Renewable and Sustainable Energy Reviews 56 (2016) 283–290

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

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

Static Security classification and Evaluation classifier design in electric power grid with presence of PV power plants using C-4.5

I.S. Saeh *, M.W. Mustafa, Y.S. Mohammed, M. Almaktar

Department of Electrical Power Engineering, Faculty of Electrical Engineering, Universiti Technolgi Malaysia, PO Box 81310, Johor Bahru, Malaysia

ARTICLE INFO

Article history:
Received 29 May 2014
Received in revised form
16 November 2015
Accepted 22 November 2015
Available online 10 December 2015

Keywords:
Photovoltaic
C4.5 Classifier design
Static Security Classification and Evaluation

ABSTRACT

Energy suppliers all over the world must expand energy in a way that is secure, clean, affordable, and environmentally responsible. Photovoltaic (PV) has been a competitive renewable-energy source for the power generation mix in the world. With the presence of solar PV technology, this paper proposes C4.5 approach for static security evaluation and classification (SSE). This paper proposes PV generators connected to the grid when bilateral energy transactions with the loads are implemented to see their impacts on the system security. To build a classifier in binary class, the process is divided into four components: data collection, pre-processing and feature selection, comparison of the techniques, best classifier selection and performance evaluation. A comprehensive comparison of four of Decision Tree's Algorithms for SSE is conducted. The study is (accomplished using) conducted on IEEE 30 bus system, which comprises 5 PV power generators deliver a total power of 40 MW. Data are generated on (30, 57, 118 and 300) bus IEEE test systems used to train and test the classifiers. Empirically, with the presence of PV power generators, the implementation results indicate that these classifiers have the capability for system security evaluation and classification. Lastly, C4.5 is an efficient and effective approach for real-time evaluation and classification classifier design.

Contents

* Corresponding author. Tel.: +60 129723949.
E-mail address: saiibrahem2@live.utm.my (I.S. Saeh).

http://dx.doi.org/10.1016/j.rser.2015.11.054
1364-0321/

1. Introduction

Determination of the power system current state, Energy Management System (EMS) and on-line performs the task of security evaluation (SE) and assessment. Through simulation, SSE assists operators to detect following a given list of contingencies such as a voltage out-of-limit or potential a system branch overloaded. Due to the large system size and deregulated power system, a steady-state security analysis becomes an impossible task due to the associated computation burden. In static security evaluation, the contingencies' severity is judged on scale performance index (PI) basis. In [1–3], numerous PI based methods have been reported. Artificial intelligence (AI) can be divided into two types of techniques, clustering techniques and classification techniques, and its power of reducing the data complexity made it to use in various areas like medical and engineering [4,5].

Back-propagation artificial neural networks [6], self-organizing map neural networks [7] and others have been suggested for evaluation of power system security. Optimal Artificial Neural Network (ANN) architecture determination is the crucial ANN problem, which is normally trial and error basis. Adaptive Neuro-Fuzzy Inference System (ANFIS) as well, has been applied for evaluating static security [8,9]. Recently, support vector machines (SVM), based on statistical learning theory have been used in the different areas of machine learning [10]. Currently, pattern recognition (PR) is gaining more reputation in numerous problems in electric power system. In this approach, to produce satisfactory training dataset, the key simulation bulk is done off-line [11]. Sometimes, Decision Tree (DT) is combined to other methods [12]. Traditional techniques to solve this problem would involve performing full AC load-flow for each contingency event followed by operating limit violations which has been reported in [13–15]. Due to computational requirements, it is hard to use conventional approaches through real time.

In general, a power system security assessment is an analysis to determine the extent a power system is realistically safe from severe interference to its operation [16]. A wide-range overview of literature work in the coverage of the subject of SSE can be found in reference [17] which exploits computational framework combining static and transient power system security evaluation using uncertainties. The paper concluded that the method is an effective way to introduce the influence of probabilistic uncertainties in power system security studies. In addition, the method establishes its capability to differentiate several types of uncertainties that pervade power systems and that are relevant to system security. Furthermore, a study on online static security assessment module using ANN [18] revealed that ANN models take loading condition and the probable contingencies as the input and assess the system security by screening the credible contingencies and ranking them in the order of severity based on composite security index. Another important literature supporting SSE is based on integrated toll for static and dynamic security assessment of large power systems [19]. Fan et al. [20] studied power system security with reference to photovoltaic energy source using AI technique. A growing need for renewable energy consumption has necessitated the need for more studies on security assessment of power plants utilizing renewable energy source. Therefore, with preference to solar PV technology, this study proposes C4.5 approach for evaluation and classification of static security evaluation (SSE). In addition, the paper also proposes PV generators connected to the grid when bilateral energy transactions with loads are implemented to see their impacts on the system security via a comprehensive comparison of four Decision Tree's Algorithms for SSE.

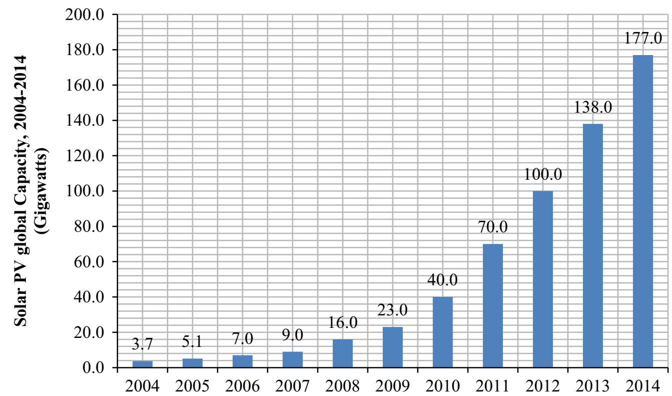

Fig. 1. Solar PV global capacity installation in Giga-watts [22].

2. Overview of photovoltaic status worldwide

Demand on electrical energy has been in the gradual increase due to the industrialization evolution and population increase. The population throughout the world doubled from 3.2 billion in 1962 to 6.4 billion in 2005 and is forecasted to grow to 9.2 billion in 2050 [21]. Department of Energy in the USA predicts that energy consumption all around the world will rise 71%, this from 2007 levels by 2030. Meanwhile available statistics revealed that global photovoltaic energy consumption is highly increasing as shown in Fig. 1. The trend is that within the span of time observed, there is continuous rise in the installed capacity of solar power generation on a global scale. However, it is most likely that the global market potential for solar power exploitation might increase beyond the 71% already predicted by the Department of Energy in the USA. This is obvious going by the analysis of the data presented in Fig. 1 showing that from 2004 to 2014 an increase of approximately 98% is recorded.

Oil prices were below US $20 per barrel prior to 2000 to nearly US $75 per barrel by the third quarter of 2006 to reach up as much as $147 by mid-2008 [23]. At the end of 2010 the prices increased from about $82 per barrel to more than $112 per barrel in 2011 and would continue in its high price for the next three decades to reach up to $125 per barrel by the year 2035 [24,25]. Hence, the liquid fuel skyrocketing price besides its depletion over time, though currently they cover almost two third of electricity demand, has led to developed and developing countries making efforts at energy sources' diversification. Adding to that the global warming emissions resulting from energy production are a serious global environmental problem. The largest contributing source of greenhouse gas is the burning of fossil fuels leading to the emission of carbon dioxide. The proof comes from direct measurements of temperatures rising of air surface and temperatures of the ocean subsurface and also, from rise in average sea levels, retreating glaciers, and some changes to several physical and biological schemes [26]. For the above reasons, the world needs to expand energy supplies in a way that is secure, clean, affordable, and environmentally responsible.

One of the promising renewable-energy sources is solar energy. Besides being free of cost, the sun radiates about 3.9 as a black body due to its high surface temperature with total energy delivered to earth ~1018 (8000x global energy consumption) [27]. By 2050, alone, solar energy is expected to source 30% of the world's energy demand and to about 64% of the electricity source in 2100, Fig. 2 [28].

Photovoltaic (PV) module is the sole mean by which the solar energy is converted into electricity. By year-end 2013, Germany had installed 35.7 GW of solar photovoltaic capacity thereby leading as the country with the largest potential capacity in solar

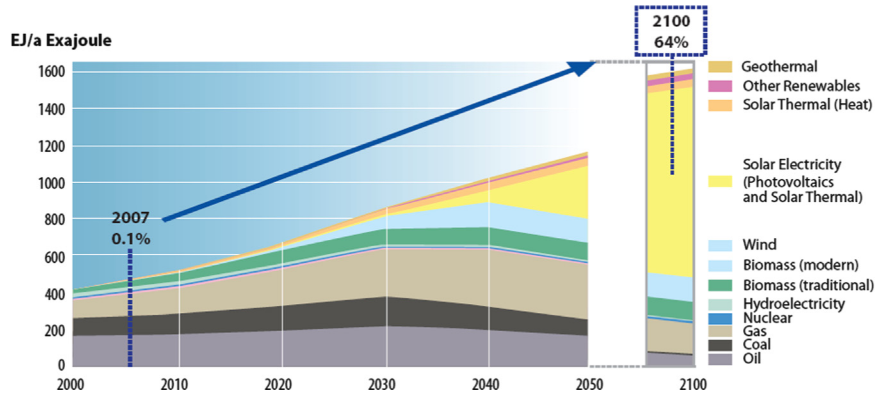

Fig. 2. Global RE technologies development until 2100.

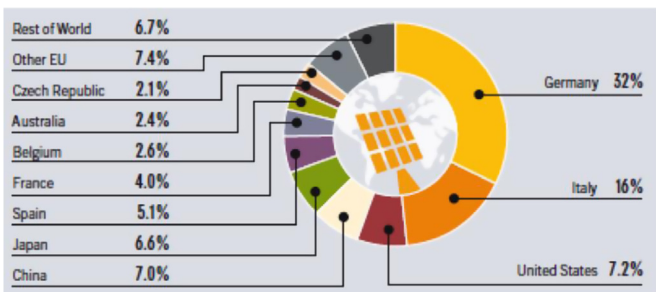

Fig. 3. Global solar PV energy based on major key market players [29].

PV power generation Fig. 3. The growth of solar PV in Germany is fundamentally changing the country's supporting systems for renewable energy development.

As presented in Fig. 1, the year 2014 marked an unprecedented demonstration of a solar PV power investment with a substantial global installed capacity of 177 GW compared to 100 GW and 138 GW for the years 2012 and 2013 respectively. The US PV market is set to witness the explosive growth over the next few years; PV installations reached 2100 MWp by 2010 under the various federal and state programs for PV with a forecast of 5 GW installed during 2013. The US industry roadmap for PV project's installations is to reach 36 GWp by 2020 and 200 GWp by 2030 [16]. There is 17,294 MW of solar PV capacity now on-line in Germany, the world leader in using PV. According to the 2010 BP Statistical Energy Survey, Germany's cumulative installed solar-energy capacity was 9677 MW in 2009, 42.2% of the world total and a change of 64.6% compared to 2008 [30]. In 2010 alone 7400 MW of Solar PV had been installed to make up along with the wind energy about 30% of total German electricity generation.

Despite its high capital cost, PV is extremely modular, easy and fast to install and accessible to the general public. Moreover, the PV system is static, quiet, and free of moving parts, and these make it have little operation and maintenance costs. Grid-connected PV system (GCPV), on the other hand, can play a vital role in lowering electricity demand and shifting peak load.

3. Static security evaluation indices selection

Power system networks are required to operate with security limits. Security is defined as promising the continuous operation of a power system capability under normal operation even next to some important contingency [16]. In the literature, several keys have been suggested as standards for Static Security Classification and Evaluation [2,31–34] which include lines overloaded or/and bus voltages collapse which let the system deviate from normal operating state limits. However, violations are not in the same level of the same significance.

Static security limits must be satisfied and can be defined as follows:

$$\sum_{i=1}^{N_g} P_{Gi} = P_D + P_{loss} \quad P_{Gi}^{min} \le P_{Gi} \le P_{Gi}^{max}, \qquad i = 1, 2, \ldots N_g \tag{1}$$

$$\left|V_i^{min}\right| \le |V_i| \le \left|V_i^{max}\right| k = 1, 2, \ldots, N_i$$

$$S_{km} \le S_{km}^{max} \tag{2}$$

where at bus i, P_{Gi} denotes real power generation and P_D is the total system load; P_{loss} is the transmission real power loss; and $|V_i|$ is the voltage magnitude.

S_{km} denotes the MVA flow in branch $k-m; N_g$ is the number of generators and N_b is the number of buses.

In the assessment process of static security, it is evaluated for several feasible contingencies via solving power flow nonlinear equations. These contingencies possibly will contain outage of a generating unit or $N-1$ transmission line or a transformer.

For numerous disturbances, the load flow is simulated and the security limitations are gauged. The operating state of power system is categorized as Static Secure (SS-Binary 1) if two of the limitations in Eqs. (1) and (2) are fulfilled. In case one limitation is identified subsequent to a contingency, the state of the system is categorized as Static Insecure (SI-Binary 0).

4. Data mining techniques

Generally, most of the data mining approaches assess information through the data-base. Nowadays, database becomes larger in size, and as result, it is very difficult to interpret complex data. Therefore, it is compulsory to develop efficient methods to deal about the complexity of data [34].

The traditional element accounts for coaching the device understanding methods for classification of static security evaluation contents. Fig. 4 presents the methodology for static security evaluation content classification approach based upon the data mining techniques. The methodology is attained through four phases: data set collection, data set preprocessing, training phase, and classifier evaluation with testing data. Consequently, the static security evaluation can be managed based on the trained machine learning classifier.

4.1. Overview of Decision Tree

The DT is a structured upside down tree and built upon a knowledge base (KB) basis which is consisting of a huge number of operating points (OPs); these OPs are covering all likely states of

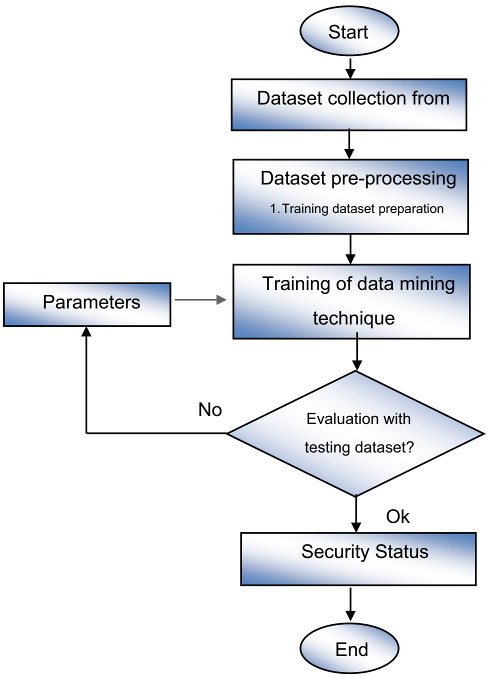

Fig. 4. Data mining procedure for static security evaluation and classification.

the power system study for its representatives' confirmation [35–38]. The KB is defined as [37], these features are the pre-disturbance steady-state characterization and variables of each OPs. The KB is divided into a learning set and test set. LS is used for deriving the structures of the classifier, while TS is used for the structure's performance evaluation on new, unobserved OPs. The DT's construction starts at the root node with the total LS of pre-classified OPs. At each stage, a tip-node of the rising tree is measured, and the system chooses whether it will be a terminal node or should be developed further. To develop a node, a proper feature is identified first, together with a contrast test on its values and the certain test is implemented to the LS of the node. Every node (subset) is considered by its security index (SI), defined as the ratio of secure OPs belonging to this node [38].

As mentioned earlier, DT's are assessed by using the T.S. The vital performance assessor of the DT is the correct classifications rate (CCR) which can clear as the proportion of correct classified OPs divided by the total number of OPs tested. The DT results and the nodes number depend upon the accurateness given from the user. Firstly, parameters of high accuracy are given to attain a large and perfect tree. Subsequently, the tree size is progressively condensed to get a tree with extra practical rules. [39] provides a systematic behavior of the DT approach.

4.2. C4.5 classifier training

C4.5 Decision Tree is one of the most broadly used and real-world approaches. In C4.5, the learned classifier is represented by a DT as sets of if-then rules to human readability improvement. Therefore, the Decision Tree is simple to be understood and interpreted. Besides, it can handle nominal and categorical data and perform well with large data set in a short time [40].

In C4.5 training, the Decision Tree is built in a top-down recursive way. Learning works of C4.5 are as follows: Primarily, all training patterns are fixed at root. These patterns are divided based on features selected based on an impurity function in recursive routine. Dividing continues till all training patterns for a certain node belong to the similar class. The parameters and their settings values were used in WEKA as shown in Table 1.

Table 1
Parameters settings for C4.5 training.

Parameter	Description	Value
ConfidenceFactor	The confidence factor used for pruning (smaller values incur more pruning).	0.25
minNumObj	The minimum number of instances per leaf.	2
Unpruned	Whether pruning is performed or not	False

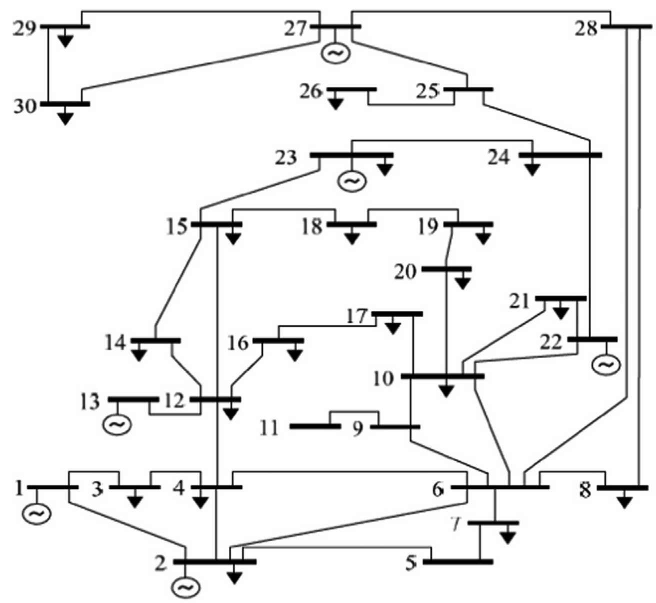

Fig. 5. Single-line diagram of IEEE 30 bus system.

5. Raw dataset collection from power system network

A bus is a node at which one or multiple lines, loads or generators are connected. Each must in a power system is associated with some electrical quantities voltage magnitude, active power, reactive power and phase angle of voltage. Two out of these four quantities are usually given while the remaining two are to be determined by means of the solution of the equations. Fig. 5 illustrates a single-line diagram of IEEE 30 bus system. Newton–Raphson Load Flow (NRLF) analysis is one of the approaches for solving problems involving load flow in buses. NRLF analysis is a load flow approach with suitable characteristics to handle most power system analysis due to its numerous advantages. Apart from the fact that the method is used for the transformation of original non-linear power flow problems to a set of linear problems, it also exhibits characteristic powerful convergence behaviors and considerable low computing times. NRLF analysis allows greater generality and flexibility, thereby enabling a wide-range of representational requirements to be accommodated without difficulty and efficiently in the development and computation of the static security evaluation.

NRLF analysis is used before implementation of Decision Tree to solve algebraic equation which is nonlinear to the system used, and collected data of all line flow and voltages of all buses. These data collected will use as input vector for training and testing the algorithms. Thus, test dataset; which is dissimilar cases from the training dataset should keep getting an acceptable accuracy results. NRLF were developed via matpower 3.0b4 program [41] and used through this study as a matrix form. In this program, the results can be shown by using the command runpf ('case Z'), where Z is the buses number. The list of attributes (features) used for the pattern vector for static security evaluation is as follows below.

$$X_{SSA} = \{|V|_i, \theta_i, S_{Gi}, S_{Li}, S_{ij}\}$$

The contingencies can include interruption on a transformer or the transmission line or maybe a generator. Performing load flow will check all the bus voltages and line thermal power limits; (1) voltage at all buses must be within their range (0.94–1.06) p.u. [13,42], and (2) all lines are not exceeding their power range as well ($S < S_{max.}$).

5.1. Dataset pre-processing

In order to attain more correct results, the datasets collected as the pattern's vectors X_{SSE} must be prepared properly. Data set pre-processing involves manipulating the data set into an appropriate training form of the Decision Tree techniques. Information pre-processing demands 2 ways: search for planning and also coaching data set planning. On the other hand, training data set preparation step requires extracting the desired information and then selecting the input/output data set or training patterns. Consequently, the actual data set will be appropriately stabilized or discretized into a specific range for better performance.

5.2. Feature selection

After we initialize a pattern vector (X_{SSE}) from data collection and data pre-processing, we initialize feature vector (Z_{SSE}) from cross validation and number of instances. Data samples generated are randomly split into training and testing process in approximately proportions of 75% and 25% respectively. Cross validation is a method applied to a data set to estimate the error. It has become quite popular because of its simplicity. When we fit a data set, we do so by minimizing some sort of loss function; most often, we will use the mean squared error for simplicity. We use the same data that was used to fit the model to test the model on a new data set to provide a better estimation of the out of sample prediction error. In this case, we turn to k-fold cross validation. In k-fold cross validation, the data set is split randomly into k partitions. We then fit our model to a data set consisting of $k-1$ of the original k parts, and use the remaining portion for validation. Based on this, we estimated the out of sample error using the portion of data left out of the fitting procedure. We repeat this k times and our estimate for the out of sample error is the average over the k validation runs. Thus, the value of $k=10$ is used for the analysis in this paper.

A training pattern (Z_{SSE} vector) takes the format $<x_1, x_2, x_3, x_4, x_5, \ldots, x_n, \ldots.., y>$ where $x_1, x_2, x_3, x_4, x_5, \ldots, x_n$ denote the input vector and y denotes the security status output vector (target). This training pattern is called instances (row) while the inputs are featured or attributes (column). The device condition is, in fact, known as 'Static Secure' (SS-Binary one) whenever all the limitations mentioned in Section 3 are often satisfied for almost any provided backup. When somebody issues break 'is identified performing a problem, the device situation is going to be known as 'Static Insecure' (SI-Binary zero).

Engineering common sense occasionally may decide on the actual enter attributes. However, this kind of choices is going to be very subjective using the chance of essential factors obtaining turned down. A typical approach to feature selection will be a consecutive feature choice, composed of two elements – a target function known as criterion and also a consecutive investigation formula. The real feature factors chosen through SFS technique can serve as an input data source regarding creating the actual classifier formula. The SFS technique utilized in the current function begins with an empty group of features and also encourages prospective client function subsets with the help of one attribute every time. For each prospective client perform component, SFS operates the actual 10-fold combined authorization through frequently contacting the actual qualifying criterion operate. The actual qualifying criterion operates is really a reduction calculation determining the amount of misclassification studies within the mix affirmation of every prospect feature part. This method has actually continued before the inclusion of many more characteristics producing absolutely no further reduction in the actual qualifying criterion operate.

5.3. Training dataset preparation

To be able to put together working out information arranged, the specified options that come with the actual system tend to be obtained from the actual ready track documents. The key functions of the power system network are extracted in order to prepare the training data set. These functions tend to be transformed into the actual input/output dataset or even coaching designs needed in the coaching stage.

When the instruction dataset is ready while described previously, the actual dataset will be stabilized appropriately in variety [0, 1] by applying Eq. (3).

$$\bar{v} = \frac{v - \min_V}{\max_V - \min_V} \tag{3}$$

where v is the attribute V original value, $\bar{v}$ is the attribute V normalized value, and $\min_V$ and $\max_V$ are the minimum and maximum attribute V values respectively.

6. Data mining performance evaluation

The correct classification rate (CCR) can be defined as a key gauge employed for analyzing one particular or even classifier. Nevertheless, CCR only can be inadequate regarding gauging a functionality of the classifier for a static security index data set. And so, the true negative rate (TNR) and true positive rate (TPR) were used to evaluate the classifier performance. Moreover, geometric mean (GM) was additionally utilized in this research to assess the actual overall performance regarding device studying techniques, as shown in Table 2.

where:

TP (true positive): the number of positive samples classified correctly.

FP (false positive): the number of negative samples classified incorrectly.

TN (true negative): the number of negative samples classified correctly.

FN (false negative): the number of positive samples classified incorrectly.

Table 2
The procedures employed for assessing the efficiency of machine learning techniques.

Measures name	Formula
Correct classification rate (CCR)	$CCR = \frac{TP+TN}{TP+FP+FN+TN}$ (%)
True positive rate (TPR)	$TPR = \frac{TP}{TP+FN}$ (%)
True negative rate (TNR)	$TNR = \frac{TN}{TN+FP}$ (%)
Geometric mean (GM)	$GM = \sqrt{TPR * TNR}$ (%)

Table 3 explains the measures widely employed for analyzing efficiency regarding supervised machine learning methods.

Subsequently training and verification, the trained classifier was kept and then used later for the performance improvement of the conventional static security evaluation policies.

7. Results and analysis

Within this research, C4.5 models were properly trained by using a WEKA tool. WEKA is truly a workbench designed to help the use of machine learning approaches to various actual difficulties. WEKA is truthfully a totally released and also free code developed in Java. In WEKA, the machine learning algorithms tend to be realists organized into programs, to allow them to become efficiently brought in and besides applied in Java's code. Right after the training, the properly trained designs had been stored just as the documents being applied in enhancing the static security stage during the test stage. About applying WEKA classifiers in Java's code, WEKA guide are available in [43].

In the steady-state, the SSE limitations are the bus voltage magnitude (V_k) and the line thermal power (S) and can be written as follows:

$1.09 > V_k > 0.91$ and $S < S_{max}$.

The outcomes of information building and show choice stages of static security evaluation are shown in Table 4. The data samples in m-dimensional feature space are randomly split into training and test sets.

*Dimensionality reduction is designated by bold values, which is feature selection essential measure of.

The normalized inputs and target's vectors are randomly divided into two sets, 60% for training the network, 20% for validating and the rest of the data (20%) for providing an independent test of network generalization to data that the network has never seen.

It is to be noted that the limitation for both voltage magnitude is ($\pm 6\%$) and the line thermal power is 100 MW maximum. IEEE 30 bus test system consists of 30 buses, 20 loads, 41 branches and 6 generators. PV solar generators are installed at buses 1, 2, 13, 22, 23, and 27 and loads are in buses 2, 3, 4, 7, 8, 10, 12, 14-21, 23, 24, 26, 29, and 30.

Firstly, we adopt the idea that load at bus no.3 requests 20 MW, load at bus no.4 requests 25 MW and load at bus no.14 requests 15 MW from generators nos.1, 2 and 13 respectively, elaborated in Table 5 as follow:

Table 3
Confusion matrix for a two-class problem.

Examples	Predicted positive	Predicted negative
Actual positive	True positive (TP)	False negative (FN)
Actual negative	False positive (FP)	True negative (TN)

7.1. Transactions implementation

Preliminarily, in the base case assessment, 30 IEEE bus test systems attain well voltage profile. Figs. 2 and 3 show the power flow result before implementing any transaction between loads and PV generators. It is obviously revealed that all voltages are in the range of their limits, as in Fig. 6 and lines thermal power are not exceeded Fig. 7. Therfore, the status of the grid is secure.

Independently, the three transactions mentioned in Table 5 were implemented to see their influences on the static security of the network.

7.2. Performance evaluation and discussions

In order to evaluate the performance of a static security evaluation approach, it is very important to measure its performance. Therefore, some common performance measures are used to evaluate the performance of a particular security status index

Table 5
Transactions of generators and load.

Trans. no.	Gens. no.	Loads no.	MW
1	1	3	20
2	2	4	25
3	13	14	15

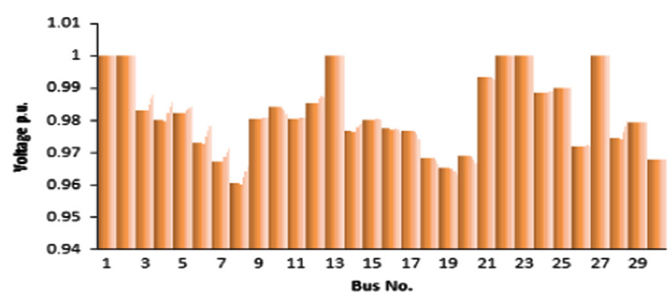

Fig. 6. Pool voltage profile.

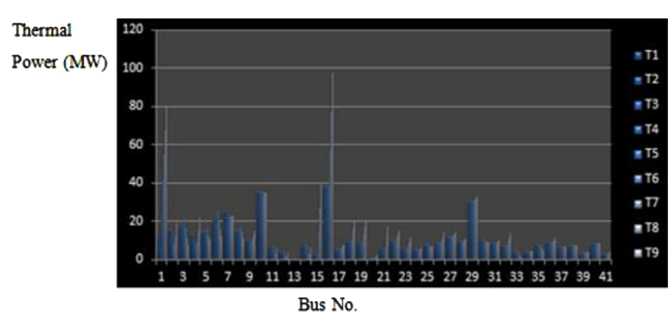

Fig. 7. Pool thermal power lines flow.

Table 4
Data generation and feature selection of different IEEE test systems.

System size	Operating scenarios	Static Secure (SS)	Static Insecure (SI)	No. of pattern variables (X_{SSE})	No. of features selected (Z_{SSE})	Dimensionality reduction
30 Bus	860	595	265	170	25	**14.70%**
57 Bus	950	630	320	185	27	**14.59%**
118 Bus	1100	750	350	210	29	**13%**
300 Bus	1330	760	570	220	26	**11.81%**

Table 6
Performance of C4.5 classifier for static security evaluation.

		Proposed	Decision Tree classifiers (DTC's)			
		C4.5 Tree	BF Tree	Stump Tree	J 48 Tree	J 48 graft
IEEE 57 bus	Total samples, 950					
Train set	Samples	630				
	CCR (%)	**98.64**	94.70	95.4	93.70	94.60
	TPR (%)	**96.30**	93.90	95.1	93.20	94.00
	TNR (%)	**97.21**	94.30	95.00	93.30	94.20
	GM (%)	**96.75**	94.09	95.049	93.25	94.10
	Time(s)	**0**	0.001	0.02	0.01	0.03
Test set	Samples	320				
	CCR (%)	**97.44**	93.50	91.2	92.50	93.40
	TPR (%)	**95.90**	93.60	95.5	93.70	94.90
	TNR (%)	**97.21**	94.15	95.70	93.30	94.20
	GM (%)	**96.55**	93.87	95.59	93.49	94.55
	Time(s)	**0**	0.003	0.04	0.02	0.05
IEEE 118 bus	Total samples, 1100					
Train set	Samples	750				
	CCR (%)	**98.44**	94.50	95.20	93.50	94.20
	TPR (%)	**96.80**	94.30	95.2	93.70	94.80
	TNR (%)	**97.5**	95.00	94.90	93.10	94.10
	GM (%)	**97.14**	94.65	95.049	93.39	94.45
	Time(s)	**0**	0.001	0.052	0.01	0.05
Test set	Samples	**350**				
	CCR (%)	**97.74**	93.80	91.50	92.80	93.70
	TPR (%)	**97.10**	93.75	94.7	94.10	94.10
	TNR (%)	**96.90**	94.30	94.90	93.20	94.30
	GM (%)	**96.99**	94.02	94.79	93.64	94.19
	Time(s)	**0.001**	0.002	0.055	0.015	0.08

compared with other approaches. Four different algorithms of DT's with same train datasets and test datasets are used in a comparison. This comparison was in terms of CCR, TNR, TPR, GM and computation time and presented in Table 6. Extra knowledge regarding the data mining algorithms used in this study is presented in [44]. Table 6 shows the comparison between the performance's measures of proposed C4.5 and other four various DT's techniques for the two network data sets (57 and 118 IEEE test systems) in both training and testing data sets. For the purpose of comparative analysis, BF Tree, Stump Tree, J 48 Tree and J 48 graft are selected because the Decision Trees are the most popular data-mining techniques and in this work they presented better results. In Table 6, the best and the worst values of the measures are highlighted in bold font and underline font, respectively. In training phase (57 bus system), BF Tree, Stump Tree, J 48 Tree and J 48 graft attained around 94.70%, 95.4%, 93.70%, and 94.60% of CCR respectively, while C4.5 Tree attained CCR of around 98.64%. In testing phase, BF Tree, Stump Tree, J 48 Tree and J 48 graft attained around 93.50%, 91.2%, 92.50%, and 93.40% of CCR respectively, while C4.5 Tree attained around 97.44% of CCR. In training phase (118 bus system), BF Tree, Stump Tree, J 48 Tree and J 48 graft attained around 94.50%, 95.2%, 93.50%, and 94.20% of CCR respectively, while C4.5 Tree attained around 98.44% of CCR. In testing phase, BF Tree, Stump Tree, J 48 Tree and J 48 graft attained around 93.80%, 91.5%, 92.80%, and 93.70% of CCR respectively, while C4.5 Tree attained around 97.74% of CCR.

Bold value validates that C4.5 provides great correct classification rate and minimum computation time to other DTC's classifiers. Finally, for train mode and test mode, Table 6 also demonstrates the computation time in seconds. Strongly, it can be observed that for both systems used, C4.5 got minimum computation time (0) second for training and testing phases. Furthermore, for the recall (test) phase where C4.5 got computation time of 0 s and 0.001 s for training and testing phase respectively.

8. Conclusion

The results and discussions of using C4.5 and other Decision Tree classifiers with presence of photovoltaic power plants for SSE the electric power grid has been presented. The implementation Decision Tree methods on several IEEE test systems involved appropriateness SSE and classification by using four algorithms of DT's. From this research work, it is observed that all these algorithms promise successful and alternative techniques for large scale power grid SSE. With the presence of solar PV technology, bilateral energy transactions between loads and PV generators are applied, and their impacts on the system security are presented too. Mentioned techniques can effectively be implemented for SSE of deregulated power system. 98.7% of CCR and zero second of computation time made C4.5 very well fit in the real-time for deregulated power systems SSE with the presence of PV solar energy.

References

[1] Singh S, Srivastava S. Improved contingency selection algorithm for voltage security analysis. Electric Machines Power Syst 1998;26:855–71.
[2] Ejebe G, Wollenberg B. Automatic contingency selection. IEEE Trans Power Appar Syst 1979:97–109.
[3] Verma K, Niazi K. Supervised learning approach to online contingency screening and ranking in power systems. Int J Electr Power Energy Syst 2012.
[4] Camara F, Samb ML, Ndiaye S, Slimani Y, Dakar S. Privacy preserving RFE-SVM for distributed gene selection; 2012.
[5] Jun S. A clustering method of highly dimensional patent data using bayesian approach; 2012.
[6] Saeh IS, Khairuddin A. Static security assessment using artificial neural network. In: Proceedings of IEEE 2nd international power and energy conference; 2008. p. 1172–8.
[7] Rosehart W, Canizares C, Quintana V. Optimal power flow incorporating voltage collapse constraints. IEEE 1999;2:820–5.
[8] Saeh I, Khairuddin A. Implementation of artificial intelligence techniques for steady state security assessment in pool market; 2009.
[9] Saeh I, Khairuddin A. Anfis and ANN comparison for static security assessment; 2008.
[10] Mohammadi M, Gharehpetian G. Power system on-line static security assessment by using multi-class support vector machines. J Appl Sci 2008;8:2226–33.
[11] Pang CK, Prabhakara FS, El-Abiad AH, Koivo AJ. Security evaluation in power systems using pattern recognition. IEEE Trans Power Appar Syst 1974;PAS-93:969–76.
[12] Singh AA, Thingujam M. Fuzzy ID3 decision tree approach for network reliability estimation; 2012.
[13] Stott B, Alsac O. Fast decoupled load flow. IEEE Trans Power Appar Syst 1974:859–69.
[14] Mori H, Tanaka H, Kanno J. A preconditioned fast decoupled power flow method for contingency screening. IEEE 1995:262–8.
[15] Brandwajn V, Lauby M. Complete bounding method for AC contingency screening. IEEE Trans Power Syst 1989;4:724–9.
[16] Morison K, Wang L, Kundur P. Power system security assessment. Power Energy Mag IEEE 2004;2:30–9.
[17] Takahata AY, dos Santos MG, Taranto GN, Schilling MT. Computational framework combining static and transient power system security evaluation using uncertainties. Int J Electr Power Energy Syst 2015;71:151–9.
[18] Sunitha R, Kumar SK, Mathew AT. Online static security assessment module using artificial neural networks. IEEE Trans Power Syst 2013;28:4328–35.
[19] Alves FR, Henriques RM, Fo P, Gomes S, Borges CL, Falcao DM, Aveleda A, Taranto GN, Assis T. Integrating solution engines under a distributed processing environment: an alternative approach for static and dynamic security assessment of large scale power systems. IEEE 2014:1–3.
[20] Zhu W, Wang Z, Yuan X, Fan R. Security assessment of distribution system with distributed photovoltaic. J Power Energy Eng 2015;3:250.
[21] WSPI, World solar power introduction; 2012. Available: 〈http://www.mbendi.com/indy/ener/sola/p0005.htm〉.
[22] Burrett R, Clini C, Dixon R, Eckhart M, El-Ashry M, Gupta D, Haddouche A, Hales D, Hamilton K, House U Chatham. Renewable Energy Policy Network for the 21st Century; 2009.
[23] Lugg A, Hong M. Energy issues in the Asia-Pacific region, vol. 2. Institute of Southeast Asian; 2010.
[24] OPEC. World Oil Outlook 2011. Organization of the Petroleum Exporting Countries (OPEC); 2011. p. 1–304.
[25] EIA. International energy outlook 2011. U.S. Energy Information Administration EIA; 2011. p. 1–301.

[26] Whitney JR, Billoski TV, Jones VR. Evidence for Triceratops in Antarctica. In: Billoski TV, editor. New directions in paleontology. New York: Academic Press; 1997. p. 24–7.
[27] Murphy WL, Andersen JM, Ebelin RM. Assessment of geology as it pertains to modeling uplift in jointed rock: a basis for inclusion of uncertainty in flow models [website]; 2002. Available ⟨http://libweb.wes.army.mil/uhtbin/hyperion/TR-02-2.pdf⟩.
[28] PTM. "PV industry handbook 2009" Malaysia building integrated photovoltaic (MBIPV) project. Pusat Tenaga Malaysia (PTM); 2009. p. 1–84.
[29] Junfeng L, Lohani B, Galàn EM, Monga P, Mubiru P, Nakicenovic N, Nassiep K, Pachauri R, Palz W, Pelosse H. Renewable energy policy network for the 21st century.
[30] Kalyani S, Swarup KS. Classification and assessment of power system security using multiclass SVM. IEEE Trans Syst Man Cybern Part C: Appl Rev 2011;41:753–8.
[31] Albuyeh F, Bose A, Heath B. Reactive power considerations in automatic contingency selection. IEEE Trans Power Appar Syst 1982:107–12.
[32] Wehenkel LA. Automatic learning techniques in power systems. Kluwer Academic Publishers; 1998.
[33] Marsadek M, Mohamed A, Nizam M, Norpiah Z. Risk based static security assessment in a practical interconnected power system. In: Proceedings of IEEE 2nd international; 2008. p. 1619–22.
[34] Mori H. State-of-the-art overview on data mining in power systems. In: Proceedings of IEEE PES; 2006. p. 33–34.
[35] Hatziargyriou N, Papathanassiou S, Papadopoulos M. Decision trees for fast security assessment of autonomous power systems with a large penetration from renewables. IEEE Trans Energy Convers 1995;10:315–25.
[36] Georgilakis P, Hatziargyriou N. On the application of artificial intelligence techniques to the quality improvement of industrial processes. In: Proceedings of methods and applications of artificial intelligence; 2002. p. 473–84.
[37] Voumvoulakis E, Gavoyiannis A, Hatziargyriou N. Decision trees for dynamic security assessment and load shedding scheme. In: Proceedings of IEEE; 2006. p. 7.
[38] Wehenkel L, Pavella M. Decision trees and transient stability of electric power systems. Automatica 1991;27:115–34.
[39] Wehenkel L, Van Cutsem T, Ribbens-Pavella M. An artificial intelligence framework for online transient stability assessment of power systems. IEEE Trans Power Syst 1989;4:789–800.
[40] Huang C-J, Wang Y-W, Huang T-H, Lin C-F, Li C-Y, Chen H-M, Chen PC, Liao J-J. Applications of machine learning techniques to a sensor-network-based prosthesis training system. Appl Soft Comput 2011;11:3229–37.
[41] Momoh J, Xia Y, Boswell GD. Locational Marginal Pricing for real and reactive power. In: Proceedings of power and energy society general meeting-conversion and delivery of electrical energy in the 21st century. IEEE; 2008. p. 1–6.
[42] Zhou W, Peng, Sun H, Probabilistic wind power penetration of power system using nonlinear predictor-corrector primal-dual interior-point method. In: Proceedings of the third international conference on electric utility deregulation and restructuring and power technologies (DRPT); 2008. pp. 2548–52.
[43] Bouckaert RR, Frank E, Hall M, Kirkby R, Reutemann P, Seewald A, Scuse D. WEKA Manual for Version 3-6-1. Hamilton, New Zealand: The University of Waikato; 2009.
[44] Witten IH, Frank E. Data mining: practical machine learning tools and techniques. Morgan Kaufmann; 2005.