IET Generation, Transmission & Distribution

Research Article

IET Journals
The Institution of Engineering and Technology

Classification of arc fault between broken conductor and high-impedance surface: an empirical mode decomposition and Stockwell transform-based approach

ISSN 1751-8687
Received on 23rd February 2020
Revised 16th July 2020
Accepted on 4th August 2020
E-First on 2nd October 2020
doi: 10.1049/iet-gtd.2020.0340
www.ietdl.org

Himadri Lala[1], Subrata Karmakar[2] ✉

[1]*School of Electrical Engineering, VIT University, Vellore, Tamil Nadu, 632014, India*
[2]*Department of Electrical Engineering, NIT Rourkela, Rourkela, Odisha, 769008, India*
✉ *E-mail: karmakar.subrata@gmail.com*

Abstract: The high-impedance arc fault (HIAF) is a result of the interaction between the current-carrying conductor and the high-impedance surface. This study presents a detailed and comparative time-frequency based analysis for the classification of different arcs, generated by the interaction of a broken conductor and different arcing surfaces. The real-time arcing voltage signals are considered as the basis of the whole time-frequency based analysis from a medium voltage distribution line. In this proposed approach, empirical mode decomposition (EMD) is used to study the real-time arc voltage signals of various surfaces. The intrinsic mode functions obtained by the application of the EMD technique is used as the input to various machine learning techniques, which later successfully classifies different HIAFs from their harmonic footprints. The Stockwell transform (ST) is additionally applied to the same test cases, and the classification of the arc is performed using similar learning algorithms to get a quantitative view of both outcomes of EMD and ST towards the selection of appropriate signal processing technique and machine learning algorithm for the same.

1 Introduction

The undesired interaction of an energised conductor and different high-impedance surfaces can lead to the formation of an electrical arc [1, 2]. It appears to be harmless as the fault current is in the milli-Ampere range. However, the arc flash heating over an extended period can cause a fire. According to the 'United States Fire Administration' report [3, 4], an estimated 23,500 residential building fire incidents registered in the United States in the year 2016 alone, causing an estimated 800 human injuries, 310 deaths, and $7.09 billion in assets loss. Several of them are caused due to short circuit and arc. On the other hand, the broken conductor and different high-impedance surfaces (concrete, tree, grass, soil, wet sand etc.) are very common causes of arc. Owing to storm in the monsoon or any other natural calamity, the energised power conductors may break apart too and create an electrical arc. The fundamental difference between the arc in these high impedance surfaces and other electrical equipment is the impedances of the grounding path for the fault current. For this purpose, arc classification and identification due to the different arcing surfaces are necessary to improve a distribution system's protection scheme.

Over the years, several real-time and simulation-based studies have been conducted on high-impedance fault (HIF), and the voltage and current signals during HIF are found to be unpredictable due to non-linear fault impedance [5–8]. Many techniques, however, such as 'power line carrier communication' [9], 'mathematical morphology' [10, 11], and 'magnetic field signature'-based study [12] are used for HIF identification. Alternatively, 'harmonic analysis techniques' [13] are often used to detect HIF. Numerous researchers have adopted the 'discrete wavelet transform (DWT)' alongside machine-learning techniques [14, 15], e.g. 'probabilistic neural network', 'artificial neural network', 'support vector machine (SVM)' etc. for detecting HIF. 'Choi–Williams distribution' is also used with 'time-frequency distribution' for detecting HIF [16]. Besides that the authors in [17] described both simulation and laboratory validation of the properties of the high-impedance arc. In [17], a DWT-based detector approach is also suggested for the arc because of the tree leaning on an electrical conductor. The computational complexity in frequency domain analysis nevertheless inspired a few researchers [18] to come up with a solution in the time domain. The contemporary literature only discussed features of high-impedance arc faults (HIAFs) with non-linear grounding paths, i.e. tree-line or line-surface, which can differ depending on its nature [19, 20]. Moreover, the different types of arc faults can produce different voltage and current characteristics and have a very distinct impact on the surrounding environment. Therefore, it is necessary to develop an algorithm to classify different HIAFs to take appropriate actions against the fault events.

In this work, the different types of surfaces, i.e. wet sand, concrete, grass, leaning tree, and soil are considered for the practical simulation to investigate the nature of different arcing phenomena in high-voltage laboratory. Furthermore, the arc voltage signals have been recorded using a digital storage oscilloscope and processed with empirical mode decomposition (EMD) and Stockwell transform (ST) to obtain the corresponding intrinsic mode function (IMF) and coefficients, respectively. Finally, for the classification of the arc, ST coefficients and IMFs are used as input to the machine learning algorithm such as SVMs and k-nearest neighbour (KNN) algorithm. The main objectives of the work are pointed below.

- Recording of different arc (i.e. wet sand, concrete, grass, leaning tree, and soil) voltage signals for the analysis in a practical environment.
- Investigation of arc fault based on their harmonic signatures.
- Selection of the proper signal processing technique for the investigation of arcing signals.
- Investigate the dissimilarities in different arc faults, if any, due to the various arcing surface.
- If there are dissimilarities in the different arc faults due to the different arcing surface, finding out the best possible machine learning tool towards arc fault classification.

2 Arc in distribution systems and the non-linear arc-simulation

Natural disasters can cause a situation of a broken conductor of distribution lines. A broken conductor may create an arc of

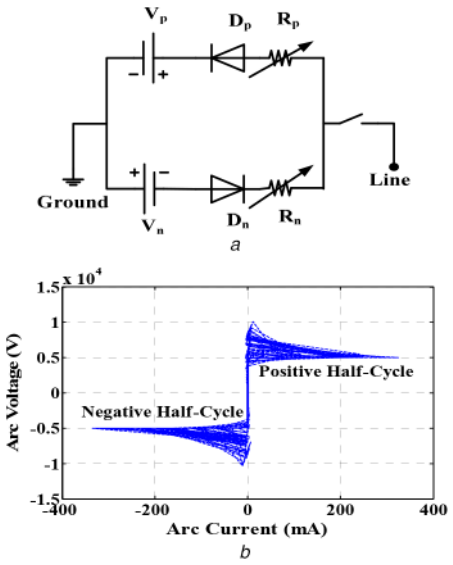

Fig. 1 *Equivalent circuit for HIAF simulation and V–I characteristics*
(a) Arc simulation model for HIAF, ***(b)*** Simulated *V–I* characteristics of HIAF

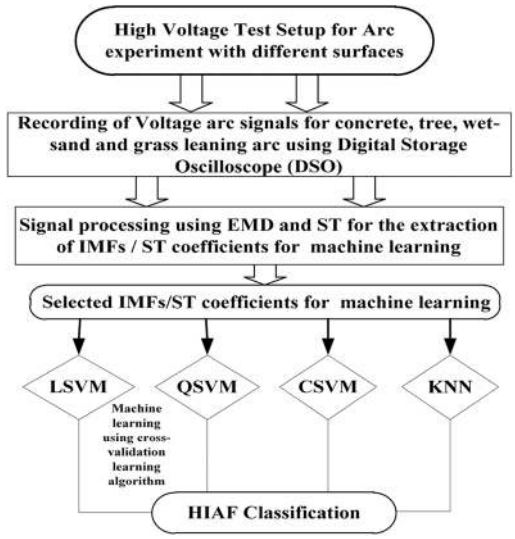

Fig. 2 *Flowchart for HIAF classification*

different types depending on the arcing surface. In an arc flash, an extensive amount of concentrated thermal energy explodes outward in the form of a blast from the electrical equipment. The blast makes high-pressure waves that by causing the fire may cause harm to both the atmosphere and the life forms around it. The arcing flash is also the result of poor connection between a conductor and the ground (soil) or grounded material (concrete, tree, wet sand, grass etc.).

In addition, the surface impedance may be linear or non-linear in nature, depending on the surface form. Similarly, the simulation of an arc can be represented by discharge through a non-linear impedance. Simulation is carried out in MATLAB environment to check the typical arcing response. Arc is simulated in two different characteristics, static and dynamic, based on its time constant. For the static arc, the rate of conductance changes in time is very low $dg/dt \rightarrow 0$ resulting time constant $\tau \simeq 0$. The time constant of the dynamic arc can be defined as the time constant of the exponential change in the voltage of the arc resulting from a step in the current. It depends not only on the test conditions and the measurement process but also on the current magnitude [21]. The thermal inertia allows the relationship between current and voltage to depend on how quickly the current changes. Hence, for dynamic arc $\tau \neq 0$ and it can be in a few microseconds depending on the medium of the arc and arcing surface. Fig. 1*a* shows the dynamic arc fault simulation model. Fig. 1*b* describes the dynamic characteristics of the arc due to the leaning tree on the conductor. The impedance model presented in Fig. 1*a* controls the variation of impedance between ground and line. From Fig. 1*b*, it is evident that the characteristics are dynamic because of the time-dependent variability of impedances between the conductor and tree. Owing to different arcing surfaces, different aspects of dynamic arcing phenomena are explored in this study.

3 Tools for HIAF classification

The non-linearity and non-stationary nature of arc signals requires an efficient signal processing tool to decode the harmonic signatures present in the arc signals. In this study, an EMD and ST-based comparative approach is adopted along with different machine learning algorithms for arc fault classification. This comparative approach is carried out to select an appropriate signal processing tool and proper machine learning algorithm for arc classification. Fig. 2 presents a flowchart for the processes of HIAF classification.

3.1 Empirical mode decomposition

The modified Hilbert–Huang (HH) transform has different enhancements from HH transform and are described in [22–25]. Furthermore, the 'masking signals' have also been prescribed to solve the problem of mode mixing [25]. The aspect of generating the masking signal during EMD [26] is also proposed. On the other hand, a 'post-processing demodulation' approach has likewise been proposed in [27]. One of the reasons is that the function has two spectrum lines. Moreover, Hilbert transform represents higher frequency better than lower frequency, and there are several complications for more complex functions. Here, the empirical envelope seems reasonable. The possibility of instantaneous frequency is the crucial rule of Huang's EMD method. The derivative of the phase of an analytic signal is portrayed as the fundamental approach of EMD [27, 28]. If the distorted signal is $s(t)$ and the IMF is $c(t)$. If c_1 is an IMF, if the number of local extrema of c_1 is equal to or differs from the number of zero crossings by one, and the average of c_1 is reasonably zero, then the residue can be calculated by

$$s(t) - c_1 = r_1 \tag{1}$$

$$r_1 - c_2 = r_2 \tag{2}$$

where c_1, r_1, and r_2 stand for IMF and residue functions, respectively. So, the nth residue function will be

$$r_{n-1} - c_n = r_n \tag{3}$$

Therefore, the final relation between all the IMF's and residue with the signal is

$$s(t) - \sum_{j=1}^{n} c_j = r_n \tag{4}$$

The EMD finds out the envelopes $e_M(t)$ and $e_m(t)$ by performing cubic spline interpolation between maxima and minima. If the mean of these envelopes is $m(t)$

$$m(t) = \frac{e_M(t) + e_m(t)}{2} \tag{5}$$

where

$$r_1 = \sum_{j=1}^{k} m_j \tag{6}$$

From (2)

$$r_1 - c_2 = r_2$$

Substituting the value of r_2 in (2) and combining with (6)

$$c_2 = r_1 - r_2 = r_1 - \sum_{j=1}^{k} m_{2j}$$
$$c_2 = \left(\sum_{j=1}^{k} m_j - \sum_{j=1}^{p} m_{2j} \right) \tag{7}$$

Adding the values of c_1, c_2, … we get

$$\sum_{j-1}^{n} c_j = s(t) - \sum_{j=1}^{k} m_j + \left(\sum_{j=1}^{k} m_j - \sum_{j=1}^{p} m_{2j} \right) + \cdots = s(t)$$

3.2 Application of ST and feature extraction

In a given time-series of a fault signal $f(t)$, the local spectrum at time $t = \tau$ can be obtained by multiplying $f(t)$ with a Gaussian function $g(t)$ located at $t = \tau$. The 'ST' used in this study is given as

$$s(f, \tau, \sigma) = \int_{-\infty}^{\infty} f(t) g(t - \tau) \mathrm{e}^{-\mathrm{i}2\pi f t} \mathrm{d}t \tag{8}$$

where

$$g(t) = \frac{|f|}{\sqrt{2\pi}} \mathrm{e}^{-\left(t^2 f^2 / 2\right)}$$

Here, σ represents the width of the Gaussian function $g(t)$. The ST of a fault signal $f(t)$ is defined as a Continuous Wavelet Transform (CWT) with a specific mother wavelet multiplied by the phase factor $p_1(t, f)$

$$p_1(t, f) = f(t) \mathrm{e}^{-\mathrm{i}2\pi f t} \tag{9}$$

Substituting in (8)

$$s(f, \tau, \sigma) = \int_{-\infty}^{\infty} p_1(t, f) g(t - \tau) \mathrm{d}t \tag{10}$$

$$= p_1(t, f) * g(t, \sigma) \tag{11}$$

Here, '$*$' denotes the convolution operation. The ST provides a time-frequency representation instead of the time scale representation developed by the wavelet transform [29]. In this work, ST coefficients are calculated from the arc signals and used as the input to the machine-learning algorithms.

3.3 Machine learning using SVM's and KNN

Support vector machines (SVMs) perform on the fundamental principle of decision planes and decision boundaries. SVM can be linear or non-linear based on the kernel trick [30]. These boundaries or planes separate different classes of data from each other using support vectors. Depending on the types of boundaries (linear or non-linear) for the different kinds of scenarios, there are linear-SVM (LSVM), quadratic-SVM (QSVM), cubic-SVM (CSVM), Gaussian-SVM etc. Apart from SVM, another promising machine learning technique that is widely used is KNN [31]. In KNN, an object is classified on the majority voting of the nearest neighbours. It is one of the simplest forms of machine learning. Therefore, it requires less computational requirements. In this work, LSVM, QSVM, CSVM, and KNN are considered for the classification.

4 Experimental arrangement for different arc faults

A 500 kV, 500 kVA power frequency test transformer is used in the high-voltage engineering laboratory at the 'National Institute of Technology, Rourkela, India' for all necessary experiments. The arc fault simulated in the laboratory includes the conductor and arcing surface/medium, i.e. concrete grass, tree, and wet sand. It is a remotely operated high-voltage (HV) laboratory with a range of experiments up to 500 kV including extensive testing and partial discharge detection (PD) with an advanced virtual instrumentation research facility [32], impulse, arcing [19, 20], lightning phenomena, insulation failure etc. The rated input/output voltage for the test transformer is 610 V/500 kV, 50 Hz, and the rated test voltage is 450 kV, respectively. Fig. 3*a* sets out a schematic diagram for the different experimental arcing configurations. Here the test transformer uses input from an auto-transformer. The auto-transformer supplies the low-voltage side of the high-voltage test transformer with a voltage of 0–610 V. The test transformer voltage is applied to the conductor by means of a capacitor divisor and Potential Transformer (PT). The actual experimental setup in the HV lab is shown in Fig. 3*b*. The voltage applied to the conductor, and the specific arcing surface is measured by a PT of 1000:1 ratio via a 1000 pF, 500 kV, and PD<5 pC capacitor (as per IEC 60270 standards). The HV test setup has a rated power output of 500 kV, 250 mA (125 kVA - continuous duty). The voltage of the arc can be seen via the voltmeter that is connected to the control panel. In addition, a compensating reactor is attached at 100 kVAR in parallel with the HV test transformer to restrict or compensate for current during the arc experiments. Various arcing surfaces, i.e. concrete, grass, tree, and wet sand, are tested in this whole experiment. Figs. 4*a* and *b* stand for voltage, current, and *V–I* characteristics captured during the experiment for the arcing in the leaning tree, respectively. The measurement of arcing voltage and current is taken using a (make: Tektronix; Model: TDS 2014C [33]) four-channel digital storage oscilloscope. Different arcing voltage signals of different conditions are obtained from the measurement units of the arcing experimental setup presented in Fig. 4*c*. During the test, all the surfaces are solidly grounded to create the scenario for the broken conductor. When a conductor breaks apart and comes into interaction with different surfaces, it creates HIF with severe arcing depending on the arcing surface and arcing voltage. During this incident, these arcing surfaces play a very crucial role in the characteristics of HIF voltage or current. Moreover, these surfaces offer a non-linear kind of high-impedance that results in arcing. On the other hand, all of this is the same category of fault, and may not cause tripping of the circuit breaker. In Fig. 5*a*, a different type of arc during different experiments is shown. The sampling rate for the acquisition of the voltage and the current signal is 100 kHz, and all the processing is performed on a PC with an Intel Core i7–3770 processor, which has a processor speed of 3.40 GHz.

5 Results and analysis

The *V–I* characteristics of an arc fault can differ based on the nature of the arc. The *V–I* characteristics are determined by the time constant of an arc. However, depending on other factors, some arc characteristics can also change. For arc experimentation, four types of surfaces are considered in this study. In the first experiment, the arc is generated between the broken conductor and the concrete surface. The arc signals are similar to other HIAFs. As the arc goes on, a conical structure of arc is formed on a concrete surface. In contrast, the arc between a medium-voltage line and an inclined tree is created in the second experiment. In both cases, the voltage of the arc and the arc current are measured by PT and current sensors, respectively. However, the leaning tree arc does not follow a consistent pattern during discharge, so as in the other two experiments involving wet sand and grass bed as the arcing surfaces. In the case of wet sand, the stress created due to the air-gap between sand and the broken conductor displaces sand and finds the low-impedance path possible. On the other hand, the grass bed offers a somewhat similar discharging arc like the tree. However, the difference was with the current. The reason behind the difference is the difference in the offered fault impedances of trees and grass. On the contrary, arc in the grass can be most dangerous compared to arc in the tree as it catches fire very quickly. Therefore, it is evident that the characteristics and

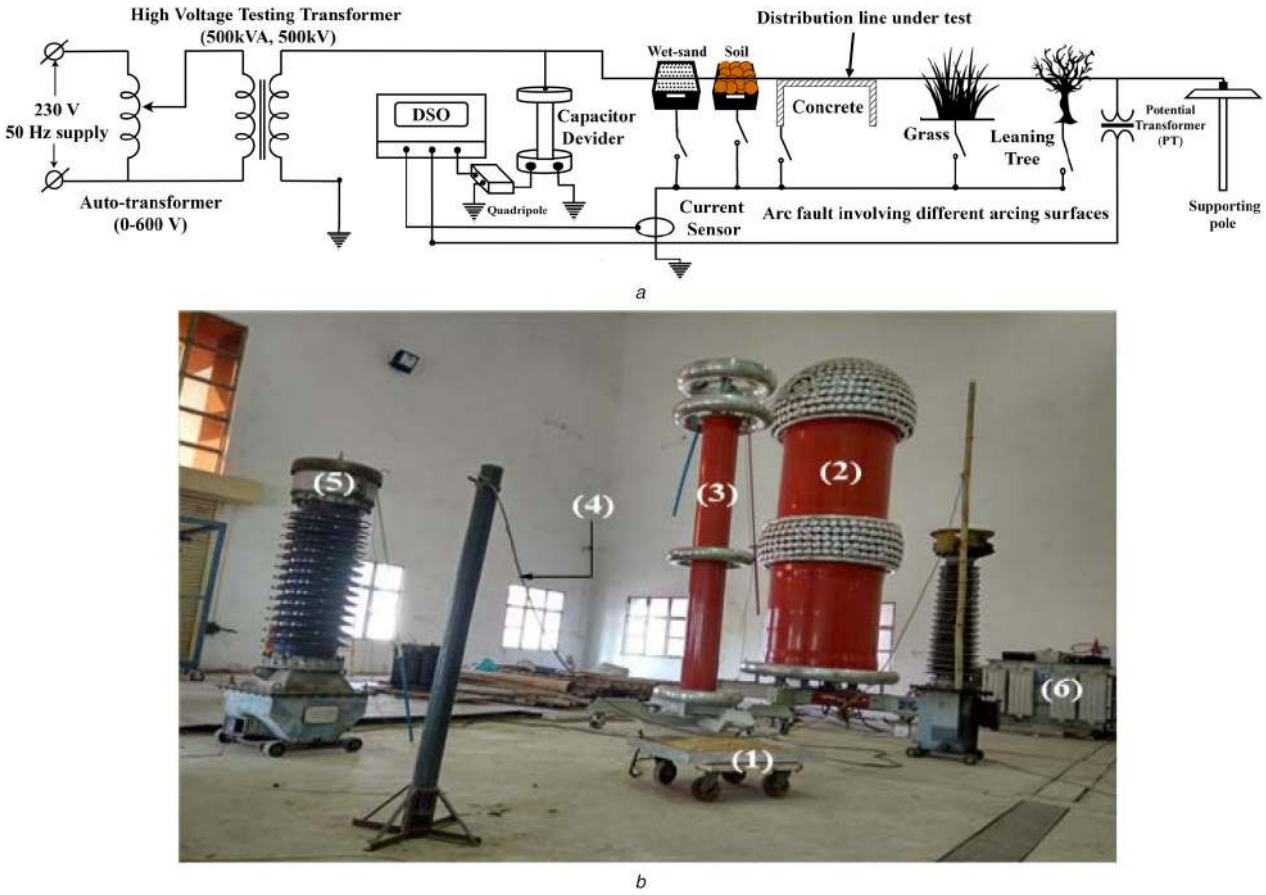

Fig. 3 *Schematic diagram and photograph of overall HIAF experimental setup*
(a) Schematic diagram, ***(b)*** Photograph of the experimental laboratory setup for the arc (1, wet sand as arcing object; 2, HV testing transformer; 3, capacitor divider; 4, conductor used for arcing experiment; 5, potential transformer; 6, auto-transformer)

consequences of different arcing conditions are very much different from each other. The voltage waveform for each arc condition is shown in Fig. 4*c*. The signals are captured in real-time using PT and sensors.

Nevertheless, the disparity between the two forms of the arc during different experiments is also prominent because of the non-linear variability in voltage. Nonetheless, the overall difference and the value of four independent experiments constitute the grounding for the discharge. However, the time-domain responses of different arc faults are not very contrary to each other. Therefore, an extensive amount of data is captured after several rounds of experiments with different voltage levels keeping in mind the voltage level 40 kV. A total of nine sets of experiments are performed, and very asymmetric test sets of arc voltage signals at different voltage levels are formed. The reason behind the asymmetric nature of the test set is to check the robustness of the classification algorithm irrespective of voltage levels. The description of different tests and the classification accuracy obtained using different learning algorithms are depicted in Table 1.

5.1 Detection of arc using EMD

The voltage waveforms captured during the arc faults depict a unique kind of characteristic compared to short circuit faults. According to the power system relaying and control committee report [1], the frequency response of a HIF must register a dominance of the second harmonic component in the voltage profile. Therefore, before concluding whether it resembles with HIF or not, the frequency profile of the voltage signals needed to be verified. Fig. 5*b* presents the fast Fourier transform of the voltage signals captured using the PT with and without arc fault. It is evident from the frequency response of the voltage signals that it completely resembles the traditional response of HIF signals. The EMD is applied to each test set, and the IMFs are calculated for those corresponding signals. The EMD breaks down the captured signals into various IMFs. Alternatively, these IMFs can be used to restore the original signal.

The EMD is used on the voltage signals in this analysis. The total number of samples taken for analysis is 10,000 (2500 for each class, i.e. concrete, tree, wet sand, and grass). Fig. 6*a* presents the power spectral density (PSD) plots of IMFs for the voltage signals of different arcing events, respectively, of a single test set as an example. Fig. 6*a* also informs the predominant frequency with a maximum spectral density for each IMF of all cases. It is evident from Fig. 6*a* that the common feature among all the arcing cases is the presence of second harmonic component and the presence of multiple numbers of fundamental frequency components with different PSDs. On the other hand, the total number of IMFs calculated from the EMD of arc voltage signals for concrete, tree, wet sand, and grass are 10, 9, 9, and 10, respectively. Furthermore, among all the IMFs calculated from test 1 and test 2 arc signals, a total of eight IMFs (IMF1–IMF8) are selected for classification after applying the ranker feature selection algorithm. For tests 3–9, a total of nine (IMF1–IMF9) IMFs from each arcing scenario are considered for further analysis as the minimum number of IMF calculated is 9. The box-plot for the selected IMF is presented in Fig. 6*b*. It clearly describes the differences between various arcing events regarding their IMFs. However, the presence of second-order harmonics is the most prominent HIF resemblance feature in the frequency responses [19]. Moreover, due to the inherent similarities between the obtained harmonic responses, it is difficult to classify different HIAF events. Few researchers [34] extracted different features from HIF. Therefore, in this study, different machine-learning techniques are used to get a comparative view of the learning tools. Table 1 presents classification results for different test sets using different machine-learning tools. The classifiers or the machine-learning tools used in this study are LSVM, QSVM, CSVM, and KNN. It is evident from Table 1 that, apart from LSVM, all other classifiers are performing according to the requirement of successful classification. Moreover, the error in

classification for QSVM, CSVM, and KNN is marginal as compared to LSVM. However, the classification accuracy of LSVM is found to be less than QSVM, CSVM, and KNN. Moreover, among QSVM, CSVM, and KNN, CSVM performs much consistently and gives the best classification result among the three. Except for test set 7, CSVM gives a classification accuracy of 100% for all other test sets. The accuracy of the classifier is shown through the receiver operating characteristic (ROC) curve is supported by the classification confusion matrix presented in Figs. 7*a* and *b*. In Figs. 7*c* and *d*, the ROCs are presented for both the learning algorithms, KNN and CSVM, respectively. It indicates the accuracy of the classifier for different classes. It describes the whole description of classified and unclassified instances of both the learning algorithms, KNN and CSVM, respectively. However, the calculated IMFs are mostly containing similar kinds of frequency contents, but their IMFs are very much different from each other. The tests are replicated and conducted at various voltage levels, and the findings reported in this analysis are validated each time. Therefore, by harmonic analysis with EMD, different HIAFs can be successfully classified using machine learning techniques such as QSVM, CSVM, and KNN.

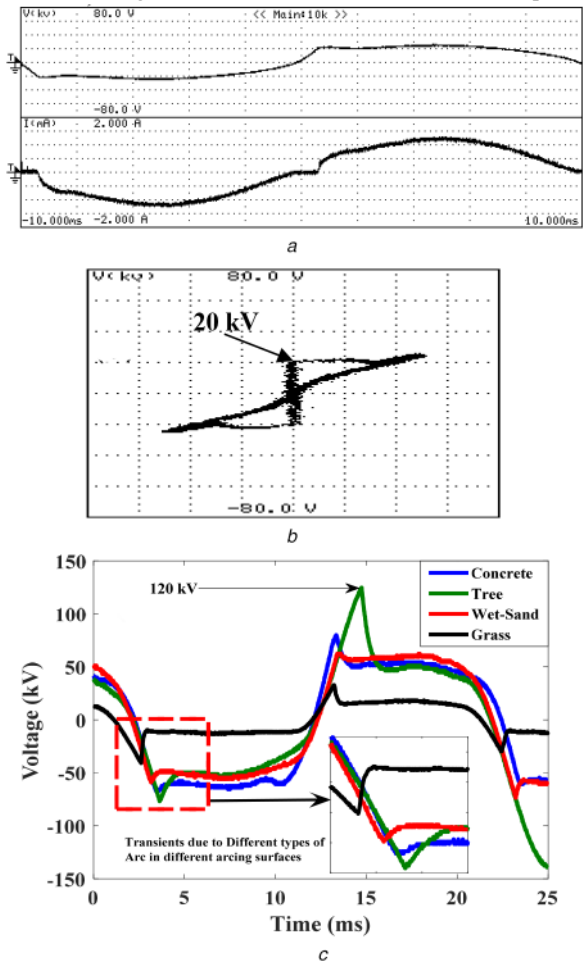

Fig. 4 *Voltage, current, and V–I characteristics of arc faults*
(a) Voltage and current of arc in the leaning tree, ***(b)*** *V–I* characteristics for leaning tree arc, ***(c)*** Voltage characteristics of different arcing surfaces

5.2 Detection of arc using ST

The ST is one of the popular signal transformation techniques. Instead of the time scale representation given by the different wavelet-based algorithms, it provides a time-frequency representation. The ST coefficients are further utilised to extract from the transient the different characteristic features. Here, the ST coefficients are obtained and used to detect the arc. Unlike EMD, the detailed time-frequency coefficients are given when ST is applied. Throughout this study, the ST is applied to the arc signals to get a comparative view of the machine-learning outcomes. Similar to the EMD, ST is applied, and the coefficients are obtained. Considering that the ST offers much more accurate coefficients, a range of 250 samples of different arc signals is used in the study. Additionally, the approximate ST coefficients for the machine-learning algorithms are used to construct the training sets. The ST coefficients are used here as inputs to similar machine-learning algorithms (i.e. LSVM, QSVM, CSVM, and KNN) used in the case of EMD. In the case of ST, the number of input features is very high than EMD. The number of features calculated from ST is 126. However, the results computed using ST coefficients are found to be reasonable for most of the SVM-based techniques. The results using KNN is not satisfactory in comparison with others. The results obtained using ST are presented in Table 2. The test sets shown in Table 2 belong to the similar test sets used in the case of EMD based analysis. However, the difference between the two

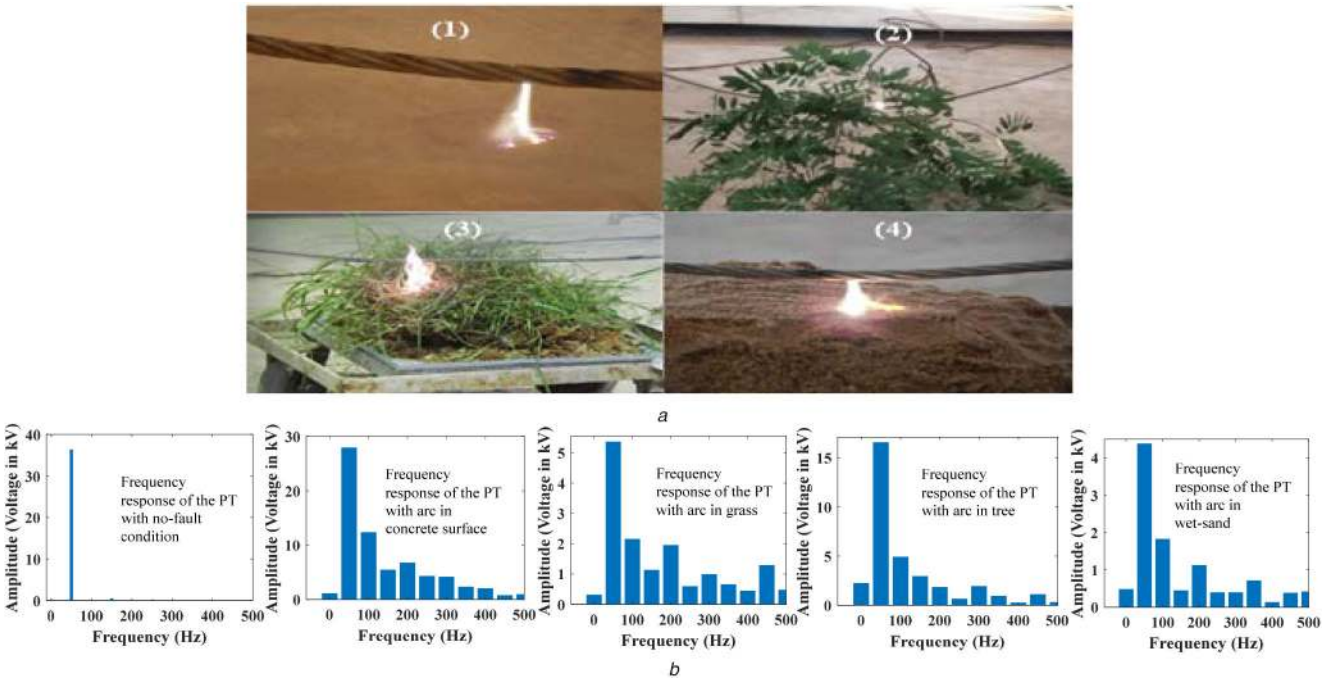

Fig. 5 *Arc in different arcing surfaces and their frequency response with or without fault*
(a) Arc between conductor and 1, concrete; 2, tree; 3, grass; 4, wet sand, ***(b)*** Frequency response of PT with and without arc faults

Table 1 Arc classification result for the different surfaces using different machine-learning algorithms and EMD

Different test sets	Arcing voltage in the different surfaces, kV				No. of input features	Classification accuracy using different classifiers, %			
	Concrete	Grass	Tree	Wet sand		LSVM	QSVM	CSVM	KNN
test 1	20	12	7	4	8	91.1	99.3	100	100
test 2	10	14	11	10	8	89.8	98.6	100	100
test 3	14	16	13	5	9	80.6	100	100	99.4
test 4	16	6	15	15	9	85.8	100	100	99.3
test 5	20	10	22	17	9	84.9	100	100	100
test 6	24	22	26	19	9	72.4	100	100	99.4
test 7	32	31	32	22	9	82.2	99.5	99.2	99.1
test 8	27	25	33	25	9	79.5	100	100	99.3
test 9	37	39	34	28	9	83.7	98.6	100	99.3

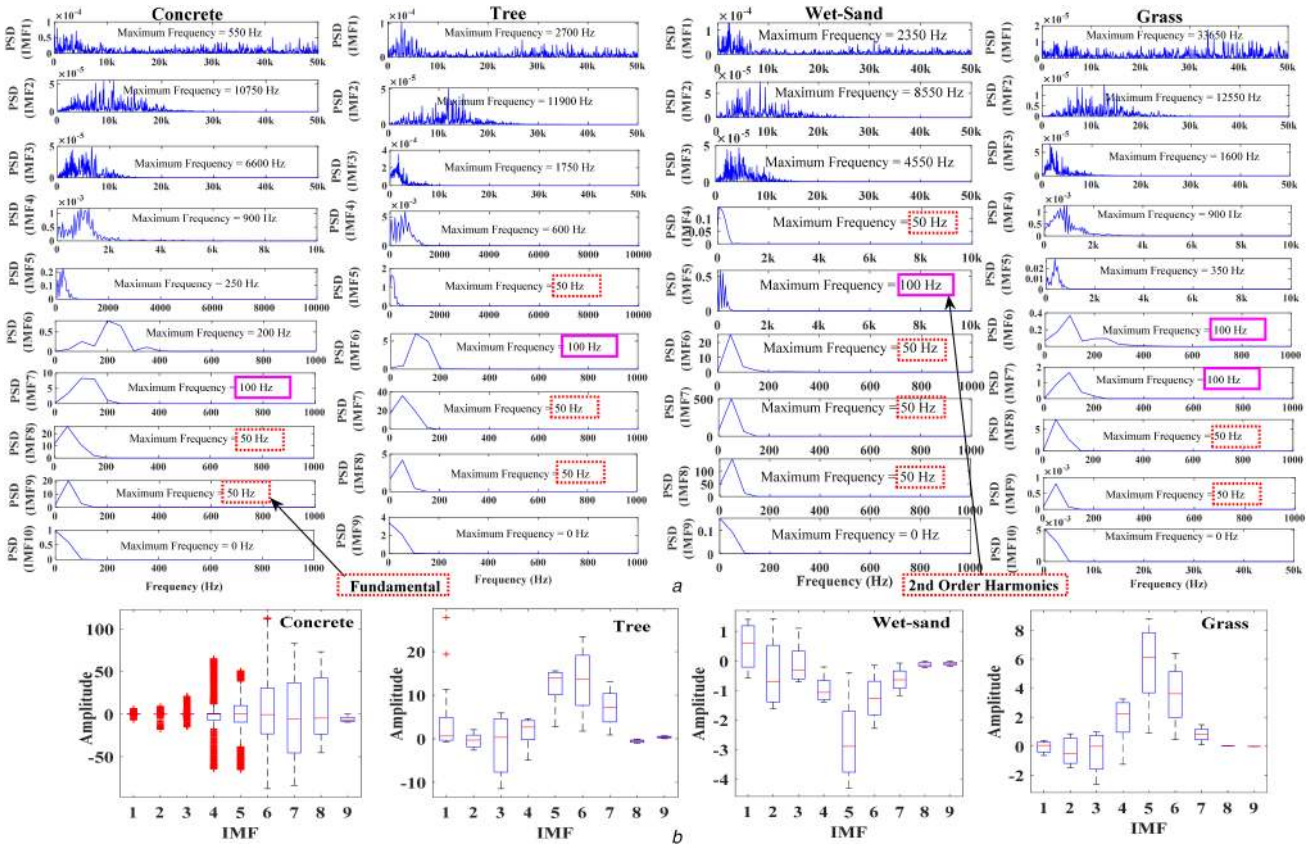

Fig. 6 *Application of EMD on the different types of arcing voltage signals and the frequency-decomposed signals (IMF)*
(a) PSD plots of the IMF, ***(b)*** Box-plot of selected IMF for different arcing scenarios

different approaches lies in the number of features considered as the input to the machine-learning algorithms. A comparative study of the results for both EMD- and ST-based approaches is shown in Fig. 8. The classification accuracy for both the signal processing techniques lies between 99 and 100%, which is an adequate level of accuracy in classification. However, the EMD (number of features: 9) based approach achieves the accuracy level using a very less number of features than ST (number of features: 126). However, there are contradictions regarding the selection of a machine learning algorithm for both ST and EMD. In the case of EMD-based approach, QSVM, CSVM, and KNN are found to be the three best classifiers. Whereas for ST, LSVM, QSVM, and CSVM are the three best classifiers concerning classification accuracy. Moreover, Fig. 8 proves the capability of CSVM for EMD-based approach and the SVM for ST-based approach as the appropriate machine-learning tool.

5.3 Classification arc fault from similar transient switching faults

The events such as harmonic loads and capacitor switching are very much similar to HIFs in terms of their frequency response. At the same time, these events can create confusion with HIF. However, none of them causes an electrical arc during the event. In view of protection strategies of a distribution system, capacitor switching and harmonic load switching can confuse with HIF. However, these events do not have any implication on living beings, as they do not cause a fire. The experiments were performed, and voltage transient signals for capacitor switching, inductive load switching and harmonic load switching are captured. The voltage signals of capacitor switching, inductive load switching, harmonic load switching etc. and their frequency response is presented in Figs. 9*a* and *b*, respectively. It is clearly visible from the results that the frequency profile of the switching events shows the dominance of second-order harmonics, which is identical to the HIAF. However, the transients are momentary and do not sustain for a long period which HIAF does. Moreover, a seven-class (four-arc event and three-switching event) classification is performed with the IMFs calculated from the

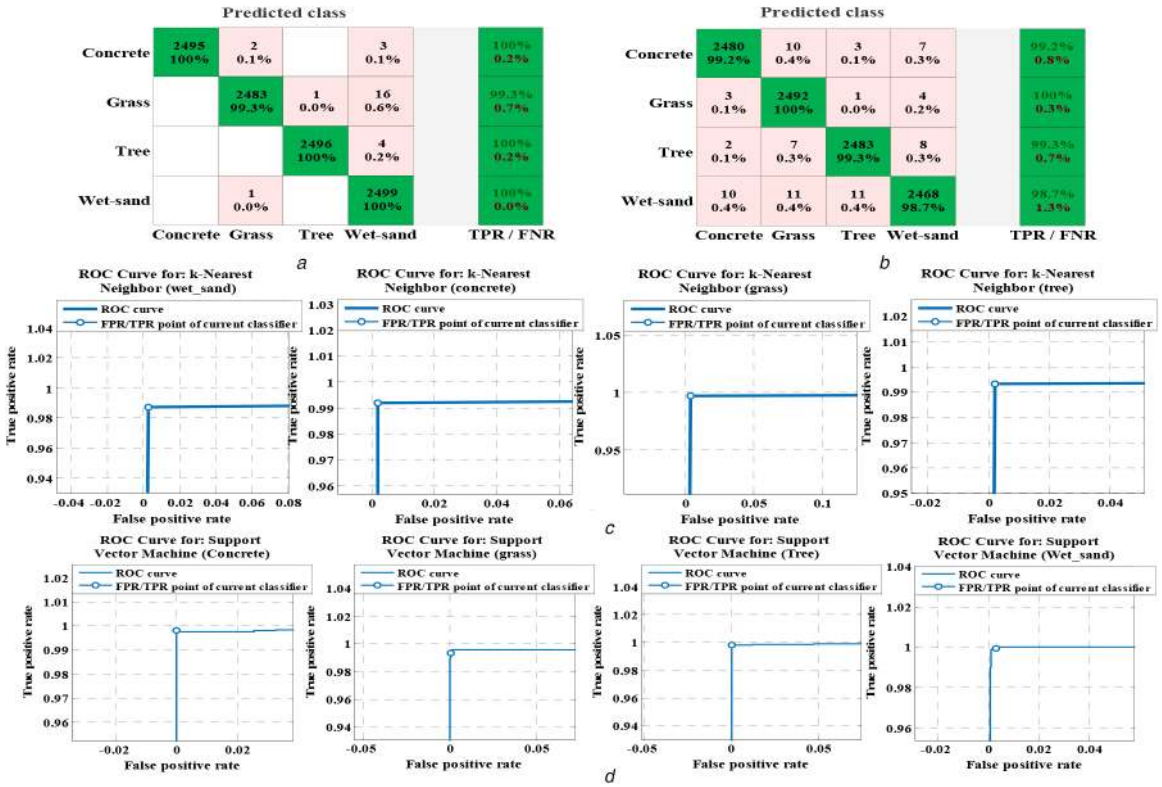

Fig. 7 *Arc fault classification results using KNN and CSVM as the classifier*

(a) Confusion matrix of the classification using KNN, ***(b)*** Confusion matrix of the classification using CSVM, ***(c)*** ROC characteristics using KNN, ***(d)*** ROC characteristics using CSVM

Table 2 Arc classification result in different test sets using different machine-learning algorithms and ST

Different test sets	Arcing voltage in the different surfaces, kV				No. of input features	Classification accuracy using different classifiers, %			
	Concrete	Grass	Tree	Wet sand		LSVM	QSVM	CSVM	KNN
test 1	20	12	7	4	126	100	100	100	100
test 2	10	14	11	10		100	100	100	99.2
test 3	14	16	13	5		100	100	100	100
test 4	16	6	15	15		100	100	100	98.8
test 5	20	10	22	17		100	100	100	95.6
test 6	24	22	26	19		100	100	99.5	90.8
test 7	32	31	32	22		100	99.4	99.2	98.2
test 8	27	25	33	25		100	100	100	98.3
test 9	37	39	34	28		100	100	100	89.1

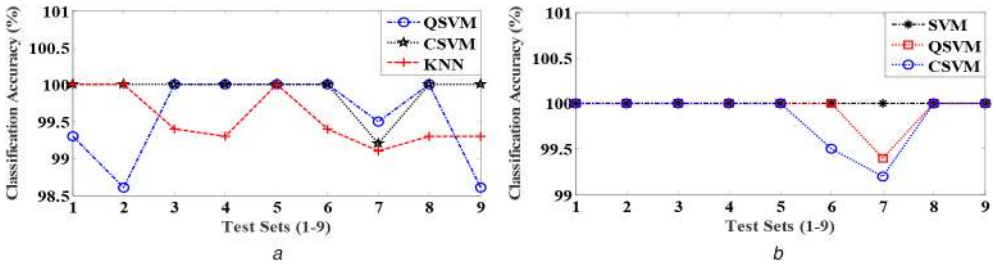

Fig. 8 *Comparative classification accuracy plot for different classifiers using EMD- and ST-based approach*

(a) EMD, ***(b)*** ST

transient signals using different SVMs and KNNs. Similar to the previous classification results for EMD and ST, the CSVM, QSVM, and KNN give an adequate level of classification accuracy and successfully classifies the transients. The classification results are depicted in Table 3.

5.4 Classification arc fault for soil

To study the behaviour of HIAF with the variation of arcing surface, more cases are considered for analysis. There are a few researchers [35] who worked on HIF arcing on sandy soil considering distribution feeders. In the first set of experiments, the arc is created on wet sand, grass, concrete, and leaning-tree whereas the arc is formed in the second set of experiments in soil with a difference in the amount of soil moisture. Photos of the experiment for the events are shown in Fig. 10. In both cases, the voltage of the arc and the current of the arc are measured as before by PT and current sensors. The schematic diagram of the experiment is presented in Fig. 3*a*. Moreover, soil bed with different thickness and moisture level is considered for analysis.

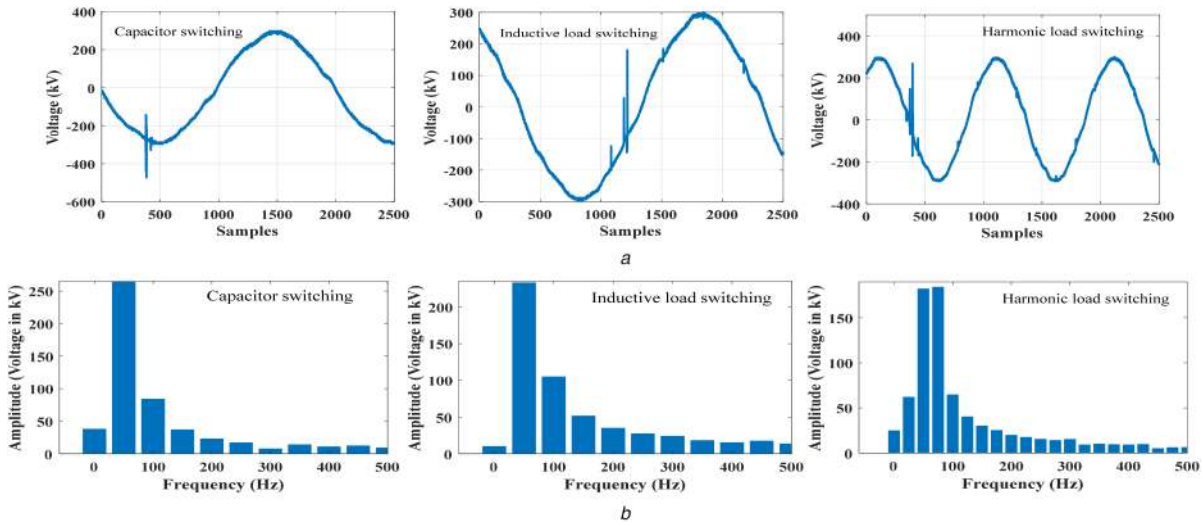

Fig. 9 *Voltage profile and frequency response of capacitor switching, inductive loads switching, and harmonic load switching*
(a) Voltage profile, ***(b)*** Frequency response

Table 3 Classification result for arc fault and similar events using different machine-learning algorithms and EMD

Test sets (classes)	Number of data in each class	No. of input features	Classification accuracy using different classifiers, %			
			LSVM	QSVM	CSVM	KNN
concrete, grass tree, wet sand, capacitor switching, load, and harmonic load switching	2500	9	86.4	99.3	97.4	99.2

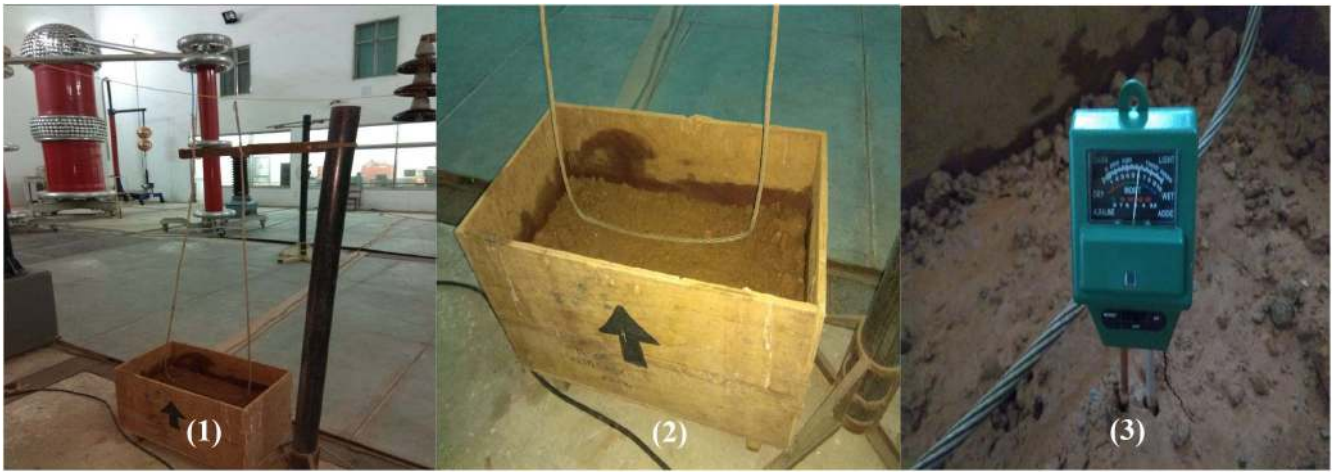

Fig. 10 *Arc in soil bed with different soil depth and moisture level (1, experiment setup; 2, soil bed; 3, moisture meter)*

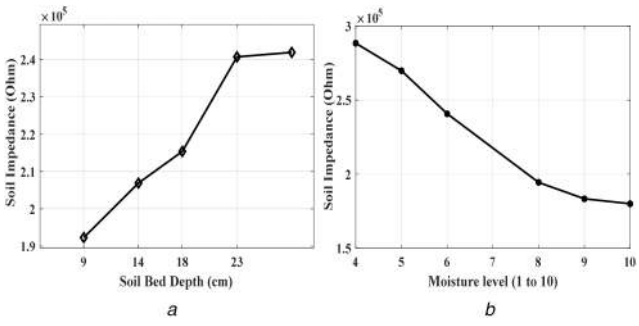

Fig. 11 *Soil impedance variation with soil bed depth and soil moisture variation*
(a) Depth, ***(b)*** Moisture level

The moisture level is varied within a scale of 1 to 10. Here, scales 1 and 2 are considered to be dry, scales 3–7 are considered to be moist, and scales 8–10 are considered as wet.

In Figs. 10*a* and *b* the variation of soil impedance with soil bed depth and soil moisture level is shown, respectively. It is clear that with the increase of soil bed depth, the impedance increases (Fig. 11). Similarly, with the increase in soil moisture level, impedance decreases. As EMD is suitable for arc signals, IMFs for the arc voltages are calculated. Tables 4 and 5 show the IMFs and their respective predominant frequencies for different experiment scenarios. The predominant frequencies of the arcing events confirm the presence of even-order harmonics of varying PSDs and multiple numbers of fundamental frequencies. The reason behind the presence of even-order harmonics being predominant in the IMF is the high-impedance arcing surfaces. These IMFs further used for the classification of arc voltages using different SVMs and KNNs. The classification results for SVMs and KNNs are presented in Tables 6 and 7, respectively. The number of IMFs, which is considered for the analysis, is 7. The results obtained after machine learning clearly shows that the arc voltages are successfully classified, and the performance of KNN and SVM in this regard is consistent.

Table 4 Predominant frequencies found from the application of EMD on soil arc voltage signals (set 1)

IMFs	Predominant frequency (Hz) for set 1					
	Depth of soil bed (in cm) and soil moisture level (on a scale of 1–10)					
	Soil moist: 6 9 cm (arc voltage: 10 kV)	Soil moist: 9 9 cm (arc voltage: 9 kV)	Soil moist: 6 14 cm (arc voltage: 12 kV)	Soil moist: 9 14 cm (arc voltage: 11 kV)	Soil moist: 6 18 cm (arc voltage: 14 kV)	Soil moist: 9 18 cm (arc voltage: 13 kV)
IMF1	32,950	1300	4800	7950	200	16,600
IMF2	2950	4800	6100	4400	8050	8200
IMF3	1450	3050	3500	200	3450	3950
IMF4	200	1700	550	750	350	2200
IMF5	300	150	550	450	250	50
IMF6	150	200	200	200	50	150
IMF7	50	100	50	50	50	50
IMF8	not found	50	not found	50	not found	50
IMF9	not found	50	not found	not found	not found	not found

Table 5 Predominant frequencies found from the application of EMD on soil arc voltage signals (set 2)

IMFs	Predominant frequency (Hz) for set 2					
	Depth of soil bed (in cm) and soil moisture level (on a scale of 1–10)					
	Soil moist: 4 23 cm (arc voltage: 15 kV)	Soil moist: 6 23 cm (arc voltage: 13 kV)	Soil moist: 10 23 cm (arc voltage: 9 kV)	Soil moist: 5 28 cm (arc voltage: 17 kV)	Soil moist: 6 28 cm (arc voltage: 15 kV)	Soil moist: 8 28 cm (arc voltage: 14 kV)
IMF1	1000	250	2700	2000	1000	2850
IMF2	5550	7250	7700	7050	5550	5700
IMF3	3900	2850	2450	4300	3900	2200
IMF4	1750	1800	1750	2000	1750	1200
IMF5	1000	850	200	200	1000	750
IMF6	500	50	300	200	500	350
IMF7	50	50	50	50	50	50
IMF8	50	50	50	50	50	not found
IMF9	not found	not found	not found	not found	not found	not found

Table 6 Arc classification result for different arcing surface using machine learning algorithms and EMD (set 1)

Different test sets	Arcing voltage in the different surfaces, kV					No. of input features	Classification accuracy using different classifiers, %			
	Concrete	Grass	Tree	Wet sand	Soil		LSVM	QSVM	CSVM	KNN
test 1	20	12	7	4	10	7	83.5	99.0	99.6	99.3
test 2	10	14	11	10	9	7	92.0	98.1	99.4	99.8
test 3	14	16	13	5	12	7	83.3	99.6	99.8	99.8
test 4	16	6	15	15	11	7	78.1	97.1	99.7	99.9
test 5	20	10	22	17	14	7	89.5	99.8	99.9	99.8
test 6	24	22	26	19	13	7	81.1	99.7	99.9	99.5

Table 7 Arc classification result for different arcing surface using machine learning algorithms and EMD (set 2)

Different test sets	Arcing voltage in the different surfaces, kV					No. of input features	Classification accuracy using different classifiers, %			
	Concrete	Grass	Tree	Wet sand	Soil		LSVM	QSVM	CSVM	KNN
test 1	20	12	7	4	15	7	93.1	99.2	95.6	99.4
test 2	10	14	11	10	13	7	92.0	99.0	95.2	99.8
test 3	14	16	13	5	9	7	83.2	99.5	99.8	99.8
test 4	16	6	15	15	17	7	88.6	99.5	99.7	99.9
test 5	20	10	22	17	15	7	89.7	99.8	99.9	99.8
test 6	24	22	26	19	14	7	81.2	99.6	99.8	99.4

6 Conclusion

The findings obtained from all the studies of different arcing surfaces clearly show the differences in the time-frequency domain analysis. Nevertheless, the predominant frequency of the entire event is quite comparable and shows behaviour close to that of HIF. In all the arcing cases, the presence of second-order harmonics is quite identical to HIF. Apart from that, different arcing surfaces show a different kind of characteristic during the experiments concerning pattern and arcing path impedance. As a result, except for the second-order harmonic, other predominant maximum frequencies are different, which is evident from the PSD plots of different IMFs. Furthermore, these differences in IMF due to inherent frequencies emerge as an essential feature for the successful classification of HIAFs. On the other hand, the classification results using ST coefficients are also found to be promising. The classification accuracy mostly lies within 99–100%

for SVM, QSVM, and CSVM classifiers. Therefore, from both the approaches, it can be concluded that both the signal processing tools are capable of classifying different arc faults. However, there is a significant amount of difference between the approaches regarding the machine-learning algorithm and the input features. EMD gives nine input features to the machine-learning tool, whereas ST provides 126 features for the learning. This small number of inputs to the network for EMD can be a factor of choosing EMD over ST for arc fault classification. The main conclusions and novelties of the work are as follows:

- In HIAF, the fundamental frequency can be found with more than one PSD, and it also changes with the arcing surface.
- EMD and ST were found to be very much effective in analysing arc signals.
- EMD along with QSVM, CSVM, and KNN successfully classified the HIAF for concrete, tree, wet sand, grass, and soil arcing. ST along with SVM, QSVM, and CSVM also present a significant classification performance.
- The number of input features to the learning algorithms for EMD is 9. Whereas, ST-based approach takes 126 features for the successful classification. Therefore, EMD is found to be more computationally viable for the classification of arc signals.
- EMD, along with QSVM, CSVM, and KNN successfully classified the HIAF and similar switching transients.
- In-depth analysis of arc in soil shows similar frequency characteristics such as other HIAF in terms of the presence of even-order harmonics of varying PSDs and multiple numbers of fundamental frequencies with different PSDs.

However, the computational burden due to ST can be optimised if required as it has shown a substantial potential of analysing arc. The future research is planned with an emphasis on the variety of other arcing surfaces (i.e. different kinds of soils and ashes, a mixture of different kinds of soils, a mixture of soil–ash and soil–sand, a compound mixture of soil–ash–sand etc.), and the possibility of optimisation in the performance of ST. Moreover, the present study takes a significant small step towards the development of an online next generation intelligent fault diagnosis system based on signal processing and machine learning. In this regard, the proposed algorithm will be implemented in general central processing units, such as high-performance array processors, digital signal processors, field programmable gate arrays, and other processors. With the help of different high-performance processors, the optimisation of expected detection delay is also planned for the future.

7 References

[1] Tengdin, J., Westfall, R., Stephan, K.: 'High impedance fault detection technology', PSRC Report, 1996, pp. 1–4

[2] Aucoin, B.M., Jones, R.H.: 'High impedance fault detection implementation issues', *IEEE Trans. Power Deliv.*, 1996, **11**, (1), pp. 139–148

[3] Yang, K., Zhang, R., Yang, J., *et al.*: 'A novel arc fault detector for early detection of electrical fires', *Sensors*, 2016, **16**, (4), p. 500

[4] Administration, USF, Fire, N.: 'Residential building fire causes', 2018, pp. 1–13

[5] Chen, J.C., Phung, B.T., Zhang, D.M., *et al.*: 'Study on high impedance fault arcing current characteristics'. 2013 Australasian Universities Power Engineering Conf. (AUPEC), Hobart, Tasmania, Australia, 2013), pp. 1–6

[6] Chen, J., Phung, T., Blackburn, T., *et al.*: 'Detection of high impedance faults using current transformers for sensing and identification based on features extracted using wavelet transform', *Transm. Distrib. IET Gener.*, 2016, **10**, (12), pp. 2990–2998

[7] Ghaderi, A., Ginn, H.L., Mohammadpour, H.A.: 'High impedance fault detection: a review', *Electr. Power Syst. Res.*, 2017, **143**, pp. 376–388

[8] Routray, P, Mishra, M, Rout, P.K: 'High impedance fault detection in radial distribution system using S-transform and neural network'. 2015 IEEE Power Communication Information Technology Conf., Bhubaneswar, Odisha, India, 2015, pp. 545–551

[9] Milioudis, A.N., Andreou, G.T., Labridis, D.P.: 'Detection and location of high impedance faults in multiconductor overhead distribution lines using power line communication devices', *IEEE Trans. Smart Grid*, 2015, **6**, (2), pp. 894–902

[10] Gautam, S., Brahma, S.M.: 'Detection of high impedance fault in power distribution systems using mathematical morphology', *IEEE Trans. Power Syst.*, 2012, **28**, (2), pp. 1226–1234

[11] Gautam, S., Brahma, S.M.: 'Application of mathematical morphology based filters to detect a power swing'. IEEE PES General Meeting, PES 2010, Minneapolis, USA, 2010, pp. 1–6

[12] Sarlak, M., Shahrtash, S.M.: 'High-impedance faulted branch identification using magnetic-field signature analysis', *IEEE Trans. Power Deliv.*, 2013, **28**, (1), pp. 67–74

[13] Yu, D.C., Khan, S.H.: 'Adaptive high and low impedance fault detection method', *IEEE Trans. Power Deliv.*, 1994, **9**, (4), pp. 1812–1818

[14] Lai, T.M., Snider, L.A., Lo, E., *et al.*: 'High-impedance fault detection using discrete wavelet transform and frequency range and RMS conversion', *IEEE Trans. Power Deliv.*, 2005, **20**, (1), pp. 397–407

[15] Livani, H., Evrenosoglu, C.Y.: 'A fault classification and localization method for three-terminal circuits using machine learning', *IEEE Trans. Power Deliv.*, 2013, **28**, (4), pp. 2282–2290

[16] Ghaderi, A., Mohammadpour, H.A., Ginn, H.L., *et al.*: 'High-impedance fault detection in the distribution network using the time-frequency-based algorithm', *IEEE Trans. Power Deliv.*, 2015, **30**, (3), pp. 1260–1268

[17] Elkalashy, N.I., Lehtonen, M., Darwish, H.A., *et al.*: 'Modeling and experimental verification of high impedance arcing fault in medium voltage networks', *IEEE Trans. Dielectr. Electr. Insul.*, 2007, **14**, (2), pp. 375–383

[18] Liu, W., Zhang, X., Ji, R., *et al.*: 'Arc fault detection for AC SSPC in MEA with HHT and ANN'. IEEE Int. Conf. on Aircraftschaft Utility Systems (AUS), Beijing, China, 2016, pp. 7–12

[19] Lala, H., Karmakar, S.: 'Characterization and classification of high impedance arc fault in distribution system involving metallic and nonmetallic surfaces using empirical mode decomposition', *Int. Trans. Electr. Energy Syst.*, 2019, **29**, (11), pp. 1–16, e12107

[20] Lala, H., Karmakar, S.: 'Classification of arc fault in sphere-gap and rod-gap using Stockwell transform and machine learning based approach'. Proc. 2019 Int. Conf. on High Voltage Engineering and Technology, ICHVET 2019', Hyderabad, India, 2019

[21] Niayesh, K., Runde, M.: '*Power switching components*' (Springer International Publishing AG, Gewerbestrasse 11, 6330 Cham, Switzerland, 2017), pp. 13–57

[22] Huang, N., Shen, Z., Long, S., *et al.*: 'The empirical mode decomposition and the Hilbert spectrum for non-linear and non-stationary time series analysis', *Proc. R. Soc. A Math. Phys. Eng. Sci.*, 1998, **454**, p. 995, 903

[23] Huang, N.E., Shen, S.S.: 'Hilbert–Huang transform and its applications', *Control*, 2005, **5**, p. 311

[24] Flandrin, P., Rilling, G., Gonc, P.: 'Empirical mode decomposition as a filter bank', *IEEE Signal Process. Lett.*, 2004, **11**, (2), pp. 112–114

[25] Deering, R., Kaiser, J.F.: 'The use of a masking signal to improve empirical mode decomposition'. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)–Proc., Philadelphia, USA, 2005, pp. 485–488

[26] Senroy, N., Suryanarayanan, S., Ribeiro, P.F.: 'An improved Hilbert–Huang method for analysis of time-varying waveforms in power quality', *IEEE Trans. Power Syst.*, 2007, **22**, (4), pp. 1843–1850

[27] Senroy, N., Suryanarayanan, S., Ribeiro, P.F.: '*Time-varying waveform distortions in power systems*' (John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom, 2009)

[28] Cohen, L.: 'Time-frequency distributions – a review', *Proc. IEEE*, 1989, **77**, pp. 941–981

[29] Stockwell, R.G., Mansinha, L., Lowe, R.P.: 'Localization of the complex spectrum: the S transform', *IEEE Trans. Signal Process.*, 1996, **44**, (4), pp. 998–1001

[30] Boser, B.E., IMG and VNV: 'Quadratic kernel-free non-linear support vector machine', *J. Glob. Optim.*, 2008, **41**, (1), pp. 15–30

[31] Tan, S.: 'Neighbor-weighted K-nearest neighbor for unbalanced text corpus', *Expert Syst. Appl.*, 2005, **28**, (4), pp. 667–671

[32] Karmakar, S.: 'Virtual-instrument-based online monitoring system for hands-on laboratory experiment of partial discharges', *IEEE Trans. Educ.*, 2017, **60**, (1), pp. 29–37

[33] Tektronix TDS2000C data-sheet: 'Digital storage oscilloscope', 2010

[34] Cui, Q., El-Arroudi, K., Weng, Y.: 'A feature selection method for high impedance fault detection', *IEEE Trans. Power Deliv.*, 2019, **34**, (3), pp. 1203–1215

[35] Emanuel, A.E.E., Cyganski, D., Orr, J.A.A., *et al.*: 'High impedance fault arcing on sandy soil in 15 kV distribution feeders: contributions to the evaluation of the low frequency spectrum', *IEEE Trans. Power Deliv.*, 1990, **5**, (2), pp. 676–686