

Received XX Month, XXXX; revised XX Month, XXXX; accepted XX Month, XXXX; Date of publication XX Month, XXXX; date of current version XX Month, XXXX.

Digital Object Identifier 10.1109/OAJPE.2023.1234567

High Impedance Fault Location Methods: Review and Harmonic Selection-based Analysis

Gabriela N. Lopes*, Thiago S. Menezes*, Douglas P. S. Gomes[†] AND Jose Carlos M. Vieira*, MEMBER, IEEE

¹Department of Electrical and Computer Engineering, São Carlos School of Engineering, University of São Paulo, São Carlos, Brazil

²Victoria University, Melbourne, Australia

(Invited Paper)

CORRESPONDING AUTHOR: Gabriela N. Lopes (e-mail: gabrielanuneslopes@usp.br).

Financed by the Coordination for the Improvement of Higher Education Personnel-Brazil (CAPES)-Finance Code 001, and by grant 2020/06935-5 São Paulo Research Foundation (FAPESP) and Brazilian National Council for Scientific and Technological Development (CNPq) grant 304900/2018-4.

ABSTRACT High Impedance Faults (HIFs) are recurring events in electrical Distribution Systems (DSs) and occur by the contact between energized conductors and high impedance surfaces. HIFs may pose hazards to living beings and cause bushfires. However, the HIF protection has not been completely solved due to the small fault current and varying impedance, inhibiting traditional protection techniques from functioning correctly. In the literature, researchers have mainly focused on detection techniques. Thus, the development of HIF Location Methods (HIFLMs) is recent, and evidences for conclusive solutions are still lacking. Moreover, to this date, no existing study reviews the main challenges concerning HIFLMs in DSs. This paper proposes a systematic analysis of the common stages to design the main existing HIFLMs. The strategy is evaluating the similar characteristics that pose a common research path regarding challenges faced in real-world conditions. Additionally, this paper proposes a case study to assess the best input signals, metrics, and machine learning-based decision algorithms of a new HIFLM. The results are promising, with high identification rates, even in noisy conditions. The methodology can help to select the datasets for supervised learning-based HIFLM. Highlighting the state-of-art of current methods and support development of HIFLMs are this paper's main contributions.

INDEX TERMS distribution system, high impedance fault location, Stockwell transform, random forest

I. Introduction

High Impedance Faults (HIFs) represent 5 to 20% of fault occurrences in Distribution Systems (DSs) [1]. They occur due to the contact between energized conductors and high impedance surfaces, such as vegetation, asphalt, sand, grass, or cement. They can occur when conductors break [2] or when the conductor contacts vegetation [3]. One of the main problems related to HIFs is the risk of exposing humans and animals to electric shock and causing bushfires [3]. HIFs can also cause interruption of the electricity supply when it results in a conductor rupture.

Due to the previously mentioned issues, it is essential to develop solutions for HIF protection on DSs. The interest in developing studies regarding this type of fault started in the 70s [3]. Since then, efforts have been mainly focused on fault detection. However, a complete protection scheme should also include the fault location, which decreases the time needed for utility maintenance teams to identify the fault location and restore the energy supply.

There are several constraints to the HIF location, such as the fault's non-linear impedance and low fault current. Typically, in low impedance faults location techniques, the error in estimation due to load current variation after the

fault can be ignored. However, for HIFs, this consideration is inaccurate, as the fault current is usually lower than the load current. Moreover, due to the nonlinearity and low amplitude of the HIF current, the uncertainty in the estimated distance is considerable, increasing the chance of errors [4].

When analyzing the existing HIF Location Methods (HIFLMs), there is no consensus regarding a new algorithm's tests. A wide range of metrics and decision algorithms are used to locate the fault. In the literature, there are comprehensive studies regarding HIF detection [5], [6]. These studies also comprise a few HIFLMs, but they do not perform a specific analysis of their challenges and development stages. Thus, there are no previous studies concerning a critical and comprehensive comparison of existing HIFLMs in DSs. Hence, this paper's extended version performs an unprecedented critical review of these techniques. The proposed approach is to identify the main steps addressed to develop each method and, by comparing them, outline the characteristics that make an algorithm more appropriate to real-world conditions. Hence, this study can contribute to establish the state-of-the-art, also performing a review of commercial solutions. Furthermore, to help researchers develop new and practical HIFLM, this paper also proposes an analysis of training and test data for a new HIFLM. The methodology aims to establish the best approach for a supervised-learning-based HIFLM regarding its training/testing data and noise levels. The analysis consists of several tests, selecting the input signals, extracting the metrics, and determining the features. The results show high identification rates, revealing the best solutions for a study case.

Accordingly, this paper has been organized as follows. Section II exhibits the critical review of HIFLMs. Section III presents the proposed analysis methodology, and Section IV shows the results. Lastly, Section V concludes the analysis.

II. A Review on the Stages to Develop HIFLMs

The selected HIFLMs discussed in this paper were published on major journals and conferences in the area. There are few existing techniques for HIF location, as it is a recent research field, starting in 2012. The selected papers aim at localizing HIFs in DSs and present comprehensive tests for their real applicability. Their main aspects are described in this paper.

After selecting the techniques, the review methodology was addressed, identifying similar characteristics among them. Overall, HIFLMs are developed considering the following stages: i) data acquisition through simulation or real tests; ii) acquired data processing, iii) metrics extraction, and iv) a decision algorithm using a mathematical technique or a classifier to locate the fault [7]. Thus, this section presents an analysis of the stages executed when developing and testing HIFLMs. The purpose is to conduct a critical study, identifying the challenges and tests required to delimit a HIFLM application.

A. Review Methodology

The first task in this study was selecting the main HIFLMs in the literature. In the sequence, similar characteristics among the techniques were identified. Overall, the main conclusions regarding the similar stages of the selected HIFLMs are:

- **Stage I - data acquisition:** Deals with the ways to obtain an HIF signal to test an algorithm. The first way is by using a simulation software to model the HIF computationally. The second one is through tests in real situations. For both approaches, it is necessary to specify how the fault is applied and the signal's measurement location.
- **Stage II - data processing:** Assesses the signals measured during HIFs in different situations, such as when varying the system loading, the existence and penetration level of Distributed Generators (DGs), the number of measurement devices, the fault location, and the noise level. In general, these are the conditions to that new HIFLMs can face in a practical application.
- **Stage III - metrics extraction:** Investigates the metrics extracted from the signals. They are the researchers' tools to locate the HIF in a given technique. Consequently, comparing and studying them can reveal the best approaches for new HIFLMs.
- **Stage IV - decision algorithm:** Defines the decision algorithm of HIFLMs. It evaluates how the location is estimated, the possibility of multiple estimations, and the threshold determination method, if applicable.

The following sections analyze each stage, and present a comprehensive comparison among different methods, which propose HIFLMs in DSs with sufficient information to make a fair comparison. The strategy is listing the elements that imply safe and effective protection and summarizing the existing HIFLMs characteristics. Hence, a critical and detailed analysis is developed.

B. Stage I: Data acquisition

Firstly, HIFLMs require input data. Thus, it is necessary to acquire real signals in a field or use a software-based model that emulates the HIFs. Performing real HIF measurements can often be difficult, especially considering different conditions, such as fault location and system loading. Nevertheless, the HIF must be adequately modeled to assess a fault location algorithm by using simulations.

The models must represent the HIF's characteristics previously established in the literature. The first characteristic is the small fault current magnitude. Also, the HIF current is distorted due to the nonlinear electrical arc resistance and implies in low order odd harmonics in the signals. The HIF currents are also asymmetric, resulting in even harmonics. Another characteristic is the build-up, given by the progressive increase in current amplitude at the fault's beginning. The shoulder characteristic represents the periods in which the fault current amplitude does not present significant changes for a few cycles. There is also modulation, which is

the progressive variation of the current amplitude over time. Moreover, HIF currents have the avalanche characteristic, which are the periods the current amplitude keeps zero when the voltage signal crosses zero until it reaches a sufficient value for the current to flow. It generates high-frequency components in HIF signals. Lastly, there is the intermittence, the period when the fault current extinguishes, with subsequent re-ignition. It is noteworthy that all these features occur randomly along the HIF due to the random nature of the electric arc. All these characteristics can be seen in a typical HIF current waveform in Figure 1 and are thoroughly described in [8].

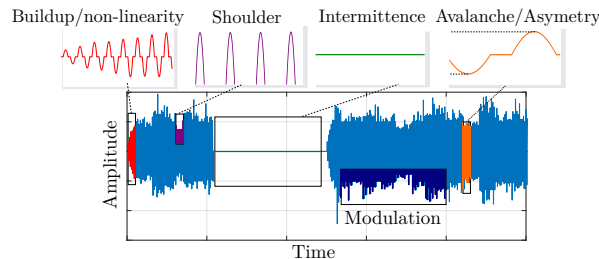


FIGURE 1 Main HIF characteristics.

Table 1 summarizes the HIF representation in the selected HIFLMs's tests. Only two techniques use real signals. In some methodologies, the HIF was simulated using only a high-value resistance. However, it may be insufficient, as it does not provide the main HIF characteristics. Furthermore, the evaluation performed by [9] and [1] used fault resistances from 60 Ω to 100 Ω , and 19 Ω , respectively. These impedance values can be considered low for HIFs in medium voltage DSs. Thus, their practical application may be affected. Table 1 also shows that some techniques were tested using a non-linear model. Despite representing resistance variation and phenomena such as avalanche, buildup and modulation, these models do not randomly vary the characteristics (which occurs in real signals). Consequently, it can affect the HIFLMs performance in real conditions. In [10], the tests used both an HIF model and real signals. The results revealed that the technique's success rate decreases when using real signals, proving that considering the actual HIF random characteristics is crucial for evaluating HIFLMs. Some techniques, such as [11] and [12], consider an analytic fault model in the methods, which makes considering the HIF randomness difficult. In this situation, using an accurate HIF model becomes even more relevant.

C. Stage II: Data processing

This section presents the main issues in DSs that can affect HIFLMs, such as the input signal, sampling frequency, electrical system size, number of measurements, noise, and DGs.

1) The influence of the input signal and sampling frequency
Table 2 summarizes the signals used as input data for the HIFLMs. Most approaches use both voltage and current

TABLE 1 HIF representation in the tests in the literature.

Reference	HIF representation		
	Pure Resistance	Non-linear (not random)	Real
[13]	✓		
[9]	✓		
[14]	✓		
[15]	✓		
[16]	✓		
[17]	✓		
[1]	✓		
[7]	✓		
[18]		✓	
[4]		✓	
[19]		✓	
[20]		✓	
[11]		✓	
[21]		✓	
[12]		✓	
[22]		✓	
[23]		✓	
[10]		✓	✓
[24]			✓

signals. Current transformers might respond as low-pass filters, resulting in the loss of part of the signal information. Thus, this aspect can restrict the use of current signals. In contrast, using voltage is only efficient when there is a sufficient voltage drop between the measurement and fault location. Therefore, as the HIF current amplitude is small, the measured voltage may not be affected [25]. Reference [24] proposes the use of magnetic field for HIF location. Nevertheless, using this signal requires specific devices for the measurements, as it is not usual in protection schemes.

TABLE 2 The input electrical signal, sampling rate, and software used in the tests of HIF location techniques.

Reference	Signal	Sampling Rate	Software
[18]	V and I	15.36, 61.44 and 107.52 kHz	ATP ^a
[4]	V and I	128 samples per cycle	EMTP ^b -RV
[20]	V and I	64 samples per cycle	ATP
[11]	V and I	256 samples per cycle	ATP
[10]	V and I	8000 samples per second	ATP/Matlab
[23]	V and I	192 samples per cycle	ATP/Matlab
[21]	V and I	*	Simulink/Matlab
[12]	V and I	*	ATP/Matlab
[14]	V and I	0 to 100 kHz	*
[7]	V and I	*	*
[13]	V and I	*	*
[9]	V	128 samples per cycle	PSCAD ^c /Matlab
[19]	V	*	PCFLO
[16]	V	*	Simulink/Matlab
[1]	V	*	Matlab/PowerWorld
[15]	I	20 kHz	Matlab
[22]	I	5, 4, 3, 2, and 1 kHz	RTDS ^d
[17]	I	*	RTDS
[24]	MF	1024 samples per cycle	Simulink/Real

*Information not mentioned; V-Voltage, I-Current, MF-Magnetic Field;
^aATP: Alternative Transients Program; ^bEMTP: Eletromagnetic Transients Program; ^cPSCAD: Power System Computer Aided Design; ^dRTDS: Real-Time Digital Simulator.

Table 2 also shows different sampling frequencies used in HIFLMs. When using a low sampling rate, only the HIF characteristics that impact the signal's low-frequency components can be analyzed. Consequently, HIFLMs based

on high-frequency characteristics require a high sampling rate. However, the measurement equipment becomes more expensive, and its computational burden increases. In general, as seen in those papers, there is no consensus on the optimal sampling frequency used to acquire the input signals of HIFLMs.

Additionally, Table 2 addresses the software used for testing the techniques. The purpose is to inform the reader on the most used simulation software for testing new HIFLMs. As shown by the table, these are transient modeling programs rather than steady state or power flow-only software. An example is the Alternative Transients Program (ATP) [26], which is widely employed for studying electromagnetic transients. It is a version of the Electromagnetic Transients Program (EMTP) [27], which is also used for this type of analysis. Another software for this purpose is PSCAD [28], which is used in [9]. There are also tests with the Real-Time Digital Simulator (RTDS) [29]. It uses real-time hardware to perform the algorithm calculation, similar to practical applications. It is worth highlighting the use of Matlab [30], as a circuit simulator via Simulink and a tool for processing the algorithms. Moreover, other software are the PCFLO and the PowerWorld [31], used in [19] and [1] respectively.

2) The influence of the meters position on HIFLMs

Usually, DSs have measurement devices in the substation. However, due to HIFs' low fault current, identifying them using a single meter becomes challenging. The current measured at the substation is the sum of the HIF and the system load current. Also, the voltage at the fault location is different from the measured voltage due to the line's voltage drop. Thus, many techniques require using multiple meters in different system locations. Nonetheless, tests concerning the HIFs incidence in several areas must be conducted to delimit the region the method can locate the fault.

Table 3 outlines the systems size and measurement devices in HIFLMs tests. Most methods were tested on small systems with less than thirteen nodes and a single measurement location. The algorithms tested on large systems used more than one measurement location. The exception is the 1436-node system [10], which uses just one meter. However, the HIF was simulated at a single location.

The higher the number of meters in relation to the system size, the easier it is to locate the fault. However, the utilities need to invest more in measurement devices, increasing the costs. Also, the measurement allocation requires using an optimization algorithm or a statistical analysis.

3) The Influence of Distributed Generators on HIFLMs

Usually, DSs have a radial topology. Therefore, protection techniques' analysis often considers a unidirectional power flow. However, there has been an increasing connection of DGs in distribution systems. The DGs influence the general performance of the DSs operation, causing a bidirectional

TABLE 3 Test system size and number of meters used in HIF location techniques in the literature.

Reference	System (number of buses)						Measurement number
	<=13	23-38	63	90-123	454	>1000	
[13]	✓						1
[19]	✓			✓	✓		20 % of the nodes
[20]	✓						1
[11]	✓						1
[10]	✓					✓*	1
[16]	✓						8
[17]	✓						1
[1]	✓						**
[12]	✓	✓					1
[22]	✓						5
[23]	✓						1
[4]		✓					1
[9]		✓					1
[21]		✓					1
[24]			✓				5, one at each 3 km
[7]				✓			15
[18]				✓			6
[14]	-	-	-	-	-	-	1
[15]	System with lines 7.5, 6.35 and 12.5 km						1

* The HIF is only in one place in the 1436 nodes system; ** The exact number of devices is not mentioned; - The authors do not mention the amount of system buses, but there is a measurement at the system substation and other along with the system.

power flow. From this perspective, the protection system structures have to change due to the modification in the fault current amplitude and direction [32]. Thus, it has led to a complete revision of the methods by which DSs should be monitored, controlled, and protected [33]. Furthermore, inverter-based DGs can inject harmonics into the system, which may increase the non-linearity on the current and voltage waveforms in DSs [34]. As there are HIFLM approaches that rely on the signals' frequencies (harmonics), they can be affected by the presence of DGs.

Fault protection devices have to deal with different types of faults, operating conditions, imbalances, load variations, and topology modifications (which may increase with the DGs) [7]. These phenomena affect techniques that use the system's power/current flow to identify the fault. Additionally, since the DGs' penetration levels can vary over time, techniques based on decision algorithms that require training may be affected. This last aspect will be addressed in Stage IV. Among the selected HIFLMs, just the approaches presented in [7] and [18] consider the presence of DGs in the system. Only [7] analyzed the influence of DG penetration level variation on the algorithm's operation. This analysis becomes increasingly important for modern DSs. In these cases, the DGs presence had little affected on the accuracy. Nonetheless, there was no analysis involving inverter-based energy resources.

4) The Noise Influence on the Measurements

The signals measured in real DSs usually have noise. Thus, it is essential to consider their influence when developing protection techniques. The noise may impose a low impact in approaches that use only the signal's fundamental component

as a metric. In contrast, techniques based on high-frequency components may be more affected by it [18]. The noise in DSs signals can be modeled as white Gaussian noise (WGN). It usually has a Signal to Noise Ratio (SNR) between 50 to 60 decibels (dB) [35]. However, some HIFLMs are tested with higher noise levels to evaluate their noise immunity.

Table 4 outlines that most HIFLMs are not tested under noisy conditions. Among the tested algorithms, there is no consensus on the noise level. An ideal approach is accomplished by [24], which evaluated the techniques in a noise range, delimiting the method applicability.

TABLE 4 Noise level in HIF location techniques' tests.

Reference	No tests	Tests with noise	Noise added
[13]	✓		
[4]	✓		
[19]	✓		
[20]	✓		
[11]	✓		
[10]	✓		
[9]	✓		
[14]	✓		
[15]	✓		
[16]	✓		
[17]	✓		
[21]	✓		
[23]	✓		
[7]	✓		
[18]		✓	60 dB
[1]		✓	0.5%
[12]		✓	1%-voltage, 2%-current
[22]		✓	20 dB
[24]		✓	From 0 to 60 dB*

* The method is tested for all the noise levels between 0 and 60 dB, and only decreases the accuracy for signal-to-noise ratios lower than 15 dB.

D. Stage III: Metrics Used in HIFLMs

Table 5 summarizes the main metrics and tools used to extract or measure them in the selected HIFLMs. In general, the metrics are based on comparing measurements among devices or signal processing techniques.

One of the signal processing techniques used in the literature is the FT. Despite being widely employed in protection relays, it is a frequency domain transform, which considers the signal stationary in a fixed window. The frequency information calculated by FT is an average over the window. Therefore, the location of a specific transient on the time axis is unknown, as it is not related to the temporal domain [36]. Additionally, there is no consensus regarding the size and shape of the window used in FT-based transforms. As a result, these issues must be addressed when using this technique in a protection algorithm. Another method used to extract the metrics in HIFLMs is the WT. It is a time scale domain technique that captures the instant in time of the signal's frequency components [25]. Signals measured during HIFs usually present discontinuities near the current's zero-crossing region due to the avalanche effect, which can be identified by changes in the coefficients obtained through WT [36]. Despite the listed benefits, it is considered a challenge to design a systematic protection technique based on

TABLE 5 Extraction technique and metrics used in HIF location methods in the literature.

Reference	Metric	Extraction
[18]	Energy of the first detail level	V*-WT ^g
[9]	The three first detail levels	V-WT
[15]	5 th detail level	I*-WT
[10]	Parameter estimation and WLS ^a	I and V-FT ^h
[23]	1 st and 3 rd harmonic orders	I and V-FT
[13]	Feeder capacitive current	I and V-Power flow
[4]	Pre and post fault nodes voltages	I and V-Power flow
[20]	Lines capacitive currents	LM ^j algorithm
[17]	ZSC ^b non-null QDC ^c	I
[22]	ZSC phase DPC ^d	I
[1]	V magnitude	V measurements
[21]	I and V magnitude	Measurement relay
[12]	R-squared of the fitted lines	I and V-lines modeling
[7]	Voltage and current RMS ^e and phase	PMU ⁱ
[11]	LSE ^f and steps descent	I and V
[14]	Impulse between devices	Power line communication
[16]	Low V unbalance	Smart meters
[24]	Magnetic field's high-frequency phase	Magnetic field sensor

* I-Current; V-Voltage; ^aWLS-Weighted Least Square; ^bZSC- Zero Sequence Current; ^cQDC- Quasi-Differential Component; ^dDPC- Declining Periodic Component; ^eRMS- Root Mean Squared; ^fLSE - Least Square Estimation; ^gWT- Wavelet Transform; ^hFT-Fourier Transform; ⁱPMU- Phasor Measurement Unit; ^jLM - Levenberg-Marquardt.

WT due to the narrow high-frequency support and subjective selection of the mother wavelet. Also, there is a resolution loss since the signal's samples are divided by two at each decomposition level [25], [36]. However, this problem can be solved by using the maximal overlap discrete wavelet transform (MODWT), for example, which does not present undersampling problems when increasing the decomposition level [8]. Table 5 also reveals that the zero-sequence current can be used as a metric to locate the HIF, as the HIFs are usually a single or two-phase event and can unbalance the system. Such methodologies must consider tests in steady-state unbalanced systems and the HIF incidence on end-of-branches nodes.

Table 5 shows that metrics based on the system's power flow are also used for HIF location. In these techniques, the metrics are given by the system's modeling and the variations in the measured current and voltage signals. It is important to carefully analyze the system and HIF modeling when using such a strategy. Furthermore, the measurement can be performed along the system using Phasor Measurement Units (PMUs), Power Line Communication (PLC), and smart meters, as in [7], [14], and [16], respectively.

The PLC-based approaches are one of the few existing active methods for HIF location. Devices are set in a transmitter-receiver scheme, where a signal is injected at one line point and received at another. Multiple transmitters can be allocated in the system. Then, the fault position is estimated by analyzing the transmission quality in different network parts. Nevertheless, installing transmitters at every lateral branch is not economically feasible. Thus, using the received signal to estimate the channel (power line) characteristics can be more suitable. As a fault can create

a discontinuity, it can be estimated by changes in the PLCs channel's characteristics. The authors of [37] elaborate on how this can be achieved, but no practical implementation was found. PLCs have ongoing and established use in high-voltage systems, but their use in medium voltage is less frequent. Nonetheless, it has potential. The application of PLCs can expand, which implies that higher sampling rates are one of the most crucial factors in enhancing the DSs' monitoring and diagnostic capabilities.

E. Stage IV: Decision Algorithms of HIFLMs

After extracting the metric, the techniques decide the HIF location using a specific algorithm, which is summarized in Table 6. Note that methods that use multiple meters can indicate differences between the devices' metrics to decide the fault location. Nevertheless, there are also HIFLMs based on machine learning techniques, which can be robust for the HIF location. They present higher identification rates than analytical techniques and can be promising for this difficult-to-solve protection. However, they require training, which demands a large database, and the HIFLMs must also consider changes in the system's topology and load variation for their evaluation. Additionally, some methodologies indicate the distance between the meter and the fault, and other methods assign the HIF occurrence region. Overall, the purpose is to ensure that utility maintenance teams quickly dislocate to the correct fault location.

TABLE 6 Decision algorithms for the determination of the region or distance of the fault in HIF location methods.

Ref.	Decision Algorithm	Local	Threshold
[18]	Highest energy among the meters	Region	Adaptive
[11]	Least error based on parameters estimation	Region	Adaptive
[16]	Unbalance among voltages in low voltage	Region	Adaptive
[17]	ZSC ^f Quasi-differential non-null	Region	Adaptive
[9]	Comparison with a database	Region	Compare ^b
[14]	Impulse between IEDs ^a	Region	Adaptive
[13]	Variation of the impedance angle	Distance	Adaptive
[19]	Buses harmonic voltage comparison	Distance	Adaptive
[22]	Decrease of the zero sequence current	Distance	Adaptive
[24]	Magnetic field high-frequency phase change	Distance	Adaptive ^c
[4]	Estimated and real voltage error	Distance	Adaptive
[10]	Distance variation between iterations	Distance	Adaptive
[1]	Error between the estimated and real voltage	Distance	Adaptive
[12]	Fitted lines R-squared comparison	Distance	Adaptive ^d
[23]	WLS ^e with a parametric error processing	Distance	Adaptive
[20]	Artificial Neural Network (ANN)	Distance	Training
[15]	ANN and Gaussian process regression	Distance	Training
[21]	Adaptive neurofuzzy inference	Distance	Training
[7]	Artificial Neural Network	Distance	Training

^aIEDs-Intelligent Electronic Devices; ^bComparison with a database; ^c Depends on curve fitting the total phase shifts at different distances obtained from finite element analysis; ^dThreshold pre-determined for the system, but it adapts; ^eWLS-Weighted least square; ^fZSC-Zero-sequence current

Additionally, Table 6 reveals that most techniques use an adaptive threshold (the threshold varies according to the situation/system without requiring direct intervention). The extensive use of adaptive thresholds occurs because most techniques compare measurements in the same system to

determine the fault location. Thus, even if changes occur in the system, by comparison among meters, the HIF location is correctly determined. However, there is a need to establish a methodology to determine the quantity and location of the measurement devices. Moreover, it is important to point out the problem of multiple estimations in the fault location. It occurs in systems with various branches, where there may be zones with the same electrical distance to the meter [12]. Consequently, there are methods that offer alternatives to prevent multiple estimations, such as [12].

F. Review on commercial solutions

Despite being elaborate, the presented approaches are not prevalent or part of a standard use from power utilities. Commercial solutions are still in early development, and most are unattractive due to their high implementation and operational costs. Standard practices such as including more selective relay schemes, reclose blocking, fast fuse curves, and sensitive ground protection are considered best practices for HIFs [38], but the fast location is still unsolved. Commercial products often are in two categories: distribution level micro-PMUs or high-frequency sensing from line sensors. Both procedures rely on multiple measurements along the network because the load current is lower at lateral feeders, making the small HIF current more detectable. The high-frequencies components get attenuated due to system reactance, requiring a measurement close to the fault.

Schweitzer Engineering Laboratories (SEL) presents a whole-system integration for addressing HIFs and locating them [38]. However, these functionalities come with crucial considerations. For HIF detection, SEL recommends their proprietary Arc Sense technology based on interharmonics, which is embedded in their recloser device [38]. The location is proposed for the downer-conductor case with sensors installed in lateral feeders that transmit current readings to a receiver closer to the substation [39]. They also present an option using private 900-MHz LTE or DNP3 wireless networks to achieve broken conductor detection before it touches the ground. Nevertheless, there is evidence that not all HIF are arcing faults presenting inter-harmonic frequencies [40], and transmitters spacing appears to be limited to 400 meters distance. Arguably, any sufficient distributed monitoring systems can be used for localizing HIFs if those result from broken conductors. Fault signals are stronger close to the fault, thus allowing a better estimation of its location. This approach of making multiple measurements financially feasible was adopted by SilverFern Power, which developed self-powered sensors to be installed in several locations in all phases [41]. The devices are effectively micro-PMUs that can record waveforms and transmit system metrics through 3-5G networks. As the devices are closer to the load, they have a higher chance of detecting HIFs. Zero sequence currents can also be recorded. However, given the latency of cellular networks, it is arguable that the trip signal will not be sent before the conductor touches the ground.

Moreover, the company does not mention any technology for HIFs generated by contact with vegetation, for example.

Another approach is using high-frequency measurements to find signatures created by such faults, which was adopted by [42], through their sensor device that sits under power lines with antennas that capture the high-frequencies components. As high-frequency signals suffer high attenuation, such devices must be installed every ten kilometers. The signals are pre-processed at the edge and then sent to the cloud for further decision-making. The main advantage is that it can be used for other goals, such as predictive maintenance and early fault detection. The main downside is that it cannot be used as a tripping device due to its processing latency.

In general, the three described approaches illustrate the state of the commercial solutions for the HIF location: devices are in an early phase of production with standard practices still to be objectively set.

G. Overall Review

This section summarizes the main conclusions regarding the performed review of HIFLMs.

The analysis revealed that most HIFLMs' tests are not performed using actual HIF signals. Consequently, HIF models are employed to emulate the fault behavior. Additionally, the test systems usually have a low number of buses or present several meters spread throughout the circuit. The study also proved the importance of considering the influence of noisy signals, as many methodologies are based on the signal's frequencies extracted by signal processing techniques. Moreover, machine learning-based decision algorithms achieved high accuracy.

Hence, the next sections aim at helping to solve some of the issues new researchers can face when developing new HIFLMs. The first one regards the input signal, whether it is best to use the neutral or phase current signals when measuring them in one location and considering different noise levels. The second analysis focuses on establishing the best metrics based on the signals' frequencies to be used as input features. Lastly, the study comprises the best use for a machine learning-based decision algorithm.

III. Proposed Methodology for HIF location

The critical review concluded that there is no consensus regarding the input features, metrics, and decision algorithms of HIFLMs. Thus, aiming to help researchers to develop efficient tools, this paper proposes a new methodology for HIF location. The approach is based on selecting the input features that represent the HIF location and using a machine learning algorithm to provide a robust decision algorithm. The goal is to indicate the best input features of a HIFLM.

Figure 2 exhibits the proposed method's main steps. In the analysis, considering the HIF incidence at different system buses, the phase and neutral current measurements are kept at the system substation. The goal is evaluating

which input signal is more relevant to HIF location. The analysis considers the input signals with different noise levels (as common for real DSs). After the measurement, the harmonics of the signals are extracted through Stockwell Transform (ST) at each cycle. Then, the harmonics' energy at each cycle is calculated. The harmonics' energies are the input features used for the HIF location. To avoid overfitting and increase efficiency, an analysis on a feature-selection algorithm to select the metrics is presented. Finally, the Random Forest algorithm (a machine learning technique) is used as a decision algorithm for the HIF location. A study for its best use is also accomplished.

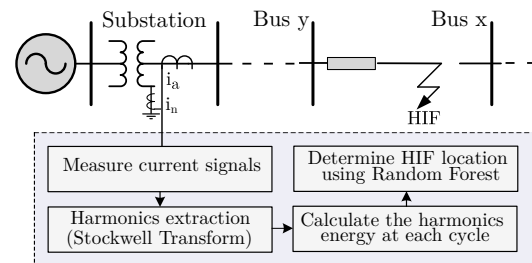


FIGURE 2 HIF location method main flowchart.

A. HIF simulation

The authors in [2] sampled real HIF currents in a medium voltage test system specially built for testing new HIF protection methods. The measurement occurred at the fault location, and the tests consisted of the conductor rupture with subsequent contact with different types of soil, such as sand, gravel, asphalt, cement, and clay, both wet and dry. The authors made the data available, which were used in this paper's tests. Some of the current waveforms are illustrated in Figure 3. Note that the currents have low amplitude, are non-linear, and have random characteristics, inhibiting the traditional fault location algorithms to proper functioning.

According to [43], current sources can be used to model harmonic-producing loads. Thus, the field-recorded HIF signals were inserted into the system using current sources controlled by Models in the ATPDraw (shown in Figure 4). The HIF simulation model also contains a switch to simulate the conductor breakage and another to connect the current source to the system.

The test system used in the study is the IEEE 34-node test system, whose single-line diagram is in Figure 5 [44]. This system was modeled through the ATP software [26]. All measurements were performed at bus 800 (the substation). The feeder has a total length of 55 km. Thus, dividing it into regions facilitates fault location. The regions were divided based on the electrical distance between the buses and the substation, considering the main three-phase branch and excluding the one-phase side branches. The main three-phase branch has a total impedance of almost 50 Ω . Thus, each region has between 5 and 10 Ω .

At each simulation, both phase and neutral currents obtained in the ATP software were transferred to MATLAB to

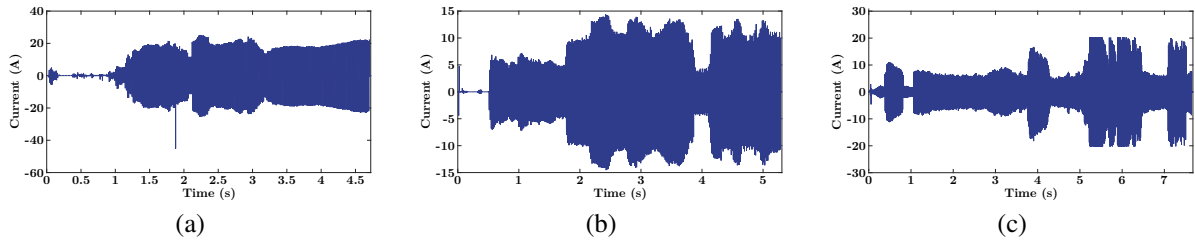


FIGURE 3 Some of the real HIF signals used in the tests, for HIFs on (a) clay, (b) gravel and (c) sand.

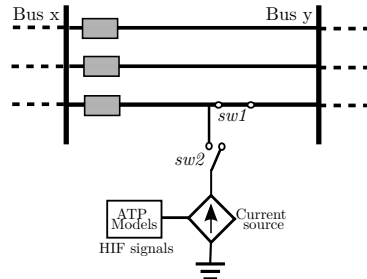


FIGURE 4 Real HIF currents insertion scheme.

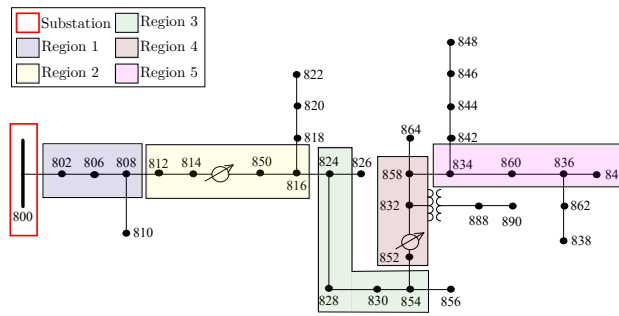


FIGURE 5 IEEE 34-node test system.

carry out the next steps of feature extraction and HIF location tests.

B. Metrics Extraction

The HIF current signals present non-linear characteristics due to the electric arc. Consequently, the frequency spectrum content extracted from the current signals measured during the HIF can be used to locate it. The metrics extraction in this analysis is accomplished by using the ST, a multiresolution spectral analysis technique that presents a great representation for time-varying harmonics [45]. It combines WT time reference ability with the FT capability of providing frequency information using a Gaussian-shaped window [45]. The inputs are the current signals. As output, it provides a matrix containing each signal sample's amplitude and phase. In this study, the ST is calculated at each one-cycle window, so the matrix dimensions are the harmonic orders per signal's samples. In the sequence, the energy of each harmonic is calculated at each cycle using Parseval's Theorem [46].

C. Features Selection Algorithm

As previously mentioned, this paper uses a machine learning technique as a decision algorithm to localize the HIF, using harmonics as input. The proposed study's metrics are the harmonics' energy at each cycle. However, not all harmonic orders may be necessary as input features to provide adequate accuracy.

In this sense, feature selection algorithms are often used to reduce the computation complexity and dimensionality in machine learning applications. In this paper, the FEAST toolbox [47] was used to select the most relevant features (the harmonics' energy) for the HIF location. Among the toolbox's feature selection algorithms, the minimal-redundancy-maximal-relevance (mRMR) is one of the best. Proposed by [48], this method selects the features with high relevance while maintaining a low redundancy among the selected features. Even though the method creates a rank of the given features, the number of features is not defined. Nevertheless, the features higher in rank are defined as the best ones. Hence, the mRMR was used to select the harmonic orders that best represent the HIF location in the system and can be used as input features for the decision algorithm. The study analyzes the real necessity of using the feature selection algorithm.

D. HIF Location Decision

The proposed HIFLM's decision technique is based on the Random Forest algorithm. It was chosen due to its positive performance after a set of tests with different machine learning-based algorithms. Random Forest is one of the most powerful machine learning techniques. It is based on an ensemble of multiple individual decision trees (thus creating a forest) [49]. Every decision tree is trained independently, using a randomly selected part of the training dataset. After training all the trees, the decision is given by the selective multi-voting method [49]. Its input data consists of the selected harmonics extracted from the HIFs' current signals. The algorithm outputs the region of the HIF location.

IV. Tests and Results

This paper's tests consider the current measurements during steady-state and the fault. In real conditions, the signals measured in DSs are noisy, varying between 50 and 70 dB [35]. Thus, the tests consisted of non-noisy and signals with a WGN with an SNR of 60 dB and 50 dB.

The HIF location algorithm assumes that the fault was previously detected. The algorithm input data consists of the first 1 s of the HIF. The signals used in the tests were acquired with 128 samples per cycle (7680 Hz). Therefore, according to the Nyquist Theorem, it is possible to obtain the up to the 64th order. Thus, for the signals acquired in the 60 Hz system during 1 s, 60 energy samples were obtained for each harmonic order as the Random Forest input features.

Overall, the analysis used 33 HIF signals applied to 18 three-phase buses with three noise levels, resulting in 1782 phase and 1782 neutral current signals. When using the energy of the amplitude of 64 harmonic orders for 1 s, 60 data samples for each harmonic result in 106,920 data for phase and neutral currents. Then, the dataset was divided into training and testing, which consisted of 55.55% and 44.44% of the data, respectively. HIFs on the terminal buses of each region were used in the training dataset, while HIFs on intermediate buses were utilized for testing. For example, in Region 2, HIFs on buses 812 and 816 are used for training, and buses 814 and 850 are used for testing.

The decision algorithm used in the proposed methodology is based on using a single Random Forest to classify the data among the five pre-defined regions (see Figure 5) or using five independent Random Forests, each one specialist in a region. For both approaches, the total number of decision trees was equal to 100. Furthermore, the decision algorithm's input data consists of using or not a feature selection algorithm to indicate the harmonics that best represent the database. As a result, the analysis is divided into four main cases to enable a comparison study:

- **Case I: Decision Algorithm-** One Random Forest; *Input Data* - The first 8, 16, 32, and 64 harmonic orders, with no feature selection;
- **Case II: Decision Algorithm-** One Random Forest; *Input Data* - The first 8, 16, and 32 selected harmonic orders using the mRMR algorithm;
- **Case III: Decision Algorithm-** Five specialist Random Forests; *Input Data* - The first 8, 16, 32, and 64 harmonic orders, with no feature selection;
- **Case IV: Decision Algorithm-** Five specialist Random Forests; *Input Data* - The first 8, 16, and 32 selected harmonic orders using the mRMR algorithm;

Thus, this section shows the results for each case, dividing the phase and neutral current analyses for each noise level.

A. Case I - One Random Forest without feature selection

Table 7 shows the accuracy achieved using the first 8, 16, 32, and 64 harmonic orders as input data for the single Random Forest to classify the HIF location. It reveals that the success rate is above 91% when training and testing with the same noise level using all the 64 harmonic orders available. The lowest accuracy was for the training with 50 dB noise and testing with no noise. When analyzing the 8, 16, and 32 first harmonic orders, the best accuracy was achieved when training with the non-noisy dataset. Using only the

first eight harmonic orders implied the worst results. Using 32 or 64 imply on similar accuracy. There was no significant difference between the phase and neutral signals.

TABLE 7 Case I - Accuracy rate when using one Random Forest, considering the first 8, 16, 32, and 64 harmonic orders as input features with no feature selection.

	Testing	Phase Current			Neutral Current		
		No noise	60 dB	50 dB	No noise	60 dB	50 dB
Training	8 th h _{arm} .	No noise	89.76%	84.66%	76.82%	89.95%	84.70%
		60 dB	84.85%	82.94%	76.59%	84.58%	83.00%
		50 dB	77.95%	77.86%	75.93%	78.23%	77.94%
	16 th h _{arm} .	No noise	97.61%	95.18%	88.46%	97.64%	95.27%
		60 dB	95.91%	95.75%	89.03%	95.85%	95.68%
		50 dB	92.41%	92.57%	91.72%	92.48%	92.59%
	32 th h _{arm} .	No noise	98.90%	92.67%	81.53%	98.92%	92.46%
		60 dB	94.81%	95.89%	82.59%	95.32%	95.93%
		50 dB	83.70%	85.38%	90.96%	83.44%	85.11%
	64 th h _{arm} .	No noise	98.64%	92.19%	82.70%	98.74%	92.20%
		60 dB	86.01%	95.91%	82.93%	85.81%	95.76%
		50 dB	72.83%	75.87%	91.65%	72.86%	75.79%

B. Case II - One Random Forest with feature selection

This case analysis used the rank of the 32 better-classified harmonics, shown in Table 8. It is important to consider the delimitation of input features. For example, if three input features are chosen, and there is no selection, the features are the DC, fundamental, and 2nd harmonic orders. However, if the ranking is used, the selected features are the DC, 12th, and 13th harmonic orders.

The results in Table 9 show the accuracy independence on the input current again. The accuracy increased compared to Case I, mainly when using the first 32 features selected. Note that, when training and testing with no noise, the accuracy was higher than 99%. Hence, the recommended situation for real applications is training the algorithm for all noise levels expected to be present, using 16 or more input features.

TABLE 8 Harmonics relevance by the feature selection algorithm.

Relevance	1	2	3	4	5	6	7	8
Harmonic	DC	12 th	13 th	14 th	15 th	16 th	21 st	31 st
Relevance	9	10	11	12	13	14	15	16
Harmonic	39 th	44 th	49 th	Fund	2 nd	3 rd	4 th	5 th
Relevance	17	18	19	20	21	22	23	24
Harmonic	6 th	7 th	8 th	9 th	10 th	37 th	17 th	22 nd
Relevance	25	26	27	28	29	30	31	32
Harmonic	40 th	45 th	50 th	11 th	32 nd	18 th	23 rd	38 th

C. Case III - Specialist Random Forests without feature selection

Table 10 shows the algorithm accuracy when using five independent Random Forests, each specialist in a region. This case used the first 8, 16, 32, and 64 harmonic orders as input without selecting their relevance. Overall, the best results were achieved using the first 16 harmonic orders as input data, presenting high accuracy even in noisy conditions. This shows that overfitting can occur in machine learning techniques when using a large number of inputs, decreasing their generalization capability. Similar to cases I and II, there

TABLE 9 Case II - Accuracy rate when using one Random Forest, considering the 8, 16 and 32 harmonics better ranked by the feature selection algorithm as input.

	Testing	Phase Current			Neutral Current		
		No noise	60 dB	50 dB	No noise	60 dB	50 dB
Training	8 harm.	No noise	95.54%	84.21%	70.42%	95.36%	84.35%
		60 dB	86.35%	88.88%	71.63%	86.31%	88.91%
		50 dB	74.09%	75.33%	81.53%	74.01%	75.65%
	16 harm.	No noise	98.86%	93.00%	84.42%	98.79%	92.97%
		60 dB	95.85%	96.35%	84.52%	95.16%	96.38%
		50 dB	88.92%	88.18%	92.18%	87.61%	87.87%
	32 harm.	No noise	99.13%	93.56%	85.73%	99.17%	93.63%
		60 dB	93.88%	96.71%	86.06%	94.70%	96.67%
		50 dB	83.02%	83.69%	92.83%	81.81%	83.16%

is no relevant difference between the accuracy of the neutral and phase currents. This analysis showed that the best results were achieved when training and testing with the same noise levels as in the previous cases.

TABLE 10 Case III - Accuracy rate when using specialists Random Forest algorithms, considering the first 8, 16, 32, and 64 harmonics as inputs with no feature selection.

	Testing	Phase Current			Neutral Current		
		No noise	60 dB	50 dB	No noise	60 dB	50 dB
Training	8 harm.	No noise	84.35%	78.04%	69.76%	83.94%	78.32%
		60 dB	77.96%	76.63%	69.93%	78.31%	76.45%
		50 dB	70.46%	70.24%	68.06%	70.89%	70.38%
	16 harm.	No noise	96.01%	92.59%	85.84%	95.88%	92.69%
		60 dB	93.39%	93.41%	86.60%	93.40%	93.37%
		50 dB	87.96%	88.45%	88.21%	88.08%	88.44%
	32 harm.	No noise	97.58%	89.91%	78.18%	97.67%	90.39%
		60 dB	91.22%	93.58%	79.38%	91.52%	93.42%
		50 dB	80.28%	81.67%	86.66%	80.88%	82.39%
	64 harm.	No noise	95.93%	88.37%	75.34%	96.15%	87.33%
		60 dB	83.88%	91.65%	77.12%	84.06%	92.28%
		50 dB	72.06%	73.14%	86.72%	71.30%	73.83%

D. Case IV - Specialist Random Forests with feature selection

The last case evaluates the HIF location using five independent Random Forests and the mRMR to select the input data (Table 8). This case resulted in higher accuracy than Case III, indicating that using the feature selection algorithm improved the performance. However, the results were still inferior to Case II, showing that using a single Random Forest presented more promising results. Also, there is still no evident difference when comparing the phase and neutral currents. The highest accuracy was achieved using the 32 best-classified harmonics listed in the mRMR ranking.

E. Results Summary

Table 12 summarizes the best accuracy achieved in the evaluated cases for the phase current, as the neutral had similar results. To analyze this table, take Case I as an example. When considering the tests with signals without noise, the best training occurs when using the 32 harmonics

TABLE 11 Case IV - Accuracy rate when using five specialists Random Forest algorithms, selecting the input features.

	Testing	Phase Current			Neutral Current		
		No noise	60 dB	50 dB	No noise	60 dB	50 dB
Training	8 harm.	No noise	93.00%	79.87%	65.25%	92.69%	79.66%
		60 dB	84.44%	84.61%	66.95%	84.19%	84.78%
		50 dB	75.19%	75.06%	76.55%	75.07%	75.26%
	16 harm.	No noise	97.57%	90.17%	79.63%	97.69%	90.09%
		60 dB	92.24%	94.26%	80.38%	91.78%	94.02%
		50 dB	82.84%	83.40%	87.83%	81.75%	83.41%
	32 harm.	No noise	98.12%	90.71%	79.77%	98.05%	90.65%
		60 dB	90.93%	94.44%	80.56%	91.67%	94.38%
		50 dB	79.24%	80.57%	89.15%	79.74%	80.24%

extracted from the non-noisy signals as input data, achieving an accuracy of 98.90%. However, when considering the overall tests, which means training with a single dataset and testing with all noise levels (no noise, 60 dB, and 50 dB), the best solution would be to train the data without noise considering the first 16 harmonic orders, achieving an average accuracy of 93.75%. The highest success rate is achieved when training the Random Forest for the same noise level included in the tests. In this situation, the Case II conditions imply the best result. Nevertheless, if only one dataset is used for training, the conditions of Case I are the best. Cases III and IV hold the lowest accuracy.

TABLE 12 Best results and overall performance.

	Testing	No noise	60 dB	50 dB	Overall
Case I	Best Train	No noise	60 dB	50 dB	No noise
	Harmonics	32	64	16	16
Case II	Best Train	No noise	60 dB	50 dB	No noise
	Harmonics	32	32	32	32
Case III	Best Train	No noise	60 dB	50 dB	No noise
	Harmonics	32	32	16	16
Case IV	Best Train	No noise	60 dB	50 dB	No noise
	Harmonics	32	32	32	32
	Accuracy	98.12%	94.44%	89.15%	89.53%

The results are still promising even if the accuracy decreases in some cases because real HIF signals were used in the tests. Additionally, there is still no consolidated solution for the HIF location. The accuracy achieved using the neutral and phase currents was similar in all cases. Thus, if there is access to the substation transformer's neutral, this measurement would be a good choice because it does not need three measurement devices. Nonetheless, using the phase currents would be recommended if the transformer's neutral is not accessible.

V. Conclusions

HIF location is an issue without evidence of a conclusive solution in DSs, and there is no thorough existing analysis of it. This paper addresses this gap by proposing a critical review on the existing methods. The critical review revealed that most tests of HIFLMs consider HIF models and not actual signals. The test systems are usually short or

present several meters spread, and noisy signals and DGs are frequently not considered. Also, the review indicated that most HIFLMs could be divided into three approaches. The first approach uses the signal's frequencies. The second one utilizes measurement devices along with the system. The third approach considers the system modeling and the signals' measurement. The presented analysis summarized these approaches. They estimate the distance between the measurement and the fault or the fault region. In summary, the strategies are based on comparing metrics (expected and measured signals) or machine learning techniques.

Additionally, this paper presented the analysis of input features and different uses of a new HIF location method using the Random Forest algorithm. The results revealed that when using a feature selection algorithm, besides decreasing the computational complexity of the method, it improves the method's accuracy when training and testing with a specific noise level. In contrast, when considering a single dataset for training, the overall accuracy when using a feature selection algorithm decreased, showing a loss of generalization. The analysis also proved that using a single Random Forest to classify the HIFs among different locations implies a higher accuracy than using specialist Random Forests. The use of phase and neutral currents implied similar results.

The authors hope that the critical review and study case will support researchers and industry personnel to establish their tests and inform decisions that must be considered in different situations when developing increasingly efficient HIFLMs.

REFERENCES

- [1] S. Hossain, H. Zhu, and T. Overbye, "Distribution high impedance fault location using localized voltage magnitude measurements," in *2014 North American Power Symposium (NAPS)*, 2014, pp. 1–6.
- [2] J. R. Macedo, J. W. Resende, C. A. Bissochi, D. Carvalho, and F. C. Castro, "Proposition of an interharmonic-based methodology for high-impedance fault detection in distribution systems," *IET Gener. Transm. Distrib.*, vol. 9, no. 16, pp. 2593–2601, 2015.
- [3] D. Gomes and C. Ozansoy, "High-impedance faults in power distribution systems: A narrative of the field's developments," *ISA Transactions*, 2021.
- [4] S. H. Mortazavi, Z. Moravej, and S. M. Shahrtash, "A searching based method for locating high impedance arcing fault in distribution networks," *IEEE Trans. Power Del.*, vol. 34, no. 2, pp. 438–447, 2019.
- [5] M. Mishra and R. R. Panigrahi, "Taxonomy of high impedance fault detection algorithm," *Measurement*, vol. 148, p. 106955, 2019.
- [6] A. Aljohani and I. Habiballah, "High-impedance fault diagnosis: A review," *Energies*, vol. 13, no. 23, 2020.
- [7] J. J. G. Ledesma, K. B. do Nascimento, L. R. de Araujo, and D. R. R. Penido, "A two-level ann-based method using synchronized measurements to locate high-impedance fault in distribution systems," *Electr. Power Syst. Res.*, vol. 188, p. 106576, 2020.
- [8] F. B. Costa, B. A. Souza, N. S. Brito, J. A. Silva, and W. C. Santos, "Real-time detection of transients induced by high-impedance faults based on the boundary wavelet transform," *IEEE Trans. Ind. Appl.*, vol. 51, no. 6, pp. 5312–5323, 2015.
- [9] M. S. Ali, A. H. Abu Bakar, H. Mokhlis, H. Aroff, H. A. Illias, and M. M. Aman, in *IEEE Int. Power Engineering and Optimization Conf.*, Melaka, Malaysia.
- [10] M. J. Ramos, A. S. Bretas, D. P. Bernardon, and L. L. Pfischer, "Distribution networks hif location: A frequency domain system model and wls parameter estimation approach," *Electr. Power Syst. Res.*, vol. 146, pp. 170–176, 2017.
- [11] L. U. Iurinic, A. R. Herrera-Orozco, R. G. Ferraz, and A. S. Bretas, "Distribution systems high-impedance fault location: A parameter estimation approach," *IEEE Trans. Power Del.*, vol. 31, no. 4, pp. 1806–1814, 2016.
- [12] M. J. Ramos, M. Resener, A. S. Bretas, D. P. Bernardon, and R. C. Leborgne, "Physics-based analytical model for high impedance fault location in distribution networks," *Electr. Power Syst. Res.*, vol. 188, p. 106577, 2020.
- [13] A. Dasco, R. Marguet, and B. Raison, "Fault distance estimation in distribution network for high impedance faults," in *2015 IEEE Eindhoven PowerTech*, 2015, pp. 1–6.
- [14] A. N. Milioudis, G. T. Andreou, and D. P. Labridis, "Detection and location of high impedance faults in multiconductor overhead distribution lines using power line communication devices," *IEEE Trans. Smart Grid*, vol. 6, no. 2, pp. 894–902, 2015.
- [15] K. Moloi, J. A. Jordaan, and Y. Hamam, "High impedance fault classification and localization method for power distribution network," in *2018 IEEE PES/IAS PowerAfrica*, 2018, pp. 84–89.
- [16] F. L. Vieira, J. M. C. Filho, P. M. Silveira, C. A. V. Guerrero, and M. P. Leite, "High impedance fault detection and location in distribution networks using smart meters," in *2018 18th Int. Conf. on Harmonics and Quality of Power (ICHQP)*, 2018, pp. 1–6.
- [17] J. T. A. Vianna, L. R. Araujo, and D. R. R. Penido, "High impedance fault area location in distribution systems based on current zero sequence component," *IEEE Lat. Am. Trans.*, vol. 14, no. 2, pp. 759–766, 2016.
- [18] W. C. Santos, F. V. Lopes, N. S. D. Brito, and B. A. Souza, "High-impedance fault identification on distribution networks," *IEEE Trans. Power Del.*, vol. 32, no. 1, pp. 23–32, 2017.
- [19] J. Zhou, B. Ayhan, C. Kwan, S. Liang, and W. Lee, "High-performance arcing-fault location in distribution networks," *IEEE Trans. Ind. Appl.*, vol. 48, no. 3, pp. 1107–1114, 2012.
- [20] P. E. Farias, A. P. de Moraes, J. P. Rossini, and G. Cardoso, "Non-linear high impedance fault distance estimation in power distribution systems: A continually online-trained neural network approach," *Electr. Power Syst. Res.*, vol. 157, pp. 20–28, 2018.
- [21] A. Bouricha, T. Bouthiba, R. Boukhari, and S. Seghir, "High impedance faults location in the distribution networks using adaptive neuro-fuzzy inference system," in *Int. Conf. on Electrical Sci.s and Technologies in Maghreb*, 2018, pp. 1–5.
- [22] J. Li, G. Wang, D. Zeng, and H. Li, "High-impedance ground faulted line-section location method for a resonant grounding system based on the zero-sequence current's declining periodic component," *Int. J. Electr. Power Energy Syst.*, vol. 119, p. 105910, 2020.
- [23] J. Nunes, A. Bretas, N. Bretas, A. Herrera-Orozco, and L. Iurinic, "Distribution systems high impedance fault location: A spectral domain model considering parametric error processing," *Int. J. Electr. Power Energy Syst.*, vol. 109, pp. 227–241, 2019.
- [24] N. Bahador, F. Namdari, and H. R. Matinfar, "Tree-related high impedance fault location using phase shift measurement of high frequency magnetic field," *Int. J. Electr. Power Energy Syst.*, vol. 100, pp. 531–539, 2018.
- [25] A. Ghaderi, H. L. Ginn, and H. A. Mohammadpour, "High impedance fault detection: A review," *Electr. Power Syst. Res.*, vol. 143, no. February, pp. 376–388, 2017.
- [26] ATP, *World-wide mostly used transients program ATP-EMTP*, 2022. [Online]. Available: <http://www.emtp.org>
- [27] EMTP, *EMTP - Electromagnetic Transients Program*, 2022. [Online]. Available: <https://www.emtp.com/>
- [28] Manitoba Hydro Int. Ltd., *PSCAD*, 2022. [Online]. Available: <https://www.pscad.com/>
- [29] RTDS, *RTDS - Real-Time Digital Simulator*, 2022. [Online]. Available: <https://www.rtds.com/>
- [30] The MathWorks Inc., *MATLAB*, Natick, Massachusetts, USA, 2022. [Online]. Available: <https://www.mathworks.com>
- [31] PowerWorld Corporation, *PowerWorld*, Champaign, IL, USA, 2022. [Online]. Available: <https://www.powerworld.com/>
- [32] P. K. Nayak, K. Sarwagya, and T. Biswal, "A novel high impedance fault detection technique in distribution systems with distributed generators," in *2016 National Power Systems Conf. (NPSC)*. Bhubaneswar, India: IEEE, 2016, pp. 1–6.
- [33] W. O'Brien, E. Udren, K. Garg, D. Haes, and B. Sridharan, "Catching falling conductors in midair: detecting and tripping broken distribution

- circuit conductors at protection speeds,” in *69th Annual Conf. for Protective Relay Engineers*, no. 69. TX, USA: IEEE, 2016, p. 11.
- [34] F. Wang, J. L. Duarte, M. A. M. Hendrix, and P. F. Ribeiro, “Modeling and analysis of grid harmonic distortion impact of aggregated dg inverters,” *IEEE Transactions on Power Electronics*, vol. 26, no. 3, pp. 786–797, 2011.
- [35] J. J. Tomic, M. D. Kusljevic, and V. V. Vujicic, “A new power system digital harmonic analyzer,” *IEEE Trans. Power Del.*, vol. 22, no. 2, pp. 772–780, Apr. 2007.
- [36] C. H. Kim and R. Aggarwal, “Wavelet transforms in power systems: Part 1 General introduction to the wavelet transforms,” *Power Engineering Journal*, 2000.
- [37] F. Passerini and A. M. Tonello, “Smart grid monitoring using power line modems: Anomaly detection and localization,” *IEEE Trans. Smart Grid*, vol. 10, no. 6, pp. 6178–6186, 2019.
- [38] *Fire Mitigation for Distribution: Achieve Quick Progress With Advanced Technology Solutions*, Schweitzer Engineering Laboratories, Inc., and GW Electric, 2021. [Online]. Available: <https://selinc.com/api/download/126445/>
- [39] *Wildfire mitigation: Detecting and Isolating Falling Conductors in Midair Using 900 MHz Private LTE at Protection Speeds*, Anterix, Inc. and Schweitzer Engineering Laboratories, Inc., 2022. [Online]. Available: 11nq.com/N8fsz
- [40] D. P. Gomes, C. Ozansoy, A. Ulhaq, and J. C. de Melo Vieira, “The effectiveness of different sampling rates in vegetation high-impedance fault classification,” *Electr. Power Syst. Res.*, vol. 174, p. 105872, 2019.
- [41] *LinePatrol Smart Grid Sensor for Distribution Grid*, Silverfern Power Technologies Co., 2022. [Online]. Available: <https://en.silverfernpower.com/products-solutions/>
- [42] *EFD SWER Trial: Final report*, IND Technology Pty Ltd, 2019. [Online]. Available: <https://m5.gs/ZVdIRj>
- [43] R. Dugan, *Electrical Power Systems Quality, Second Edition*. McGraw-Hill, 2003.
- [44] IEEE Distribution System Analysis Subcommittee, *IEEE 34 Node Test Feeder*, EUA, 2010. [Online]. Available: <http://sites.ieee.org/pes-testfeeders/resources/>
- [45] S. Samantaray, B. Panigrahi, and P. Dash, “High impedance fault detection in power distribution networks using time–frequency transform and probabilistic neural network,” *IET Gener. Transm. Distrib.*, vol. 2, no. 2, p. 261, 2008.
- [46] B. P. Lathi and R. A. Green, *Linear systems and signals*. EUA: Oxford University Press New York, 2005, vol. 2.
- [47] G. Brown, A. Pocock, M.-J. Zhao, and M. Luján, “Conditional Likelihood Maximisation: A Unifying Framework for Information Theoretic Feature Selection,” *Journal of Machine Learning Research*, vol. 53, pp. 46–58, may 2016.
- [48] H. Peng, F. Long, and C. Ding, “Feature selection based on mutual information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, 2005.
- [49] L. Breiman, “Random forests,” *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

Thiago S. Menezes received a B.Sc. degree in Electrical Engineering from the Federal University of Triângulo Mineiro, Brazil in 2017, and an M.Sc. degree in Electrical Engineering from the University of São Paulo, São Carlos, Brazil in 2019. He is currently pursuing the Ph.D. degree at the Department of Electrical and Computer Engineering, São Carlos School of Engineering, University of São Paulo, Brazil. His research interests include microgrids, phasor estimation, power system protection, and intelligent systems.

Douglas P. S. Gomes received the B.E. degree in electrical engineering from the Federal University of Mato Grosso, Cuiabá, Brazil, in 2013, and the M.Sc. degree in power systems from the University of São Paulo, São Paulo, Brazil, in 2016. He received the Ph.D. degree from Victoria University, Footscray, VIC, Australia. His research interests include power systems, protection, and artificial intelligent systems.

José C. M. Vieira (S’98–M’07) received a M.Sc. and Ph.D. degrees in Electrical Engineering from the University of Campinas, Campinas, Brazil, in 1999 and 2006, respectively. From 1999 to 2003, he was a Consulting Engineer with Figener, São Paulo, Brazil. Currently, he is an Associate Professor at the São Carlos School of Engineering, São Carlos, Brazil. His research interests are power distribution systems, integration of distributed energy resources and power system protection.

Gabriela N. Lopes received a B.Sc. in Electrical Engineering from Federal University of Mato Grosso, Cuiabá, Brazil, in 2018 and M.Sc degree from EESC - University of São Paulo (USP), Brazil, in 2020. She is currently pursuing the PhD degree in Electrical Engineering from University of São Paulo (USP), São Carlos, Brazil. Her areas of research interest are high impedance faults, signal processing, and power system protection.