Accurate Single-Phase Fault-Location Method for Transmission Lines Based on K-Nearest Neighbor Algorithm Using One-End Voltage

Mohammad Farshad, *Student Member, IEEE*, and Javad Sadeh, *Member, IEEE*

***Abstract*—In this paper, some useful features are extracted from voltage signals measured at one terminal of the transmission line, which are highly efficient for accurate fault locating. These features are the amplitude of harmonic components, which are extracted after fault inception through applying discrete Fourier transform on one cycle of three-phase voltage signals and then are normalized by a transformation. In this paper, the location of single-line-to-ground faults as the most probable type of fault in the transmission networks is considered. The SLG fault locator, which is designed based on the simple algorithm of k-nearest neighbor (k-NN) in regression mode, estimates the location of fault related to the new input pattern based on existing available patterns. The proposed approach only needs the measured data from one terminal; hence, data communication between both ends of the line and synchronization are not required. In addition, current signals are not used; therefore, the proposed approach is immune against current-transformer saturation and its related errors. Tests conducted on an untransposed transmission line indicate that the proposed fault locator has accurate performance despite simultaneous changes in fault location, fault inception angle, fault resistance, and magnitude and direction of load current.**

***Index Terms*—Fault location, Fourier transform, k-nearest neighbor, single-line-to-ground fault, transmission line.**

I. Introduction

ACCURATE fault locating of permanent and temporary faults is of high importance from the aspects of quick repairs and troubleshooting, identification of weak points of the transmission line, and adoption of required measures for decreasing fault occurrence probability at those locations [1]. Existing approaches for fault locating in transmission lines can be categorized into two main groups of approaches based on hard computing and soft computing. In recent years, there have been considerable efforts to improve and develop the fault-locating approaches based on hard computing using analytical models or traveling-wave theory. Parallel to these efforts, there has been a substantial amount of research on implementing soft computing techniques to solve the fault-location problem, due to the flexibility and capability of these techniques in facing the complexity of the problem. Learning machines are considered as the tools of performing soft computations in the fault-location problem. In the learning-based approaches, it is possible to train machine according to existing real patterns or patterns generated using reality-based simulation techniques. In this mode, the learning algorithm is responsible for the task of discovering hidden rules and complicated relationships between features of the patterns.

Selecting or extracting appropriate features and implementing an efficient learning algorithm are two pivotal issues in the fault-locating approaches based on machine learning. Low sensitivity to changes in the effective parameters, such as magnitude and direction of prefault current, fault resistance and fault inception angle, and high correlation with the fault location can be mentioned as some positive characteristics of the extracted features for fault locating. Furthermore, using only measured data from one end of the line can bring about a decrease in expenses resulting primarily from requirements for transmitting and synchronizing measured data of both terminals. In addition, among single-ended measurement data, voltage signals have some advantages over current signals. Using the current signals may be associated with a decrease in fault locating accuracy level caused by the existence of a significant dc component, saturation of current transformer, and high sensitivity of the extracted features to magnitude and direction of prefault current. In [2], an approach was presented for ground fault locating in transmission lines, which was based on high-frequency voltage transients measured at one end of the lines. The presented approach in [2] requires a very high sampling frequency and its fault locating accuracy decreases noticeably with a decrease in the sampling frequency.

Several learning tools and methods have been implemented for fault locating, including the multilayer perceptron neural network (MLPNN) [3]–[12]; radial basis function neural network (RBFNN) [13]–[15]; support vector machine (SVM) [16], [17]; extreme learning machine (ELM) [17]; Elman recurrent network [18]; fuzzy inference system (FIS) [19]–[22]; fuzzy neural network (FNN) [20]–[25]; and adaptive structural neural network (ASNN) [26]. These tools have a limitation on the dimension of input space, which can be trouble when a large number of potential features like high-frequency components of current or/and voltage signals are used. They are not practically capable of appropriate learning patterns with a high number of features due to enlargement of the structure and an extreme increase of the number of learning parameters [10]. On the other hand, dimension reduction using linear transforms or experimental selection of some features may result in elimination of some useful information. Using energies of limited

Manuscript received March 14, 2012; revised June 11, 2012; accepted July 27, 2012. Date of publication September 11, 2012; date of current version September 19, 2012. Paper no. TPWRD-00267-2012.

The authors are with the Electrical Engineering Department, Faculty of Engineering, Ferdowsi University of Mashhad, Mashhad 91779-48944, Iran (e-mail: m.farshad@ieee.org; sadeh@um.ac.ir).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TPWRD.2012.2211898

frequency bands out of the output spectrum resulting from applying Fourier transform [13], [23] or the application of energy and entropy of coefficients obtained from wavelet transform (WT) [10], [18] may conceal some useful information. Furthermore, applying principal components analysis (PCA), which was suggested in [26] and [27] for reducing the input space dimension and for identifying dominant patterns of processed signals by WT, is only appropriate for linear-separable data and can cause the elimination of existing nonlinear relationships between the features [28].

The single-line-to-ground (SLG) fault is the most probable type of fault, which can occur in transmission lines. Its occurrence probability at the voltage level of 300–500 kV is about 14 times more than that of the double-phase-to-ground fault [1].

In this paper, an approach based on machine learning is proposed for locating the SLG faults in transmission lines. The fault-locating patterns comprise features which are extracted from the single-ended voltage data. The extracted features do not require very high sampling frequency and in addition to low sensitivity to changes in the effective parameters, such as fault resistance, fault inception angle, and load current, they are highly dependent on the fault location. The harmonic components are extracted by applying discrete Fourier transform (DFT) on the three-phase voltage signals and are utilized for generating the input features of the fault locator. Since a wide range of frequency components is considered for generating the patterns, there are relatively large numbers of patterns' features. It is the first time, where such a large number of features have been used directly for fault locating. In this paper, the simple k-nearest neighbor (k-NN) algorithm is utilized in regression mode, which can manage the large number of patterns' features. The k-NN algorithm was implemented in regression mode for fault locating in radial distribution networks [29].

The rest of this paper is organized as follows. In Section II, a brief explanation is presented regarding the k-NN algorithm in regression mode. In Section III, the main ideas of the proposed approach for fault locating are explained. In Section IV, training and test patterns, which consist of various combinations of fault conditions in a study system, are generated. The simulations are carried out using PSCAD/EMTDC software [30]. After that, the proposed method is examined using the generated training and test patterns, and the fault locating test results are presented. The evaluation is followed by the conclusion in Section V.

II. Regressional k-NN Algorithm

The k-NN algorithm is a simple and efficient algorithm for estimating class or target variable of input patterns based on existing available patterns. In this algorithm, only saving the training patterns is required and the estimation is achieved locally based on some near patterns. The k-NN algorithm, despite its simplicity, can demonstrate suitable performance in the estimation of functions [29]. The k-NN estimator in regression problems can be defined based on the average of the target variables in k near neighborhood of the pattern in question, as the following equation [31]–[33]:

$$\widehat{f}(x') = \frac{1}{k} \sum_{x \in N(x')} f(x) \tag{1}$$

where $\widehat{f}(x')$ is the estimation of the target variable of the new pattern x', $f(x)$ is the target variable of the existing pattern x, and $N(x')$ is a set of k existing patterns in the near neighborhood of the new pattern x'. Generally, the patterns which are closest to the input pattern have higher importance in estimation of the target variable of that input pattern. Therefore, taking the average can be carried out based on the weights dependent on distances between the neighbors and the new input pattern, as the following equation [32]:

$$\widehat{f}(x') = \frac{\sum_{x \in N(x')} f(x) \cdot KL(d(x, x'))}{\sum_{x \in N(x')} KL(d(x, x'))} \tag{2}$$

where $d(x, x')$ is the distance between the existing pattern x and the new pattern x', and $KL(\cdot)$ is the weighting function or kernel function. The formulation of Euclidean distance and Gaussian function as one of the most common applicable weighting functions are expressed as follows [32]:

$$d(x, x') = ||x - x'||_2 = \sqrt{(x - x')^T (x - x')} \tag{3}$$

$$KL(d) = \exp(-d^2). \tag{4}$$

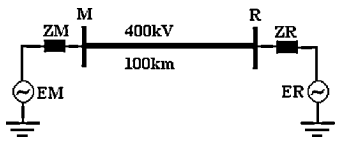

Fig. 1. Schematic of the transmission line in the system under study.

III. Main Idea and Generalities

A. Harmonic Components of Voltage Signals

Fig. 1 indicates the schematic of a 400-kV, 100-km-long untransposed double-end fed single-circuit transmission line. The source impedance of measuring and remote ends for the system frequency of 50 Hz are $\mathrm{Z_M} = 0.469 + \mathrm{j}6.283\ \Omega$ and $\mathrm{Z_R} = 0.391 + \mathrm{j}5.215\ \Omega$, respectively. The specification data of the transmission line, which are simulated under the frequency-dependent model, are presented in Table IX. The sampling frequency is adopted to be 80 kHz. The Nyquist criterion dictates the use of a low-pass antialiasing filter for the data-acquisition system to avoid aliasing.

Fig. 2 shows three-phase voltage signals for an SLG fault in about 0.163 s and at 25 km from the measuring end. As can be seen in the voltage signals after fault inception, there are some transients that result from fault-induced traveling waves.

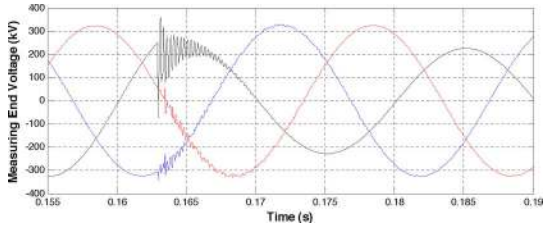

Fig. 2. Three-phase voltage signals measured at the measuring end under the SLG fault (A-G) at 25 km.

These fault-generated transient signals can offer useful information about the fault location.

In this paper, harmonic components of voltage signals are extracted by applying DFT on one cycle of data after fault inception and are used for fault locating. A transform with the following equation is applied for normalizing and producing more efficient features [34]

$$V'(f_i) = \frac{V(f_i) - \frac{1}{H}\sum_{j=2}^{H} V(f_j)}{\max_{f_p}\left(\left|V(f_p) - \frac{1}{H}\sum_{j=2}^{H} V(f_j)\right|\right)}, \quad i = 2, 3, \ldots, H. \quad (5)$$

In transform (5), H is the selected maximum harmonic order, and p is an index which can take values from 2 to H. In addition, $V(f_i)$ and $V'(f_i)$ are the amplitude of the harmonic order at frequency f_i before and after transformation, respectively. The harmonic orders of the phase voltage signal from second up to Hth are considered in the transformation. It is worth noticing that this transformation is carried out on the harmonic spectrum obtained from each phase signal separately.

Figs. 3 and 4 illustrate normalized harmonic spectra (from 2nd to 300th orders) of the voltage signal of faulted phase under SLG faults (A-G) at distances of 25 km and 75 km from the measuring end, respectively. In Figs. 3(a) and 4(a), the fault inception angle is assumed to be 45° and the fault resistance (R_f) is changed from 0.01 to 30 Ω. Furthermore, in Figs. 3(b) and 4(b), while the fault resistance is taken 10 Ω, the fault inception angle (FIA) varies from 18 to 90°. As can be seen in Figs. 3 and 4, the normalized harmonic spectra follow a unique pattern for each specific fault location. Moreover, it can be observed that these patterns have nearly low sensitivity to variations of fault resistance and fault inception angle.

B. k-NN Fault Locator

The SLG fault locator is designed based on the k-NN algorithm. Each of the patterns includes the normalized harmonic spectra of voltage of the faulted and healthy phases. The harmonic spectra are obtained through application of DFT on one cycle of the measured voltage signals after fault inception. The fault location related to each new pattern can be estimated through the k-NN algorithm based on existing patterns. The general procedure of the method is presented in Fig. 5.

If the line is untransposed, then separating k-NN fault locators should be considered for each type of SLG fault (A-G, B-G, and C-G) for obtaining better accuracy. The time of fault signature appearance at the measuring end and the type of fault are

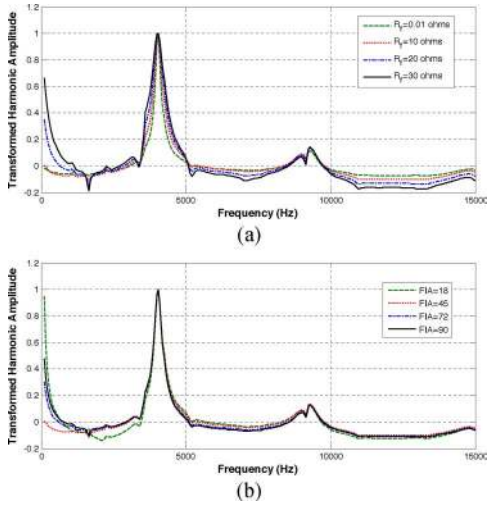

Fig. 3. Normalized harmonic spectra of faulted phase voltage under the A-G fault at a distance of 25 km. (a) Different fault resistances. (b) Different fault inception angles.

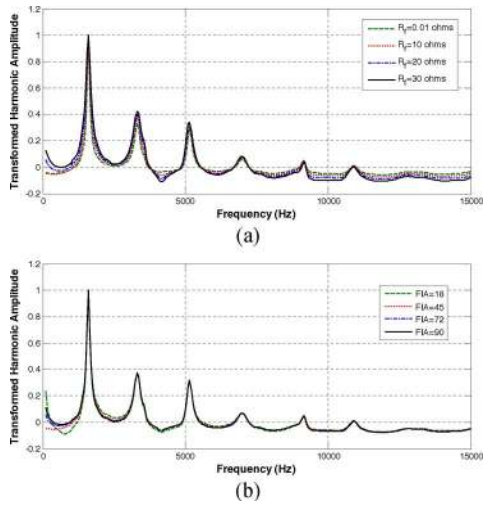

Fig. 4. Normalized harmonic spectra of the faulted phase voltage under the A-G fault at a distance of 75 km. (a) Different fault resistances. (b) Different fault inception angles.

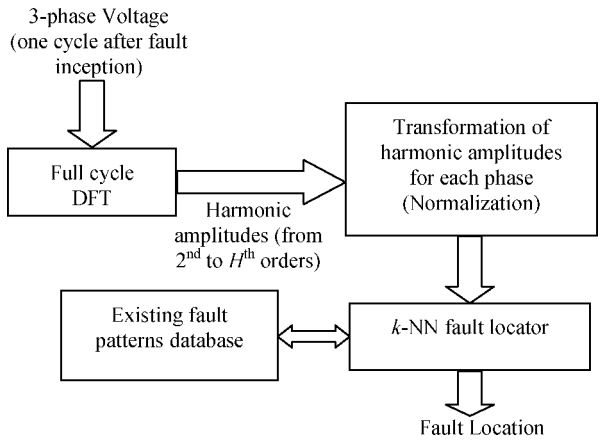

Fig. 5. General procedure of the proposed method for fault locating.

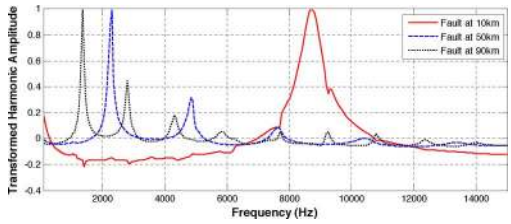

Fig. 6. Normalized harmonic spectra of faulted phase voltage under the A-G fault at distances of 10, 50, and 90 km.

considered as known information which can be obtained via a variety of methods proposed in this field. Based on the methods like the ones presented in [35] and [36], the fault transients can be detected within a one-sample delay.

C. Determination of Maximum Frequency Level

For the selection of desired maximum frequency level, the harmonic spectrum of voltage under the fault at a point close to the measuring end should be inspected. Fig. 6 indicates the normalized harmonic spectra of voltage of the faulted phase under the A-G fault at distances of 10, 50, and 90 km from the measuring end. As can be observed in this figure, the maximum amplitude of harmonic components related to the fault located at 10 km occurred in a higher frequency than those of the two other fault locations. Therefore, considering Fig. 6, it seems that in the system under study, the harmonic components up to frequency level of 10 kHz $(H = 200)$ are sufficient for generating efficient features for fault locating.

IV. Numerical Study

Simulations are carried out regarding the system of Fig. 1 through PSCAD/EMTDC software [30]. Here, the A-G type is adopted out of a variety of SLG faults. The Matlab environment is used for application of DFT and for implementation of the k-NN algorithm.

A. Generation of Training and Test Patterns

Training and test patterns are generated through changes of fault location, fault resistance, fault inception angle, and magnitude and direction of the prefault load current. Conditions for generating the patterns are based on a mixture of various states of the A-G fault incident. These conditions for training and test patterns are accessible in Tables I and II, respectively. According to the analysis carried out in Fig. 6, harmonic components up to 200th order are considered for generating the patterns. Hence, regarding the normalized spectra of three-phase voltages (from 2nd to 200th orders), each pattern contains 597 features. Based on the conditions of Table I, a total number of 8050 training patterns are generated, and according to Table II, a total number of 3840 test patterns are generated.

B. Parameter Adjustment for the k-NN Fault Locator

The only parameter of the introduced k-NN algorithm is the number of near neighborhoods (k) used for the estimation of target variables. For adjusting this parameter, first, the existing training patterns are divided randomly into 10 subsets. Then, the

TABLE I
Generation Conditions of Training Patterns

Fault Location	10% to 90% of the line length, with step 0.5%
Fault Inception Angle (Reference: Voltage of phase A at the fault point) (Degree)	4.5, 9, 18, 36, 72, 108, 144, 162, 171, 175.5
Fault Resistance (Ω)	0.01, 10, 30, 50, 100
Power Flow Angle (Degree)	20

TABLE II
Generation Conditions of Test Patterns

Fault Location	20 Random places
Fault Inception Angle (Reference: Voltage of phase A at the fault point) (Degree)	6.75, 13.5, 22.5, 54, 90, 126, 157.5, 173.25
Fault Resistance (Ω)	2, 15, 20, 40, 60, 80
Power Flow Angle (Degree)	-30, -10, 10, 30

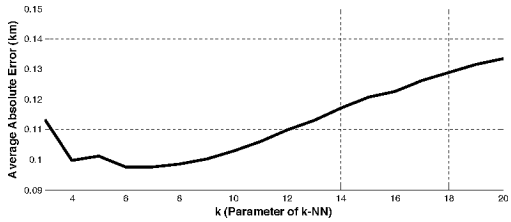

Fig. 7. Mean absolute error of fault locating in terms of k neighbors.

k-NN algorithm is repeated 10 times for each k value; in such a way that in each execution, one of the subsets is adopted as the test patterns set and all of the remaining subsets are considered as a training pattern set (10-fold cross-validation). Fig. 7 indicates the mean absolute error of fault locating under different values of parameter k. Based on the results presented in this figure, parameter k is selected to be 6.

C. Results and Discussion

The fault locations related to the test patterns generated based on the conditions of Table II are estimated by the k-NN algorithm and using the available training patterns. Absolute error in estimating the fault location can be calculated as

$$e = |L_{\text{est}} - L_{\text{act}}| \tag{6}$$

where e is the fault location absolute error, L_{est} is the estimated fault distance from the measuring end, and L_{act} is the actual fault distance from the measuring end. The percentage of error in estimating the fault location based on the total line length can be calculated as

$$e\% = \frac{|L_{\text{est}} - L_{\text{act}}|}{L_T} \times 100 \tag{7}$$

where L_T is the total length of the transmission line. The obtained error values by using (6) can be adopted in terms of percentage, since the length of the transmission line under study is 100 km. The results of fault locating for different fault distances are shown in Table III. There are a total number of 192 test patterns for each fault distance. By examining the results presented in Table III, it can be concluded that the proposed approach has sufficient accuracy with respect to the fact that the generation

TABLE III
RESULTS OF A-G FAULT LOCATING FOR DIFFERENT FAULT DISTANCES

Fault location (km)	Min error (km)	Max error (km)	Average error (km)	Fraction of prediction errors >0.5 km
10.70	0.0144	0.2036	0.0796	0%
13.25	0.0015	0.2356	0.0596	0%
16.30	0.0034	0.2000	0.0599	0%
21.75	0.0000	0.2500	0.0556	0%
26.20	0.0035	0.2100	0.0647	0%
31.80	0.0170	0.2705	0.0676	0%
39.30	0.0203	0.2000	0.0829	0%
45.25	0.0002	0.4621	0.0666	0%
49.80	0.0141	0.4146	0.1214	0%
53.40	0.0118	0.3474	0.0788	0%
55.75	0.0002	0.2424	0.0406	0%
57.30	0.0248	0.2877	0.0709	0%
60.80	0.0219	0.2355	0.0909	0%
63.25	0.0002	0.2500	0.1269	0%
68.75	0.0005	0.3252	0.0553	0%
71.80	0.0311	0.3931	0.1109	0%
77.20	0.0100	0.2000	0.0928	0%
80.75	0.0012	0.3056	0.1047	0%
87.30	0.0261	0.2856	0.0740	0%
89.25	0.0001	0.2500	0.0630	0%
All	0.0000	0.4621	0.0783	0%

TABLE IV
RESULTS OF A-G FAULT LOCATING FOR DIFFERENT FAULT INCEPTION ANGLES

Fault Inception Angle (Degree)	Min error (km)	Max error (km)	Average error (km)	Fraction of prediction errors >0.5 km
6.75	0.0005	0.4146	0.0917	0%
13.5	0.0001	0.3285	0.0827	0%
22.5	0.0002	0.2500	0.0728	0%
54	0.0002	0.2500	0.0684	0%
90	0.0000	0.2500	0.0732	0%
126	0.0002	0.2500	0.0681	0%
157.5	0.0001	0.2411	0.0663	0%
173.25	0.0002	0.4621	0.1035	0%

conditions of the test patterns have differed from those of the training patterns. The average prediction errors for all of the test patterns are equal to 78.3 m; in other words, it is equivalent to 0.0783%. It is worth noting that the error value did not exceed 0.5 km in any case.

For investigating the effect of important parameters on the accuracy, the related fault-location results as a function of fault inception angle, fault resistance, and load current are presented in Tables IV–VI, respectively. For each fault inception angle, fault resistance, and load current, there are a total number of 480, 640, and 960 test patterns, respectively.

Considering Table IV, it can be concluded that the prediction accuracy of the fault-locating approach at fault inception angles close to zero has a slight decline. However, it shows remarkable accuracy at other angles. At the worst case scenario, at angle 173.25°, the average prediction error is equal to 103.5 m.

Through examination of Table V, it can be noticed that the maximum prediction error occurred in the fault resistance equivalent to 2 Ω. However, at this value of fault resistance, the average prediction error is less than 125 m.

TABLE V
RESULTS OF A-G FAULT LOCATING FOR DIFFERENT FAULT RESISTANCES

Fault Resistance (Ω)	Min error (km)	Max error (km)	Average error (km)	Fraction of prediction errors >0.5 km
2	0.0003	0.2500	0.1245	0%
15	0.0001	0.4146	0.0615	0%
20	0.0003	0.3277	0.0690	0%
40	0.0001	0.3931	0.0571	0%
60	0.0003	0.2769	0.0662	0%
80	0.0000	0.4621	0.0917	0%

TABLE VI
RESULTS OF A-G FAULT LOCATING FOR DIFFERENT LOAD CURRENTS

Power Flow Angle (Degree)	Min error (km)	Max error (km)	Average error (km)	Fraction of prediction errors >0.5 km
-30	0.0001	0.4621	0.0787	0%
-10	0.0001	0.4618	0.0783	0%
10	0.0000	0.4613	0.0785	0%
30	0.0001	0.4610	0.0778	0%

TABLE VII
RESULTS OF A-G FAULT LOCATING USING DIFFERENT FREQUENCY LEVELS

Used voltage harmonic orders	Selected Parameter k	Min error (km)	Max error (km)	Average error (km)	Fraction of prediction errors >0.5 km
2^{nd} to 100^{th} order	4	0.0000	2.0494	0.1327	3.7240%
2^{nd} to 150^{th} order	6	0.0000	1.0634	0.0931	0.9375%
2^{nd} to 200^{th} order	6	0.0000	0.4621	0.0783	0%
2^{nd} to 250^{th} order	7	0.0001	0.5048	0.0792	0.0781%
2^{nd} to 300^{th} order	7	0.0001	0.5031	0.0777	0.0521%

According to the results presented in Table VI, it can be inferred that the proposed approach has approximately no dependence on the magnitude and direction of prefault current.

D. *Effect of the Selected Number of Harmonics*

As mentioned previously, harmonics up to the 200th order are adopted for generating the patterns. For ensuring the validity of this selection, the generation of the patterns is considered based on harmonics up to 100th, 150th, 250th, and 300th order as well. The related fault-locating test results are shown in Table VII for comparison. As can be observed from this table, using harmonics higher than the 200th order causes no major changes in prediction accuracy of the fault-locating approach. Notwithstanding, using fewer harmonics reduces the fault-location prediction accuracy to some extent. The results of Table VII confirm the validity of selecting 10 kHz as the maximum level of harmonic frequencies for the system under study, while the minimum fault distance of 10 km is considered.

E. *Effect of Instrument Transformer*

Recently, resistive-capacitive voltage transformers (RCVTs) are made available for high-voltage levels from 24 to 765 kV. RCVTs have an adequate frequency characteristic and their gain of frequency response is close to unity in the wide range of frequencies of interest [37]. Using these devices, the proposed

method can be implemented without concern for the distortion of measured voltage harmonic spectra, which may exist when using conventional capacitive voltage transformers.

Some analyses are performed considering possible harmonic spectra measurement errors. Before the normalizing process, $\pm 10\%$ random measurement errors are applied to the voltage harmonic spectra of the test cases, and the fault locations are estimated through application of the proposed method. The obtained overall average estimation error is 0.0906%, which does not show a significant increase compared to the accurate measurement mode of 0.0783%.

F. Fault Inception Angles Close to the Zero Crossing Point

The fault-locating approach is examined for fault inception angles close to the voltage zero crossing point. Some new test patterns are generated based on the fault inception angles of 2.25° and 177.75°. The other conditions for generating these test patterns are based on Table II. The obtained average and maximum errors for the fault inception angle of 2.25° are 0.2018% and 0.6333%, respectively. Also, 0.2911% and 2.6403% are the results for the average and maximum errors of the cases with the fault inception angle of 177.75°. It can be seen that approaching the zero crossing point of voltage at the moment of fault inception results in a decrease in prediction accuracy of the fault-locating approach.

If an SLG fault occurs when the voltage of the faulted phase at the fault location approaches zero, then the voltage signals measured at the terminal will not be rich in terms of harmonic contents. Under such circumstances, the proposed approach may not provide desirable performance. The examinations show that in the system under study, if the fault inception angle distance from the zero crossing point of the fault-location voltage is less than about 2.25°, then the proposed approach will not exhibit desirable performance. If these areas are taken in 360° of a cycle, their lengths add up to 9°. Consequently, fault occurrence probability in these areas is 2.5%. In other words, the proposed approach demonstrates acceptable performance 97.5% of the time.

G. Importance of Using Healthy Phases' Data

To investigate the importance of using the healthy phases' data, the training and test patterns are reconstructed based on the normalized voltage harmonic components of only faulted phase and only healthy phases. A comparison of the related fault-location results shows that faulted phases' spectrum suffices for accurate locating of the faults occurring in relatively high instantaneous voltage magnitudes; and healthy phases' spectra are more useful in cases with fault inception angles close to the zero crossing point. Therefore, simultaneous use of faulted and healthy phases' data gives an appropriate performance for a wide range of cases.

H. Fault Locations Close to Terminals

Some new train cases are simulated with the same conditions of Table I, except for the fault location which is changed from 1% to 9.5% and from 90.5% to 99% of the line. Some new test cases are simulated, with 12 random places in the first and last 10 km of the line as fault locations, while the other conditions follow Table II. The new train and test patterns are added primarily to train and test sets, and the proposed method is applied to estimate the fault location of the test patterns. The results of fault locating for the new test patterns for different fault distances are presented in Table VIII. As can be observed from this table, the estimation errors for the faults of the last 10 km of the line do not increase significantly, except for the faults occurring at distances very close to the remote bus. On the other hand, for faults at the first 10 km, the estimation errors are much higher for the fault locations close to the measuring end. Despite the fault-location estimation errors being in a relatively acceptable range, if the maximum frequency level used in generating the patterns is increased up to about 80 kHz, which is obtained from the inspection of voltage spectrum of a fault at 1 km via a higher sampling rate, lower errors are expected for the close-in faults.

TABLE VIII
RESULTS OF A-G FAULT LOCATING FOR DIFFERENT FAULT DISTANCES AT THE FIRST AND LAST 10 km OF THE LINE

Fault location (km)	Min error (km)	Max error (km)	Average error (km)	Fraction of prediction errors >0.5 km
1.20	0.0346	1.7970	0.4482	35.4167%
1.65	0.0137	1.8458	0.3568	22.3958%
4.35	0.0135	1.6937	0.3468	29.1667%
5.75	0.0009	1.9375	0.2853	16.6667%
8.20	0.0084	0.4406	0.1620	0%
9.25	0.0001	0.1545	0.0513	0%
91.75	0.0000	0.4432	0.0976	0%
93.30	0.0196	0.3850	0.1045	0%
94.25	0.0000	0.3280	0.1032	0%
96.80	0.0011	0.3010	0.1047	0%
98.10	0.0119	0.3507	0.1516	0%
98.80	0.0317	0.6401	0.1828	6.2500%

I. Effect of Source Impedances

Based on the conditions of Table I, some new train cases are simulated, while Z_M is decreased by 16.5% and Z_R is increased by 20%. Then, the new train patterns are primarily added to the train patterns set. Considering the conditions of Table II, some new test cases are simulated, while Z_M is decreased by 10% and Z_R is increased by 10%. The overall average of prediction error obtained through applying the method to the new patterns is 0.1746%, which shows an increase compared to the condition of unchanged source impedances, 0.0783%. Despite the fault-location errors being in a fairly acceptable range, it can be concluded that the method is relatively sensitive to changes of source impedances.

J. Highlighting Distinctive Aspects

In the proposed approach, the accurate SLG fault-locating algorithm is conducted using single-ended voltage measurements. In most of the existing methods, application of one-end or two-end current and voltage data were suggested, which are associated with problems, such as error occurrence resulting from saturation of CTs, or problems and expenses caused by transmitting and synchronizing measured data of both ends of the line. In addition, in the proposed method, there is no need for very high sampling frequency. In [2], an approach was

TABLE IX
SPECIFICATION DATA OF THE TRANSMISSION LINE UNDER STUDY

Conductors configuration	3 conductors, Triangular
Height of lowest conductors	32.8 m
Vertical distance of centre conductor above outer conductors	1.65 m
Horizontal spacing between phases	9.15 m
Conductor name	pheasant
Conductor geometric mean radius	0.0175514 m
Conductor DC resistance	0.0442913 Ω/km
Number of sub-conductors in a bundle	2
Bundle spacing	0.4572 m
Number of ground wires	2
Ground wire name	1/2" High Strength Steel
Ground wire radius	0.0055245 m
Ground wire DC resistance	2.8645 Ω/km
Height of ground wires above lowest conductor	7.2 m
Spacing between ground wires	13.6 m
Ground resistivity	60 Ω.m

suggested for ground fault locating in transmission lines, which was based on high-frequency transients of voltage signals measured at one end of line. The method proposed in [2] requires very high sampling frequency and its locating accuracy decreases with decreasing sampling frequency. The fault-locating result of approach [2] for SLG faults at the middle of a 230-kV, 100-km-long transmission line in sampling frequencies of 6700, 2200, 1300, 1000, and 670 kHz was associated with the average estimation errors of 0.360%, 1.440%, 2.610%, 3.890%, and 6.700%, respectively. In contrast, for implementing the novel proposed method in the study system, regarding the implemented harmonic components up to a frequency level of 10 kHz ($H = 200$), the lower band of the required sampling frequency is 20 kHz based on the Nyquist sampling theorem. In this case, the average fault-location prediction error for the primarily generated test patterns is less than 0.080%. Furthermore, when the maximum frequency level of used harmonic components is lowered to 5 kHz ($H = 100$), while the lower band of required sampling frequency is 10 kHz, the average prediction error for the test patterns is less than 0.135%. Moreover, the novel-proposed method has lower fault-locating errors in comparison with recently published machine-learning-based methods, which use current and voltage signals measurements, such as the valuable one proposed in [26].

None of the usual learning tools and methods used for fault locating, such as artificial neural networks, is practically capable of appropriate learning of the patterns due to the high number of used features (597 features for the system under study); in addition, the selection of structure and/or optimal adjustment of parameters are the challenges in most of the machine-learning algorithms; but they are not acute problems in the k-NN algorithm. The used k-NN algorithm has only one parameter which can be set easily. The main reason for the good performance of k-NN as the simplest learning method is high performance of the selected and generated patterns' features.

V. CONCLUSION

In this paper, some useful features are extracted for SLG fault locating in transmission lines using the harmonic spectrum of just one-end voltage, which has a low sensitivity to the changes in influential parameters, such as fault resistance, fault inception angle, and magnitude and direction of prefault load current. Furthermore, the generated features show a high level of correlation with the fault location, which leads to accurate fault locating. In the suggested method, the simplest form of learning algorithm k-NN is implemented, which demonstrates suitable capability despite the large number of patterns' features. The sole utilization of single-ended voltage measurements and no need for transmitting and synchronizing measured data of both ends of the line, no need for current measurements, and subsequently no measuring errors caused by current transformers, no need for very high sampling frequency, and, finally, simplicity of the utilized learning algorithm are some of the salient advantages of the proposed approach. The results of tests conducted on the untransposed transmission line in the study system are really promising despite variations in influential parameters on fault locating.

REFERENCES

[1] M. M. Saha, J. Izykowski, and E. Rosolowski, *Power Systems-Fault Location on Power Networks*, 1st ed. Berlin, Germany: Springer-Verlag, 2010.

[2] R. Mardiana, H. Al Motairy, and C. Q. Su, "Ground fault location on a transmission line using high-frequency transient voltages," *IEEE Trans. Power Del.*, vol. 26, no. 2, pp. 1298–1299, Apr. 2011.

[3] A. J. Mazon, I. Zamora, J. F. Minambres, M. A. Zorrozua, J. J. Barandiaran, and K. Sagastabeitia, "A new approach to fault location in two-terminal transmission lines using artificial neural networks," *Elect. Power Syst. Res.*, vol. 56, no. 3, pp. 261–266, Dec. 2000.

[4] Z. Chen and J.-C. Maun, "Artificial neural network approach to single-ended fault locator for transmission lines," *IEEE Trans. Power Syst.*, vol. 15, no. 1, pp. 370–375, Feb. 2000.

[5] G. K. Purushothama, A. U. Narendranath, D. Thukaram, and K. Parthasarathy, "ANN applications in fault locators," *Int. J. Elect. Power Energy Syst.*, vol. 23, no. 6, pp. 491–506, Aug. 2001.

[6] M. M. Tawfik and M. M. Morcos, "ANN-based techniques for estimating fault location on transmission lines using prony method," *IEEE Trans. Power Del.*, vol. 16, no. 2, pp. 219–224, Apr. 2001.

[7] S. Osowski and R. Salat, "Fault location in transmission line using hybrid neural network," *COMPEL*, vol. 21, no. 1, pp. 18–30, 2002.

[8] T. Bouthiba, "Fault location in EHV transmission lines using artificial neural networks," *Int. J. Appl. Math. Comput. Sci.*, vol. 14, no. 1, pp. 69–78, 2004.

[9] J. Gracia, A. J. Mazon, and I. Zamora, "Best ANN structures for fault location in single- and double-circuit transmission lines," *IEEE Trans. Power Del.*, vol. 20, no. 4, pp. 2389–2395, Oct. 2005.

[10] S. Ekici, S. Yildirim, and M. Poyraz, "Energy and entropy-based feature extraction for locating fault on transmission lines by using neural network and wavelet packet decomposition," *Expert Syst. Appl.*, vol. 34, no. 4, pp. 2937–2944, May 2008.

[11] P. S. Bhowmik, P. Purkait, and K. Bhattacharya, "A novel wavelet transform aided neural network based transmission line fault analysis method," *Int. J. Elect. Power Energy Syst.*, vol. 31, no. 5, pp. 213–219, Jun. 2009.

[12] J. Ezquerra, V. Valverde, A. J. Mazon, I. Zamora, and J. J. Zamora, "Field programmable gate array implementation of a fault location system in transmission lines based on artificial neural networks," *Inst. Eng. Technol. Gen. Transm. Distrib.*, vol. 5, no. 2, pp. 191–198, Feb. 2011.

[13] M. Joorabian, S. M. A. Taleghani-Asl, and R. K. Aggarwal, "Accurate fault locator for EHV transmission lines based on radial basis function neural networks," *Elect. Power Syst. Res.*, vol. 71, no. 3, pp. 195–202, Nov. 2004.

[14] R. N. Mahanty and P. B. Dutta Gupta, "Application of RBF neural network to fault classification and location in transmission lines," in *Proc. Inst. Elect. Eng., Gen. Transm. Distrib.*, Mar. 2004, vol. 151, no. 2, pp. 201–212.

[15] S. R. Samantaray, P. K. Dash, and G. Panda, "Fault classification and location using HS-transform and radial basis function neural network," *Elect. Power Syst. Res.*, vol. 76, no. 9–10, pp. 897–905, Jun. 2006.

[16] R. Salat and S. Osowski, "Accurate fault location in the power transmission line using support vector machine approach," *IEEE Trans. Power Syst.*, vol. 19, no. 2, pp. 979–986, May 2004.
[17] V. Malathi, N. S. Marimuthu, and S. Baskar, "Intelligent approaches using support vector machine and extreme learning machine for transmission line protection," *Neurocomputing*, vol. 73, no. 10–12, pp. 2160–2167, Jun. 2010.
[18] S. Ekici, S. Yildirim, and M. Poyraz, "A transmission line fault locator based on Elman recurrent networks," *Appl. Soft Comput.*, vol. 9, no. 1, pp. 341–347, Jan. 2009.
[19] M. J. Reddy and D. K. Mohanta, "A wavelet-fuzzy combined approach for classification and location of transmission line faults," *Int. J. Elect. Power Energy Syst.*, vol. 29, no. 9, pp. 669–678, Nov. 2007.
[20] M. J. Reddy and D. K. Mohanta, "A wavelet-neuro-fuzzy combined approach for digital relaying of transmission line faults," *Elect. Power Compon. Syst.*, vol. 35, no. 12, pp. 1385–1407, Sep. 2007.
[21] M. J. Reddy and D. K. Mohanta, "Adaptive-neuro-fuzzy inference system approach for transmission line fault classification and location incorporating effects of power swings," *Inst. Eng. Technol. Gen. Transm. Distrib.*, vol. 2, no. 2, pp. 235–244, Mar. 2008.
[22] M. J. Reddy and D. K. Mohanta, "Performance evaluation of an adaptive-network-based fuzzy inference system approach for location of faults on transmission lines using monte carlo simulation," *IEEE Trans. Fuzzy Syst.*, vol. 16, no. 4, pp. 909–919, Aug. 2008.
[23] M. Joorabian, "Artificial intelligent based fault location technique for EHV series-compensated lines," in *Proc. Int. Conf. Energy Manage. Power Del.*, 1998, vol. 2, pp. 479–484.
[24] C. K. Jung, K. H. Kim, J. B. Lee, and B. Klockl, "Wavelet and neuro-fuzzy based fault location for combined transmission systems," *Int. J. Elect. Power Energy Syst.*, vol. 29, no. 6, pp. 445–454, Jul. 2007.
[25] J. Sadeh and H. Afradi, "A new and accurate fault location algorithm for combined transmission lines using adaptive network-based fuzzy inference system," *Elect. Power Syst. Res.*, vol. 79, no. 11, pp. 1538–1545, Nov. 2009.
[26] J.-A. Jiang, C.-L. Chuang, Y.-C. Wang, C.-H. Hung, J.-Y. Wang, C.-H. Lee, and Y.-T. Hsiao, "A hybrid framework for fault detection, classification, and location—Part I: Concept, structure, and methodology," *IEEE Trans. Power Del.*, vol. 26, no. 3, pp. 1988–1998, Jul. 2011.
[27] P. Jafarian and M. Sanaye-Pasand, "A traveling-wave-based protection technique using wavelet/PCA analysis," *IEEE Trans. Power Del.*, vol. 25, no. 2, pp. 588–599, Apr. 2010.
[28] J. Mora-Florez, G. Morales-Espana, and S. Perez-Londono, "Learning-based strategy for reducing the multiple estimation problem of fault zone location in radial power systems," *Inst. Eng. Technol. Gen. Transm. Distrib.*, vol. 3, no. 4, pp. 346–356, 2009.
[29] G. Morales-Espana, J. Mora-Florez, and G. Carrillo-Caicedo, "A complete fault location formulation for distribution systems using the k-nearest neighbors for regression and classification," in *Proc. IEEE/PES Transm. Distrib. Conf. Expo.: Latin America (T&D-LA)*, Sao Paulo, Brazil, pp. 810–815.
[30] *"PSCAD/EMTDC User's Guide,"* Manitoba HVDC Research Ctr., Winnipeg, MB, Canada, 2005.
[31] L. P. Devroye, "The uniform convergence of nearest neighbor regression function estimators and their application in optimization," *IEEE Trans. Inf. Theory*, vol. IT-24, no. 2, pp. 142–151, Mar. 1978.
[32] C. G. Atkeson, A. W. Moore, and S. Schaal, "Locally weighted learning," *Artif. Intell. Rev.*, vol. 11, no. 1, pp. 11–73, Feb. 1997.
[33] A. Navot, L. Shpigelman, N. Tishby, and E. Vaadia, , Y. Weiss, B. Scholkopf, and J. Platt, Eds., "Nearest neighbor based feature selection for regression and its application to neural activity," in *Advances in Neural Information Processing Systems*. Cambridge, MA: MIT Press, 2006, vol. 18, pp. 995–1002.
[34] T. J. Hestilow and Y. Huang, "Clustering of gene expression data based on shape similarity," *EURASIP J. Bioinf. Syst. Biol.*, pp. 1–12, 2009.
[35] F. B. Costa, B. A. Souza, and N. S. D. Brito, "Real-time detection of fault-induced transients in transmission lines," *Electron. Lett.*, vol. 46, no. 11, pp. 753–755, May 2010.
[36] F. V. Lopes, W. C. Santos, and D. Fernandes, "An adaptive fault location method for smart distribution and transmission grids," in *Proc. IEEE Power Eng. Soc. Conf. Innovative Smart Grid Technol.*, pp. 1–7.
[37] J. Schmid and K. Kunde, "Application of non conventional voltage and currents sensors in high voltage transmission and distribution systems," in *Proc. IEEE Int. Conf. Smart Measure. Future Grids*, Bologna, Italy, 2011, pp. 64–68.

Mohammad Farshad (S'11) was born in Gonbad-e-Qabus, Iran, in 1981. He received the B.Sc. degree in power transmission and distribution networks engineering from Power and Water University of Technology (PWUT), Tehran, Iran, in 2003 and the M.Sc. degree in power system engineering from Ferdowsi University of Mashhad, Mashhad, Iran, in 2006, where he is currently pursuing the Ph.D. degree in power system engineering.

His main research interest is the application of intelligent systems in power system protection and operation.

Javad Sadeh (M'08) was born in Mashhad, Iran, in 1968. He received the B.Sc. and M.Sc. degrees in electrical engineering (Hons.) from Ferdowsi University of Mashhad, Mashhad, Iran, in 1990 and 1994, respectively, and the Ph.D. degree in electrical engineering from Sharif University of Technology, Tehran, Iran, with the collaboration of the electrical engineering laboratory of the Institut National Polytechnique de Grenoble (INPG), Grenoble, France.

Currently, he is an Associate Professor in the Department of Electrical Engineering, Ferdowsi University of Mashhad. His research interests are power system protection, dynamics, and operation.