CIGRE-517

2022 CIGRE Canada Conference & Expo
Calgary, Alberta, Oct.31 – Nov. 3, 2022

Automated Data Labeling for Machine Learning Based Short-Term Voltage Stability Classification

K.D DHARMAPALA[1] and A.D RAJAPAKSE[1]
[1]University of Manitoba
Canada

SUMMARY

Power systems encounter more short-term voltage instability events due to changing dynamics of the loads and generation. It is desirable to have a system to predict the Short-term Voltage Stability (SVS) status of a power system after a disturbance using a short observation window in real-time. Predicting the SVS status within sufficiently short time allows opportunity for activating remedial measures. Recent approaches to develop online SVS status prediction methods based on Machine Learning (ML) techniques have shown promising results. However, most of these attempts employed ML models which need training using time-series classifier based supervised learning algorithms. Supervised learning requires a large amount of data instances and these instances should be labeled as stable/unstable by observing the time-series voltage trajectories. Manual labeling process is cumbersome and consumes a significant amount of time. The objective of this study is to explore the use of a semi-supervised learning algorithm called label propagation to generate labels for unlabeled data instances using a small portion of manually labeled data instances as a guide.

The proposed scheme is initiated by generating a database of input data (in the form of simulated voltage trajectories for a large number of contingencies) for which labels must be generated. These input data is generated by automating PSSE® dynamic simulation using a Python program. The time series voltage trajectories cannot be directly used as input to the label propagation algorithm. Therefore, several well-established SVS indices based on transient energy theory and voltage curve method are considered to use as the input features to increase the accuracy and robustness of this semi supervised learner. In order to facilitate semi-supervised learning process, a small portion of the input data are observed considering a longer time window and the corresponding labels are assigned manually. Then, the labeled and unlabeled voltage trajectories are fed to calculate a set of SVS indices. These indices are used as inputs for a graph-based label propagation algorithm, in which the labeled data instances propagate labels to non-labeled data instances based on the weightages computed via input features.

The proposed semi- supervised learning method was applied to generate labels for 700 data instances generated by simulating IEEE Nordic 32 bus system in PSSE®. The original test system contains static load models. However, dynamic loads such as induction motors are the main contributor of short-term voltage instability after faults. Therefore, some of the static loads at random locations are replaced by dynamic loads to make the simulation data more practical. Using 3 SVS indices as input features, labels could be assigned with an acceptable accuracy of 99.5%; the results show the labeling accuracy increases when the percentage of manually labeled data instances increases, and the labeling accuracy changes with the type and the number of input features used.

KEYWORDS

Short-term Voltage Stability (SVS), semi-supervised learner, label propagation, machine learning classifier, Short-term Voltage Stability Indices (SVSI)

dharmapk@myumanitoba.ca

1. INTRODUCTION

Voltage stability is the ability of a power system to restore an acceptable steady state voltage at all the buses after the system been subjected to a disturbance. Based on the timeframe, Short-term Voltage Stability (SVS) events extend from one second to several seconds. Therefore, SVS involves dynamics of loads (eg: induction motors, power electronic loads) and dynamics associated with fast acting power system equipment [1]. Lately, power systems encounter more Short-term Voltage Stability (SVS) issues than ever due to increasing demand of dynamic loads and high penetration of Inverter Based Resources (IBR) [1]-[4].

Therefore, researchers have drawn their attention to assess SVS status through different approaches. These approaches can be categorized based on control theory, voltage curve, stability boundary and Machine Learning (ML). The control theory-based assessing methods are based either transient energy functions or non-linear dynamic approaches [13]. The Lyapunov Exponent (LE) is a well-known transient energy function which is used to assess SVS. This method is used in [9] to assess SVS using real-time PMU measurements. Bifurcation theory can be used to assess the stability of non-linear dynamic systems. The possibility of applying this method in SVS analysis is discussed in [8] and [13]. Nevertheless, this method is difficult to use in real-time for large-scale power systems [11]. The most practical method of assessing the SVS is the curve analysis method proposed in [12]. Low computational burden and less execution time makes this method more favorable to real-time analysis. There are several approaches found in literature based on voltage curve analysis [5]-[7], [12]. These methods propose threshold settings based on the probable voltage trajectory at a short-term voltage instability event, but these thresholds are defined based on operator experiences, which will not accurate under all operating conditions and will be unique for a specific power system.

In recent years, researchers have attempted to improve the SVS assessment based on ML. Various ML techniques are used to assess SVS ex: Decision Tree (DT) [14]-[16], Random Forest (RF) [18], Support Vector Machine (SVM) [17] and Artificial Neural Network (ANN) in form of Extreme Learning Machine (ELM) [19] and Recurrent Neural Network (RNN) [20]. The ML techniques proposed [14]-[18] are based on analyzing time series of voltage trajectories which enable them to capture the temporal features. In contrast, the assessment methods proposed in [19] and [20] based on an index computed at different time instances which losses sequential features.

Even though ML classification models shows promising results, these models use supervised learning and therefore require a large set of training data in the form of inputs and the corresponding output. The input data is typically generated through dynamic simulations and the output, which is referred to as the label, is the classification result. The label is typically manually generated, by observing the simulated voltage trajectory, by a human expert. Label generation according to domain knowledge by an expert is a cumbersome process which inevitably consumes significant amount of engineering time, and it may cause human errors while labeling. In order to make the label generation more efficient, a semi-supervised learning algorithm can be adopted to generate labels where unlabeled data instances can be labeled using a small portion of labeled data instances as a guide. There are several approaches proposed in the literature [14] and [21]. However, these approaches have used simple assumption to create a constraint in the COP k-mean algorithm [23] which is not sufficient to generate accurate labels. Furthermore, the accuracy of these generated labels has not been validated. This paper proposes a method to automatically generate consequence labels for the data instances of a ML training database for SVS assessment. Section II presents the process of label generation by observing voltage trajectories. In Section III proposed automatic label propagation scheme is discussed and input feature computation is introduced. Section IV contains the results of the case study carried out using IEEE Nordic system. Finally, Section V presents the conclusions.

2. LABEL GENERATION FOR SVS ASSESSMENT

When considering the SVS classification, voltage trajectories can be categorized under fast voltage recovery, Fault Induced Delayed Voltage Recovery (FIDVR), sustained low voltage without recovery and voltage collapse associated with rotor angle instability. In both fast recover and FIDVR scenarios eventually all the bus voltages are recovered to an acceptable voltage level. Therefore, those scenarios are considered as stable. On the other hand, if all the bus voltages cannot recover to an acceptable operating equilibrium point, system will be considered as unstable. Therefore, scenarios such as voltage

collapse and sustained low voltage without recovery scenarios are considered as unstable [20]. Figure 1(a) illustrates a sustained lower voltage scenario and Figure 1(b) shows the voltage collapse scenario followed by a rotor angle instability. Therefore, during manual data labeling voltage trajectories with similar to fast recovery and FIDVR are labeled as stable and voltage trajectories showing sustained low voltages without recovery and voltage collapse are labeled as unstable.
It is also important to differentiate the challenges in real-time SVS assessment and data labeling. The motive of the real-time SVS assessment using ML based classifiers is to recognize the potential instabilities as early as possible using a short observation window after a disturbance. While the off-line data labeling process has the liberty to use much longer observation window to determine the SVS status after the disturbance, as shown in Figure 1 (c).

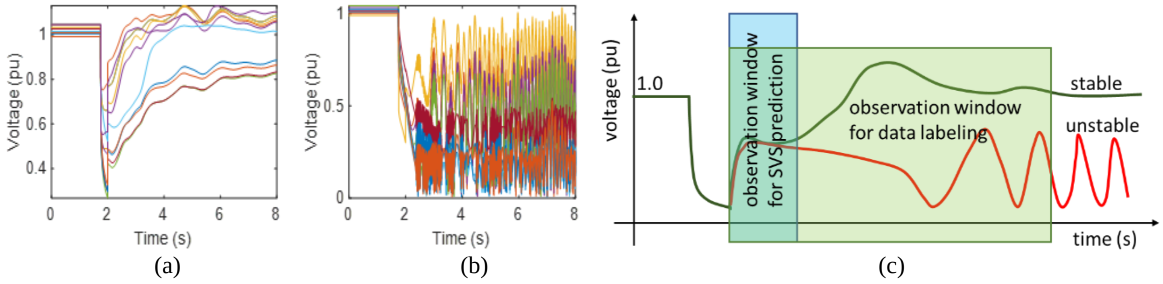


Figure 1. (a) sustained low voltage scenario (unstable), (b) voltage collapse associated with rotor angle instability (unstable) (c) Observation window for SVS prediction and data labelling

3. PROPOSED AUTOMATIC LABEL GENERATION SCHEME

The proposed scheme is initiated by generating a database of input data (in the form of simulated post-disturbance voltage trajectories for a large number of contingencies) for which labels (SVS status) must be generated. In order to facilitate semi-supervised learning process, a small portion of the input data are manually observed and the corresponding labels are assigned. Then, the labeled and unlabeled voltage trajectories are fed to calculate a set of SVS indices, which are used as inputs for a graph-based label propagation algorithm. The labeled nodes propagate labels to non-labeled nodes based on the weightages computed via input features. The proposed scheme is shown in Figure 2.

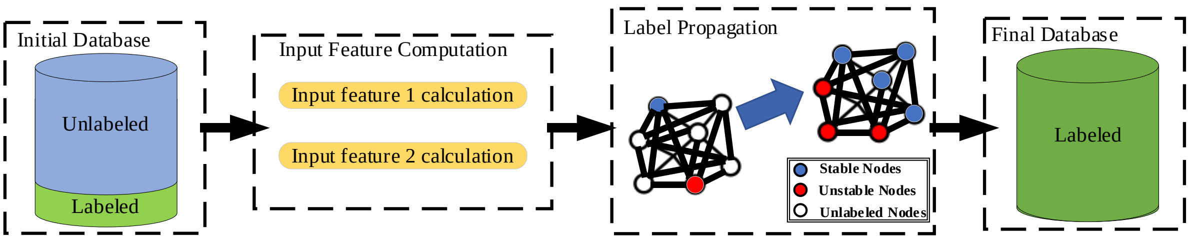


Figure 2. Proposed automatic database labeling scheme.

3.1 Voltage Trajectory Database Generation and Labeling

The power system model used for data generation must be as accurate as possible, and important features for SVS such as dynamics loads need to be included. The validity of the proposed labeling process must be proven for credible contingencies and acceptable operational conditions. Therefore, it is required to generate voltage trajectories under such conditions. Power systems are typically designed to tolerate contingencies up to a certain level, and therefore scenarios that make the post-event voltage unstable involves extreme contingencies such as faults cleared after a long delay (clearing by backup, breaker failure, etc.) or cascading faults. In this work, the voltage trajectories under different conditions were generated using automated PSSE® dynamic simulations. Afterwards a small portion of voltage trajectories are labeled manually by observing the plots of voltage trajectories. Therefore, the final dataset contains time series bus voltage magnitudes of test system for different conditions. The labeled data instances have logical variable which indicates whether the data instance is stable or unstable (-1 for unstable / 1 for stable)) and unlabeled data instances have null values as the logical variable.

3.2 Input Feature Computation

The time series voltage magnitudes of all the buses cannot be directly used as input to the label propagation algorithm because these raw data cannot generate a relationship between the observations. Therefore, three well-established SVS indices which are based on transient energy theories and voltage curve method are considered to use as the input features to increase the accuracy of this semi supervised learner. Considering the feasibility of implementing this method to a large system, SVS assessment indices which uses only voltage magnitude are considered.

3.2.1 NERC Contingency Severity Index (NERC CSI)

The NERC/WECC Planning Standard [5] has introduced these indices to measure the severity of power system contingencies. Standard proposed two indices which analyze two different parameters under two different time windows of interest as shown in Figure 3. The first index SI_V analyze the level of voltage dip after a contingency, and it is computed using (1). The second index SI_t analyze the time during which the voltage trajectory fallen below 80% of the steady state voltage level. It is computed as defined in (2) and t_f is the time at the end of observation window of 20 cycles from the fault clearing time (t_{cl}).

$$SI_v = \begin{cases} \dfrac{|V_0 - V_{dev}|}{V_0}, & \dfrac{|V_0 - V_{dev}|}{V_0} \geq \gamma \\ 0, & \dfrac{|V_0 - V_{dev}|}{V_0} < \gamma \end{cases} \tag{1}$$

$$Where\ \gamma = \begin{cases} 0.25 & : load\ buses \\ 0.3 & : other\ buses. \end{cases}$$

$$SI_t = \frac{\sum_{i=1} \tau_i}{t_f - t_{cl}} \tag{2}$$

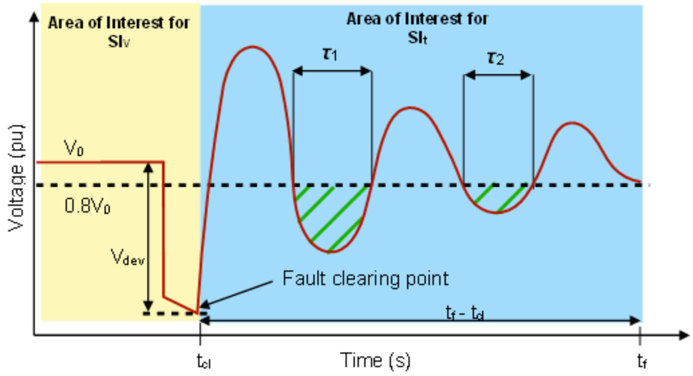


Figure 3. NERC CSI areas of interest

3.2.2 Lyapunov Exponent (LE)

The Lyapunov Exponent (LE) is adopted from ergodic theory. LE has the potential to determine the chaotic nature of the system at any moment by analyzing the rate of divergence of the relevant dynamic system variables. Power systems are dynamic systems therefore, the power system voltage can be considered as a as dynamic variable and LE can be used to determine the SVS [9]. The computation of maximum LE at a given instance can be obtained from (3).

$$\lambda(k\Delta t) = \frac{1}{Nk\Delta t} \times \sum_{m=1}^{N} log \frac{\left\| V_{(k+m)\Delta t} - V_{(k+m-1)\Delta t} \right\|}{\left\| V_{m\Delta t} - V_{(m-1)\Delta t} \right\|}, k > N \tag{3}$$

$$Where\ integer\ N\ is\ chosen\ such\ that \quad \epsilon_1 < \left\| V_{m\Delta t} - V_{(m-1)\Delta t} \right\| < \epsilon_2\ for\ m = 1,2,\dots,N.$$

The λ represents the maximum LE at k^{th} instance of the V data series where Δt is the sampling period. N represent the number of data points of the considered initial window. If the maximum LE of the system is negative (positive) the voltage trajectory tends to converge (diverge). The mean of the series of maximum LE values generated for each data instance is considered as the input feature of label propagation algorithm.

3.3 Label Propagation Algorithm

After computing SVS indices. Theses indices are fed to label propagation algorithm. Generally, for a semi-supervised learner such as label propagation, a data set X={(x_1^1.. x_N^1),..,(x_1^n.. x_N^n), .. ,(x_1^M.. x_N^M)} which have N data instances and M number of features have two subsets X_L and X_U which represents labeled and unlabeled data respectively. Label Propagation is a graph based semi supervised learning algorithm [10]. Therefore, the first step is to generate a fully connected similarity graph (G). This graph represents the local relationship of labeled and unlabeled data instances. Therefore, the nodes or the edges (E) of the similarity graph represent the data instances and vertices (V) represents the relationship between data instances. The weightage of edges which connect *i* and *j* is computed using (4) for a kernel scale value of σ:

$$S_{ij} = \exp\left(-\frac{\sum_{n=1}^{M}\left(x_i^n - x_j^n\right)^2}{\sigma^2}\right). \tag{4}$$

The larger edge weights easily allow labels to propagate to none labeled nodes. In order to generalize the weights throughout the graph, the probability transition matrix P defined in (5) is computed.

$$P_{ij} = P(j \to i) = \frac{S_{ij}}{\sum_{k=1}^{L+U} S_{kj}} \tag{5}$$

where P is the probability of propagating from node *j* to *i*. Initially, all nodes have soft labels ($Y^{(0)}$). Taking the product of these probability transition matrix and soft labels, the new labels are obtained. In label propagation the labeled outputs should be reinitiated to the original values. This process is iterated until the all the node labels converges. The label propagation algorithm is outlined in Figure 4 .

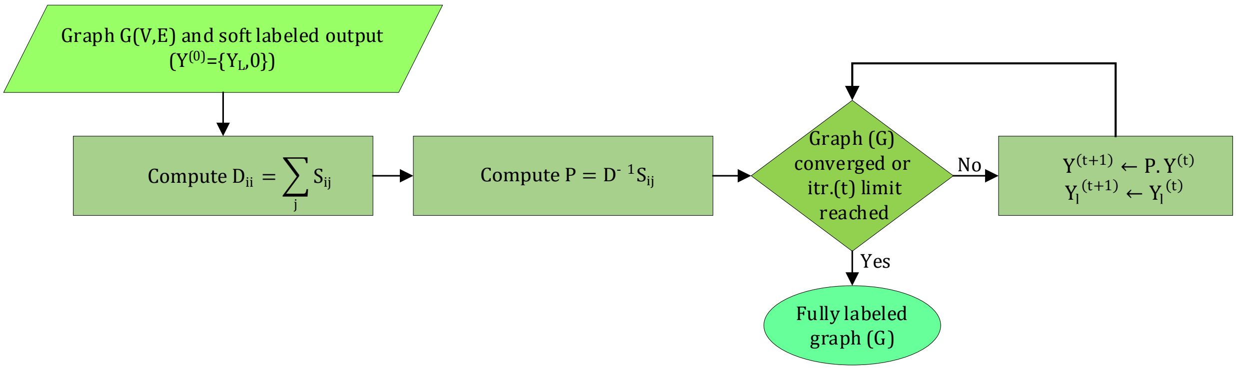


Figure 4. Label propagation algorithm

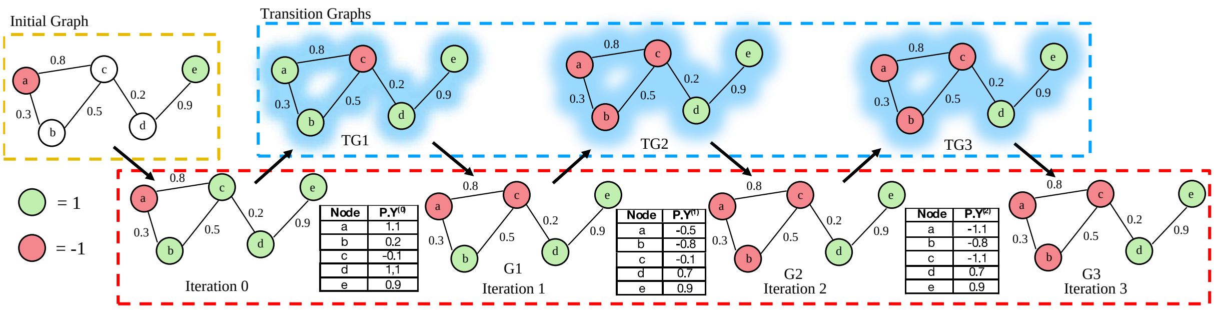


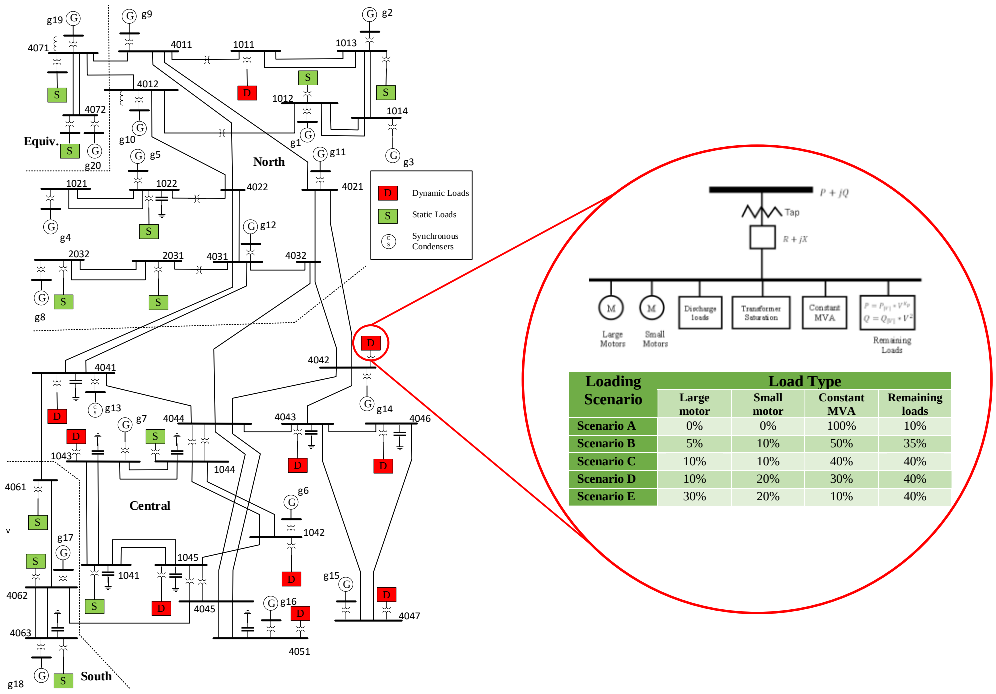
Figure 5. An example case to illustrate the process of label propagation

The process of label propagation is further explained using the example shown in Figure 5. Assume an initial graph with nodes a, b, c, d and e. Node a and e are labeled with class “red” and “green” which are numerically represented by “-1” and “1” respectively and other nodes are unlabeled. The respective P values are denoted on each vertex. P values of nodes which are not connected are assumed to be negligible. At iteration 0, unlabeled nodes are soft labeled with class “green”. The product of P matrix and the node label vector for the first iteration is shown in the table adjacent to the graph. The results of the product are approximated to closest label value, −1 or 1. For the first iteration, node c is a negative value while others are positive values, and therefore, node c is labeled as class “red” while others are labeled as “green” as shown in transition graph TG1. However, the originally labeled nodes are reinitiated to the original class and graph G1 is obtained after the first iteration. This process is continued until consecutive graphs contain same node labels (as in the example) or the algorithm reaches the iteration limit. The final graph is considered the fully labeled graph.

4. CASE STUDY

4.1 Test System

The proposed scheme is tested on IEEE 32 bus Nordic test system [22]. The original system contains static load models. However, dynamic loads such as induction motors are the main contributor of short-term voltage instability after faults [20]. Therefore, some of the static loads at random locations were replaced by dynamic loads. Complex Load Model or CLOD is a widely used dynamic load model. The architecture of CLOD is denoted in Figure 6. In this study, static loads were converted to CLOD model at locations marked in Figure 6.



Loading Scenario	Load Type			
	Large motor	Small motor	Constant MVA	Remaining loads
Scenario A	0%	0%	100%	10%
Scenario B	5%	10%	50%	35%
Scenario C	10%	10%	40%	40%
Scenario D	10%	20%	30%	40%
Scenario E	30%	20%	10%	40%

Figure 6. Test system with CLOD load replacements and loading scenarios

4.2 Training Database Generation

In order to generate voltage magnitudes under different dynamic loading levels and contingencies, automated PSSE® dynamic simulation program is utilized. Large and small motors of CLOD models were loaded under different loading scenarios shown in Figure 6. In all simulation cases a value of 0.01 pu is used for the branch resistance (R) and reactance (X) of CLOD model and the Kp value of the remaining loads is defined as 1. Voltage trajectories under different contingences such as three phase temporary line faults which cleared after 5 cycles,10 cycles, 15 cycles ,20 cycles and 24 cycles were generated. Additionally, three phase faults which cleared after 5 cycles were simulated at each generator and load bus. Under different contingencies and loading conditions 700 cases were generated. Then, the bus voltage magnitudes of each case were used to compute the considered indices. The IEEE Nordic test system contains 32 buses and since 3 SVS indices were considered, the total of 96 features were generated as inputs. Finally, a database of size 700×96 is generated to train the semi-supervised learner.

4.3 Proposed Method to Validate Semi-supervised Learner

The accuracy of this semi-supervised learner is validated using the generated database. All the 700 data instances are labeled manually for the purpose of validation. Randomly selected data instances were fed to the label propagation algorithm where the labels of the other data instances are deleted. Then the labels generated through the label propagation is cross validated with true labels and the accuracy is calculated as the percentage of correctly labeled data instances to unlabeled data instances. Training Accuracy (TA) is defined as in (6).

$$TA = \frac{Number\ of\ correctly\ labled\ data\ instances}{Number\ of\ unlabeled\ data\ instances} \times 100\% \tag{6}$$

4.4 Impact of Different Input Features and Portion of Initial Labeled Data

The effect of different indices towards accurate automatic labeling of database is analyzed using 10%, 20% and 30% of labeled data. Percentages higher than 30% are not considered since the objective of this study is to obtain an accuracy using a minimum amount of sampled data. The accuracy levels under these conditions are tabulated in TABLE 1.

TABLE 1. Case study results

SVS Indices used as input features	**TA (%) under different percentages of labeled data**		
	10%	***20%***	***30%***
NERC CSI (SI_v & SI_t)	80.6	82.5	82.8
Lyapunov Exponent (λ)	94.4	98.5	98.5
All indices (SI_v , SI_t & λ)	98.2	99.5	99.5

The impact of different sets of SVS indices can be observed in the results. The Lyapunov Exponent makes the highest accuracy of labeled data from an induvial index. When all the indices are used the automatic labeling process becomes more accurate. It can be seen that with 20% of manually labeled data, 99.5% accuracy can be obtained with the proposed label propagation method, when both indices are used as inputs.

5. CONCLUSIONS

The proposed label propagation method was applied to generate labels for 700 data instances generated by simulating IEEE Nordic power system using automated PSSE®. The results show the labeling accuracy changes with the percentage of manually labeled data instances provided as initial guidance and with the type of input features used. When the percentage of manually labeled data instances increased the accuracy of labeling process has increased. Furthermore, when all the indices are considered as input features the level of accuracy increases since the contribution of each index will be ensembled. For the considered problem, using three SVS indices (SI_v , SI_t & λ) as input features, labels are assigned with an accuracy of 99.5 % when only 20% of the data are manually labeled. This proposed method provide fully labeled database for the purpose of training machine learning based SVS applications. This semi-supervised learning process can be used for other applications, for example to to screen the results of automated contingency simulations done for large networks which are currently done manually.

BIBLIOGRAPHY

[1] P. Kundur, J. Paserba, M. Electric, P. Products, and N. D. Hatziargyriou, "Definition and Classification of Power System Stability IEEE/CIGRE Joint Task Force on Stability Terms and Definitions," IEEE Trans. Power Syst., vol. 19, no. 3, pp. 1387–1401, 2004.

[2] J. A. Diaz de Leon and C. W. Taylor, "Understanding and solving short-term voltage stability problems," IEEE Power Engineering Society Summer Meeting, 2002, pp. 745-752 vol.2, PESS.2002

[3] C. Dwivedi, "Literature Survey on Short-Term Voltage Stability Effect, Cause and Control," 2018 IEEE Green Technologies Conference (GreenTech), 2018, pp. 15-20.
[4] A. Alshareef, R. Shah, N. Mithulananthan and S. Alzahrani, "A New Global Index for Short Term Voltage Stability Assessment," in IEEE Access, vol. 9, pp. 36114-36124, 2021.
[5] NERC/WECC. Planning Standards. 2003. Available online: https://www.scribd.com/document/81304318/WECC-NERC-Planning-Standards
[6] A. Tiwari and V. Ajjarapu, "Optimal Allocation of Dynamic VAR Support Using Mixed Integer Dynamic Optimization," in IEEE Transactions on Power Systems, vol. 26, no. 1, pp. 305-314, Feb. 2011, TPWRS.2010.
[7] H. Ge et al., "An Improved Real-Time Short-Term Voltage Stability Monitoring Method Based on Phase Rectification," in IEEE Transactions on Power Systems, vol. 33, no. 1, pp. 1068-1070, Jan. 2018, TPWRS.2017.
[8] E. A. Tapia, J. D. Pinzón and D. G. Colomé, "Load Dynamic Impact on Short-Term Voltage Stability," 2019 FISE-IEEE/CIGRE Conference - Living the energy Transition (FISE/CIGRE), 2019, pp. 1-6.
[9] Dasgupta, S.; Paramasivam, M.; Vaidya, U.; Ajjarapu, V. Real-time monitoring of short-term voltage stability using PMU data. IEEE Trans. Power Syst.2013,28, pp. 3702–371.
[10] Xiaojin Zhu and Zoubin Ghahramani. 2002. Learning from labeled and unlabeled data with label propagation.
[11] W. Zhao et al., "Real-time analysis of transient voltage security based on off-line database and data fitting,"IEEE Innovative Smart Grid Technologies - Asia (ISGT-Asia), 2016, pp. 670-674.
[12] Zhao, W., Guo, Q., Sun, H., Ge, H. and Li, H. (2018), Practical short-term voltage stability index based on voltage curves: definition, verification and case studies. IET Gener. Transm. Distrib., 12: pp. 4292-4300.
[13] M. You-jie, L. Xiao-shuang, Z. Xue-song and L. Ji, "The Comments on Dynamic Bifurcation of Voltage Stability in Power System," 2010 WASE International Conference on Information Engineering, 2010, pp. 272-275.
[14] L. Zhu, C. Lu, and Y. Sun, “Time series shapelet classification based online short-term voltage stability assessment,” IEEE Trans. Power Syst., vol. 31, no. 2, pp. 1430–1439, Mar. 2016
[15] L. Zhu, C. Lu, Z. Y. Dong, and C. Hong, “Imbalance learning machine-based power system short-term voltage stability assessment” IEEE Trans.Ind. Informat., vol. 13, no. 5, pp. 2533–2543, Oct. 2017.
[16] L. Zhu, C. Lu, Y. Liu, W. Wu, and C. Hong, “Wordbook-based light-dutytime series learning machine for short-term voltage stability assessment,” IET Gener. Transmiss. Distrib., vol. 11, no. 18, pp. 4492–4499, 2017.
[17] H. Yang, W. Zhang, J. Chen, and L. Wang, “PMU-based voltage stability prediction using least square support vector machine with online learning”. Elect. Power Syst. Res., vol. 160, pp. 234–242, 2018.
[18] J. D. Pinzón and D. G. Colomé, “Real-time multi-state classification of short-term voltage stability based on multivariate time series machine learning”. Int. J. Elect. Power Energy Syst., vol. 108, pp. 402–414, 2019.
[19] Y. Zhang, Y. Xu, Z. Dong, and R. Zhang, “A hierarchical self-adaptive data-analytics method for power system short-term voltage stability assessment”. IEEE Trans. Ind. Informat., vol. 15, no. 1, pp. 74–84, Jan. 2019.
[20] Y. Zhang, Y. Xu, R. Zhang, and Z. Y. Dong, “A missing data tolerant method for data-driven short-term voltage stability assessment of power systems,” IEEE Trans. Smart Grid, vol. 10, no. 5, pp. 5663–5674, Sep. 2019.
[21] M. Zhang, J. Li, Y. Li and R. Xu, "Deep Learning for Short-Term Voltage Stability Assessment of Power Systems," in IEEE Access, vol. 9, pp. 29711-29718, 2021.
[22] Test Systems for Voltage Stability Analysis and Security Assessment. Accessed: Jun. 2020. [Online]. Available:https://resourcecenter.ieee-pes.org/publications/technicalreports/PESTR19.html
[23] K. Wagstaff, S. Rogers, and S. Schroedl, “Constrained-means clustering with background knowledge,” in Proc. 8th Int. Conf. Machine Learning, 2001, pp. 577–584.