

Risk assessment of power system transient instability incorporating renewable energy sources

Ayman Hoballah, Salah Kamal El-Sayed, Sattam Al Otaibi, Essam Hendawi, Nagy Elkalashy, Yasser Ahmed

Department of Electrical Engineering, College of Engineering, Taif University, Taif, Saudi Arabia

Article Info

Article history:

Received Dec 7, 2021

Revised Apr 7, 2022

Accepted May 5, 2022

Keywords:

Feature selection

Gaussian process regression

Renewable energy sources

Statistical analysis

Transient stability assessment

ABSTRACT

Transient stability affected by renewable energy sources integration due to reductions of system inertia and uncertainties associated with the expected generation. The ability to manage relation between the available big data and transient stability assessment (TSA) enables fast and accurate monitoring of TSA to prepare the required actions for secure operation. This work aims to build a predictive model using Gaussian process regression for online TSA utilizing selected features. The critical fault clearing time (CCT) is used as TSA index. The selected features map the system dynamics to reduce the burden of data collection and the computation time. The required data were collected offline from power flow calculations at different operating conditions. Therefore, CCT was calculated using electromagnetic transient simulation at each operating point by applying self-clearance three phase short circuit at prespecified locations. The features selection was implemented using the neighborhood component analysis, the Minimum Redundancy Maximum Relevance algorithm, and K-means clustering algorithm. The vulnerability of selected features tends to result great variation on the best features from the three methods. Hybrid collection of the best common features was used to enhance the TSA by refining the final selected features. The proposed model was investigated over 66-bus system.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Ayman Hoballah

Department of Electrical Engineering, College of Engineering, Taif University

P.O. Box 11099, Taif 21944, Saudi Arabia

Email: ayman.h@tu.edu.sa

1. INTRODUCTION

Fast transient stability assessment (TSA) enables the system operator to initiate the necessary remedial action for enhancing system security during abnormal operating conditions. The load variations and system topology changes as well as uncertain of generation levels of different renewable energy sources (RES) may impulse the system towards the stability boundary following large disturbances [1], [2]. TSA of large-scale power system depends on the synchronization among generators. However, it is necessary to continuously evaluate TSA during online operation to prevent serious electromagnetic oscillations. The computation burden of online TSA is a great challenge [3], [4]. Accurate TSA requires step by step time domain simulation (TDS) of large number of nonlinear equations at pre-fault and post-fault trajectories. TSA can be evaluated by monitoring the deviation between the rotor angles of synchronous generators which required handling with collected big data. The generators are considered out of step when the rotor angle deviations exceed the pre-specified limits following faults [5]. Therefore, the critical fault clearing time (CCT) is considered as accurate indicator of the system transient stability [6], [7].

Artificial intelligent and statistical analysis methods were applied to reduce the TSA computation time using phasor measurement units (PMUs) [8]. The collected information from PMUs was utilized to evaluate the deviation among rotor angles to design the out-of-step protection system avoiding system collapse [9]. In research of Gomez *et al.* [10], support vector machines were used to predict TSA using voltage magnitude measurements following faulty condition at balanced or unbalanced fault conditions. The decision tree (DT) along with artificial neural network (ANN) were implemented to specify the required counter-measurement to assess and enhance the power system transient stability. TDS was used to calculate the CCT and the power flow information where DT was used to estimate the CCT based on selected predictors to map the system dynamics. ANN was used to estimate the generating levels for economic dispatch considering the system stability or initiate a prespecified correction actions based on the historical or excremental information [11]. ANN was implemented for TSA by monitoring the rotor angle oscillations among generators. The inputs were the phase angle differences and their rate of variations [12]. The success of these schemes encourages the researchers to develop more systematic approaches for TSA tools to account the continuous variation of operating conditions, uncertainties associated with the RES generating levels and reduction in system inertia due to RES replacing traditional generating units [13]–[15].

In this paper, hybrid analytical method for TSA using predictive model based Gaussian process regression (GPR). Therefore, the random variation of generation and loads were considered. The method depends on the offline collected data from applying optimal power flow (OPF) for variety of operating conditions where feature selection algorithms were used to reduce the data dimension. The GPR predictive model is a nonparametric algorithm that depends on the calculation of the probability distribution over the assumptive possible functions to fit the input and output data. GPR has the capability of cub complex relationships by approximating the target function. GPR is being employed in many engineering applications.

2. PROPOSED METHOD AND CONTRIBUTION

The main steps of the proposed method for online TSA evaluation by using selected features based on GPR predictive model can be summarized as follow:

- Step 1: The description of the photovoltaic (PV) systems and wind system dynamic models within the DigSilent simulation software which was used to evaluate OPF calculations and the corresponding CCT according to set of contingencies. Therefore, large number of datasets were collected during offline.
- Step 2: Different features selection algorithms (the neighborhood component analysis, the Minimum Redundancy Maximum Relevance algorithm, and K-means clustering algorithm) were applied for data mining to select the best features to map the system dynamics for constructing of TSA predictive model. The results were compared to improve the accuracy of TSA predictive model.
- Step 3: GPR predictive model was built to estimate the CCT based on selected features. The GPR predictive model was trained offline based on the selected features to predict the CCT as indicator for TSA during online applications. The strong correlation between selected features and TSA indicator reflects the importance of system stability monitoring to move away from stability boundary. The GPR predictive model maps the relationship between the selected features and the CCT to predict the system state of stability based on new values of selected features. The process of evaluations can be summarized as follows: i) random variation of loads according to the expected loading levels and RES generation levels, ii) collect the Data by applying the OPF at each operating point to specify the generation rescheduling, iii) evaluate the minimum CCT as index for TSA each operating point following the expected set of contingencies (self-clearance three-phase short circuit at a preselected set of critical locations), iv) apply different feature selection algorithms (NCA, MRMR, and K-means algorithms) to select the best correlated features with TSA indicator, v) build GPR predictive models based on the selected features from three feature selection algorithms, and vi) evaluate the predictive models using performance indices to enhance the accuracy.

2.1. Transient stability assessment

TSA is influenced by the initial operating states as well as the severity of applied disturbance. The most sever contingency is the self-cleared three-phase short circuit which is used in this study [16]. The dynamic response of generators depends on the fault duration and location. The synchronization among generators is governed by the swings between generator rotor angles where the angular deviation between generators should not exceed the predefined accepted limit to consider the system stable. The CCT represents the minimum fault duration where the system remains stable without loss of synchronization following the clearance of fault. The first generator starts to out of step is called critical generator. The CCT is specified by the system operator according to the settings of protection system and dynamic behavior of generators which usually 150 or 200 milliseconds [17]. The fault duration beyond this limit makes the system loss of the ability

to preserve the system stability. The CCT can be calculated using TDS by increasing the fault duration till one of the generators out of synchronization. TDS solves the power system differential and algebraic equations using step by step calculations during pre-fault, during fault, and post-fault to simulate the system dynamics. If the applied fault makes the rotor angle deviation reaching this limit, the duration of the applied fault is called CCT which depends on the net kinetic energy of all generators and the produced electromechanical oscillations [16]. Fast online TSA tool enable the system operator to activate the required countermeasures to force the system to stable region. The application of fast TSA tools such as GPR predictive model is significantly reducing the required computation time as well as the burden of data collection. The minimum CCT was considered as 0.15 second in this study according to the commission regulation (EU) 2016/631 of 14 April 2016 which was established to specify the network code requirements for grid connection of generators [17]. Therefore, every generator should have CCT longer than the specified operating time limit of circuit breaker to avoid out of synchronism. DigSilent software is used to simulate the test system and perform the necessary calculations.

2.2. Data collection

The analytical investigation of online TSA was conducted using 66-bus test system in Figure 1. The system consists of 16-machine, 54-transmission line and 42 constant impedance loads [17]. The system is divided into three areas (A, B and C) connected through three double circuit tie lines. The system was developed to investigate several stability problems based on the relevant characteristic parameters of European power system. The test system was modifying by adding four RES stations at the tie lines connecting different areas. The original test system was designed to supply 16.516 gigavolt-ampere (GVA) total demand with power exchange of 1000 MVA from area A to each of area B and area C. RES stations were installed with 200 MVA wind system and 50 MVA PV system at the terminals of the tie lines connecting different areas.

Variety of operating points were collected by random varying of loads and applying of OPF within acceptable limits [17]. At each operating point, CCT was calculated using the electromechanical transient's evaluation. Table 1 presents the offline collected variables using OPF. Figure 2 shows the classification of system states with fault duration less than 0.5 second. Accordingly, if fault duration less than 150 millisecond leads to loss of synchronism, the system was considered as transiently unstable. The collected data was divided randomly into training set of 600 operating points and testing set of 150 operating points.

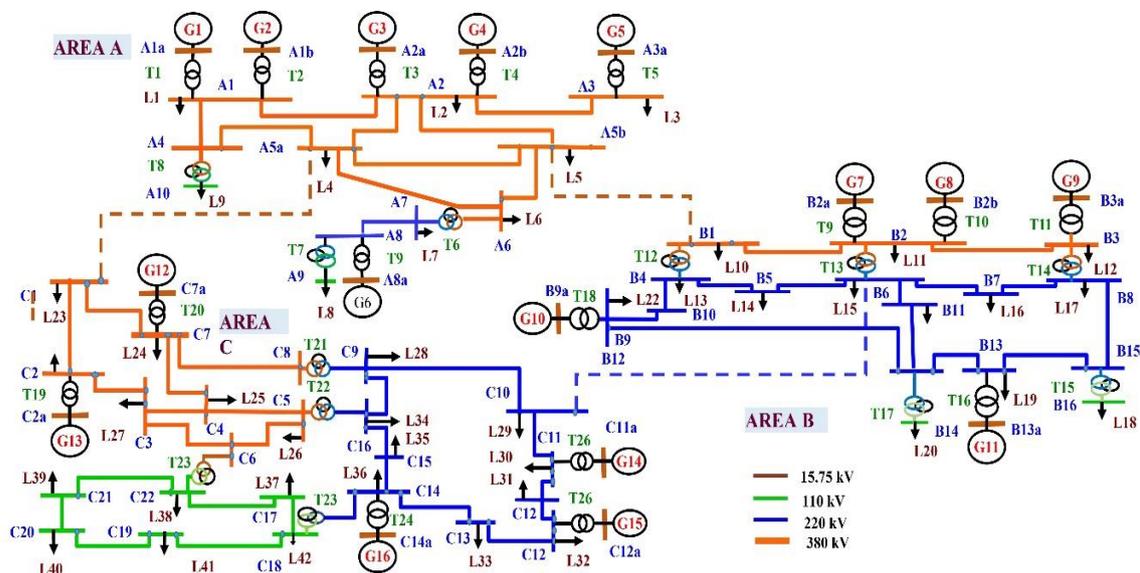


Figure 1. Single line diagram of 66-bus test system

2.3. PV system modelling

The PV array consists of many modules which are connected in series and parallel according to the desired power. The PV module consists of solar cells, the capacitor at direct current (DC) bus for voltage control, power electronic devices, integrated controller, and energy storage system. Figure 3 presents the block diagram of the PV composite model as described in DigSilent software [18].

Table 1. Offline collected variables using OPF

Variable	Name	No.	Variable	Name	No.
PQ-Area	Area active and reactive power	6	Tap-T	Transformer tap changer setting	28
PQ-G	Generator active and reactive power	32	PQ-Load	Load active and reactive power	84
PQ-Line	Lines active and reactive power	108	PQ-RES	Active and reactive power of RES	6
V-Buss	Bus voltage	132	Total number of variables		396

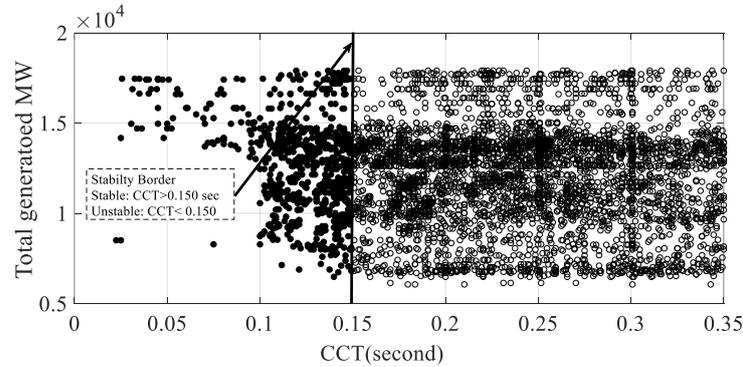


Figure 2. Classification of operating points stability using CCT

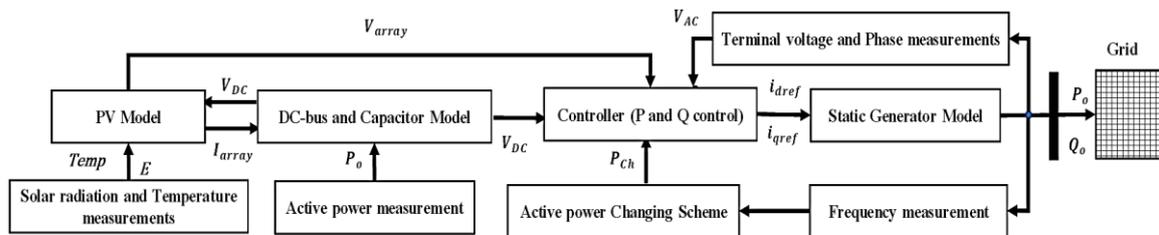


Figure 3. Composed model for PV system in DigSILENT

Solar cell is represented by an ideal single-diode module. The output current is defined as in (1):

$$I = I_{pv} - I_o \left[\exp\left(\frac{q(V+IR_s)}{k T a}\right) - 1 \right] - \frac{V+IR_s}{R_p} \tag{1}$$

where I_{pv} and I_o are photogenerated and saturation currents, V is terminal voltage, R_p and R_s are parallel and series resistances, q is electron charge, T is temperature, and a is the diode's ideality factor.

The output voltage is connected across DC-bus capacitance. the capacitor protects for the PV array during abnormal conditions providing isolation between the PV array and the grid. The charging and discharging process enables the capacitor to operate as energy storage device. This improves the ability of maximum power point tracking (MPPT) and active power control to inject the schedule output power from the PV system to the grid. The approximating linear prediction algorithm is used to evaluate the DC-link voltage for achieving MPPT operation [19]. The accomplished variation in dc-link voltage and control signal of frequency stability are used to control the d-component of reference current through PI-controller. The PI-controller is used to regulate the DC voltage across the capacitor terminals by comparing to the PV array reference voltage and the voltage across the capacitor terminals. The output of the PI controller is evaluated based on deviation of the array output voltage from the required DC voltage across the capacitor. Additional input signal can be added to compensate the fluctuation of grid frequency from the reference value. The output power injected to the grid through static generator which simulates the inverter behavior generating the AC signal based its input d-q components of reference current controlled signals ($i_{d,ref}$, $i_{q,ref}$). The static generator represents a current source model with output current (I_g) as in (2) at the grid voltage ($V_g=V_r+jV_i$) and frequency. They are synchronized with grid using the d-q reference angle.

$$I_g = (v_r * i_{d,ref}/|V_g| - v_i * i_{q,ref}/|V_g|) + j(v_i * i_{d,ref}/|V_g| + v_r * i_{q,ref}/|V_g|) \tag{2}$$

2.4. Wind system modelling

The doubly-feed induction generator (DFIG) is implemented within the DigSilent as static generator. Figure 4 presents the main parts of 6 MVA, 0.69 kV wind system which can be explained as follows: i) DFIG generates the power based on the input mechanical power and the controlled rotor voltage; ii) the rotor model includes the aerodynamic model of turbine, shaft, and pitch angle control models. The output mechanical power to DFIG depends on the wind speed and the specified value of reference speed; iii) compensation block calculates the rotor voltage based on the calculated rotor input current signal from current controller and output active and reactive power from DFIG. The current controller controls the output power and limits of rotor current and frequency deviation; iv) PQ control model evaluates the active and reactive reference currents for rotor side converter (RSC) to control the variation on rotor voltage according to the target output active and reactive power. The inputs are the terminal voltage, reference speed, over-frequency control signals and under-frequency control signals; and v) rotor protection inserts crow-bar circuit during faults and under-frequency controller. Current, voltage and frequency measurement devices are used for transformation into stator voltage-oriented reference frame.

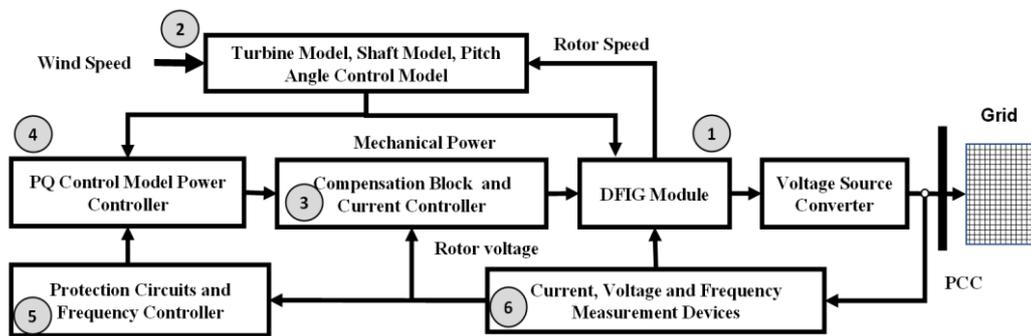


Figure 4. Block diagram of 6 MW wind turbine model in DigSilent

2.5. TSA based GPR protective model

The GPR predictive model was implemented to relate the selected features and the CCT for TSA. GPR is accurate prediction algorithm which was used for measuring the goodness of the selected features with the associated response in many applications [20]. GPR is used to predict the CCT using the selected features. GPR is a supervised learning algorithm which classifies the response using covariance relationship to represent the similarity of predictor. The *fitrgp* function is used to fit the GPR model in MATLAB software and is used in this study. Various standard kernel functions can be used to represent the effect of the response one point, x_i by other one, x_j . The default one in the *fitrgp* function is the squared exponential kernel function. The selection of the best kernel function was obtained iteratively to improve the prediction accuracy based on the correlation between selected features and the CCT. The accuracy of correct system state of stability prediction was measured using the relation between the number of correct assessments of system stability to the total number of operating points. The correct assessment was considered when the absolute deviation between predicted and calculated CCT less than 5 milliseconds. The obtained best results were obtained using the rational quadratic kernel function in (3). The collected 750 data sets were divided into a training set $T_1 = \{(x_i, y_i) | i = 1, 2, 3, \dots, M_1\}$, and testing set $T_2 = \{(x_i, y_i) | i = 1, 2, 3, \dots, M_2\}$.

$$k(x_i, x_j | \theta) = \sigma_f^2 \left(1 + (x_i - x_j)^T (x_i - x_j) / 2\alpha\sigma_l^2 \right)^{-\alpha} \quad (3)$$

Where, α is a positive scale-mixture parameter, σ_l is the characteristic length scale and σ_f is standard deviation.

Many methods used to reduce the number of variables by selecting the best correlated ones with TSA. Applying different method may determine different promising features from one algorithm to another as well as using the actual values rather than using normalized values [21], [22]. The performance evaluation of the GPR prediction model was performed in terms of indexes as presented in (4) to (6). The indexes include the accuracy of true classifying the system state into stable and unstable (N_{true}/N) with error less than 5 milliseconds in CCT prediction, the mean absolute error (MAE), root mean square error (RMSE), and the goodness of the regression based on the ratio of variation ($0 \leq R^2 \leq 1$). The closest R^2 to one is an indication to the healthy regression process.

$$MAE = \frac{1}{N} \sum_{k=1}^N \left| \frac{\bar{y}_k - y_k}{\bar{y}_k} \right| \tag{4}$$

$$RMSE = \sqrt{\sum_{k=1}^N (\bar{y}_k - y_k)^2 / N} \tag{5}$$

$$R^2 = 1 - \frac{RMSE}{\text{variance}(\bar{y})^2} \tag{6}$$

where \bar{y}_k is the predicted CCT value using GPR model and y_k is the calculated value.

3. RESULTS AND DISCUSSION

In this section, NCA, MRMR, and K-means clustering algorithms are used for data mining and features selection. Furthermore, GPR was used to build the corresponding three predictive model to select the best one. The results are compared with the TSA based on TDS results.

3.1. GPR based neighborhood component analysis features selection

The neighborhood component analysis (NCA) is a filter type feature selection algorithm which depends on the features' similarities and correlations [23]. NCA is considered as a robust feature selection technique which can be applied for features ranking and selection. The method works on the diagonal adaptation of NCA with regularization to minimize a loss function such as RMSE or MAE. The regularization term tends to reduce the weights of the irrelevant features to zero [24], [25]. Figure 5(a) presents the fitted values of CCT using the NCA algorithm relative to the actual values of CCT which explains the ability of NCA algorithm to predict the response values and presents fitted CCT and the actual values. The mean squared error as the measure of accuracy of the predict relative to the actual values of CCT is 0.005. Figure 5(b) presents the features weight where the small correlation features with the CCT have nearly zero weights.

The features with high weight factors were collected to be used in predictive model implementation. The selected 30 features which have high correlations with the CCT are presented in Table 2. The data sets of the selected 30 features and corresponding actual CCT were used to build GPR predictive model.

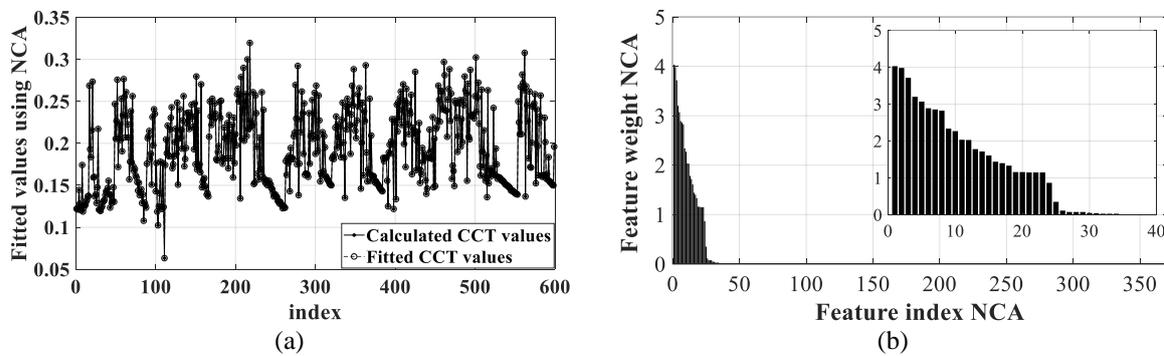


Figure 5. The fitted values of (a) CCT and (b) the features weight using NCA

Table 2. The selected 30 features using NCA algorithm.

Variable	High score 30 features by NCA	No
1	PQ-Area Sum PA-Sum QA-Sum QB-Sum PC-Sum QC	5
2	PQ-G Pg1- Pg2 - Pg3 -Pg4- Pg6-Qg8 -Pg8-Qg9-Pg10-Pg11- Pg12-Pg13	12
4	PQ-Line PA1-4/QA1-4/PA1-2/QA1-2/PA4a-5/PA2-5a/PA2-B5b/QB1-2/PB2-3/QB2-3/PB5-9/PB7-8/PC1-2	13
Total number of selected features		30

The performance evaluation of GPR is presented in Table 3. The results show the goodness of the regression process during the training process to classify the system states into stable or unstable correctly where the error between the predicted CCT relative to the calculated CCT is almost less than 2 milliseconds and standard deviation of 0.34 milliseconds as shown in Figure 6(a). Figure 6(b) presents the CCT obtained during the testing stage for 50 out of 150 unforeseen operating points. The GPR protective model was able to

classify the system state of 128 out of 150 operating point correctly. The results indicate the difficulty to exact mapping of system dynamics by the selected features at some points. The maximum error is 30 milliseconds with standard deviation of 6 millisecond. This is due to the sensitivity of the CCT for the variation in operating conditions. However, the GPR model classifies the system states into stable or unstable correctly where most relative values are in the same side from the border line of 150 millisecond.

Table 3. The performance indexes for GPR model based NCA evaluation

Data	%Acc	RMSE	R2	MAE
Training	100	0.0003	0.999	0.0011
Testing	85	0.0172	0.877	0.0126
All Data	96.3	0.0077	0.973	0.0113

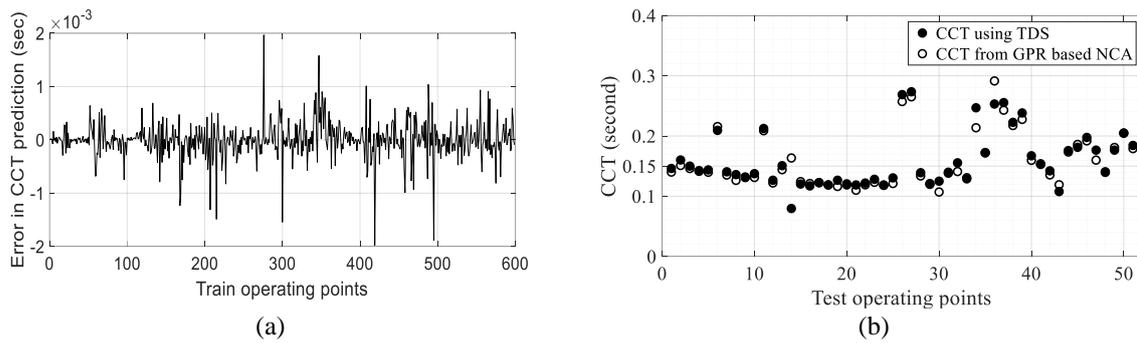


Figure 6. The error between calculated and predicted CCT using NCA during (a) training and (b) testing

3.2. GPR based minimum redundancy maximization algorithm features selection

The minimum redundancy maximum relevance (MRMR) algorithm is used to rank the features according to their importance with respect to the target response. The basic idea of MRMR depends on maximizing the mutual information which relate the different discrete variables. MRMR considers the mutual information between variables and the target classes of response. The mutual information represents the independence of any two variables and is defined as in (7) [26], [27];

$$I(X; Z) = \sum_{z \in X} \sum_{x \in X} p(x, z) \log \left(\frac{p(x, z)}{p(x)p(z)} \right) \quad (7)$$

where $p(x, z)$ represents the joint probability distribution function (PDF) of x and z . $p(x)$ and $p(z)$ are the marginal PDF of x and z respectively.

MRMR algorithm aims to maximize the relevance (V_x) between selected feature (x_i) and the target class C . The relevance value is defined as in (8).

$$V_x = I(x_i; C) \quad (8)$$

Maximizing the redundancy (Mutual distance between variables, W_x) of variable x_i with respect to the set of S variables enhances the classification or regression process. The redundancy is defined in (9).

$$W_x = \frac{1}{S} \sum_{x_i, x_j \in S} I(x_i, x_j) \quad (9)$$

The MRMR algorithm ranks the features using the mutual information quotient (MIQ) value which involves relevance maximizing and redundancy minimizing simultaneously as in (10).

$$\text{Max}_{x \in S} \text{MIQ}_x = \text{Max}_{x \in S} \left(\frac{V_x}{W_x} \right) \quad (10)$$

Figure 7 shows the features ranking based on their importance with respected to the CCT. The 30 features with higher score are selected and tabulated in Table 4. The ranking of the selected features shows the variation among the NCA and MRMR algorithms in features selection. The high ranked 30 features using MRMR algorithm were selected to build GPR predictive model.

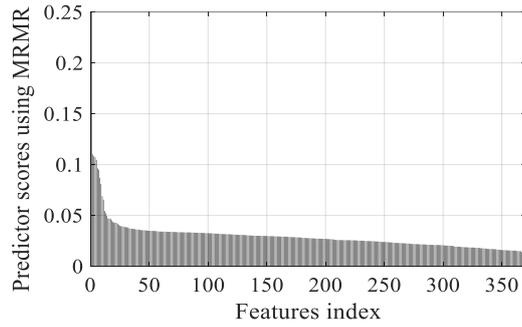


Figure 7. The score of features using MRMR algorithm

Table 4. The selected 30 features using MRMR algorithm

Variable	Higher weight 30 features by MRMR	No	
1	Sum PQ-Area	Sum PA-Sum QB-	2
2	PQ-G	Qg5-Qg6-Pg13	3
3	V-Bus	VmA3 -VmA6-VmA7-VmB1-VaB8-VmC6-VmC12	7
4	PQ-Line	QA1-4/QA1-2/PA4a-5/PA2-5a/QA5a-5b/PA6-7/QB1-2/QB2-5/QB6-C10/QB3-11/PC1-2/QC2-3/QC3-6/PC5-6/QC5-16/PC7-8/QC7-8/PC9-10/QC11-12/PC13-14	18
Total number of selected features		30	

Figure 8(a) presents the error in CCT prediction using GPR predictive model during training process. The model was able to predict the system state of all training data sets with error less than 2 milliseconds and standard deviation of 0.27 milliseconds. Figure 8(b) presents the CCT obtained during the testing stage for 50 out of 150 unforeseen operating points. The GPR protective model was able to classify the system state of 120 out of 150 operating point correctly. The results show that the GPR predictive model based NCA features selection algorithm was slightly accurate than the GPR model based MRMR selected features. The maximum error is 30 milliseconds at five operating points with standard deviation of 7.8 milliseconds. The variation is due to the strategy of each algorithm to pike up the preferred feature from each correlated group. Also, the reduction of the accuracy occurs due to the nonzero weights corresponding to the other features. Only six features were common between NCA and MRMR model. For more investigation, third feature selection-based K-means algorithm was used to classify the features into 30 clusters. The performance indexes of the GRP predictive model based MRMR is tabulated in Table 5.

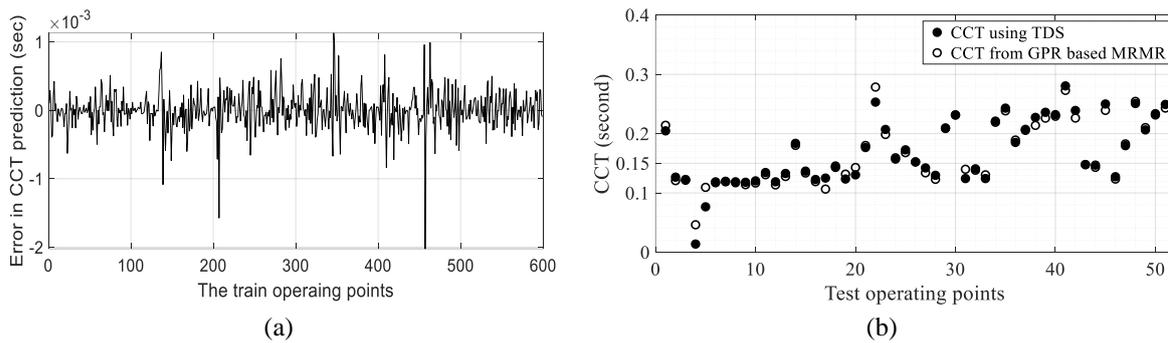


Figure 8. The error between calculated and predicted CCT using MRMR during (a) training and (b) testing

3.3. GPR based K-means clustering algorithm

K-means clustering algorithm is recursive, sequential, and heuristic search algorithms that add or/and remove features using selection criterion into subsets of variables [26]. K-means algorithm depends on the variable’s allocation into an arbitrary number of clusters based on the minimization of the average squared Euclidean distance between the centroid of the cluster and its observation. The allocation process is repeated iteratively to positioning the variables closed to the k centroids to separate variables into k clusters (groups). In this study, K-means algorithm was used to categorize the collected variables based on the

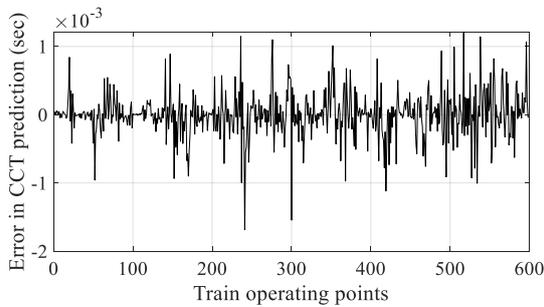
observations to 30 clusters containing the nearest variables to the centroids. The variables in each group have similar characteristics which keep the Euclidean distance to the corresponding centroid minimal. The separation depends on the actual data instead of the dissimilarity between each two variables. Therefore, the closest variables to the centroids are selected to represent the groups. The shortage of K-means clustering algorithm in features selection is that the clusters have the same distance from the centroids are treated equally and the selected best ones depends on the method of ranking not the correlation with CCT. Table 6 presents the selected 30 features using K-means algorithms. The selected 30 features were used to build GPR predictive model. The results show that there are 15 features are common between NCA and K-means features selection algorithms where only 6 features are common between MRMR and K-means features selection algorithms. Figure 9(a) shows the error in CCT for the GPR predictive model during training process. The maximum error of GPR model was 1.4 milliseconds with standard deviation of 0.3 milliseconds. Figure 9(b) presents the predicted CCT of unforeseen 50 operating points out of 150 test operating points using GPR model relative to the calculated CCT using TDS during the testing stages. The GPR model was able to predict 100 out of 150 operating points. The performance indices of the model for training, testing and all data sets are tabulated in Table 7. The performance indices are less quality than the performance of the GPR models based NCA and MRMR features selections.

Table 5. The performance indexes for GPR model based MRMR evaluation

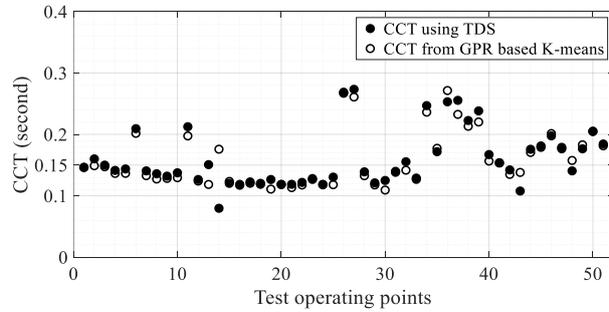
Data	%Acc	RMSE	R2	MAE
Training	100	0.0003	1.000	0.001
Testing	79.3	0.0176	0.871	0.1126
All Data	95.8	0.0079	0.971	0.0233

Table 6. The selected 30 features using *k-means* algorithm

Variable	Selected 30 features by K-means	No.	
1	Sum PQ-Area	Sum PA-Sum PB-Sum QB-Sum Pc-Sum Qc	3
2	PQ-G	Pg1 -Pg2- Pg3-Qg8 - Pg8 -Pg9-Pg12-Pg13-Qg13-Pg14-Pg15-Qg16	10
3	V-Bus	VmA4-VmB11-VmC5-VmC13	4
	PQ-Load	P_L10-P_L24-P_L20	3
4	PQ-Line	PA4a-5/PA2-5a/QB1-2/PB2-3/QC2-3/PC18-19	9
	Total number of selected features		30



(a)



(b)

Figure 9. The error between calculated and predicted CCT using K-means during (a) training and (b) testing

Table 7. The performance indexes for GPR model-based k-means evaluation

Data	%Acc	RMSE	R2	MAE
Training	99.7	0.0003	0.999	0.0011
Testing	66.34	0.0181	0.863	0.1398
All Data	92.2	0.0081	0.969	0.0288

3.4. Identification of hybrid features selection

To verify accuracy of the GPR predictive model, the results obtained from the three investigated features selection algorithm are compared with the accurately obtained results from TDS. Figure 10 presents the comparison between the CCT corresponding to randomly selected 20 operating points using TDS and the GPR predictive model based the 30 selected features using NCA, MRMR and K-means algorithms. The

results show the ability of GPR models to state the system stability. The investigation is investigated based on the selection of best features as well as testing the GPR predictive model using test operating points.

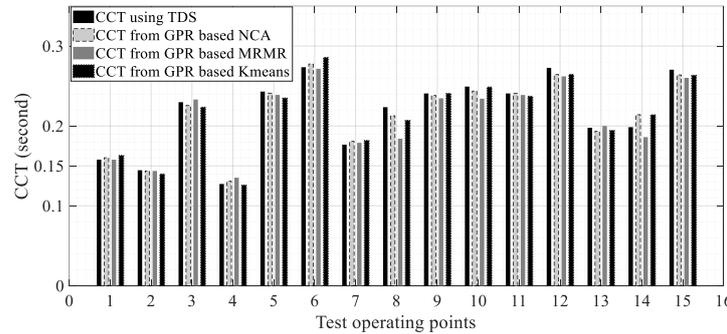


Figure 10. The CCT using TDS and GPR predictive models for 20 unforeseen operating points

The results from the three algorithms show that 6 features are common as presented in Tables 3, 5, and 7. The selected features depend on the used method of algorithms during ranking and correlation process. In this step, GPR predictive model was built using the common and best features from the three methods. The features with higher scores using NSA and higher weights using MRMR were collected sequentially and used to build new hybrid GPR models. The accuracy is enhanced from 83, 79.3 and 63.34 for NCA, MRMR and K-means respectively to 89.23% with 26 selected features during training process. The selected features with high accuracy are tabulated in Table 8. Table 9 displays the performance indices of the GPR model using the combined features from the three techniques where the number of the selected features are 26. The results show the enhancement not only in the testing data but also in all data sets. Therefore, the application of different feature selection algorithms can be used to discover hidden characteristics and correlation within collected big data.

Table 8. The selected features with high accuracy of the hybrid GPR predictive model

Variable	Common features	Features (NCA)	Features (MRMR)	No.
1 PQ-Area	Sum_PA -Sum QB	Sum PC		3
2 PQ-G	Pg13	Pg1-Pg2-Pg8-Qg8-Pg3-Pg12-Pg10	Qg5	9
3 PQ-Line	P_A4aA5-P_A2A5a-Q_B1B2	P_C1C2-P_A1A2	P_A6A7-Q_C2C3-Q_C5C16-Q_B2B5-Q_C11C12-P_C5C6	11
4 V-Buss			VmC6-VmA7-VmA6	3
Total number of selected features				26

Table 9. The evaluation of the hybrid GPR model based combined selection

% Acc	Training stage			% Acc	Testing stage			% Acc	All Data		
	RMSE	R2	MAE		RMSE	R2	MAE		RMSE	R2	MAE
100	0.0011	0.999	0.0026	89.23	0.0053	0.985	0.013	0.971	0.0028	0.996	0.0053

4. CONCLUSION

This work presents transient stability assessment of power system using analytical methods-based feature selection techniques. The effect of the RES was considered during data collection through random variation of load levels and the penetration level of RES. Minimum CCT is considered as indicator for TSA which represents the system dynamic stability following self-clearance three-phase faults at critical fault locations. GPR model was built for online monitoring of the TSA using group of selected features which can be collected using PMU units. The features were selected using NCA, MRMR and K-means algorithms. The application of the different feature selection algorithms presents different correlations between the selected features and CCT. The selection of the common features and the features with high correlations with CCT from different feature selection algorithms enhances the performance of the GPR model. The results show the high accuracy of the GPR predictive model (97.1%) to estimate CCT for TSA over a wide range on operating points. The proposed method can be used to build GPR predictive model for TSA in large scale power systems.

ACKNOWLEDGEMENTS

This research was funded by Deanship of Scientific Research, Taif University, grant no. 1-441-99.

REFERENCES

- [1] P. Kundur *et al.*, "Definition and classification of power system stability IEEE/CIGRE joint task force on stability terms and definitions," *IEEE Transactions on Power Systems*, vol. 19, no. 3, pp. 1387–1401, Aug. 2004, doi: 10.1109/TPWRS.2004.825981.
- [2] B. Wang, B. Fang, Y. Wang, H. Liu, and Y. Liu, "Power system transient stability assessment based on big data and the core vector machine," *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2561–2570, Sep. 2016, doi: 10.1109/TSG.2016.2549063.
- [3] M. Oluic, M. Ghandhari, and B. Berggren, "Methodology for rotor angle transient stability assessment in parameter space," in *2017 IEEE Power & Energy Society General Meeting*, Jul. 2017, pp. 1–1, doi: 10.1109/PESGM.2017.8274261.
- [4] D. R. Gurusingham and A. D. Rajapakse, "Post-disturbance transient stability status prediction using synchrophasor measurements," *IEEE Transactions on Power Systems*, vol. 31, no. 5, pp. 3656–3664, Sep. 2016, doi: 10.1109/TPWRS.2015.2496302.
- [5] X. He, L. Chu, R. C. Qiu, Q. Ai, and Z. Ling, "A novel data-driven situation awareness approach for future grids—using large random matrices for big data modeling," *IEEE Access*, vol. 6, pp. 13855–13865, 2018, doi: 10.1109/ACCESS.2018.2805815.
- [6] N. Yorino, A. Priyadi, H. Kakui, and M. Takeshita, "A new method for obtaining critical clearing time for transient stability," *IEEE Transactions on Power Systems*, vol. 25, no. 3, pp. 1620–1626, Aug. 2010, doi: 10.1109/TPWRS.2009.2040003.
- [7] J. D. Morales and J. V. Milanovic, "Methodology for optimal deployment of corrective control measures to ensure transient stability of uncertain power systems," *IEEE Transactions on Power Systems*, vol. 36, no. 3, pp. 1677–1687, May 2021, doi: 10.1109/TPWRS.2020.3035113.
- [8] R. Yousefian, R. Bhattarai, and S. Kamalasan, "Transient stability enhancement of power grid with integrated wide area control of wind farms and synchronous generators," *IEEE Transactions on Power Systems*, vol. 32, no. 6, pp. 4818–4831, Nov. 2017, doi: 10.1109/TPWRS.2017.2676138.
- [9] Y. Ohura *et al.*, "A predictive out-of-step protection system based on observation of the phase difference between substations," *IEEE Transactions on Power Delivery*, vol. 5, no. 4, pp. 1695–1704, 1990, doi: 10.1109/61.103664.
- [10] F. R. Gomez, A. D. Rajapakse, U. D. Annakkage, and I. T. Fernando, "Support vector machine-based algorithm for post-fault transient stability status prediction using synchronized measurements," *IEEE Transactions on Power Systems*, vol. 26, no. 3, pp. 1474–1483, Aug. 2011, doi: 10.1109/TPWRS.2010.2082575.
- [11] J. W. Shim, G. Verbic, and K. Hur, "Stochastic eigen-analysis of electric power system with high renewable penetration: Impact of changing inertia on oscillatory modes," *IEEE Transactions on Power Systems*, vol. 35, no. 6, pp. 4655–4665, Nov. 2020, doi: 10.1109/TPWRS.2020.3000577.
- [12] N. Amjady and S. F. Majedi, "Transient stability prediction by a hybrid intelligent system," *IEEE Transactions on Power Systems*, vol. 22, no. 3, pp. 1275–1283, Aug. 2007, doi: 10.1109/TPWRS.2007.901667.
- [13] I. B. Sulistiawati, A. Priyadi, O. A. Qudsi, A. Soeprijanto, and N. Yorino, "Critical clearing time prediction within various loads for transient stability assessment by means of the extreme learning machine method," *International Journal of Electrical Power & Energy Systems*, vol. 77, pp. 345–352, May 2016, doi: 10.1016/j.ijepes.2015.11.034.
- [14] M. Basu, V. R. Mahindara, J. Kim, R. M. Nelms, and E. Muljadi, "Comparison of active and reactive power oscillation damping with PV plants," *IEEE Transactions on Industry Applications*, vol. 57, no. 3, pp. 2178–2186, May 2021, doi: 10.1109/TIA.2021.3059810.
- [15] S. Harasis, H. Abdelgaber, Y. Sozer, M. Kisacikoglu, and A. Elrayyah, "A center of mass determination for optimum placement of renewable energy sources in microgrids," *IEEE Transactions on Industry Applications*, vol. 57, no. 5, pp. 5274–5284, Sep. 2021, doi: 10.1109/TIA.2021.3085674.
- [16] A. Hoballah and I. Erlich, "Online market-based rescheduling strategy to enhance power system stability," *IET Generation, Transmission & Distribution*, vol. 6, no. 1, pp. 30–38, 2012, doi: 10.1049/iet-gtd.2011.0265.
- [17] European Commission, "Commission Regulation (EU) 2016/631 of 14 April 2016 establishing a network code on requirements for grid connection of generators (Text with EEA relevance)," *Official Journal of the European Union*, 2016.
- [18] M. Z. Oskouei, B. Mohammadi-Ivatloo, O. Erdinc, and F. G. Erdinc, "Optimal allocation of renewable sources and energy storage systems in partitioned power networks to create supply-sufficient areas," *IEEE Transactions on Sustainable Energy*, vol. 12, no. 2, pp. 999–1008, Apr. 2021, doi: 10.1109/TSTE.2020.3029104.
- [19] W. Xu, C. Mu, and J. Jin, "Novel linear iteration maximum power point tracking algorithm for photovoltaic power generation," *IEEE Transactions on Applied Superconductivity*, vol. 24, no. 5, pp. 1–6, Oct. 2014, doi: 10.1109/TASC.2014.2333534.
- [20] M. Tsili and S. Papathanassiou, "A review of grid code technical requirements for wind farms," *IET Renewable Power Generation*, vol. 3, no. 3, p. 308, 2009, doi: 10.1049/iet-rpg.2008.0070.
- [21] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Computers & Electrical Engineering*, vol. 40, no. 1, pp. 16–28, Jan. 2014, doi: 10.1016/j.compeleceng.2013.11.024.
- [22] A. T. Eseye, M. Lehtonen, T. Tukia, S. Uimonen, and R. John Millar, "Machine learning based integrated feature selection approach for improved electricity demand forecasting in decentralized energy systems," *IEEE Access*, vol. 7, pp. 91463–91475, 2019, doi: 10.1109/ACCESS.2019.2924685.
- [23] W. Yang, K. Wang, and W. Zuo, "Neighborhood component feature selection for high-dimensional data," *Journal of Computers*, vol. 7, no. 1, pp. 162–168, Jan. 2012, doi: 10.4304/jcp.7.1.161-168.
- [24] D. Wang and X. Tan, "Bayesian neighborhood component analysis," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 7, pp. 3140–3151, Jul. 2018, doi: 10.1109/TNNLS.2017.2712823.
- [25] M. Alzaqebah, N. Alrefai, E. A. E. Ahmed, S. Jawameh, and M. K. Alsmadi, "Neighborhood search methods with moth optimization algorithm as a wrapper method for feature selection problems," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 4, pp. 3672–3684, Aug. 2020, doi: 10.11591/ijece.v10i4.pp3672-3684.
- [26] D. Wang, F. Nie, and H. Huang, "Feature selection via Global redundancy minimization," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 10, pp. 2743–2755, Oct. 2015, doi: 10.1109/TKDE.2015.2426703.
- [27] M.-S. Yang and K. P. Sinaga, "A feature-reduction multi-view k-means clustering algorithm," *IEEE Access*, vol. 7, pp. 114472–114486, 2019, doi: 10.1109/ACCESS.2019.2934179.

BIOGRAPHIES OF AUTHORS



Ayman Hoballah    received the B.Sc. and M.Sc. degrees in Electrical Engineering from Tanta University, Egypt in 1996 and 2003. Since 1998, he has been with the Electrical Power and Machines department, Faculty of Engineering, University of Tanta/Egypt. He completed his Ph.D. in electrical engineering department from the university Duisburg-Essen, Germany in 2011. His Ph.D. thesis focuses on the power system dynamic stability. Enhancement utilizing artificial intelligent techniques. His current reach interests include power system stability, DGs, smart grid, grounding systems and optimization techniques. He is an ass. professor at Taif University. Email: ayman.h@tu.edu.sa.



Salah K. Elsayed    was born in 1982 He obtained his B.Sc. and M.Sc. and PhD degrees from the Electrical Engineering Department, Faculty of Engineering- AL-Azhar University, Cairo, Egypt in 2005, 2009, and 2012 respectively. From 2012 to 2017, he was a lecturer with the Electrical Engineering Department AL-Azhar University, then he promoted to an associate professor, in 2017. He joined the Electrical Engineering Department, Faculty of Engineering, Taif University, KSA. His research interests include power systems analysis and operation, Power Systems Stability and control, power system optimization techniques, Artificial Intelligence Systems Applications in Power Systems and renewable energy sources. Email: sabdelhamid@tu.edu.sa



Sattam Al Otaibi    Department of Electrical Engineering, Taif University, Ta'if, Saudi Arabia. Sattam Al Otaibi was the Head of the public relation Center, Taif University, Saudi Arabia. He is a researcher and an academician specializing in electrical engineering and nanotechnology. His practical experience in the field of industry, education, and scientific research has been formed through his research work and through his mobility among many companies, institutions, and universities as well as active participation in research centers that resulted in much scientific research published in refereed scientific bodies. Email: srotaibi@tu.edu.sa



Essam Hendawi    was born in Egypt, 1968. He received his B.Sc., M.Sc., and Ph.D. degrees in Electrical Power and Machines Engineering from the Faculty of Engineering, Cairo University, Egypt, in 1992, 1998, and 2009 respectively. He is a Researcher in the Electronics Research Institute (ERI), Egypt, and he is currently working as an ass. Professor in the Electrical Engineering Department, College of Engineering, Taif University, Saudi Arabia. His research interests include electrical machine drives, converters, microcontrollers, and renewable energy. Email: essam@tu.edu.sa



Nagy I. Elkalashy    received the B.Sc. and M.Sc. degrees from the Electrical Engineering Department, Faculty of Engineering, Menoufia University, Shebin Elkom, Egypt, in 1997 and 2002, respectively. He received the Doctoral of Science in Technology (DSc. with Distinction) in Dec. 2007 from Helsinki University of Technology (TKK), Otaniemi, Finland. However, he was with Taif University, Taif, Saudi Arabia from 2015 to 2020. Currently, he is the head of the Electrical Engineering Department, Faculty of Engineering, Menoufia University. His research interests include protection, fault location determination, smart grids, system transients, HV engineering, switchgear technology, digital signal processing for power system applications. Email: n.elkalashy@tu.edu.sa



Yasser Ahmed    obtained his B.Sc. from Tanta University, Egypt, Faculty of Engineering, Electrical Power and Machines department in 1999. He received M.Sc. and Ph.D. from Cairo University, Egypt, Faculty of Engineering in 2006 and 2014, respectively. He works at Electronic Research Institute (ERI), Egypt, Power Electronics and Energy Conversion department since 2001. Currently, he is working at Electrical Engineering Department, College of Engineering, Taif University, Saudi Arabia. His major interests are electric drives, electric and hybrid electric vehicles, and modeling and simulation of electrical systems. Email: y.abdelsalam@tu.edu.sa.