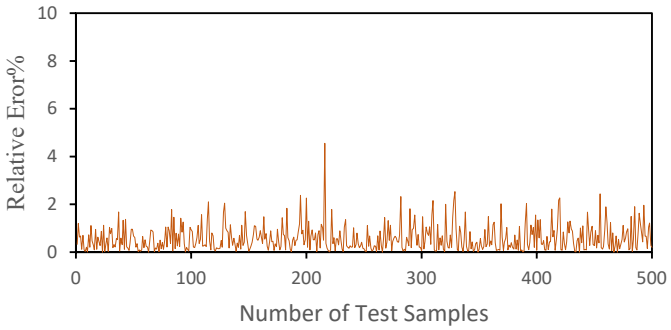
(**a**) SVR



(**b**) RF



(**c**) GBRT



(**d**) XGBoost

Figure 3. Error distribution of different machine learning algorithms.

Table 4. Training and predicting time comparison of different models in IEEE-14 system.

Model	T_1/s	T_2/s
SVR	2.73	0.022
RF	46.62	0.095
GBRT	47.10	0.007
XGBoost	16.89	0.004

4.2.2. Comparative Analysis of Different Models in IEEE-118 System

The IEEE-118 system is used to demonstrate the universal applicability of the proposed method. The method generated dataset for the IEEE-118 system is the same as that of the IEEE-14 dataset. This dataset consists of 5000 groups of training samples and 500 groups of test samples. Table 5 shows the comparison results of the prediction performances of different methods. The results show that the XGBoost algorithm performs the best in predicting the L-index among the four methods.

Table 5. Comparison of prediction performance of IEEE-118 dataset in different machine learning models.

Model	MAPE	RMSE
SVR	1.462	0.00399
RF	1.531	0.00678
GBRT	0.833	0.00317
XGBoost	0.795	0.00283

As illustrated in Table 6, the training and prediction duration of the XGBoost model applied to the IEEE-118 system is comparatively shorter than that of alternative models. Furthermore, the prediction time is recorded to be under 0.1 s, thereby satisfying the criteria for online prediction within the system. The evaluation of XGBoost's performance on the IEEE-14 and IEEE-118 systems indicates that the model demonstrates a robust capacity for generalization in the prediction of voltage stability within power systems.

Table 6. Training and predicting time comparison of different models in IEEE-118 system.

Model	T_1/s	T_2/s
SVR	20.17	0.177
RF	108.31	0.233
GBRT	112.59	0.093
XGBoost	47.89	0.089

4.3. Performance Test and Analysis of Correction Control Model

The power system's operational state is currently undergoing dynamic fluctuations. When the L-index surpasses the predetermined early warning threshold, the system is susceptible to voltage collapse. Some eigenvalues of the system need to be adjusted to restore the system to a secure state. The characteristic parameter of the IEEE-14 system, as determined by the XGBoost algorithm, is illustrated in Figure 4. As shown in Figure 4, the x-axis represents the system features, and the y-axis represents the influence of the features on the L-index. The Q_{L14} feature has the greatest impact on the L-index, while the P_{L14} and P_{L9} features also significantly influence the L-index. Due to its superior predictive accuracy, XGBoost enables the estimation of the sensitivity approximation for each feature in relation to the L-index by utilizing the sensitivity definition. Therefore, when the L-index value violates the predetermined threshold, the system can be adjusted based on the sensitivity of critical features. This greatly reduces the time required for system correction and provides significant assistance to the grid operators.

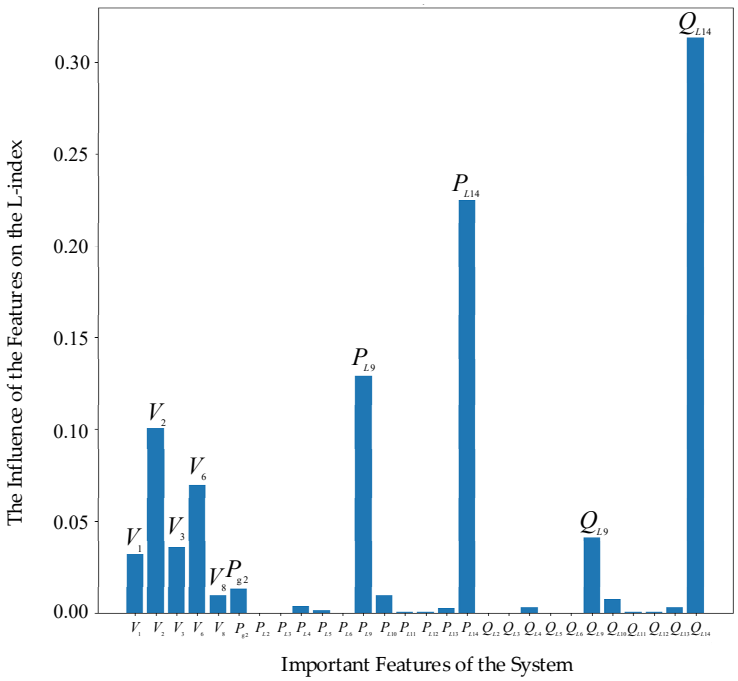


Figure 4. Feature parameter importance of IEEE-14 system node.

During the system correction procedure, it is essential to give precedence to the regulation of both the voltage and active power output of the generator to guarantee the reliability of the system. In the first stage, the generator terminal voltage and output power are adjusted; the second stage is to adjust the output of the load. To verify the efficiency of the proposed approach, the threshold value of t is 0.5, and the L-index is assumed to be 0.957. The system is modified according to the sensitivity values of V_1, V_2, V_3, V_6, V_8, Q_{L14}, P_{L14}, and P_{L9}, ensuring that the L index value falls below the designated threshold t.

The system correction process is illustrated in Tables 7 and 8. Table 7 displays the approximate sensitivity values of the characteristics concerning the L-index at various L values. Table 8 displays the results of each system correction. The values in the table represent the normalized values. L_{ture} and L_{pre} are the real and predicted values of the L-index, respectively. The upper limitation of the generator voltage is set to 1.1, and during the adjustment of the load node, the fluctuation range is within 20%. As illustrated in Table 7, the procedure involves initially adjusting the generator terminal voltage followed by adjusting the load node once the generator voltage reaches its upper limit. The nonlinearity of the model error results in a nonlinear response of input features to the variation in the L value. Therefore, even after the initial calibration, the L value remains higher than the safety threshold, albeit reduced compared to the previous value. Consequently, the calibration process is repeated until the stability index of the system voltage reaches a value within the safety threshold. The final optimized solution to be obtained can ensure the stable operation of the power grid. Thus, the efficacy of the system correction approach outlined in this study is validated.

To assess the universal applicability of the proposed methodology, the calibration model is employed to rectify the operational mode of the system across four different cases. Table 9 displays the correction results under four different cases, and the results show that the proposed approach can modify the node feature of the system, thereby ensuring that the voltage stability index remains within a safe and stable range. In instances where the predicted L value fails to satisfy the threshold condition, the model will continue to correct on the basis of this correction until the predicted L value meets the threshold condition. As illustrated in Table 10, the relative errors between the predicted values and the actual values of the L-index, following the correction, fall within an acceptable range. This finding substantiates the efficacy of the optimization model presented in this study.

Table 7. Sensitivity of indexes V_1, V_2, V_3, V_6, V_8, Q_{L14}, P_{L14}, P_{L9}, P_{L14}, and Q_{L14} corresponding to different values of the L-index during system calibration of the IEEE-14 bus system.

Correction	L_{pre}	$\frac{\partial L}{\partial V_1}$	$\frac{\partial L}{\partial V_2}$	$\frac{\partial L}{\partial V_3}$	$\frac{\partial L}{\partial V_6}$	$\frac{\partial L}{\partial V_8}$	$\frac{\partial L}{\partial P_{L9}}$	$\frac{\partial L}{\partial P_{L14}}$	$\frac{\partial L}{\partial Q_{L14}}$
Initial	0.957	0.779	1.400	0.950	2.954	1.520	0.328	2.684	6.237
1st	0.618	—	—	—	—	—	0.215	0.523	1.554
2nd	0.519	—	—	—	—	—	0.153	0.359	1.029

Table 8. The correction outcomes are observed when the L-index equals 0.957.

Correction	L_{ture}	L_{pre}	V_1	V_2	V_3	V_6	V_8	P_{L9}	P_{L14}	Q_{L14}
Initial	0.946	0.957	1.070	1.049	1.019	1.072	1.097	1.397	0.703	0.236
1st	0.621	0.618	1. 100	1.100	1.100	1.100	1.100	1.397	0.703	0.204
2nd	0.524	0.519	1.100	1.100	1.100	1.100	1.100	1.397	0.599	0.163
3rd	0.516	0.499	1.100	1.100	1.100	1.100	1.100	1.397	0.599	0.144

Table 9. Correction results in different cases in the IEEE-14 bus system.

	Initial		1st		2nd		3rd	
	L_{ture}	L_{pre}	L_{ture}	L_{pre}	L_{ture}	L_{pre}	L_{ture}	L_{pre}
Case1	0.946	0.957	0.621	0.618	0.524	0.519	0.516	0.499
Case2	0.837	0.843	0.601	0.588	0.522	0.513	0.509	0.498
Case3	0.773	0,764	0.541	0.532	0.517	0.499	—	—
Case4	0.637	0.641	0.518	0.505	0.511	0.497	—	—

Table 10. An analysis of the correction outcomes within the IEEE-14 bus system.

	Threshold	L_{ture}	Relative Error
Case1	0.500	0.516	3.20%
Case2	0.500	0.509	1.80%
Case3	0.500	0.517	3.40%
Case4	0.500	0.511	2.20%

This model demonstrates the ability to propose effective corrective actions by precisely determining both the direction and magnitude of adjustments required for each node feature within the system. This capability enhances correction efficiency, aligns the correction outcomes more closely with the anticipated values, and alleviates the workload for frontline personnel.

5. Conclusions

Accurate prediction of the static voltage stability index in the power grid is crucial for maintaining the system's operation within the voltage stability range. In this paper, a prediction and correction control model based on XGBoost is proposed to improve voltage stability. The evaluation was conducted on the IEEE 14 and 118 systems. The predictive outcomes indicate that the RMSE and MAPE metrics of the XGBoost model surpass those of the alternative algorithms under comparison. Moreover, the XGBoost model is faster than tree-based models in training speed and quicker than other algorithms in prediction speed, as well as the time required to predict the actual power system's online forecasting needs. Additionally, this paper also proposes a method to correct the system by utilizing feature sensitivity, integrating the XGBoost model with a calibration control framework. Correction results in the IEEE-14 bus system show that when the L index of a specific operational mode exceeds the established safety threshold, the model effectively recalibrates the L index to align with the safety threshold, thereby enhancing the operational reliability of the system and demonstrating considerable practical applicability.

This study exclusively focuses on conventional energy generation methods. The integration of renewable energy sources into the power grid has introduced significant volatility when compared to traditional energy generation. Consequently, ensuring the safe and stable operation of the power system presents a considerable challenge for the grid. Future research will incorporate wind and solar power generation to investigate the effects of these new energy sources on voltage stability, as well as the feasibility of the proposed model.

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

Funding: This research was funded by Guangxi Power Grid Corporation. This research was funded by "Research on distribution system planning integrating large-scale distributed PV system", grant number "GXKJXM20222144".

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Conflicts of Interest: Authors Huiling Qin, Juncheng Zhang, Chengyu He and Zhijun Chen were employed by the Guangxi Power Grid Corporation. Authors Shuang Li and Zhi Rao were employed by the China Southern Power Grid Co., Ltd. The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

References

1. Chandra, A.; Pradhan, A.K. A comparative study of voltage stability indices used for power system operation. In Proceedings of the 2016 21st Century Energy Needs—Materials, Systems and Applications (ICTFCEN), Kharagpur, India, 17–19 November 2016; pp. 1–4.
2. Shi, Z.; Yao, W.; Li, Z.; Zeng, L.; Zeng, L.; Zhao, Y.; Zhang, R.; Tang, Y.; Wen, J. Artificial intelligence techniques for stability analysis and control in smart grids: Methodologies, applications, challenges and future directions. *Appl. Energy* **2020**, *278*, 115733. [CrossRef]
3. Glavic, M.; Greene, S. Voltage stability in future power systems. In *Encyclopedia of Electrical and Electronic Power Engineering*; García, J., Ed.; Elsevier: Oxford, UK, 2023; pp. 209–223.
4. Moulin, L.S.; da Silva, A.P.A.; El-Sharkawi, M.A.; Marks, R.J. Support vector machines for transient stability analysis of large-scale power systems. *IEEE Trans. Power Syst.* **2004**, *19*, 818–825. [CrossRef]
5. Zhou, D.Q.; Annakkage, U.D.; Rajapakse, A.D. Online Monitoring of Voltage Stability Margin Using an Artificial Neural Network. *IEEE Trans. Power Syst.* **2010**, *25*, 1566–1574. [CrossRef]
6. Malbasa, V.; Zheng, C.; Chen, P.; Popovic, T.; Kezunovic, M. Voltage Stability Prediction Using Active Machine Learning. *IEEE Trans. Smart Grid* **2017**, *8*, 3117–3124. [CrossRef]
7. Su, H.; Liu, T. Enhanced-Online-Random-Forest Model for Static Voltage Stability Assessment Using Wide Area Measurements. *IEEE Trans. Power Syst.* **2018**, *33*, 6696–6704. [CrossRef]
8. Devaraj, D.; Preetha Roselyn, J. On-line voltage stability assessment using radial basis function network model with reduced input features. *Int. J. Electr. Power Energy Syst.* **2011**, *33*, 1550–1555. [CrossRef]
9. Chakrabarti, S. Voltage stability monitoring by artificial neural network using a regression-based feature selection method. *Expert Syst. Appl.* **2008**, *35*, 1802–1808. [CrossRef]
10. Gupta, A.; Gurrala, G.; Sastry, P.S. An Online Power System Stability Monitoring System Using Convolutional Neural Networks. *IEEE Trans. Power Syst.* **2019**, *34*, 864–872. [CrossRef]
11. Sajan, K.S.; Kumar, V.; Tyagi, B. Genetic algorithm based support vector machine for on-line voltage stability monitoring. *Int. J. Electr. Power Energy Syst.* **2015**, *73*, 200–208. [CrossRef]
12. Guo, X.; Gao, Y.; Zheng, D.; Ning, Y.; Zhao, Q. Study on short-term photovoltaic power prediction model based on the Stacking ensemble learning. *Energy Rep.* **2020**, *6*, 1424–1431. [CrossRef]
13. Cao, L.; Li, Y.; Zhang, J.; Jiang, Y.; Han, Y.; Wei, J. Electrical load prediction of healthcare buildings through single and ensemble learning. *Energy Rep.* **2020**, *6*, 2751–2767. [CrossRef]
14. Li, W.; Peng, X.; Cheng, K.; Wang, H.; Xu, Q.; Wang, B.; Che, J. A Short-Term Regional Wind Power Prediction Method Based on XGBoost and Multi-Stage Features Selection. In Proceedings of the 2020 IEEE 3rd Student Conference on Electrical Machines and Systems (SCEMS), Jinan, China, 4–6 December 2020; pp. 614–618.

15. Yu, L.; Liu, W.; Si, R.; Xing, P.; Huang, M.; Wen, Y. Short-Term Voltage Stability Assessment of Multi-infeed HVDC Systems Based on JMIM and XGBoost. In Proceedings of the 2021 3rd Asia Energy and Electrical Engineering Symposium (AEEES), Chengdu, China, 26–29 March 2021; pp. 752–758.
16. Shahriyari, M.; Safari, A.; Quteishat, A.; Afsharirad, H. A short-term voltage stability online assessment based on multi-layer perceptron learning. *Electr. Power Syst. Res.* **2023**, *223*, 109562. [CrossRef]
17. Gao, H.; Cai, G.; Yang, D.; Wang, L. Real-time long-term voltage stability assessment based on eGBDT for large-scale power system with high renewables penetration. *Electr. Power Syst. Res.* **2023**, *214*, 108915. [CrossRef]
18. Abbass, M.J.; Lis, R.; Awais, M.; Nguyen, T.X. Convolutional Long Short-Term Memory (ConvLSTM)-Based Prediction of Voltage Stability in a Microgrid. *Energies* **2024**, *17*, 1999. [CrossRef]
19. Anthony, K.; Arunachalam, V. Application of cascaded neural network for prediction of voltage stability margin in a solar and wind integrated power system. *Eng. Appl. Artif. Intell.* **2024**, *138*, 109368. [CrossRef]
20. Kessel, P.; Glavitsch, H. Estimating the Voltage Stability of a Power System. *IEEE Trans. Power Deliv.* **1986**, *1*, 346–354. [CrossRef]
21. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
22. Yang, Y.; Huang, Q.; Li, P. Online prediction and correction control of static voltage stability index based on Broad Learning System. *Expert Syst. Appl.* **2022**, *199*, 117184. [CrossRef]
23. Yang, Y.; Long, J.; Yang, L.; Mo, S.; Wu, X. Correction Control Model of L-Index Based on VSC-OPF and BLS Method. *Sustainability* **2024**, *16*, 3621. [CrossRef]