Cleaner Engineering and Technology 15 (2023) 100654

Contents lists available at ScienceDirect

Cleaner Engineering and Technology

journal homepage: www.sciencedirect.com/journal/cleaner-engineering-and-technology

Real-time computing of power flows and node voltages in electrical energy network using decision trees

Sonali Nandanwar [a], Narayan Prasad Patidar [b], M. Deva Brinda [c], Mohan Lal Kolhe [d,*]

[a] Electrical & Electronics Engineering Department, IES College of Technology Bhopal, India
[b] Electrical Engineering Department, Maulana Azad National Institute of Technology, Bhopal, India
[c] Norwegian Solar India Private Limited, Chennai, India
[d] Faculty of Engineering and Science, University of Agder, PO Box 422, NO 4604, Kristiansand, Norway

ARTICLE INFO

Keywords:
Sustainable electrical energy network
Power flows
Voltage security
Binary decision tree
Regression tree

ABSTRACT

In sustainable operation of electrical energy network, it is necessary to compute in real-time power flows and voltages at nodes for prioritizing power injection from clean energy resources. Intermittent renewable energy sources are likely to create voltage and power balancing issues and to maintain the voltage security of electrical network, real-time information of network power flows and bus voltages are required accurately and instantaneously. This paper presents an approach based on decision trees (DT) for real-time estimation of power flows within the electrical energy network and node voltages. A single tree structure is built for estimation of discrete (or categorical) as well as continuous values of line flows and node voltages of each line and node separately. A simple binary decision tree (BDT) and regression tree (RT) are used for estimation of discrete values and continuous values respectively. The training and testing patterns are generated by performing power flow analysis on an electrical energy network. Once the DT is trained, it estimates the line power flows and bus voltages with desired accuracy. The accuracy of the DT model is tested on a typical IEEE 30-bus system, using test patterns. Result shows that mean absolute error in case of line flow estimation for line number 1 and 10 are found to be 0.0028 p. u. and 0.0017 p. u. Also mean absolute error in case of bus voltage estimation for bus number 3 and 10 are found to be 0.0019 p. u. and 0.0016 p. u. Above results are suggestive of instantaneous estimationwith desired accuracy of line flow and bus voltages, which is the need of the hour for sustainable electrical energy network with integration of cleaner energy resources.Since, DT gives instantaneous result therefore suitable for real-time applications in sustainable electrical energy management system.

1. Introduction

The grid integration of clean energy sources is increasing exponentially, and it is impacting sustainable operation of electrical energy network. It demands a high degree of power network security for sustainable operation. Thus, there is a pressing need to develop fast real-time power flow management for managing the electrical energy network in secure manner by controlling line power flows, voltage, etc. The real-time monitoring methods, which will help power system operators to analyse the level of network security and to apply possible control actions in case there is a violation of the system security constraints. As a result of disturbances and integration of clean energy resources, the power system operating state may move into an undesirable emergency state, whenever there is a violation of security constraints. Power system voltage security assessment needs databases that are created by load flow simulation study. Practically instabilities in the system are very small for most of the systems. Creation of databases are laborious task but it is advantageous, because it is possible to create large number of scenarios which is required to solve a certain task. Under these conditions fast computation of line power flows and bus voltage profiles is very important to enhance the power system steady state security assessment (Wood and Wollenberge, 1984). The computation of bus voltage and line flows by load flow analysis time consuming as it should run for any change in load/generations. In past several approaches are proposed such as distribution factors method (Illic-spong and Phadke, 1986; Ching and Nanming, 1992), bounding

* Corresponding author.
E-mail addresses: sonalinandanwar@gmail.com (S. Nandanwar), patidarnp@manit.ac.in (N.P. Patidar), arshhiyadeepak@gmail.com (M.D. Brinda), Mohan.L.Kolhe@uia.no (M.L. Kolhe).

https://doi.org/10.1016/j.clet.2023.100654
Received 30 April 2023; Received in revised form 6 June 2023; Accepted 18 June 2023
Available online 27 June 2023

method (Brandwajn and Lauby, 1989) concentric relaxation method (Zaborszky et al., 1980) pattern recognition method (Chang, 1989) etc., but they suffer from large computational time. Also,an artificial neural network (ANN) based method (Srivastav et al., 1998) is proposed but it takes long training time and knowledge to solve the problems is represented by the synoptic weights of ANN and difficult for system operators to understand the insights.Decision tree method (Yang and Hsu, 1994) is applied in past for line flows and bus voltage estimations in which all input attributes and desired output were taken in terms of levels (discrete values). Since input attributes are defined in terms of levels, hence splitting level of attribute deviate from true threshold of the attribute value, which introduces more classification errors. In this approach the number of branches from each node of decision tree is equal to the number of levels of the feature in that node hence tree size becomes larger as the number of levels are increases. S. Nandanwar et al. (Nandanwar and Warkad, 2016) have proposed probabilistic fuzzy decision tree (PFDT) and classification and regression tree (CART) algorithm for voltage security event classification. Both methods were compared for classification accuracy which prove the PFDT has more accurate. S. Nandanwar et al. (2018) proposed probabilistic fuzzy decision tree for voltage security assessment, in which PFDT is used for case-based reasoning approach to learn new cases into the case-base, which is further used for security events classification. S.K. Jain et al. (2022a) proposed adaptive fuzzified decision tree for voltage security classification which only classify in terms of 'secure' and 'insecure' states with fuzzified decision variable but no real values are predicted.S. K. Jain et al. (2022b) also proposed power system voltage security event classification based on basic decision tree C4.5 model, in which security classes are classified secure and insecure operating states. All the above approaches discussed were used only for prediction in terms of classes like secure or insecure, but not prediction of continuous values of output parameter. Xiangfeiet et al. (Meng et al., 2020) proposed a DT approach for assessment of voltage security margin using DT algorithm C4.5 and CART which gives output in terms of the classes (categorical values) only. Meng Mahdi Kioumarsi et al. (Kioumarsi and Dabir-iAmirrezaKandiri, 2023) proposed a decision tree method for estimation of compressive strength of concrete, which give continuous values in output. Shahid Husain et al. (Husain and Khan, 2021) proposed decision tree approach for assessment of monthly average diffuse solar radiation with solitary input forecaster as clearness index. Vasudev Dehalwar et al. (Dehalwar et al., 2022) proposed a blockchain-based self-sovereign identification method which has been tested by analysing to avert the identity theft and masquerading in which IoT devices based on Blockchain is presented. This approach can minimize the chances of identity-based security breaches in the smart grid.

In this paper, a decision tree-based approach is presented to compute transmission line power flows and node voltages of the power system. In this approach using the combination of BDT and RT, all the input attributes are taken in terms of their continuous values and output (line flows and node voltages) are also estimated in terms of real or continuous values. BDT and RT are embedded into the single tree structure referred as decision tree to estimate continuous values of line flow and node (i.e., bus) voltages. In a BDT all the input attributes are taken in terms of their continuous values and only output i.e., transmission line power flows and node voltages are taken in terms of levels (discrete values), hence it estimates the output values in terms of levels. BDT is useful for categorical classifications like states of power system 'secure' and 'insecure'. In RT all the inputs as well as output are considered in terms of their real or continuous values. RT would be more useful, as the power system operation and control require real values of decision variables. The accuracy of decision tree is improved because of splitting is done at true threshold values of input attributes. Also, the tree structure becomes simple, because each node is associated with two branches only. Training time of the decision tree is very less (negligible); therefore, decision trees tend to be more useful in circumstances where the training is required regularly when new transmission lines and generators are added. This approach would also be useful with integration of clean energy resources which has intermittent nature of power generation like solar and wind generation.

2. Decision tree

In this approach simple binary decision tree (BDT) and regression tree (RT) are embedded into the single tree structure referred as decision tree. Here BDT is by-product of RT because RT is generated by first generating BDT. RT predict real values, are similar to binary decision tree (Pao, 1989; Quinlan, 1986; Wehenkel and Pavella, 1993). RT stores the numeric class values in the terminal nodes, whereas in case of binary decision tree, the output is obtained in terms of levels. RTs based on CART are binary trees, which gives constant real values in the terminal nodes and measure of impurity is determined by variance (Breiman et al., 1984). In this approach RT created by first creating binary tree and measure of impurity is given by entropy of learning data set. The average of the output values of instances that reaches to the particular leaf of binary decision tree is defined as numeric class value of RT. When a new instance is presented to the regression tree for testing, it travels along testing nodes and finally reaches to the leaf to which it belongs. The output to new instance is the class value of the leaf to which it reaches. If the number of levels is increased in the output, then the output values of the instances belong to that level of data becomes more closure to the average value of them. Hence accuracy of predictions is increased.

DT required a reach data set, therefore in this work data set is generated by simulation under varying load conditions using the AC load flow analysis. A separate DT is built for each bus and transmission line. And all the DT are configured in parallel. When a testing pattern is presented, it will reach to the related DT only. The whole scheme is presented and discussed in sub section 3.1.

2.1. DT Vs other approaches

Most important features of DT are its inherent ability of selecting important input features (real and reactive load at PQ buses) based on information gain which has large impact on bus voltage and line flows, which is the outcome of this work. But in case of ANN, input feature selection needs separate algorithm. DT has taken about very small time (in the range of msec)to train and build the decision tree, whereas ANN requires large training time due to iterative process used in ANN training. After the DTs are trained, the execution time for computing line flows and bus voltage is almost instantaneous, because DT has only to make few comparisons (if-then rules). As the decision trees generate if-then rules to estimate the desired parameters, hence more transparent and easier for human user to understand insights, whereas in ANN the knowledge is represented by synoptic weights which are impossible to understand by operators. Real-time performance of decision tree is almost instantaneous, whereas load flow study and analytical methods, such as the distribution factor and bounding methods etc. are time consuming and not feasible for real time applications. In view of above DT are most suitable for real time applications and development of future smart grids and sustainable energy networks.

3. Line flows and bus voltage estimation

3.1. Methodology

The block diagram that describes the transmission line power flow and bus voltage computation problem is shown in Fig. 1. First the database is generated by performing optimal power flow analysis. Data base consist of real power P_d, and reactive power Q_d, at each PQ bus, real power generation P_g and voltage V_g of PV buses, transmission line power flows and bus voltages at all the lines and PQ buses respectively. DTs are generated to extract the hidden knowledge from database. Transmission line power flows and bus voltages (p.u. values on the base of 100MVA)

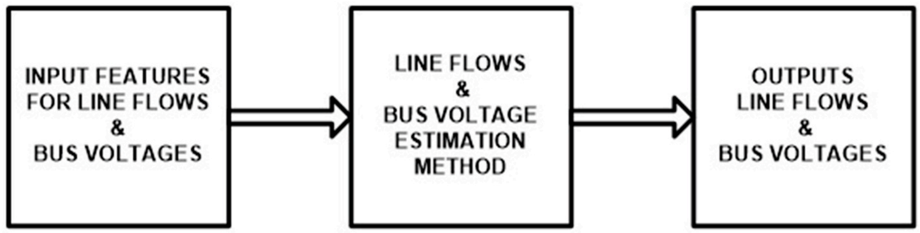

Fig. 1. Line flow and bus voltage computation framework.

are target quantities to be estimated. To identify the pattern and its respective bus or transmission line of the power system, an additional attribute 'K' is introduced into the input pattern vector.

Thus, the input vector becomes

$$X = [P_1, P_2, ...P_n, Q_{1,}Q_2, ...Q_n, P_{gi}, V_{gi}, K]$$

where P_n is real and Q_n is reactive demand on n^{th} PQ bus respectively. P_{gi} and V_{gi} are real power generation and voltage of i^{th} PV bus.

Fig. 2 shows the generalized diagram of proposed approach. DTs are required for all the transmission lines or buses separately for computation of line power flow and bus voltage profiles. DTs are configured in parallel to estimate line power flows and bus voltages profiles separately. Identifier K is required to identify the DT to which test pattern belongs to. Where K is PQ bus or transmission line number.

3.2. Algorithm

Detailed procedure for creating the DTs is summarized as follows.

Step 1 Performing AC load flow to generate large number of load patterns by varying the load randomly at each bus. Each pattern comprises N (N = 59 in this work) input features (Active load P_d, reactive load Q_d, active generation P_g, and voltage of PV buses) and an output (line flows and bus voltages). These patterns are generated for the same topology.

Step 2 Transmission line power flows and bus voltages (outputs) are discretized in to a suitable number of levels between their maximum and minimum values (in this work 5 number of levels is found suitable to get desired accuracy). Output can be discretized into a greater number of levels (in case of large variation in output values) to obtain output more closure to actual value.

Step 3 Determining the root node by computing prior classification entropy (A3) of all the attributes and selecting the attribute, which has highest entropy.

Step 4 Dichotomy test (A1) is performed on each test node to split so that maximum information gain (A2) can be obtained.

Step 5 Determining the leaves and dead ends by applying stop splitting rule as:

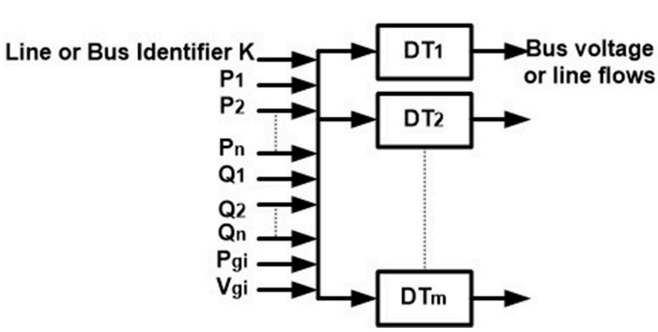

Fig. 2. Conceptual diagram of decision tree model for bus voltage and line flow estimation.

(a) If the subset S is containing (almost completely) those states, which belongs to any one class, then the node is a leaf, i.e. If Hc (s) $\leq H_{min}$, minimum value of entropy at the node then the node is a leaf node

(b) If $H_c(s) \leq H_{min}$ in step 5(a) is not true, then test on score of best attribute has to be done. If the best attribute has a too low score (smaller than some minimum value, S_{cmin}), the node is declared to be a dead-end.

Step 6 Determine numeric class values for all the leaves of the tree using (A7).

4. Results and discussions

An IEEE 30-bus test system (Power system test cases, 2023) is used to test effectiveness of proposed approach. The 30-bus system is having 6 PV buses, 24 PQ buses and 41 transmission lines. For each bus and transmission line a DT is required separately to be trained to compute bus voltage profiles and transmission line power flows, under varying demand. Since, for real time voltage security assessment we need to find voltage at all the buses and power flows through all the lines under varying loads and contingencies to monitor any over/under voltage at the buses and overloading of the transmission lines. Since AC load flow is an iterative process and takes long times to estimate the bus voltage and line flows which is not feasible in practical size power networks. Therefore, data mining methods like DT which gives prompt response in real time applications and useful for sustainable energy networks. System operators has to monitor any violation of the line loadings which may result into blackouts if not identified and appropriate corrective action taken in advanced. Since, clean energy resources or intermittent in nature which may cause the load flow study more tedious and time consuming which is not desirable in real time applications. Therefore, DT approach is proposed to estimate line flows and bus voltages in real time for sustainable energy networks. In this work, 300 load patterns were generated randomly for each bus voltage and lines flows. Out of which 200 patterns were used for training purposes and the remaining 100 patterns are used for testing purposes.

4.1. Line flow estimation

Decision trees are built using training set for all the 41 lines. A Decision trees for each line has tested by100 test patterns, which were unseen to the decision tree. Figs. 3 and 4 shows the decision trees for line flows of line number 1 & 10 respectively. Variables inside the node (ovals shapes in the figure of the tree) are input attributes i.e., real and reactive load at PQ buses, which leads that node (e.g., in Fig. 3, Pd18 is the real power demand at bus No. 18, etc.). In the leaves (terminal node) of the tree, range of line flows are shown, which were defined as levels in their ascending orders and it is the class value of BDT. Constant value in the leaves represents the class value of RT. Number appears below the leaf's indicate number of training patterns, falling in to that leaf.

Because of space limitation, testing results of two lines 1&10, for 10 samples testing patterns are listed in Tables 1 and 2 respectively. Tables 1 and 2 shows predicted and actual values in terms of levels and continuous values (p.u.) of line flows. The mean absolute error for all the

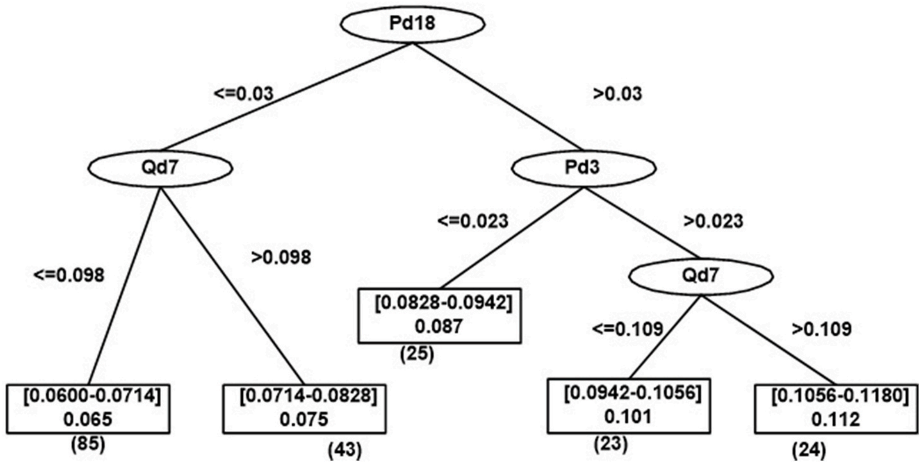

Fig. 3. Decision tree for line-1 flow.

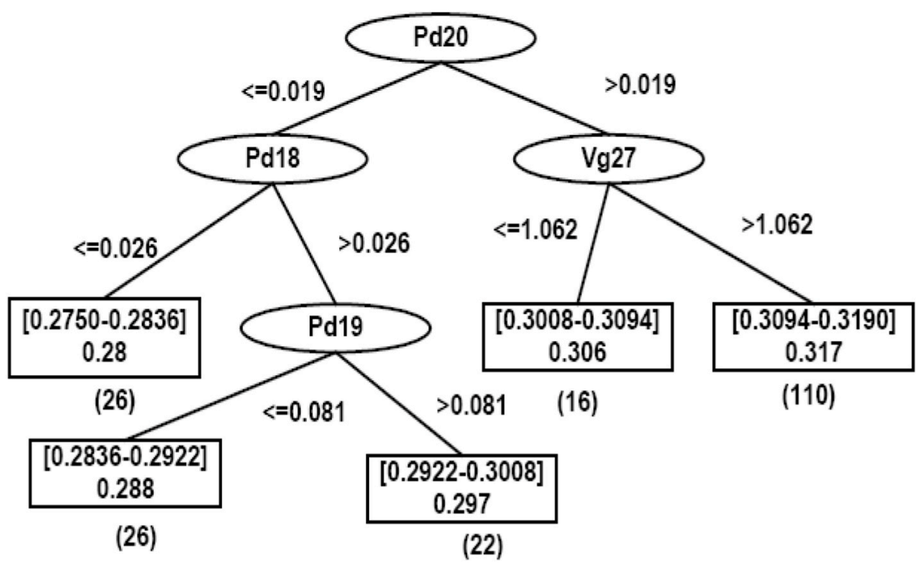

Fig. 4. Decision tree for line-10 flow.

Table 1
Line flow estimation of line-1.

By Binary decision Tree			By Regression Tree		
Actual Flow Level	Predicted Flow Level	Error	Actual Line Flow	Predicted Line Flow	Error
2	2	0	0.074	0.075	0.001
2	2	0	0.072	0.075	0.003
2	2	0	0.074	0.075	0.001
1	1	0	0.066	0.065	−0.001
2	2	0	0.074	0.075	0.001
2	2	0	0.074	0.075	0.001
4	4	0	0.099	0.101	0.002
4	4	0	0.104	0.101	−0.003
3	3	0	0.089	0.087	−0.002
1	1	0	0.071	0.065	−0.006
No of test pattern = 100			No of test pattern = 100		
Correctly classified pattern = 100			Mean absolute error = 0.0028		

Table 2
Line flow estimation of line-10.

By Binary decision Tree			By Regression Tree		
Actual Flow Level	Predicted Flow Level	Error	Actual Line Flow	Predicted Line Flow	Error
4	4	0	0.305	0.306	0.001
1	1	0	0.277	0.280	0.003
5	5	0	0.318	0.317	−0.001
3	3	0	0.299	0.297	−0.002
1	1	0	0.280	0.280	0.0
5	5	0	0.310	0.317	0.007
1	1	0	0.283	0.280	−0.003
5	5	0	0.318	0.317	−0.001
3	3	0	0.296	0.297	0.001
2	2	0	0.292	0.288	−0.004
No of test pattern = 100			No of test pattern = 100		
Correctly classified pattern = 98			Mean absolute error = 0.0017		

100 testing patterns was found 0.0028 and 0.0017 for lines 1 & 10 respectively. Line flow estimation using BDT, for lines 1 & 10, respectively, in terms of levels is also given in Tables 1 and 2 in which errors was found zero. Testing results for both the lines 1 & 10 for 50 testing patterns are graphically presented in Figs. 5 and 6. It can be seen that estimated line flows by RT closely matches with the values calculated by optimal power flow analysis.

It is found that discretizing the line flows into 5 numbers of levels are enough to get the desired accuracy. Error analysis of BDT for all the 41 lines is given in Table 5. In which each line is tested for 100 load patterns

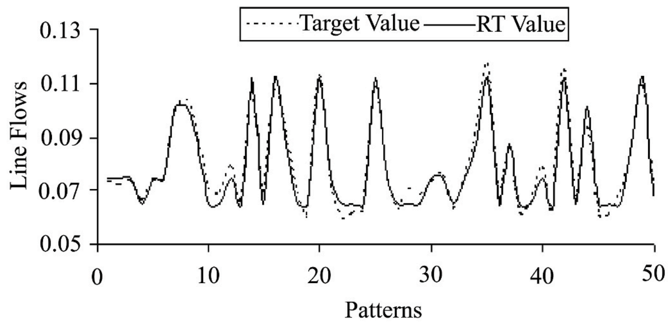

Fig. 5. Testing result of line-1.

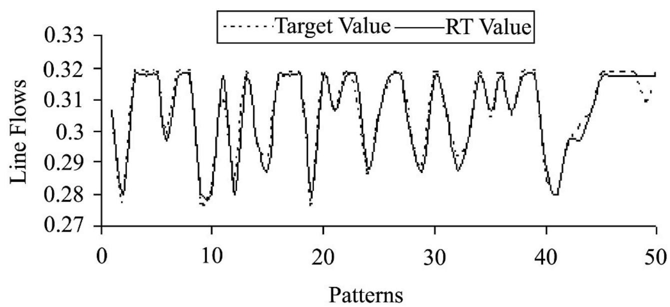

Fig. 6. Testing result of line-10.

(total test are 4100) out of 4100 tests performed, 4035 are having no error. Percent of correctly classified patterns for line flows is 98.41%. Percent of patterns in which, there is error of one level is 1.57%. Percent of patterns in which, there is error of two level is 0.0732%. Percentage expected error per test is 1.66%.

DT has self-contained ability to select important features only, which impacts the line flows most. In all the cases of line flows total input features were 59 (P_d,& Q_d at PQ buses, P_g &V_g at PV buses). Figs. 3 and 4 shows the tree for line flows of line 1&10, in which out of 59 features only 4 input features are selected for building the tree. Input features of least importance, which are not significantly influencing the line flow, are discarded on the basis of information measure.

4.1.1. Decision process

Taking example of estimation of line flow for line number-10 given in Fig. 4. For this line, during random pattern generation for 300 patterns, minimum and maximum p.u. value of line loading is divided into 5 levels defined as Level-1 [0.2750–0.2836], Level-2 [0.2836–0.2922], Level-3 [0.2922–0.3008], Level-4 [0.3008–0.3094] and Level-5 [0.3094–0.3190] as mentioned in Fig. 4. Now during training session 200 patterns were presented to build the DT. Then these 200 patterns were divided in five levels from Level-1 to Level-5 as 26, 26, 22, 16, 110 respectively. Using Eq. (A7)(in Appendix) the average values of line flow for instances falling in each individual levels are mentioned in terminal nodes of DT as constant values viz. 0.28, 0.288, 0.297, 0.306, 0.317 in Fig. 4.

When testing pattern (which is unseen to DT) is presented to DT, then DT made logical comparisons (using if-then rules) with input attributes. Input attributes which are selected during training as mentioned in DT of Fig. 4 viz. Pd18, Pd19, Pd20 (i.e. real load on bus no. 18, 19 & 20) and Vg27 (voltage at bus no. 27). Based on the values of input attributes, given testing pattern finally travels to any of the terminal node to which it belongs. Once the testing pattern reached to particular terminal node, the average value given in that node is assigned as outcome of the testing pattern i.e. line flow corresponding to that loading pattern. Results for line flow for line-10 are given in Table-2, in which predicted values are in terms of average value of instances in terminal nodes of DT as given in Fig. 4. Results reveals that absolute mean error for line-10 flow is 0.0017 which is very small and acceptable for real-time contingency analysis. For other cases of line flow and bus voltages, same process of decision making is followed.

4.2. Bus voltages estimation

Figs. 7 and 8 shows the decision trees for bus voltages of bus number 3 & 10 respectively, in which out of 59 input features only 4 input features in each are selected for building the tree. Because of space limitation, testing results of two buses 3 & 10, for 10 sample testing patterns are listed in Tables 3 and 4 respectively. Tables 3 and 4 shows the mean absolute error for all the 100 testing patterns, which was 0.0019 and 0.0016 for buses 3 & 10 respectively. Tables 3 and 4 also show the bus voltage estimation in terms of levels using BDT, for buses 3 & 10, respectively, in which errors was found zero.

Testing results for both the buses 3 & 10 for 50 testing patterns are graphically presented in Figs. 9 and 10. It can be seen that estimated bus voltages by RT closely matches with the values calculated by optimal power flow analysis. It is found that discretizing the bus voltages for RT into 5 number of levels are enough to get the desired accuracy. Error analysis of BDT for all the 24 buses is given in Table 5. In which each bus is tested for 100 load patterns (total test are 2400) out of 2400 test performed, 2369 are having no error. Percent of correctly classified patterns for bus voltages is 98.72%. Percent of patterns in which, there is error of one level is 1.29%. In bus voltage estimations only error of one level is found, hence percentage expected error per test is same as percent of patterns in which error of one level is found. In real-time contingency ranking of we need only relative severity. Therefore, without performing load flow study in real-time (which is very time consuming) proposed approach is most useful and gives results almost instantaneously. Objective of this work is to estimate line flow and bus voltages in real time, and also the same approach can be extended for contingencies ranking and selection using suitable performance indices for severity measurement, which will be done in the future work.

5. Conclusion

In this paper, a BDT and RT approach has been presented to compute transmission line power flow and bus voltages of a 30-bus IEEE test system. The BDT gives the accurate results in terms of the pre-defined levels. These levels can be analysed and defined as secure or insecure for power system security assessment, which is very useful in real time environment, where this information is required almost instantaneously. RT predicts the values of transmission line power flows and bus voltages in terms of continuous values with the desired accuracy. Integration of clean energy resources with grid may cause large fluctuation in voltages.

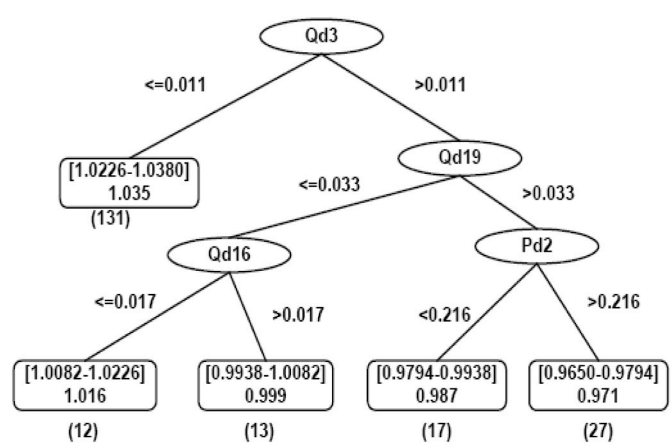

Fig. 7. Decision tree for Voltage of Bus-3.

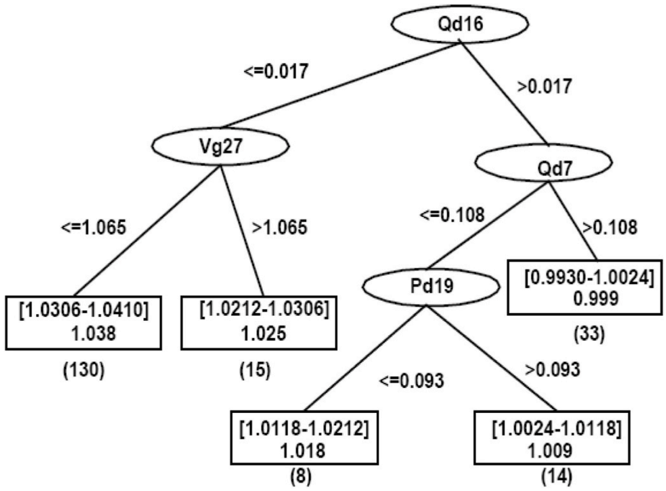

Fig. 8. Decision tree for voltage of bus-10.

Table 3
Bus voltage estimation of bus-3.

By Binary decision Tree			By Regression Tree		
Actual Voltage Level	Predicted Voltage Level	Error	Actual Bus Voltage	Predicted Bus Voltage	Error
5	5	0	1.036	1.035	−0.001
5	5	0	1.037	1.035	−0.002
5	5	0	1.025	1.035	0.01
1	1	0	0.972	0.971	−0.001
4	4	0	1.017	1.016	−0.001
5	5	0	1.036	1.035	−0.001
1	1	0	0.973	0.971	−0.002
2	2	0	0.984	0.987	0.003
5	5	0	1.037	1.035	−0.002
5	5	0	1.037	1.035	0.002
No of test pattern = 100			No of test pattern = 100		
Correctly classified pattern = 100			Mean absolute error = 0.0019		

Table 4
Bus voltage estimation of bus −10.

By Binary decision Tree			By Regression Tree		
Actual Voltage Level	Predicted Voltage Level	Error	Actual Bus Voltage	Predicted Bus Voltage	Error
1	1	0	0.998	0.999	0.001
5	5	0	1.038	1.038	0.0
5	5	0	1.038	1.038	0.0
1	1	0	1.0	0.999	−0.001
5	5	0	1.037	1.038	0.001
5	5	0	1.038	1.038	0.0
1	1	0	1.0	0.999	−0.001
4	4	0	1.03	1.025	−0.005
5	5	0	1.037	1.038	0.001
3	3	0	1.012	1.009	−0.003
No of test pattern = 100			No of test pattern = 100		
Correctly classified pattern = 100			Mean absolute error = 0.0016		

Therefore, fast computation of real time line flow and bus voltages are having paramount importance in a power grid with integration of clean energy resources like solar and wind energy. Since, RTs estimates continuous values in real time, therefore this approach would equally be useful for practical power systems with exponentially increasing clean energy resources. The estimated value can be achieved more closure to actual value by dividing the output into large number of levels. Most important features of DT are its inherent feature selection ability, based

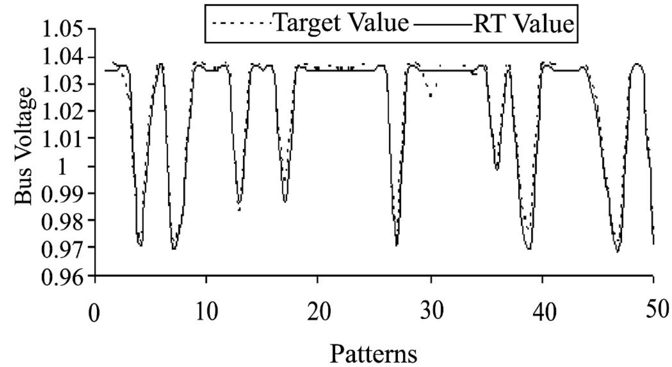

Fig. 9. Testing result of bus-3.

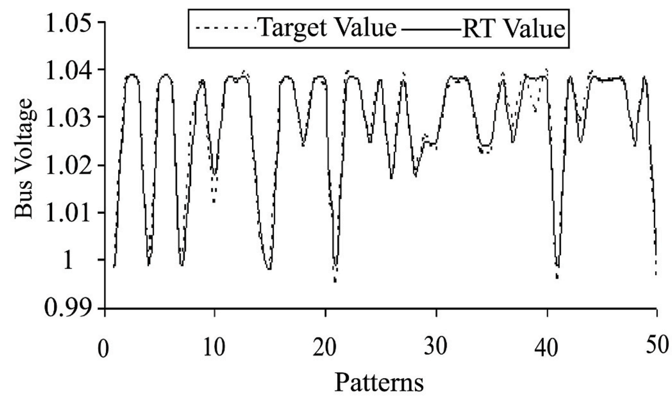

Fig. 10. Testing result of bus-10.

Table 5
Testing results for all the lines and buses.

Absolute error level	For line flows		For bus voltages	
	Number of tests	Percentage (%)	Number of tests	Percentage (%)
0	4035	98.41	2369	98.72
1	65	1.57	31	1.29
2	3	0.0732	0	–
3	0	–	0	–
4	0	–	0	–
5	0	–	0	–
	Total no. of test performed for line flows = 4100		Total no. of test performed for bus voltages = 2400	

on information gain it selects only those attributes (real and reactive load at PQ buses) for building DT, which influence the transmission line flow and bus voltage most. On the other hand, in case of ANN, feature selection is to be done separate algorithm to reduce size of architecture of ANN. In this work DT has taken about 0.09 s to build the decision tree, whereas ANN requires large training time due to iterative process used in ANN training. After the DTs are trained, the execution time for computing line flows and bus voltage is almost negligible, because DT has only to make few comparisons. Due to its inherent feature selection capability, DT selects only the features which influence the line flow and bus voltage largely, as system size increases, training and execution time would not be affected significantly. As the decision trees generate if-then rules to estimate the desired parameters, hence more transparent and easier for human user to understand insights, whereas in ANN the knowledge is represented by synoptic weights which are impossible to understand by operators. Real-time performance of decision tree is almost instantaneous, whereas load flow study and analytical methods, such as the distribution factor and bounding methods etc. are time consuming. The proposed approach would support to develop sustainable electrical energy networks which may deliver the clean energy efficiently and reliably to the consumers. Also due to intermittent nature of renewable energy resources, fluctuation in bus voltage and line flow may be large and need instantaneous estimation of them, which can be done using DT approach. This work can be further extended by considering contingencies for real time voltage security assessment. Therefore, proposed approach is going to be useful for sustainable operation of future electrical energy networks.

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.

Data availability

The authors are unable or have chosen not to specify which data has been used.

Appendix

Dichotomy

Then dichotomy test ‘T’ is defined as (Quinlan, 1986)

$$a_i \leq V_{is}? \tag{A1}$$

Test is performed on each node progressively which chooses an attribute a_i and cutoff value as V_{is}. Test is executed on all the values of a_i to decompose the training set into subsets at each node. Threshold value is selected in such a way to get the maximum information gain.

To measure information gain

According to information gain theory, information gain in set of pattern ‘S’ of the training set at the current node by the test ‘T’ is given by (Quinlan, 1986)

$$I_C^T(S) = H_C(S) - H_{C/T}(S) \tag{A2}$$

where Hc(s) is prior classification entropy measures the impurity of set ‘S’ given by (Quinlan, 1986)

$$H_C(S) = -\left[\left(\frac{N_{c1}}{S}\right)Log_2\left(\frac{N_{c1}}{S}\right) + \left(\frac{N_{c2}}{S}\right)Log_2\left(\frac{N_{c2}}{S}\right)\right] \tag{A3}$$

where Nc_1 and N_{c2} are the number of training patterns pertaining to classes say ‘yes’ and ‘no’ respectively.

Also,$H_{c/T}$(S) is mean posterior classification entropy of subset ‘S’. If test ‘T’ comes true, then number of patterns are S_y and if test ‘T’ is false then ‘S’ breaks into subset having states is S_n (Quinlan, 1986).

$$H_{C/T}(S) = -\left[\left(\frac{S_y}{S}\right) \times H_C(S_y) + \left(\frac{S_n}{S}\right) \times H_C(S_n)\right] \tag{A4}$$

Normalized form of information gain (called score) is described as (Quinlan, 1986)

$$G_C^T(S) = 2 \times \frac{I_C^T(S)}{H_C(S) + H_T(S)} \tag{A5}$$

here H_T(s) is the quantify the uncertainty in the outcome of test ‘T’ for the subset ‘S’ at current node. (Quinlan, 1986)

$$H_T(S) = -\left[\left(\frac{S_y}{S}\right)Log_2\left(\frac{S_y}{S}\right)\right] + \left(\frac{S_n}{S}\right)Log_2\left(\frac{S_n}{S}\right) \tag{A6}$$

Regression

$$C = \frac{1}{N}\sum_{i=1}^{N}\left(c_i\right) \tag{A7}$$

where.

C = Class value (numeric) of the leaf.
N = Number of instances reaches to that leaf.
c_i = Class value of i^{th} instance reaches to the leaf.

References

Brandwajn, V., Lauby, M.G., 1989. Complete bounding method for ac contingency screening. IEEE Trans. Power Syst. 4 (2), 724–728.

Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A., 1984. Classification and Regression Trees", first ed. Routledge, New York, USA.

Chang, C.S., 1989. Fast power system voltage prediction using knowledge-based approach and on-line box data creation. IEE proc 136, 87–99.

Ching, Y.L., Nanming, C., 1992. Distribution factors of reactive power flow in transmission line and transformer Outage studies. Trans. on power syst 7 (1), 194–200.

Dehalwar, V., Kolhe, M.L., Deoli, S., Jhariya, M.K., 2022. "Blockchain-based trust management and authentication of devices in smart grid". Journal of Cleaner Engineering and Technology 8.

Husain, S., Khan, U.A., 2021. Machine learning models to predict diffuse solar radiation based on diffuse fraction and diffusion coefficient models for humid-subtropical climatic zone of India. Journal of Cleaner Engineering and Technology 13.

Illic-spong, M., Phadke, A., 1986. Redistribution of reactive power flow in contingency studies. IEEE Trans. Power Syst. PWRS-1 (3), 266–275.

Jain, S.K., Patidar, N.P., Kumar, Y., Agrawal, S., 2022a. Real-time voltage security assessment using adaptive fuzzified decision tree algorithm. International Journal of Engineering Systems Modelling and Simulation, Inderscience 13 (1), 85–95.

Jain, S.K., Agrawal, S., Shukla, P.K., Shukla, P.K., 2022b. A decision tree C4.5-based voltage security events classifier for electric power systems. International Journal of Engineering Systems Modelling and Simulation, Inderscience 13 (4), 268–276.

Kioumarsi, Mahdi, Dabiri, Hamed, AmirrezaKandiri, VisarFarhangi, 2023. "Compressive strength of concrete containing furnace blast slag; optimized machine learning-based models". Journal of Cleaner Engineering and Technology 13.

Meng, Xiangfei, Zhang, Pei, Zhang, Dahai, 2020. Decision tree for online voltage stability margin assessment using C4.5 and relief-F algorithms. Energies, MDPI publisher 13 (15), 1–13, 3824.

Nandanwar, S.R., Warkad, S.B., 2016. Classification of VSA by probabilistic fuzzy decision tree approach. Int. J. Control Theor. Appl. 9 (20), 115–122.

Nandanwar, S.R., Kolhe, M.L., Warkad, S.B., Patidar, N.P., Singh, V.K., 2018. Voltage security assessment by using PFDT and CBR methods in emerging power system. Energy Proc. 144, 170–181. Elsevier.

Pao, Y.H., 1989. Adaptive Patterns Recognition and Neural Networks". Addison-Wesley Longman Publishing Co., Boston, MA United States.

Power system test cases. http://www.ee.washington.edu/research/pstca/. (Accessed 15 April 2023).

Quinlan, J.R., 1986. Induction of decision trees. Mach. Learn. 1 (1), 81–106.

Srivastav, L., Singh, S.N., Sharma, J., 1998. Parallel self-organising hierarchical neural network based fast voltage estimation. IEEE Proceedings Gener. Trasm. and Distrib. 145 (No.1), pp98–104.

Wehenkel, L., Pavella, M., 1993. Decision tree approach to power system security assessment. Int. J. Electr. Power Energy Syst. 15 (1), 13–36. Feb.

Wood, J., Wollenberge, B.F., 1984. Power Generation, Operation and Control", second ed. John Wiley & Sons, Newyork.

Yang, C.-C., Hsu, Y.-Y., 1994. Estimation of line flows and bus voltages using decision trees. IEEE Trans. Power Syst. 9 (2), 1569–1574.

Zaborszky, J., Hwang, K.W., Prasad, K., 1980. Fast contingency evaluation using concentric relaxation. IEEE Trans PAS-99, 28–36.