Electrical Power and Energy Systems 123 (2020) 106251

Contents lists available at ScienceDirect

Electrical Power and Energy Systems

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

# Decision tree-based classification of multiple operating conditions for power system voltage stability assessment

Luigi Vanfretti[a,*], V.S. Narasimham Arava[b]

[a] Electrical, Computer and Systems Engineering Department, Rensselaer Polytechnic Institute, Troy, NY 12180, United States
[b] WAMS Center of Excellence, GE Grid Solutions, 1 Tanfield, Edinburgh EH3 5DA, United Kingdom

ARTICLE INFO

*Keywords:*
Voltage stability assessment
Decision tree
Optimal power flow
Saddle node bifurcation
Machine Learning

ABSTRACT

This paper presents a method that performs classification of thousands of operating conditions w.r.t. power system voltage stability by using decision trees. The proposed method uses a new and flexible classification criterion that allows to identify operating conditions that are near or within the region for which the system is voltage unstable, and more importantly, that can consider operational requirements. The method creates both training and test data sets when building and validating the decision trees. To minimize computational burden, a sampling method is proposed, this method exploits the Saddle Node Bifurcation conditions to explore the operational space used to train the decision trees. Case studies were performed using the IEEE 9-bus system for several operating conditions and different network configurations. This paper also proposes the use of time domain simulations to assess the prediction accuracy of decision trees. Decision trees were created for network configurations involving outage of the line were tested on test sets and also using time domain simulations results from PSS/E. The ability to classify the degree of voltage stability of a multitude of operation conditions could be useful to aid operators in selecting and applying preventive measures to steer away the system from unstable conditions or conditions that are close to breaching operational requirements w.r.t. voltage stability.

## 1. Introduction

### 1.1. Motivation

During the recent years the challenge brought by the ongoing energy transition has led electric utilities to operate closer to their operating-limits which has made voltage instability a major concern for power systems. One of the great challenges for electric utilities and regional transmission organization is being able to meet system-wide voltage security. Voltage stability is the ability of a power system to sustain acceptable voltage at all buses under the normal condition after being subjected to a disturbance [2]. Voltage instability usually occurs in power systems that are heavily loaded or faulted or has shortage of reactive power.

Broadly speaking, two types of situations may lead to voltage instability. The first type is associated with the demand not being met by the available generation due to transmission or reactive power limitations. This situation may result from unexpected large load increase and/or an earlier weakening of the system, such as low voltages and increased losses. The second type of incident is a major event affecting the generation or transmission system in such a way that the demand, which is the pre-fault consumption, cannot be satisfied with the available generation or transmission capacity.

Voltage Stability Indices (VSI) were developed to deal with the first type of situation. These indices help to foresee unacceptable effects of load increments. Moreover, this anticipation capability along with inherent delay (in some cases) of load increments (because of the type of behavior of certain loads) gives the operator some time to take remedial actions such as switching capacitor banks, changing the generator voltage set points, etc. However, the picture is quite different for voltage instabilities that can follow major incidents such as outage of a large capacity generator that is producing its maximum rated power or the disconnection of heavily loaded transmission lines. The time left to take remedial actions for the second type of situation is relatively shorter than the first. This short time is very important and early detection of a critical state can prevent the system from collapsing.

These above considerations motivate the development of approaches that can help in early identification of voltage instability and suggest remedial actions to bring back the system to stable state. Machine learning techniques like Decision Trees (DTs) can offer useful tools to handle the early identification of voltage instability by performing off-line analysis of thousands of potential operating conditions

* Corresponding author.
*E-mail address:* VANFRL@RPI.EDU (L. Vanfretti).

https://doi.org/10.1016/j.ijepes.2020.106251
Received 4 October 2019; Received in revised form 23 April 2020; Accepted 2 June 2020
Available online 25 June 2020
0142-0615/ 

ahead of time.

### 1.2. Literature review

Identification of the voltage stability boundary (VSB) plays a vital role in the reliable operation of a power system. Although the voltage stability margin depends on numerous possible system stresses in practical real-time applications, only several selected stress directions are computed and checked in practice, especially, real-time static or dynamic security assessment (DSA) tools. However, with the increasing variability and uncertainty in today's power systems, it is becoming increasingly clear that the stability margin assessment should be broadened to multiple stresses covering various sources and ranges of uncertainty and variability. Therefore, an accurate and fast estimation of the available voltage stability margin is of paramount importance for the secure operation and control of electric power systems.

The voltage stability boundary surrounds the region of feasible and stable operating points in power system parameter space. These operating points cannot cross the VSB without losing their stability [4]. The voltage stability region (VSR) is a safe region for guaranteeing local stability at the equilibrium under slow parametric variations [5]. Voltage stability conditions are usually considered as power flow feasibility conditions; and the VSB is associated with singularity conditions of the power flow Jacobian matrix and saddle-node bifurcation (SNB). References [4] and [5] provide a comprehensive discussion on feasibility boundaries and regions in state and parameter space in the power system domain and summarize some recent development on the stability analysis of large-scale systems. Some publications that address the voltage stability problems include [6–9].

The traditional methods for calculating SNBs traditionally employ iterative procedures. An extensive review of these methods is provided in [10]. The two commonly used iterative methods are Continuation Power Flow (CPF) and Direct methods [11]. The purpose of CPF is to find a series of power flow solutions for a given load/generation change scenario [12]. The CPF method provides reliable convergence due to its predictor-corrector approach but is computationally intensive. Direct methods were proposed for assessing the VSB, in which augmented power flow equations are solved [13–16]. These methods provide the left or right eigenvectors corresponding to the zero Jacobian matrix eigenvalue at the point of voltage collapse. Direct methods are sensitive to the initial guess, i.e., the initial guess affects the speed of convergence and may even cause divergence of the iterative process [16]. This iterative process requires a considerable number of calculations to find a single VSB point. To obtain the full VSB, the computational effort becomes prohibitively significant for large-scale power systems. Computational time becomes critically important for real-time analyses, massive contingency screenings, and time domain simulations.

Machine learning techniques, such as Decision Trees (DTs), clustering algorithms, neural networks and statistical methods have been considered for voltage stability assessment [17,18]. These methods can create/use a model, which is based on the knowledge of the operator decision or historical data. The DT is a white-box model that can be applied when functioning/working of a system is unknown or complex, but there is plenty of data available. These models do not explicitly model the physical system but establish a mathematical relationship between many input-output pairs measured from the system. The mathematical relationship is a model of the system, which can be computed numerically from the measurements or simulated outputs. The accuracy of the model may vary depending on accuracy of the simulated outputs replicating the behavior of original system. The investigation of DTs for voltage security assessment sparked interest in the early 90s. The DT based approach for power system security assessment was presented in [19,20]. Due to the wide deployment of PMUs in the recent years, real time security assessment combining DT and synchrophasor measurements became possible [21–27].

### 1.3. Paper contributions

This paper proposes the use of DTs for voltage stability assessment, which in turn can aid in deriving preventive actions that can be given as recommendations to system operators or automatic load shedding schemes. The idea behind this approach is to enlarge and generalize the existing security boundary method of "stable" and "unstable" region to classify the operating space based on the distance from the nearest Saddle-Node Bifurcation (SNB). This approach, along with DTs trained on the load parameter space for voltage security assessment, offers the following contributions:

- To propose a new classification criterion that enlarges and generalizes the existing security boundary method of "stable" and "unstable" regions to classify the operating space based on the distance from the nearest Saddle-Node Bifurcation (SNB); thus, allowing to consider operational requirements w.r.t. voltage stability.
- To propose a sampling approach that reduces computational burden by exploiting Saddle Node Bifurcation conditions when exploring the operational space used to train the decision trees.
- To propose a "workflow" (i.e. algorithm) based on the two previous points above to create a database with power flow outputs of several operating points for different network configurations and use them to train the decision trees.
- To propose a new approach in testing the prediction accuracy of decision trees w.r.t. random time domain simulation outputs (voltage magnitude, active power consumption at load buses).

## 2. Voltage stability assessment and operating condition classification using decision trees

A Decision Tree (DT) is a form of inductive learning. For a given data set, the objective is to build a model that captures the mechanism that gave rise to the data. The process of constructing the model is a "Supervised learning" problem because the training is supervised by an outcome variable called the target. Decision Trees are grown through a systematic method known as recursive binary partitioning; where the successive questions with yes/no answers are asked in order to partition the sample space [19].

The nodes are the points in a tree where a test is done on the attribute; branches are the outcomes of the test that lead to another node. There are three kinds of nodes: root node, internal node and leaf node. The root node is the topmost node, internal nodes are in-between and leaf node and end nodes. The completion of the test is decided by the purity of each node. If a node attains a certain pre-defined level of class purity (i.e. having only one type of output in that node), then the node is terminated. In order to classify a new sample, the attribute values are tested against the decision tree. A path is traced from root node to the leaf node that holds the class prediction for that sample. A schematic view of the DT is shown below in Fig. 1. The basic task in building a DT is to repeatedly find an attribute to be tested on a node and then branch to another node. The process of finding an attribute for testing and branching is called splitting.

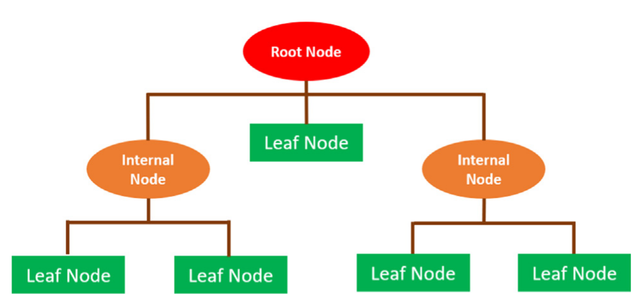


**Fig. 1.** Schematic view of a decision tree.

The objective of a split in a tree is to reduce the impurity (uncertainty) in the dataset w.r.t. class in the next stage, which can be accomplished by calculating the information gain. This calculation is done in two stages. First, the Entropy (ENT) of the dataset is measured as

$$Entropy(S) = -\sum_{i=1}^{c} p_i \log_2(p_i) \quad (1)$$

where $c$ is the number of classes, S is the training data/instances, and $p$ is the proportion of S classified as $i$. The expected information gain is calculated by using ENT as follows

$$Gain(S, a) = Entropy(S) - \sum_{v \in values} \frac{|S_v|}{|S|} Entropy(S_v) \quad (2)$$

Where $S_v = \{s \in S: a(s) = v\}$ with $v$ being the value of the attribute, $Gain(S, a)$ is the expected information gain obtained from the knowledge of the attribute $a$.

A. Attribute selection for the DTs

In this paper, load active powers and voltages are considered as attributes for splitting the data because they are one of the important factors in assessing the voltage stability of the system. From the definition of voltage stability, it can be observed that systems inability to cater the power demand of the load is one important factor that may lead a system to voltage instability. Therefore, the chosen attribute should be such that it can discriminate between different system conditions. For example, the voltage of a voltage-controlled bus is a bad attribute as it is tightly controlled, while voltages and angles of the buses that are electrically distant from the generators (e.g. radially connected loads with high impedance) are good attributes for classification. Hence, load active powers and voltages are used as attributes to build the decision trees that can aid the power system operator with voltage stability assessment for different load power consumptions and network configurations.

B. Workflow

Data is required to train and test a decision tree. The training set is used to build the decision tree and the test set is used to check accuracy of the decision tree. Depending on the availability of data, there are various procedures as cross validation, leave one out and bootstrap to use for model validity [21].

The workflow (i.e. algorithm) proposed in this work to build a decision tree for the selected network is shown in Fig. 2. For each topology, such as the base case and for different contingencies, a database is created with the power flow results for different load power consumptions. These databases are then used to train the decision trees that are used to predict the voltage stability of the considered system using measured load powers and voltages.

The advantages of using simulated data is the ease of obtaining a variety of data within a very short span of time which is not the case for real time data, for example when obtaining the response of the system in case of a fault (which happens seldom in real system), and the low cost and low risk environment that it provides to assess all possible operation scenarios.

C. Inputs to train the DTs

The DTs are created w.r.t. every load bus for different network configurations. The number of DTs created is therefore proportional to number of load buses and topologies considered. The number of branches for a tree increases with an increase in data. Creating one tree for a network configuration increases the size of the tree that further complicates interpretation. Moreover, creating one tree for a network configuration increases the computational burden and lookup time. For these reasons every network configuration will have a tree w.r.t. every load bus. Next, a Continuation Power Flow (CPF) is carried out for different network configurations with several load variations that are then used to train and test the decision tree for those network configurations.

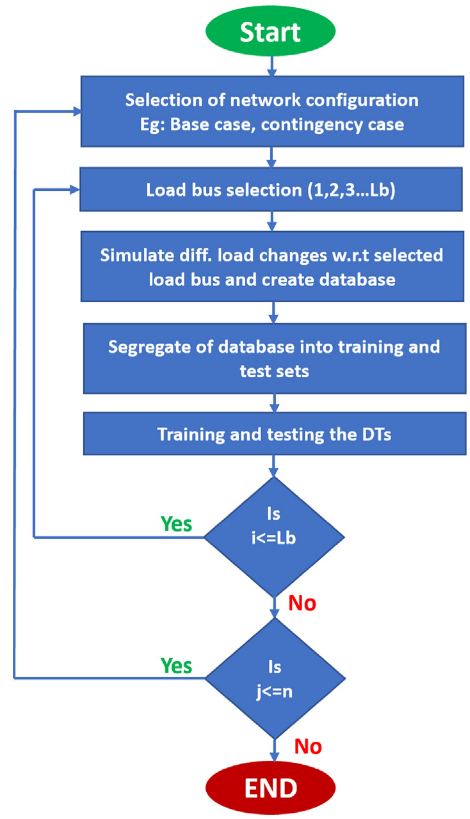


**Fig. 2.** Proposed workflow for the creation of decision trees.

D. Sampling

Initially, SNB points are calculated using CPF [12]. The purpose of CPF is to find a series of power flow solutions for a given load/generation change scenario. The CPF method provides reliable convergence due to its predictor-corrector approach but is computationally intensive. Therefore, in order to lessen the computation burden, this paper proposes the use of a direct method to calculate the SNB points. Direct methods are sensitive to the initial guess [13,14]. Consequently, the SNB direct method is initialized by running a CPF routine once to provide good initial guess [16]. Direct methods provide the left or right eigenvectors corresponding to the zero Jacobian matrix eigenvalue at the point of voltage collapse. The conditions for a SNB point are as follows:

$$\begin{aligned} g(y, \lambda) &= 0 \\ g_y v &= 0 \\ |v| &= 1 \end{aligned} \quad (3)$$

or

$$\begin{aligned} g(y, \lambda) &= 0 \\ g_y^T w &= 0 \\ |w| &= 1 \end{aligned} \quad (4)$$

where $v$ and $w$ are the right and left eigen vectors.

In this paper, instead of running CPF method for all the operating points to trace the unstable boundary, the CPF is executed to provide good initial guess for the direct method. The direction of load variations in CPF is fixed and it is along the load bus of interest (note that this would be repeated for every bus of interest). Initially, the CPF method is run for the load bus of interest (load P1 as shown by the blue star). The CPF method finds the SNB bifurcation point. The direct method uses this SNB and set direction as initial guess to calculate the eigen vectors (as shown by the blue arrows). From here, the direct method is applied sequentially in the same direction to calculate and trace the SNB boundary (as shown by the blue arrows and the red stars). Thus, the direct method calculates the boundary of SNB points as shown in the Fig. 3. Using equations (3) and (4), the workflow Fig. 2. is modified to sample a reduced number of operation conditions are obtained when

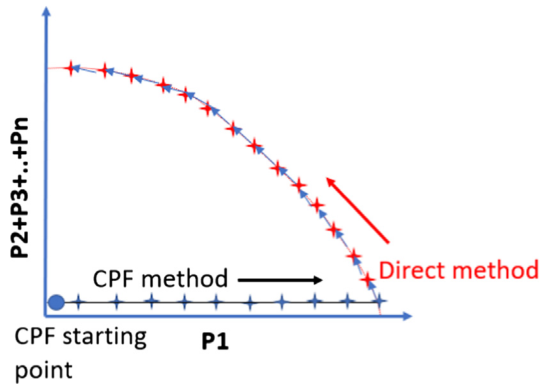


**Fig. 3.** Illustration of the applied method.

exploring the operational space used to train the decision trees, instead of using the entire space.

In the proposed method, the CPF routine is executed only once followed by a direct method to calculate the boundary of unstable points. The runtime burden for using CPF routine to calculate the boundary of unstable points for $N$ operating points is given by $O(Nn)$(where $n$ is the number times the CPF simulations needs to be executed). The computational burden for the propsed method is $O(1)$ because it runs CPF only once for one operating point followed by less computationally intensive direct method that calculates the boundary of unstable points. The same can be observed from Fig. 4 that as the number of CPF routines increased, a runtine burden on $O(Nn)$(e.g. N = 3 - for 3 operating points) is much higher than $O(1)$. The operating points for which CPF routine need to run will be in the order of hundred or thousand based on the network. Thus, making the proposed method computationally more efficient.

E. Voltage Stability classification for training the DTs

DT applications to power systems have been studied in [17–19]. According to [19,29], there is no standard universal approach for voltage stability classification using DTs. The region of operation is classified to "stable" and "unstable". The disadvantage of classifying the region of operation to only "stable" and "unstable" is that if the system is operating in the boundary of these regions, the decision tree trained on this data identifies the current operating point in "stable" region. In order to avoid this problem, the region of operation in this paper is broadened and generalized, providing a classification in four regions. They are the "stable within grid limits", "stable outside grid limits", "marginally stable" and "unstable" regions. Initially, the Euclidean distance is calculated for the given load operating point $i$ w.r.t. load bus $a$ (considering loads at bus $a$, $b$, $c$….$k$) ($Pa_i$, $Pbc$…$k_i$) from nearest unstable point ($Pa_{nSNB}$, $Pbc$….$k_{nSNB}$) using

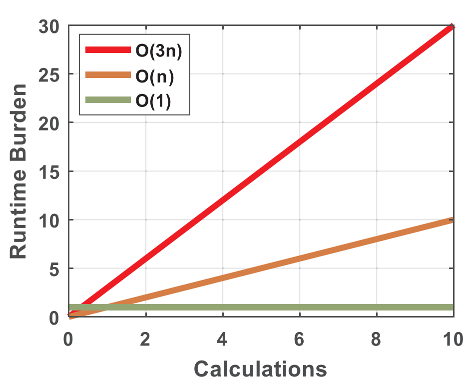


**Fig. 4.** Comparison of runtime characterstics of $O(Nn)$ and $O(1)$.

$$\text{Distance} = \sqrt{(Pa_{nSNB} - Pa_i)^2 + (Pb_{nSNB} - Pb_i)^2 + \ldots + (Pk_{nSNB} - Pk_i)^2} \quad (5)$$

The margin is calculated as given by

$$\text{Margin} = \frac{\text{Distance}}{Pa_i + Pb_i + \ldots + Pk_i} \quad (6)$$

The nearest unstable point ($Pa_{nSNB}$, $Pbc$….$k_{nSNB}$) is selected based on the distance calculated to all the unstable point on the boundary, which are obtained using the Sampling method discussed in Section II.D above. Note that other methods that could be used for this purpose [36].

If (6) is less than a given percentage (e.g. 25%), then the region of operation is classified as a "marginally stable" region. If the available margin is greater than or equal to the given percentage (e.g. 25%) with voltages at all the load buses being greater than a given threshold (e.g. 0.95 pu), then the region is classified as "stable within grid limits" region, otherwise it is classified as "stable outside grid limits" region. If the given load operating point is the saddle node bifurcation point or it exists further away from the given saddle node bifurcation point, then the region is classified as being in the "unstable" region. The fact that the margins can be customized depending on the power system and how conservative an operator makes these criteria general. The classification criteria are visualized in Fig. 3. The given load flow outputs are classified into the regions based on the conditions explained above. The trained decision trees are tested with the test set and the accuracy of the classification is calculated. Low accuracy in classification indicates that the decision tree is not trained properly, and it is needed to be retrained.

## 3. Case studies and results

The main objective of the studies herein was to demonstrate the use of decision trees for voltage stability classification. The proposed workflow in Fig. 2 and the sampling method described in the previous section was implemented in MATLAB. The reason for this choice is because access to the model's Jacobian to perform the SNB condition computations is necessary. In addition, access to a CPF for the initialization of the proposed sampling method was also necessary, and to this end the CPF from PSAT [30], was used. This proposed approach was tested on the IEEE 9 bus system. The time domain simulations of IEEE 9 bus system were carried out for different loading conditions and network configurations using PSS/E. The simulations in PSS/E were automated by a Python script. Later, the machine learning toolbox available in MATLAB was used to train and test the decision trees on this simulation results obtained from PSS/E. Finally, a MATLAB script was written to validate the created decision trees by generating random load powers.

A. Test network:

The IEEE 9 bus system shown in Fig. 4 is used for case studies. The power system data for the case study was in the formats used by PSS/E and PSAT. Therefore, power system data (transmission line impedances, transformer impedances, generator ratings and exciter types) in both the simulation environments were verified manually and care was taken to make the power flow converge to the same bus voltage magnitudes and angles. With regards to load modeling, the following choices were made:

- For all power flow and CPF computations, the loads are assumed to have no dynamic components and are modeled as consuming constant power under a constant power factor. However, for time domain simulations these loads were replaced with exponential recovery load models.
- Simulations were performed for different loading conditions w.r.t. every load bus for different network configurations.

*Simulated network configurations:*

Power flow studies were conducted by varying the demand at the load buses. These simulations were automated using a Python script. Data was generated for different network configurations w.r.t. every load. Apart from the base case where all the transmission lines are in service, the following network configurations were also simulated:

a) Outage of line between Bus 4 and 5,
b) Outage of line between Bus 5 and 7,
c) Outage of line between Bus 4 and 6,
d) Outage of line between Bus 6 and 9,
e) Outage of line between Bus 7 and 8,
f) Outage of line between Bus 8 and 9.

For the above-mentioned network configurations, power flow simulations were conducted for load variations w.r.t. every load, i.e. w.r.t. load 5, load 6 and load 8. The DT for each network configuration is trained on over 100,000 simulated operational points. The generated data was stored in .mat files.

B. Results for voltage stability assessment using DTs:

The set of simulated load data points for the base case network configuration is shown in Fig. 5. w.r.t. load 5, i.e. for a fixed change in demand at load 6 and load 8, demand at load 5 is changed in small increments. Positive load growth w.r.t all the loads (load 5, load 6 and load 8) are considered to generate data. The load power at the bus 5 is increased in fixed steps of 0.1pu along with the rest of the loads in the system.

These points were split into a training set (65% of the data) and testing set (35% of the data). Training and test sets are randomly selected from the simulated data to discretize the formulation of decision trees. The points in the testing set were then classified to regions, based on the classification rules provided in the previous section. Fig. 6. shows the data in Fig. 5. classified as per defined classification rules. This classified data was used to create the decision tree for the base case network configuration w.r.t. load 5.

*Training voltage stability assessment DTs:*

For a network configuration, decision trees were created for variations w.r.t. every load. So, every network configuration of the IEEE 9 bus system has three decision trees. The MATLAB function fitctree in the Statistics and Machine Learning Toolbox was used to create the decision trees. It should be noted that the MATLAB machine learning toolbox divides the training set data into training and cross validation sets in order to deal with the overfitting problem [34]. This function performs the following steps [34]:

1. Start with all the input data and examine all possible binary splits on every predictor.
2. Select a split based best optimization criterion subject to minimum leaf size constraint.
3. Impose the split
4. Repeat steps 1–4 recursively for the two child nodes

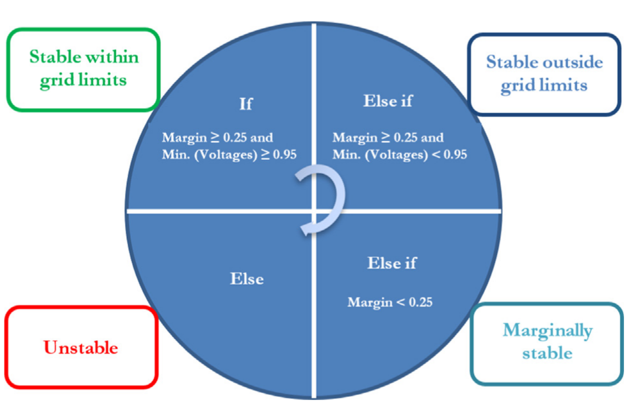


**Fig. 5.** Proposed classification criteria.

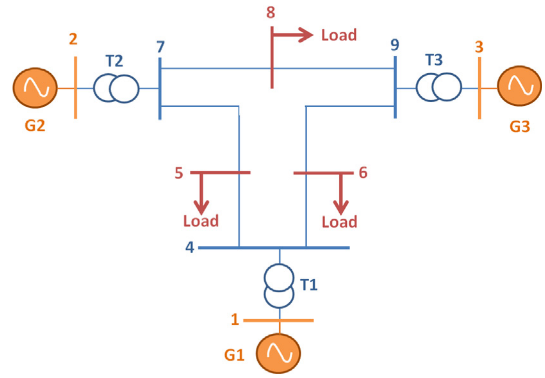


**Fig. 6.** Single Line Diagram of IEEE 9 bus system.

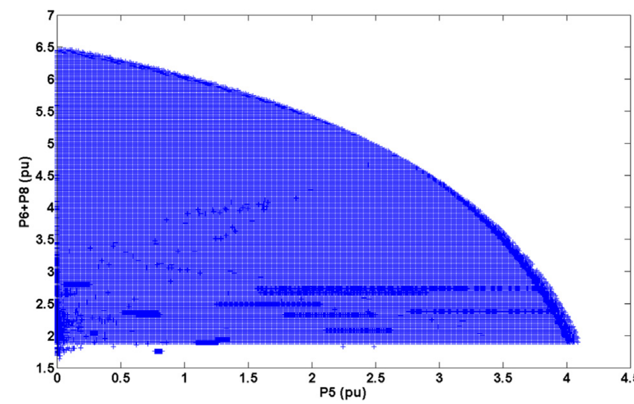


**Fig. 7.** Data of different loading conditions for base case w.r.t load 5.

The function stops when it cannot make any more splits or if further splits would not improve classification accuracy. The optimal value for minimum leaf size was chosen by calculating the prediction error using the kfoldloss function from the same toolbox. This function calculates the error by checking the prediction of the decision trees on trained on a percentage of the training set. For example, the training set is divided to two sets with 75% and 25% data. The decision trees are trained on the 75% of the data and tested on the 25% of data to calculate an error.

The imposed minimum leaf size versus the cross validated error for base case network configuration is shown in Fig. 7. from where it can be observed that the cross-validated error remains constant for the DT w.r.t. load 5 for minimum leaf size greater than 8000. This helps to give insight whether the training set has enough data points to train the DTs. It can also be observed that the DTs w.r.t. load 6 and 8 have a constant cross-validated error for minimum leaf size ranging from 3500 to 7500. Based on this cross-validated error the function fitctree sets an optimum value for minimum leaf size when training the DTs. When the leaf size is less than the selected, it stops splitting the nodes. The created tree has an average of 35 branches (size). The obtained size is the optimal size based in the cross-validation error as shown in Fig. 7.

*Testing voltage stability assessment DTs for base case network configuration:*

The created DTs for different network configurations were tested on the test set (35% of the data). The test set data was given as input to the created DTs and these outputs were verified with the classification

**Table 1**
Accuracy of the created DTs w.r.t. load 5 on test set data for the base case network configuration.

| Train/Test | Stable | Out of grid limits | Marginally stable | Unstable |
|---|---|---|---|---|
| Stable | 100% (4434/4434) | – | – | – |
| Out of grid limits | – | 99.96% (2706/2707) | 1 | – |
| Marginally stable | – | – | 99.96% (8812/8819) | – |
| Unstable | – | – | 2 | 99.97% (64/66) |

**Table 2**
Accuracy of the created DTs w.r.t. load 6 on test set data for the base case network configuration.

| Train/Test | Stable | Out of grid limits | Marginally stable | Unstable |
|---|---|---|---|---|
| Stable | 99.96% (3037/3036) | – | – | – |
| Out of grid limits | – | 99.69% (2592/2600) | 1 | – |
| Marginally stable | – | – | 99.88% (8121/812) | – |
| Unstable | – | – | 2 | 99.96% (51/53) |

**Table 3**
Accuracy of the created DTs w.r.t. load 8 on test set data for the base case network configuration.

| Train/Test | Stable | Out of grid limits | Marginally stable | Unstable |
|---|---|---|---|---|
| Stable | 100% (3240/3240) | – | – | – |
| Out of grid limits | – | 99.79% (3385/3392) | 7 | – |
| Marginally stable | – | – | 99.89% (9191/9181) | – |
| Unstable | – | – | 3 | 99.95% (60/53) |

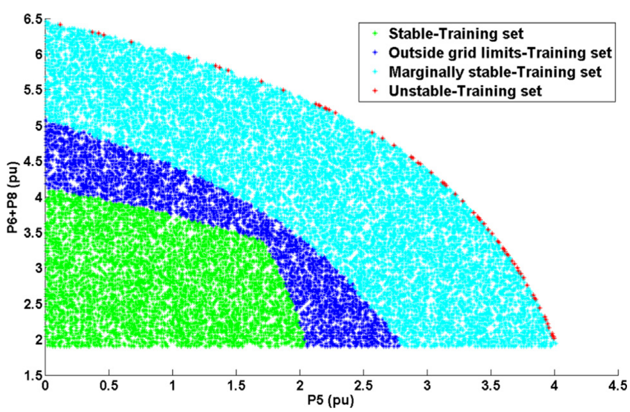


**Fig. 8.** Data of different loading conditions for base case w.r.t load 5 classified based on classification rules.

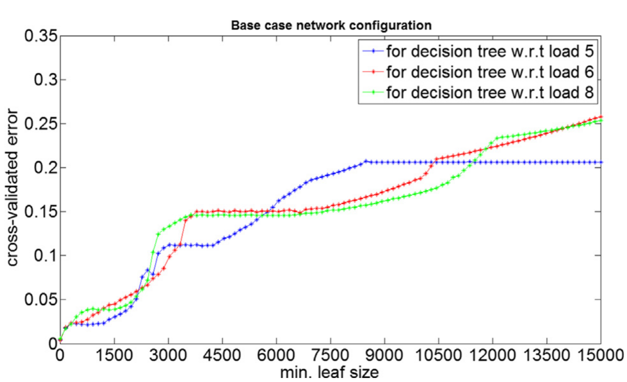


**Fig. 9.** Worst Case Approach (WCA) DT.

criteria used to create the decision trees. Based on the number of misclassifications by the created DTs, the accuracy was calculated the base case network configuration w.r.t. load 5 is provided in

Table 1. The values that are given under the percentage in the table are in the format of (estimated/actual). It can be observed from

Table 1 that two operating points in "unstable" region are misclassified as "marginally stable" region and one operating point in "outside grid limits" region is misclassified as "marginally stable". The prediction accuracy is low in "outside grid limits" and "marginally stable" region.

The classification accuracy of the decision tree for base case network configuration w.r.t. load 6 is provided in

Table 2. It can be observed from the tables that the prediction accuracy is low in "outside grid limits" region because seven operating points are misclassified as "marginally stable".

The classification accuracy of the decision tree for base case network configuration w.r.t. load 8 is provided in Table 3. It can be observed from the above table that the prediction accuracy is low in "outside grid limits" region since seven operating points are misclassified to "marginally stable" region. It can be observed from Table 1, Table 2 and Table 3 that even though there are misclassifications, each DT has an average prediction accuracy of 99.9%.

*Testing voltage stability assessment DTs for outage of line between bus 5 and 4:*

The prediction accuracy of decision trees created for network configurations involving outage of the line were evaluated on test sets and using time domain simulations results from PSS/E. The time domain simulations were run for t = 25 s and the change in network configuration was applied at t = 10 s. An example is shown in Fig. 8. and Fig. 9. In all the time domain-simulations, all loads were modeled as exponential recovery loads given by

$\dot{x}_p = \frac{-x_p}{T_p} + P_s + P_t$(7)where $x_p$ is the state variable for the load active power, $P_s$ and $P_t$ are the static and transient real power absorptions, which depend on load voltage as given in (8). $T_p$ is the active power time constant in

$P_s = P^0\left(\frac{V}{V^0}\right)^{\alpha_s}$ and $P_t = P^0\left(\frac{V}{V^0}\right)^{\alpha_t}$(8)where $\alpha_s$ and $\alpha_t$ are the static and transient active power exponents and $V^0$ is the voltage at the load bus from the load flow solution. Similar equations hold for reactive power. The response of the load after the event (overshoot/undershoot) depends on $\alpha_t$[1]. The time taken to recover the load power consumption

[1] If $\alpha_t = 0$, there will be load undershoot. For the values of $\alpha_t < 0$, there will be load overshoot and for the values of $\alpha_t > 0$, there will be load undershoot. The undershoot for $\alpha_t > 0$ is more than the undershoot for $\alpha_t = 0$.

**Table 4**
Accuracy of the created DTs for network configuration with outage of line between bus 5 and 4.

| Train/Test | Stable | Out of grid limits | Marginally stable | Unstable |
|---|---|---|---|---|
| Stable | 99.93% (1517/1516) | – | – | – |
| Out of grid limits | – | 99.98% (250/246) | – | – |
| Marginally stable | 1 | 4 | 99.92% (3922/3925) | – |
| Unstable | – | – | 2 | 96.42% (54/56) |

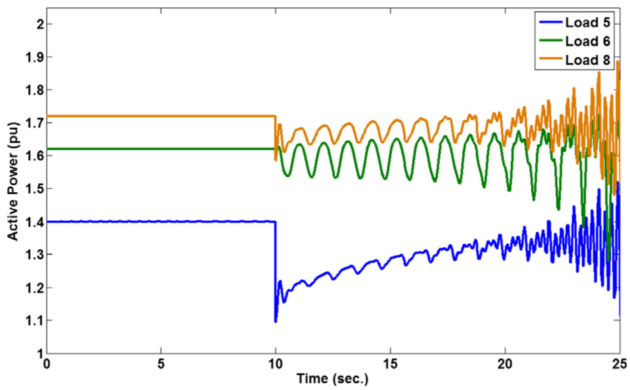


**Fig. 10.** Active power consumption at the load buses for outage of line between bus 5 and 4.

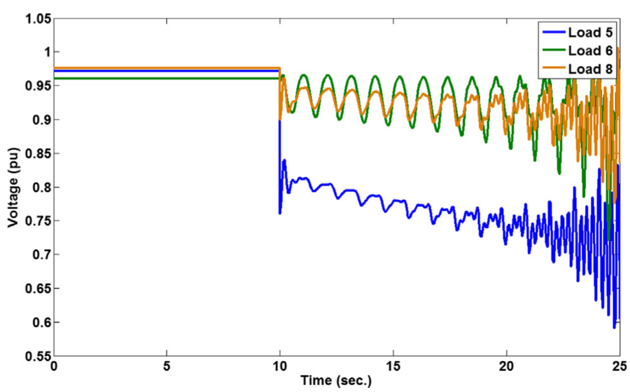


**Fig. 11.** Voltage magnitude at the load buses for outage of line between bus 5 and 4.

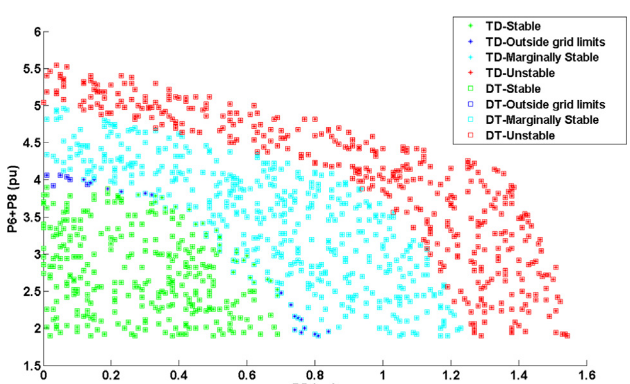


**Fig. 12.** Classification of time domain simulation outputs by classification criteria and created DTs w.r.t load 5 for outage of line between bus 5 and 4. TD indicates the use of Time Domain simulation for verification.

to the value (load power consumption) before the occurrence of the event depends active power time constant.

To illustrate why this is relevant for the prediction accuracy analysis, an operating point chosen such that the system is heavily loaded is considered below. When the outage of the line between bus 5 and 4 applied (see Fig. 8) the system becomes unstable (after the contingency) at the end of simulation. Fig. 8. also shows that after the contingency the load consumption at bus 5 is recovering to the pre-contingency load power consumption with exponential recovery. The load power consumption recovers to the pre-contingency load power consumption levels, the voltage at the load buses reduces (from (8), when $\alpha_t = 1$ then $P_t = P^0\left(\frac{V}{V^0}\right)$) as shown in Fig. 8. As the load power consumption recovers, the voltage at the load buses drop leading to oscillations and further making the system unstable. The load bus voltage magnitudes for this contingency are shown in Fig. 9. The dynamic load model used in this case study has the same properties as the load model used for training the decision trees.

The details of prediction accuracy on the test set of the created DTs w.r.t. load 5 for outage of line between bus 5 and 4 is given in Table 4. Because the decision trees were created using power flow solutions, it was tested on the time domain simulation outputs from PSS/E, with simulations like those illustrated above, for cases involving a line outage.

These decision trees were tested with the output from time domain simulations in the following steps.

1. A set of random load power values were generated.
2. These load powers were applied on the base case network configuration.
3. Time domain simulations were simulated with these load powers on the base case network configuration up to t = 10 s.
4. At the end of t = 10 s the change in network configuration was applied and the simulation was continued up to 25 s. For example, in the case discussed above, the time domain simulations were simulated for 25 s with line outage between bus 5 and 4 at end of t = 10 s.
5. Mean load consumptions values and voltages at the load buses from t = 18 s to end of simulation (t = 20 s) were considered as input to the trained DTs.
6. The classification criterion was applied on these operating points to verify them with the trained DTs classification.
7. The classification accuracy is calculated, and the outputs are plotted to visualize the misclassified operating points.
8. Classification of time domain simulation outputs using classification criteria and decision tree w.r.t. load 5 for outage of line between bus 5 and 4 are shown in Fig. 10.

It can be observed from Fig. 11. that misclassification occurs in the boundary region because of the decimal values of the load powers and voltages from time domain simulations. For example, if 0.25 is the boundary value between two regions (“outside grid limits”, “marginally stable”). It was observed that misclassification happened for the values 0.251, 0.252 … 0.254, etc., that are classified as “outside grid limits” instead of “marginally stable”.

The purpose of this verification using dynamic simulations is to determine the range of validity of our classification results, i.e. the accuracy of the trained decision trees (that were computed using a static analysis) when the system undergoes a contingency where the load recovery mechanism can lead to lower voltage values than those for which the training has been conducted. These results show that most of the training can be carried out using the type of static analysis

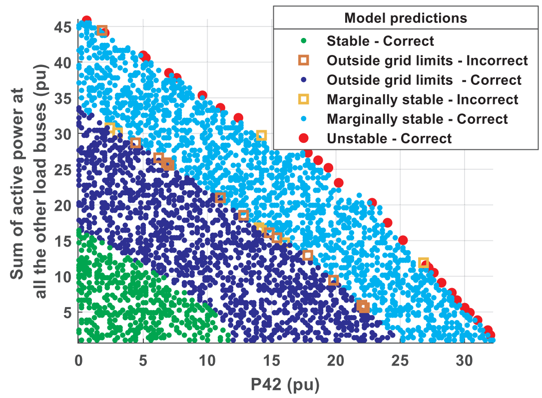


**Fig. 13.** Predicted states by the trained DTs w.r.t load bus 42.

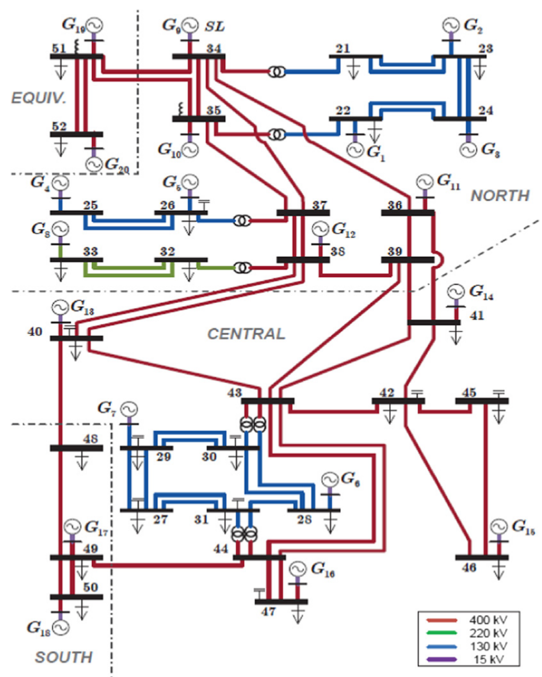


**Fig. 14.** Single Line Diagram of KTH Nordic 32 bus system [35].

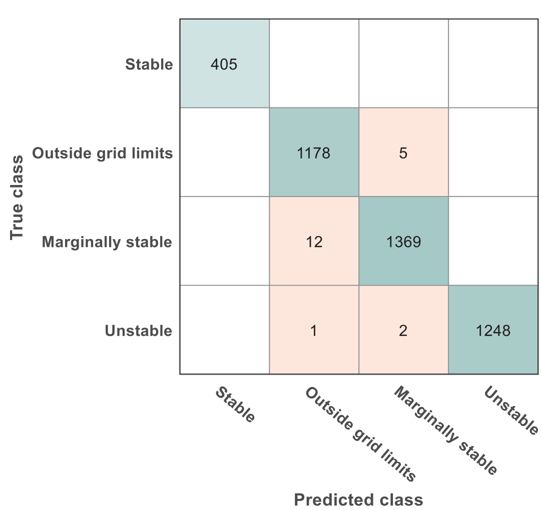


**Fig. 15.** Accuracy of the trained DTs w.r.t load bus 42.

proposed in the paper, however, at the boundary some misclassification can occur when considering the system's dynamic response. Hence, the second purpose of this exercise is to show the limitations of the proposed approach. Ultimately, to improve the prediction accuracy of the decision trees would require the use time-domain simulation for the training of the decision trees themselves, which would be computationally more expensive than the proposed approach herein. This is a subject for future research.

C. KTH Nordic 32 bus test system:

In order to check the performance of this method on a bigger system, it is tested on KTH Nordic 32 bus system as shown in Fig. 14. It can be observed that the trained DTs predicted the states of the operating points with 99% accuracy. It can be observed from Fig. 12. that the DTs incorrect prediction is confined to boundary regions. Increasing the sampling of the data at the boundary region can reduce this problem but this will be the computational intense.

The performance of the trained DTs is shown below in Fig. 13. It can be observed that the operating points in "Stable" region are predicted with 100% accuracy but the operating points in "Outside grid limits", "Marginally stable" and "Unstable" regions are predicted with bit less accuracy.

It can be observed from Fig. 15. that misclassification occurs in the boundary region because of the decimal values of the load powers and voltages. This misclassification can be reduced by increasing the sampling in the boundary region.

## 4. Conclusions

The average accuracy of classification by the created decision trees for random time domain simulations was 99.06%. It was observed that most of the misclassified operating points lie on the boundary of regions. Therefore, more operating points are required in the boundaries of the regions when training the decision trees in order to reduce the misclassification of operating points at the boundary.

The idea of this approach to use decision trees to classify the operating regions ("unstable", "outside grid limits", "marginally stable", "unstable") based on distance from the nearest SNB point has proved it to be fruitful in training the DTs, providing excellent prediction accuracy. However, further study on a larger power system model is essential to further identify any anomaly behaviors such as impact of voltage control devices like FACTS devices, tap changing transformer or any neglected conditions such as impact of intermittency of the renewable energy generation. The proposed time domain simulation-based verification can be of great value for DT accuracy verification in such cases.

## CRediT authorship contribution statement

**Luigi Vanfretti:** Conceptualization, Methodology, Writing - review & editing, Supervision, Project administration, Funding acquisition. **V.S. Narasimham Arava:** Methodology, Software, Validation, Writing - original draft, Visualization.

Declaration of Competing Interest

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

## Acknowledgement

This work was supported in part by the European Union Seventh Framework Programme (FP7/2007-2013) under Grant 283012, within Innovative Tools for Electrical System Security within Large Areas (iTESLA), in part by the New York State Energy Research and Development Authority (NYSERDA) through the Electric Power

Transmission and Distribution (EPTD) High Performing Grid Program under agreement number 137951, in part by the Engineering Research Center Program of the National Science Foundation and the Department of Energy under Award EEC-1041877 and in part by the CURENT Industry Partnership Program, and in part by the Center of Excellence for NEOM Research at King Abdullah University of Science and Technology.

## Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijepes.2020.106251.

## References

[2] IEEE CIGRE joint task force on stability terms and definitions. "Definition and classification of power system stability." International conference on large high voltage electric systems; 2003.
[4] Venkatasubramanian V, Schattler H, Zaborsky J. Dynamics of large constrained nonlinear systems-a taxonomy theory [power system stability]. Proc IEEE Nov. 1995;83(11):1530–61.
[5] Venkatasubramanian V, Schattler H, Zaborszky J. Local bifurcations and feasibility regions in differential-algebraic systems. Autom Control IEEE Trans Dec. 1995;40(12):1992–2013.
[6] Zhang YS, Chiang HD. Local bifurcation boundary and steady-state security boundary in large electric power systems: Numerical studies. Int J Bifurc Chaos Mar. 2011;21(03):647–62.
[7] Dong X, Liang J, Zhang X, Sun H. Computation of closest steady state voltage stability bifurcation using PSO approach. Innovative Smart Grid Technologies - Asia (ISGT Asia) IEEE 2012;2012:1–4.
[8] Nguyen TV, Nguyen YM, Yoon YT. A new method for static voltage stability assessment based on the local loadability boundary. Int J Emerg Elec Power Sys 2012;13:1–14.
[9] Perninge M. Approximating the loadability surface in the presence of SNB–SLL corner points. Int J Elec Power Energy Sys 2013;22:313–23.
[10] Parashar M, Agarwal A, Makarov YV, Dobson I. Real-time voltage security assessment (RTVSA) algorithms & framework. Prepared For: California Independent System Operator, prepared by: Consortium for Electric Reliability Technology Solutions, Funded By: California Public Interest Energy Research Transmission Research Program, Contract No. 500-02-004, MR-041; December 2007.
[11] Eto JH, Parashar M, Lewis NJ. Real-time system operations (RTSO) 2006-2007. Consortium for electric reliability technology solutions (CERTS). California Energy Commission, PIER Transmission Research Program; 2008.
[12] Ibsais A, Ajjarapu V. Voltage stability-limited interchange flow. Elec Power Sys Res 1996;38(2):91–5.
[13] Dobson I, Lu L. Computing an optimum direction in control space to avoid stable node bifurcation and voltage collapse in electric power systems. Autom Control IEEE Trans Oct. 1992;37(10):1616–20.
[14] Dobson I, Lu L. New methods for computing a closest saddle node bifurcation and worst-case load power margin for voltage collapse. Power Syst IEEE Trans Aug. 1993;8(3):905–13.
[15] Kwatny HG, Fischl RF, Nwankpa CO. Local bifurcation in power systems: theory, computation, and application. Proc IEEE Nov. 1995;83(11):1456–83.
[16] Makarov YV, Hill DJ, Dong Z. Computation of bifurcation boundaries for power systems: a new Delta-plane method. Circuits Syst I Fundam Theory Appl IEEE Trans Apr. 2000;47(4):536–44.
[17] Wehenkel L, Pavella M. Advances in decision trees applied to power system security assessment. IEE 2nd international conference on advances in power system control, operation and management; December 1993.
[18] Wehenkel L, Van Cutsem T, Pavella M, Jacquement Y, Heilbornn B, Pruvot P. Machine Learning, neural networks and statistical pattern recognition for voltage security: a comparative study. Eng Intell Syst December 1994;2.
[19] Van Cutsem T, Wehenkel L, Pavella M, Heilbornn B, Goubin M. Decision tree approach to voltage security assessment. IEEE Proc-C May 1993;140(3).
[20] Wehenkel Louis A. Automatic learning techniques in power systems. Springer Science & Business Media; 2012.
[21] Sun K, Likhate S, Vittal V, et al. An online dynamic security assessment scheme using phasor measurements and decision trees. IEEE Trans Power Syst 2007;22(4):1935–43.
[22] Diao R, Sun K, Vittal V, et al. Decision tree-based online voltage security assessment using PMU measurements. IEEE Trans Power Syst 2009;24(2):832–9.
[23] Diao R, Vittal V, Logic N. Design of a real-time security assessment tool for situational awareness enhancement in modern power systems. IEEE Trans Power Syst 2010;25(2):957–65.
[24] Kamwa I, Samantaray SR, Joos G. Development of rule-based classifiers for rapid stability assessment of wide-area post disturbance records. IEEE Trans Power Syst 2009;24(1):258–70.
[25] Karapidakis ES, Hatziargyriou ND. Online preventive dynamic security of isolated power systems using decision trees. IEEE Trans Power Syst 2002;17(2):297–304.
[26] Hatziargyriou HD, Contaxis GC, Sideris NC. A decision tree method for on-line steady state security assessment. IEEE Trans Power Syst 1994;9(2):1052–61.
[27] Gao Q, Rovnyak SM. Decision trees using synchronized phasor measurements for wide-area response-based control. IEEE Trans Power Syst 2011;26(2):855–61.
[29] Chengxi Liu, Zakir Hussain Rather, Zhe Chen, Claus Leth Bak. An overview of decision tree applied to power systems. Int J Smart Grid Clean Energy; 2013.
[30] Federico Milano. Power system analysis toolbox: documentation for PSAT version 2.1.8; January 6, 2013.
[34] The MathWorks, Inc., Statistics and machine learning toolbox™ user's guide; March 2016.
[35] Chompoobutrgool Y, Li W, Vanfretti L. Development and implementation of a Nordic grid model for power system small-signal and transient stability studies in a free and open source software. In: Proc IEEE PES Gen Meeting; Jul. 2012. p. 1–8.
[36] Danish MSS, Senjyu T, Danish SMS, Sabory NR, Narayanan K, Mandal P. A recap of voltage stability indices in the past three decades. Energies 2019;12(8):1544.