High-Sensitivity Vegetation High Impedance Fault Detection Based on Signal's High-Frequency Contents

Douglas P. S. Gomes, Cagil Ozansoy, *Member, IEEE*, and Anwaar Ulhaq, *Member, IEEE*

***Abstract*—High Impedance Faults (HIFs) are linked to enduring unaddressed knowledge gaps due to their diverse and complex behavior, despite being extensively researched disturbances. Vegetation HIFs, for instance, are a particular type of fault that can lead to great fire hazards and life risks. They have unique fault signatures and should receive special attention if fire risk mitigation is desired. This paper focuses on the detection of these distinct, very small current faults. As the main correlational features, the proposed methodology uses the vegetation fault signatures' high-frequency content. Different from many previous works that rely on HIF models, the approach validation is performed using a real dataset comprising a large number of experiments, sampled in a functioning network in the presence of noise. The classification is performed by boosted decision trees, which showed high dependability and security in the classification of small phase-to-earth and phase-to-phase HIFs.**

***Index Terms*—Decision trees, High-Frequency, High Impedance Fault, Vegetation faults.**

I. Introduction

HIGH Impedance Faults (HIFs) are widely researched power systems disturbances [1] due to their dangerous characteristics. These faults are traditionally defined by the resulted contact of a powerline with a high impedance fault surface. When in contact with vegetation, HIFs can ignite fires, posing a threat to human life and the possibility of major financial damages [2].

Despite much progress, HIF detection remains a field with many knowledge gaps. One of the main reasons is the complexity and costs associated with real experiments, which leads to fewer phenomena observations resulting in a poor understanding of the various HIF types. In this manner, the aforementioned traditional description for HIF is a condensed definition for an intricate problem. Considering that fault signatures may vary between HIFs types [3], it is reasonable to say that they should also be investigated in sub-classes given by parameters such as the type of the fault, surface of contact, type of network, and etc.

This work was supported in part by the Victorian Government Department of Economic Development, Jobs, Transport and Resources.

Douglas P. S. Gomes is working towards his PhD in the College of Engineering and Science, Victoria University, PO Box 14428, Melbourne, Vic 8001, Australia (email: douglas.pintosampaiogomes@live.vu.edu.au).

Cagil Ozansoy is a Senior Lecturer in the College of Engineering and Science, Victoria University, PO Box 14428, Melbourne, Vic 8001, Australia (email: cagil.ozansoy@vu.edu.au).

Anwaar Ulhaq is Lecturer, Information technology in the College of Engineering and Science, Victoria University, PO Box 14428, Melbourne, Vic 8001, Australia (email: anwaar.ulhaq@vu.edu.au).

A sift into the past ten years of research can shed light on common correlational features used in modern detection approaches. Among the most popular, wavelet features have been gaining presence and popularity in much of the works [4]–[7]. The detailed coefficients, outputs of the wavelet transformation, are usually used in two main detection approaches. Either by applying an arbitrary threshold in a selected parameter given by the detailed coefficients [6], or by feeding the same parameters into classifiers based on machine/deep learning methods [4]. Mathematical morphology-based methods that rely on irregularities in the current waveforms as HIF inception features were proposed in [8] and [9]. The outputs of the mathematical morphology transform, dilation and erosion were used to establish specific thresholds in [9], and as inputs to an Artificial Neural Network (ANN) classification approach in [8]. The energy, steadiness, and randomness of the current signals were used as features in [10]. The work relied on tunable arbitrary thresholds for each of those features, forming a systematical method that indicates fault occurrence when all the threshold conditions were met. Additionally, a time-frequency transform analysis approach was proposed in [11], where the detection relied on the energy of determined low-frequency components analysis as reliable features. The transformed and adjusted outputs were fed into an SVM classifier that distinguished between normal system conditions and HIF occurrences.

Although valuable and promising, most approaches share similar knowledge gaps and constraints. They often neglect the network noise effect by using models or isolated experiments, such neglection has been shown to diminish the detection security [12]. Few approaches that do not solely rely on models, make use of a small number of real tests, associating ambiguity to the validation. None of the cited work relates an effective current threshold for detection. Rather, methods are conceptualized to detect all the large range of fault current amplitudes that are lower than the normal pick up values of overcurrent protection devices. Time delay to detection is often not a discussed criterion either. This constraint, when added to the challenge of current detection threshold, create further concerns regarding fire risk since even a small fault current, sustained for a long enough period, can create fire ignition [2].

Fairly touching on the mentioned constraints, this work aims at distinguishing vegetation HIFs from a system's normal state by relying on the fault signal's High-Frequency (HF) content. Relating to mentioned validation gaps, the lack of precision regarding different current levels is addressed by the effective

current thresholds settled between 0.5 and 4 A. The proposed method has been tested using a real large dataset, reducing validation ambiguity when compared to works that rely solely on HIF simulations. The feeder used in the experiments was dedicated but part of a functioning 22 kV network. This led to the capturing of the effect of Electromagnetic Interference (EMI) and noise generated by connected loads, associating real noise resilience to the solution. The classification was performed with a single sampled signal of 20 ms (sweep sampling further discussed), indicating a reasonable time delay for vegetation HIF detection regarding a fire mitigation scenario. Another feature of this approach, separating it from comparable works in the literature, is the large bandwidth utilized, from 10 kHz to 1 MHz. Past methods commonly investigate a narrow bandwidth (mainly a few thousand of Hz) of the electric signals' frequency spectrum.

The paper is structured as follows: Section II presents the Vegetation Ignition Tests, conducted as part of the Powerline Bushfire Safety Program. Section III meticulously describes the proposed methodology, with the feature extraction and classification method. Results and discussions regarding the outcome interpretation and future research are given in Section IV, followed by the overall conclusions in Section V.

II. Bushfire scenario and vegetation ignition

Faults in electric assets were directly responsible for five of the Black Saturday fires in Victoria, Australia (2009) [13]. These resulted in billions of economic damages, more than 170 fatalities, and injury of many others. The investigation of the causes by the Victorian government led to the creation of the Powerline Bushfire Safety Program (PBSP). It was an initiative to reduce the risk of bushfires caused by faulty electric assets that granted funding to R&D projects to identify cost-effective risk reduction technologies and procedures. The 'Vegetation Conduction Ignition Testing' [2], one of the funded research projects, tested a variety of vegetation species on a real unearthed 22-kV feeder under various fault scenarios. These tests produced a large database of HIF fault signatures with high sampling resolution, low-noise, and wide-band recordings.

A. Vegetation Conduction Ignition Testing

The program was conceptualized around two main goals. The first was to identify which of the native species, that could possibly come into contact with powerlines, are most and least likely to start a fire. The second was to deliver a reference database of fault signatures to foster the development of fault detection technology. The test rig was built inside two shipping containers at a local distribution substation, and more than twenty plant species were tested in 1038 experiments. These can be categorized into three main types of faults: 'Branch touching wires' (phase to earth), 'Branch across wires' (phase to phase), and 'Wire into vegetation' (phase to earth) faults.

The first fault type refers to a tree branch laid across two conductors, one earthed and one with the nominal phase voltage (12.7 kV). The second type followed the same geometry, but with both conductors energized (22 kV). The third test was conducted by dropping the High Voltage (HV) conductor into vegetation, either grass or bush. To simulate extreme weather conditions, the ignition test area was maintained at 45° C, less than 25% relative humidity, and a pedestal fan was used to simulate a light hot wind. The temperature in the container was controlled and operated continuously throughout the test program.

Having several phase-to-phase experiments in the dataset is of great importance for this work. HIFs are disturbances often modeled by a scenario where a conductor breaks and falls to a surface such as asphalt or gravel. These are phase to earth tests that, consequently, involve the neutral current. Despite their relevance, they have different arcing characteristics from a vegetation phase-to-phase fault that only involves phase currents. Datasets from these phase-to-phase fault tests were largely used in training and testing the classification method proposed in this paper, which is a noticeable novelty when compared to other works in the area.

B. Tests characteristics and data sampling

Given the main objective of relating the likelihood of different species igniting fires versus distinct current levels, current conduction was stopped once it reached a specified value. In the great majority of the experiments (99%+), thresholds were set between 0.5 and 4 A. Despite being a constraint, it introduces the novelty of investigating a specific set of HIFs characterized by a very low current amplitude, differentiating it from most related works. Often neglected, very low fault currents represent an important scenario to be considered since they can ignite fires despite having small amplitudes [2]. If fire ignition mitigation is a goal of any given approach, such current levels need to be properly addressed.

In respect to the sampling of the electric fault signals, the project team decided to sample current and voltage waveforms simultaneously in two channels to ensure wideband, and low sampling noise. The LF channel was responsible for sampling the electric signals at 100 kS/s, continuously, with suitable anti-aliasing filters in the 0 to 50 kHz frequency range. The HF channel outputs were fed through filters with 10 kHz corner frequency, sampling both signals in the sweep mode at 2 MS/s rate, with anti-aliasing filters, resulting in a 10 kHz to 1 MHz bandwidth.

It is important to note that after initial investigations, correlations between extracted features and fault occurrences were only found when investigating the latter channel. In fact, this is the main reason why the proposed method focuses on the HF spectrum of the sampled signals. At the same time, it must be acknowledged that such an approach can also generate justifiable and relevant concerns related to the usage of this wideband spectrum in a real implementation scenario. This is related to the challenge of acquiring such HF, wideband data in a real-life setting, which is unquestionably a valid concern. The key objective herein is to propose the theoretical aspects of a HIF classification method based on non-traditional (non-current based) features and validation of the developed method through the use of feeder voltage HF data. Nonetheless, topics such as novel capacitive sensor-based sampling technologies, the tendency of having more distributed measurements along

the feeders, or the probable inevitability of this approach for fire ignition mitigation, can be cited as relevant motives to consider such solution.

The sweep mode used by this channel was given by the recording of 20 ms windows at every second interval. Such choice was made by the PBSP team due to the unmanageable amount of data that would have resulted from continuous sampling in the related sampling rate. Also, continuous recording was already present in the LF channel. Though considered a drawback, this channel was responsible for analyzing only frequency components higher than 10 kHz, which relates to at least 200 cycles at the mentioned sampling period.

The second drawback is related to the fact that measurements were only recorded at the test site. The test rig supplying feeder had no consumer loads connected to it, despite connected to a loaded network. For this reason, there was a impossibility to analyse the fault signatures created by HIF occurrence in the normal load current. This was the main reason why the present research focused on investigating the voltage signals.

Yet, the most severe constraint in the data comes from the execution set-up. The branches were put in place between the conductors prior to their energization, i.e. conduction began immediately after the voltage source was turned on. By adopting such a method, almost no pre-fault voltage signals were recorded, except in few test scenarios. This lead to an adoption of a particular strategy regarding the 'Non-fault' observations, further detailed discussed in the next section.

Conversely, despite described constraints, after a comprehensive investigation, reliable correlational features were found even after just one (first) voltage HF sweep. They were sufficient to indicate the occurrence of a vegetation HIF with high associated accuracy and security.

Appendix A covers further specific information, for the more interested reader, regarding test rig configuration, system under analysis, and sampling equipment.

C. Findings

The set-up, analysis, and findings from the experiments were described in the 'Vegetation Conduction Ignition Test' Final Report by Marxsen [2].

Interesting results regarding the fault current threshold versus fire risk were observed in [2]. It was stated that in 'branch touching wire' faults, ignition risk greatly increase for currents higher than 0.5 A. Threshold of 1 A translated to 33% chance of ignition, and 2 A to 53% change of fire risk. Such information corroborates for the need of high sensitivity protection devices for vegetation areas in Victoria, where protection devices are usually configured in the 5-10 A range.

As many tests resulted in flashovers that bridged the HV conductors, it was concluded that most faults would eventually be detected by traditional protection systems. However, getting to flashover state means that the vegetation sample went through the phases of expulsion of moisture and progressive charring of the bark, which are the most responsible phases regarding fire risk [2]. This implies that protection equipment would only respond after fire creating phases (see ignition phases and definition in [2]) would already occur.

A well-known commercial protection relay, developed for the North-American market, with an embedded HIF detection function, was also tested. None of the 1038 faults were detected by the device. At the same time, it is worth remembering that the network used for experiments was a three-wire unearthed system that tends to have smaller fault currents.

When assessing the fire risk of different vegetation species, the most dangerous were shown to be the ones that took longer for fault current to fully develop. It was observed that if the fault current rapidly grew to the detection rate, then ember formation was less probable than cases of sustained conduction. Such observations also led to detection time delay considerations. In 'branch between wires' faults, it was determined that detection times longer than 20 s would be unlikely to dramatically reduce fire risk. A significant decrease in such risk can be achieved by responding in 5 s or less. In the case of 'branch touching wires', it was also stated that event detection in 2 s with 0.5 A sensitivity could reduce fire risk up to tenfold.

III. Methodology

Adopting the sweep sampling mode and making use of just one power cycle at each second can certainly heavily impact any methodology to detect a system's disturbances. However, despite being a small part of the signal, 20 ms is sufficient time to detect the presence of the HF components in the HIF signatures (10 kHz+). If fault detection is proven to be possible with such a small amount of information, sampling signals in this mode can actually come in favor when considering real-time implications. This possibility is especially important for methods relying on signal's HF content since they should be able to deal with a lot of samples and complex calculations that translate to relatively long processing times. By choosing a sweep sampling period of 20 ms, 980 ms remains to be dedicated towards sampling delays, calculations, and decision making processes before the next sampling cycle.

As the present paper focuses mainly on attesting a solid and strong correlation between the extracted features and HIF vegetation occurrences, the steps followed in the methodology can be summarized as:

- Data importing, cleansing and pre-processing;
- Partition and labeling of fault and non-fault sweeps;
- Feature extraction; and
- Classification/validation.

It is worth mentioning that all the steps, including data management, feature extraction, classification algorithms, and validation were performed in the MATLAB environment with the respective toolbooxes: Signal Processing Toolbox™ [14], Statistics and Machine Learning Toolbox™ [15], and Wavelet Toolbox™ [16].

A. Data importing, cleansing and pre-processing

After importing the data, the labeling of corrupted files and data cleansing took place as the first step of the procedure.

Tests that fell into the following categories were excluded from the classification analysis:

- Files missing or corrupted;
- Tests that did not show current conduction;
- Tests with max. effective current lesser than 0.5 A;
- Tests that presented problems in sampling;
- Tests with high intermittency in current conduction; and
- Tests marked as invalid by the program.

B. Non-fault recordings

The pre-processing part also relates to the extraction of sweeps of interest in each recording. However, due to the lack of pre-fault sampling limitation, as described in Section II, a different strategy was adopted while considering non-HIF states. Despite the absence of voltage sampling before fault inception took place (in the majority of the tests), background voltage noise was recorded in different parts of some test days. The main objective of such recordings was to capture normal operation states of the network (with standard noise and connected loads) so they could possibly aid the fault signature analysis. In this manner, these background noise recording sweeps were extracted from the tests data and labeled as non-fault observations.

Fig. 1 compares the Power Spectral Density (PSD) of a random 'Non-fault' sweep and a fault sweep from Test #552. Surely, the same clearly distinct behavior between observations was not present at every test, but Fig. 1 clearly illustrates how a fault occurrence can impact the HF spectrum of the voltage signals. The peaks greater than 400 kHz can all be related to the local AM radio stations and shared EMI sources.

C. Database of observations

Both types (HIF and Non-HIF observations) of sweeps were processed to be the same length, i.e. 40k samples, considering a 2 MS/s sampling rate. It is important to remember that fault experiments had distinct duration times. Some faults lasted for few seconds and others more than a minute. For this reason, a choice was made to perform the classification relying only on one unique sweep given by the first recording after the fault inception. In fact, by doing so, it is guaranteed that the signal considered when testing the classifier came from the first second of the fault occurrence.

A fault inception time was defined to be the moment when the root mean square value of fault current was greater than 0.5 A. The separation of the analyzed fault sweep was done by comparing sweep trigger times with the calculated fault inception time. Consequently, the recording considered for classification was the closest sweep after the calculated fault inception time.

D. Feature extraction

Primarily, spectrograms and signal processing techniques, as the one shown in Fig. 1, revealed noticeable fault signatures in the voltage's HF content. Furthermore, an investigation of the most relevant HF features proved that a combination of the wavelet detailed coefficients, its peaks, energy, and

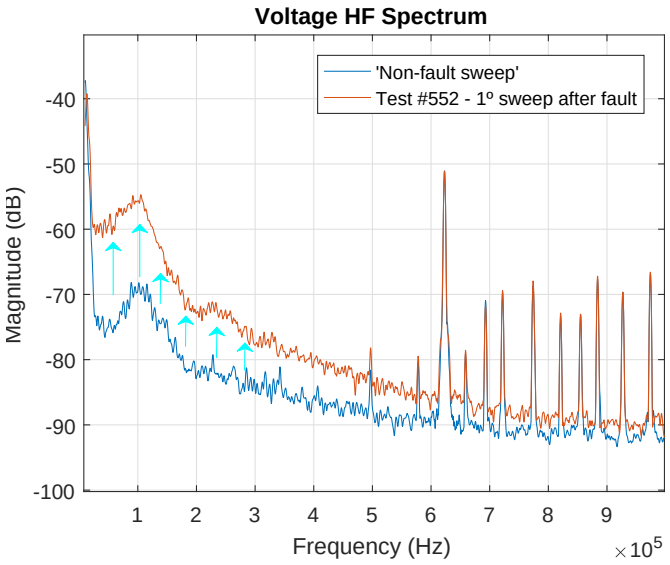

Fig. 1. PSD comparison between a 'Non-fault' state and a Test #552 sweep.

Power Spectral Density (PSD) measurements yielded the best classification results. Their extraction and considerations are described in the subsequent sections.

1) Wavelet features: When analyzing the fault signals, fast step discontinuities in the fault's current waveform in the form of rapid spikes were observed. This behavior was translated to the voltage signals as bursts of high-frequency noise in the whole range of the analyzed spectrum.

Although the contribution of some frequency ranges increased more than others, HF responses could not be correlated to a narrow well-defined bandwidth. On this condition, a tool such as the Discrete Wavelet Transform (DWT) comes as a powerful feature extractor, as it can accurately represent transients in different frequency ranges. The DWT transformation of a signal x(t) is given by (1).

$$DWT(x,m,n) = \frac{1}{\sqrt{a_o^m}} \sum_l x(k)\psi(\frac{n - la_o^m}{a_o^m}) \qquad (1)$$

Where, $x(k)$ is the sampled signal, $\psi(n)$ represents the mother wavelet, a_o^m is the dilation coefficient, and la_o^m is the translation coefficient [17].

The coefficients are calculated by the Multi-Resolution Analysis (MRA) approach that can be seen as an extension of DWT. In this process, the decomposition is given by the iterative application of the transform in the successive approximation output coefficients [18]. The MRA procedure filters the input signal, separating the frequency components present in the signal by passing it through a series of low-pass and high-pass filters [19]. In the first iteration, the output is given by the convolution of the original time domain signal with the impulse response function of the low- and high-pass filters. From that, two signals known as, approximation (y_a) and detail (y_d) coefficients, will result. Moreover, the iteratively process is performed by using the last calculated approximation coefficient as new inputs of the filtering process. The i_{th}level can be generalized as in (2) and (3).

$$y_d^i[n] = \sum_{k=-\infty}^{\infty} y_a^{i-1}[k] \times h[2n-k] \tag{2}$$

$$y_a^i[n] = \sum_{k=-\infty}^{\infty} y_a^{i-1}[k] \times g[2n-k] \tag{3}$$

Where, $h[t]$ and $g[t]$ are, respectively, the low-pass and high-pass impulse response function, and y_d^0 and y_a^0 are defined by x(t).

The detailed coefficients, however, are not single measurements that can be directly used as features. Approaches in the literature usually calculate the energy [6], entropy [20], or standard deviation [5] of these signals to use it as the main correlational features. In the proposed approach, with the intention of increasing the methods' reliability, several possibilities of wavelet features details were tested: energy percentile (4), sum of the absolute values (5), peaks (6), standard deviation (7), and Shannon's entropy (8).

The 'peaks' feature, given by S_{peaks}, defined by (9), to the best of the author's knowledge, was never presented before in this regard, characterizing an original feature of this work. In simpler words, the feature represents the mean of the top M (sorted) peaks found in the signal. The integer M can be a fixed value or, as used in this paper, a proportion of the length of the detail vector.

$$e_i = \frac{\sum_{n=1}^{N} \left|y^i[n]\right|^2}{Et} \tag{4}$$

$$sum_i = \sum_{n=1}^{N} \left|y^i[n]\right| \tag{5}$$

$$pks_i = \frac{1}{M} \sum_{n=1}^{M} (S_{peaks}(n)) \tag{6}$$

$$std_i = \sqrt{\frac{1}{N-1} \sum_{n=1}^{N} \left|y^i[n] - \overline{y^i}\right|} \tag{7}$$

$$ent_i = -\sum_{n=1}^{N} y^i[n]^2 log(y^i[n]^2) \tag{8}$$

$$S_{peaks} = sort(\{\left|y^i[n]\right| > \left|y^i[n-1]\right| \wedge \left|y^i[n]\right| \geq \left|y^i[n+1]\right|\}) \tag{9}$$

Where, Et is the sum of the energy of the approximation and detail coefficients, and $\overline{y^i}$ is the mean of the detail vector.

2) Periodograms features: Although wavelet features alone resulted in high accuracy (+90%) in the classification, further analysis showed that the feature group could still be enhanced with frequency domain measurements. Features extracted from the PSD analysis, first preference in the search for reliable features, produced the best result in accuracy when combined with wavelet features. The periodogram was calculated by the Welch's method and frequency ranges used as features were further described in the Results section.

As first presented in [21], Welch's method is a popular technique for estimating the power spectral density. It consists of breaking the analysed time series into overlapping segments where each has its spectral estimation calculated and averaged. The result is a modified nonparametric periodogram with reduced variance that can be directly calculated by (10).

$$\hat{P}(f_n) = \frac{1}{K} \sum_{k=1}^{K} I_k(f_n) \tag{10}$$

Where, K is the number of segments in the time series, f_n is the analyzed frequency and I_k is the periodogram of each segment given by (11).

$$I_k(f_n) = c \left| \frac{1}{L} \sum_{j=0}^{L-1} X_k(j) W(j) e^{-2kijn/L} \right|^2 \tag{11}$$

Where, L is the length of the segment, X_k is the analyzed k-segment, j is the sample number, W is the window function (Hamming window in this paper), i is the imaginary unit, and c is a constant dependent on the length of the segment and window used.

E. Classifier - Boosted decision trees

The choice of the type of classifier was made by comparing the performance of various classifiers with standard MATLAB default parameters. Machine learning techniques such as discriminant analysis, support vector machines, k-nearest neighbors, decision trees (and ensembles) were considered. In respect to such comparison, the best result was given by the boosting the decision trees technique.

Decision trees split the data points strategically in binary decision nodes with indicator functions that evaluate each feature to classify an observation. They can handle big datasets, work with quantitative and qualitative predictors, easily ignore redundant variables, and have relatively high levels of interpretability. The standard CART (Classification And Regression Trees) algorithm was first presented in [22] and was soon followed by many variations.

Although they are adaptive and robust, decisions trees are weak learners that have high variance, they do not generalize well. Nevertheless, recent applications such as the use of an ensemble of trees can help overcome these constraints. The main goal of this approach is to increase the accuracy of individual trees by usually randomizing (random forests) or/and averaging (bootstrap aggregation) the results of many learned trees. Commonly known strategies are random forests [23], tree bagging [24], and tree boosting [25].

Boosting is an averaging and weighting technique in which many learned trees are used to greatly improve the classification accuracy. Between the boosting strategies tested in the present methodology, the one associated with best results is known as AdaBoost [25]. It trains learners sequentially, and for every learner with index t, it computes a weighted classification error as in (12).

$$\varepsilon_t = \sum_{n=-1}^{N} d_n^{(t)} I(y_n \neq h_t(x_n)) \tag{12}$$

Where, x_n is the feature vector from the observations, y_n is the classification label response vector, h_t is the prediction of the learner with index t, I is the indicator function, and $d_n^{(t)}$ is the weight of observation n at step t. Training such classifier can the thought as the stagewise minimization of the exponential loss E given by (13).

$$E = \sum_{n=-1}^{N} w_n e^{-y_n f(x_n)} \tag{13}$$

Where, w_n are the observation weights normalized to add up to 1, and $f(x_n)$ is the predicted classification score.

IV. Results and Discussions

This section details the outcomes of the proposed methodology regarding the feature extraction and HIF classification. It is then followed by discussions on the implications of observed results and the course action for future works.

A. Considered dataset

Although data cleansing resulted in a large purge, due to the high number of experiments, an extensive quantity of tests prevailed for analysis. In total, 568 different tests were used as observations for fault events. 351 were 'phase-to-earth tests', 193 were 'phase-to-phase tests', 23 were 'bush tests' and 1 was a 'grass test'. The small number of the latter is due to the fact that only 24 'grass tests' were performed and the majority did not show any current conduction. In respect to the effective current threshold value of these tests, 5.11% of the experiments were limited to 0.5 A, 50% to 1 A, 39.96% to 2 A, 4.58% to 4 A, and 0.35% were not classified.

To avoid bias in the classification, a balanced number of observations (HIF and non-HIF) was considered. The analyzed non-HIF observations originated from 11 voltage background recordings made between February 24$^{\text{th}}$ to 27$^{\text{th}}$ of 2015. Each run was performed in three different periods of each day, resulting in 719 non-fault observations. 548 sweeps were randomly picked from these observations for balancing and added to 20 sweeps of white Gaussian noise. The latter was used to help teach the classifier to not mislabel a situation where there is no connection to the voltage source (supply off) as a fault occurrence.

B. Feature extraction

When utilizing the MRA, the output decomposition relates each level to a certain frequency range of the sampled signal. The upper limit of the frequency pass-band in Hertz of each detail coefficient is approximately given by (14), and the lower limit by (15). Where, Fs is the sampling frequency and n is the detail level.

$$Fb_{up} = \frac{Fs}{2^n} \tag{14}$$

$$Fb_{down} = \frac{Fs}{2^n + 1} \tag{15}$$

Finding the optimal level of decomposition is not always an easy task. One needs to understand how much information the next level of decomposition might give and the implications for the problem at hand. However, the present classification problem is bounded by the fact that the signals passed through a high-pass filter with a 10 kHz corner frequency before their recording. This implicates that detailed coefficients greater than the 7$^{\text{th}}$ level (7.81 to 15.62 kHz) would not provide reliable information. Given that, a choice was made to do an exhaustive search from level 1 to 7. By associating the detailed decomposition levels with the accuracy given by the classifier, it was found that no accuracy was gained for levels of decomposition greater than 4 levels.

At this point, it is worth to make a note regarding the choice of the mother wavelet used. This is mainly due to the common argument stating that the DWT efficiency, at representing transients, may be heavily influenced by such choice [1], [3]. Hence, aiming at investigating such claim, a prior comparison was executed concerning possible choices of many mother wavelets. The evaluation compared the performance of different wavelet families such as the Haar, Daubechies, Symlets, Coiflets, BiorSplines, ReverseBior and DMeyer, in their different scales. The performance indeed changed regarding different choices, with the sym4 (Symlets) giving the best overall accuracy. A reasonable explanation for that is given by the similarity between the mother wavelet waveform and the transients created in the HF voltages signals, which have origins on the fast step discontinuities in the HF current. However, the maximum difference in overall accuracy was about 1% between different wavelet families. Thus, not corroborating greatly with the critiques made by researchers in this particular case.

In regards to the feature selection, the class 'Tree Bagger' from the Statistics and Machine Learning Toolbox, previously cited, was used. More precisely, its built-in function for measuring features importance. Such process is given by the permutation of the values of each feature across every observation in the dataset and the calculation of how accuracy changes after each permutation. Based on it, the features were selected and can be listed as the following: sum of absolute coefficients for 1$^{\text{st}}$, 3$^{\text{rd}}$ and 4$^{\text{th}}$ level; 1% of the top peaks from 3$^{\text{rd}}$ level; and energy percentile of 3$^{\text{rd}}$ level.

As whole HF spectrum components showed relative increases in the presence of a fault, an investigation of reliable and simple features from the PSD results was considered using the same procedure described above. To that end, the Welch's method periodogram technique was performed with 450 frequency bins (0 to 1 MHz), a window size of 10k samples, and 50% of window overlap. Peaks in three different ranges showed a strong correlation with fault occurrences: approximately 350-370 kHz, 770-775 kHz, and 890-901 kHz. These density power values were used together with the wavelet features, adding to the total of eight features.

C. Classification

In regards to the used classifier, the boosted decision trees approach choice was made based on a performance comparison between various supervised learning techniques. The procedure took place before the feature selection part where all the calculated features were given as predictors and overall accuracy was translated to performance evaluation of each technique. The tested classifiers, using 491 predictors, can be listed in ascending performance order: Quadratic discriminant (68.2%), Linear discriminant (77.2%), Weighted KNN (K-Nearest Neighbour) (86.7%), Fine Gaussian SVM (87.1%), Fine KNN (87.7%), Linear SVM (88.7%), Quadratic SVM (91%), Complex decision tree (94.9%), and Boosted Trees (98.06%).

After the classifier selection, adjustments of parameters, and feature curation, the resulting performance can be expressed by the confusion matrix shown in Fig. 2. This matrix is commonly used in the machine learning field to describe important features from a proposed method. In HIF detection research, these numbers can infer important parameters such as dependability, security, and overall accuracy. The diagonal terms of the matrix, given by the green blocks, represent accurate classification, while the remaining terms show the occurrence of mislabelling. In this case, the first term (1,1) is related to Non-HIF observations being labeled as such, i. e., it translates the security of the classifier (99.47%). The second diagonal term also represents accurate classification but now for HIF observations, i. e., the dependability of the classifier (96.65%). 19 HIF observations were misclassified as Non-HIF and 3 Non-HIF as HIF observations. If aggregated as the total overall accuracy, a result of 98.06% will be achieved.

As shown by these results, security comes as the most relevant parameter presented by the method. It is fair to assume that, for an HIF detection method, security should come as a priority over dependability since HIFs occurrences do not cause great stresses in the system's equipment, and false trips may lead to severe consequences in terms of load importance.

An example of the classifier in performance is illustrated in Fig. 3. This experiment was labeled as a 'bush test' and was used for exemplification due to two main reasons: it was one of the few tests that conduction started after the energization, thus making it possible visualize results for the pre-fault data; and it was misclassified in one sweep. The algorithm detected great relative changes in the features at the energization, mislabelling as a fault, but adjusted quickly in the next sweep. It is worth remembering that the classification is made based on one sweep (20 ms) and lasts for the second related to its recording. Examples like these inspired the idea of only registering HIF occurrences when consequent sweeps were labeled as faults. Even though the method already achieved high security (99%+), such a scheme could further enhance security and create greater resilience to quick like-fault transients. As implications regarding speed of the method would be an obvious consequent concern, its evaluation will be reserved for subsequent investigations.

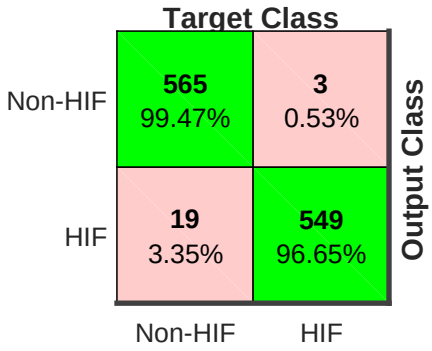

Fig. 2. Confusion matrix of the proposed classifier.

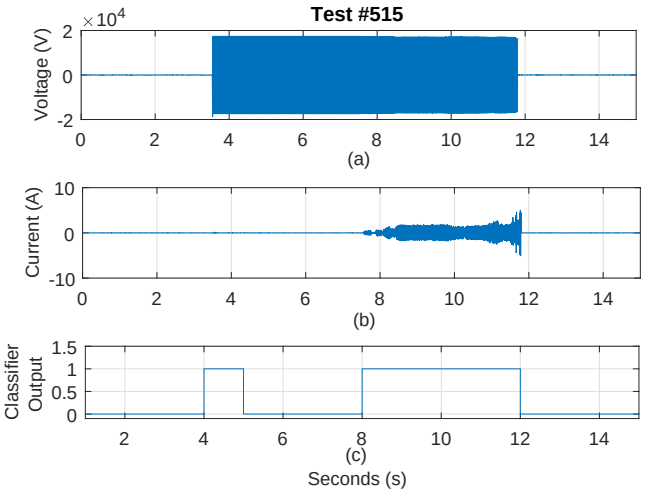

Fig. 3. Exemplification of the classifier response. (a) Voltage (V) and (b) Current (A) from from LF channel. (c) Classifier response (logical).

D. Validation

Given the relatively large dataset, the chosen method for validating the classifier model was carried out by a 10-fold cross validation procedure. Doing so means randomly partitioning the dataset into ten equal sized samples groups and validating it ten times in an iterative process. At each iteration, nine of the ten parts were used for training the classifier and one was used for testing. The process ends in the tenth iteration when all the partitions were used for testing. The main advantage of such approach is that all the observations are used for both training and testing. Additionally, as every observation is tested exactly once, the overall accuracy translates to an evaluation for the whole dataset, differently from holdout validation methods, for example.

A common practice to validate HIF detection methods is to test the proposed algorithm against data from simulated network switching transients. However, it is important to remember at this point that such validation scheme has great value due to the fact that most of proposed HIF detection algorithms are current, low-frequency methods based which may greatly suffer from such transients. The presented approach is a voltage and high-frequency based method, slightly diverging from the main goal of such validations. Nonetheless, although deviating from the scope of the present paper, i.e

dealing with real sampled data, these simulations can still be useful to demonstrate the classifier's security towards simplistic transients models. To that end, simulations of the given disturbances in Simulink™ environment were performed.

To that end, the 4-bus IEEE test node feeder was used. All its details and characteristics can be found comprehensively described in [26]. Some relevant reasons such as the fact that the feeder where the real tests took place was a dedicated feeder inside the substation (short feeder), and that the 4-bus feeder was a system made public to test different transformer connections can be cited as most pertinent ones. The latter has great relevance given the fact that the real feeder is a part of a three-wire distribution system. Such transients and faults characteristics can severely change when considering the existence or absence of solid grounding.

The simulation adopted the step-up transformer (12.47/24.9 kV) set-up, connected in Delta-Delta configuration. The connected load is linear, with 6 MVA, and 0.8 lagging power factor. The simulated transients were the normal switching events that usually concern HIF algorithm's security: transformer energization (24.9/415 kV, no load), capacitor energization (1 and 2/3 of the system's Q), load switching (1.5 MVA, overloading the transformer in 25%), and non-linear load switching (also 1.5 MVA). The time step in the discrete simulation was the same as the sampling frequency in the tests, i.e. 5.10^{-7} s or 2 MHz, with data also fed through a 10 kHz corner frequency filter. Furthermore, in regards to noise consideration, white Gaussian noise was added based on the average power of noise in the background noise recordings from the real tests (same noise power). The switching times were simulated considering eight equidistant angles, from 0 to 315°, and the most severe transients are illustrated in Fig. 4.

After simulated, filtered, and added noise, the signals were sliced in sweep length sizes and fed into the classifier for testing. None of the 40 different experiments were labeled as faults.

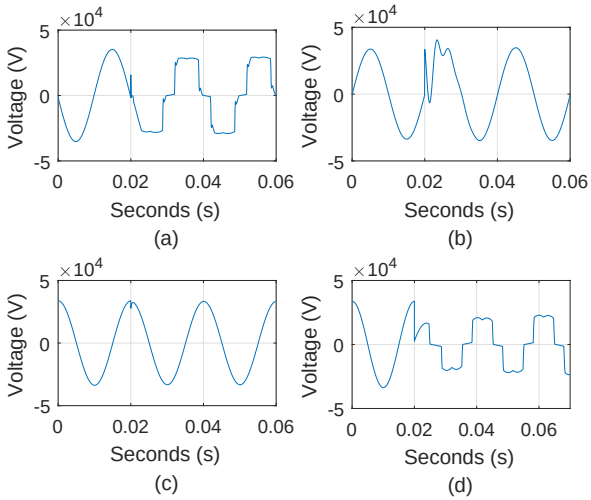

Fig. 4. Illustration of the simulated transients. (a) Transformer energization (b) Capacitor energization. (c) Load switching. (d) Non-linear load switching.

E. Discussions

Despite resulting in a classifier with high security, one limitation of the presented methodology is the fact that Non-faults observations came from background noise recordings, rather than pre-fault signals. Nonetheless, since the used signals were recorded in normal, steady state, it is expected that the presented performance would not strongly differ in the absence of this limitation. In fact, it is reasonable to assume that, if pre-fault data was accessible, better training could be done and even higher classification parameters would be achieved.

Other probable concern, given by the calculation of numerous features and classification process, is regarding the speed and/or computation burden of this method. It is worth reiterating that the decision regarding a second of classification is given by a single sweep of 20 ms duration. This means that approximately 2% of a second is dedicated to sampling, while the rest for calculating and labeling. As an example of these implications, when testing an experiment of 21 s duration, the code that loaded the data, calculated the features, and had its sweeps labeled, took less than nine seconds to perform these actions. All taking place inside MATLAB environment, with code not optimized for speed. Hence, it is fair to assume that such an algorithm built in hardware with signal processing functions would not have problems regarding computational burden.

Having said that, the same sweep mode of sampling and decision making greatly affects the speed of the classifier. In a worst-case scenario, a vegetation HIF may begin right after the recording of the last sample of a given sweep, leading to a longer detection delay. This means that detection could possibly take longer than one second, though unlikely. Nonetheless, as the main goal of this approach is the mitigation of bushfires ignited by vegetation HIFs. If findings given by the report [2] were taken as guidelines for the required detection speed, 2 s event detection and 0.5 A current sensitivity would be enough for significantly reduce fire risk.

It may be important to mention that the proposed method does not affirm to have higher accuracy than previously existing approaches. The key novelty of this work is related to its specificity regarding fault type (vegetation HIFs), its sensitivity (0.5 A), and the fact that it uses the feeder's high-frequency voltage signals. Also, that the results were consistent for either phase-earth and phase-phase faults since HIF is mainly discussed by a conductor breakage scenario (phase to ground surface). Consequently, considering that existing methods having investigated these characteristics only separately, comparing them with the proposed method would not be a fair estimate. Nevertheless, good literature reviews citing the performance of existing methods can be found in [1] and [27].

In respect to future works, the guidelines for further investigations will be taken from the constraints of this presented approach. Voltage signals and their respective HF content can be hard and expensive to obtain considering network topologies and existing hardware. However, the authors believe that the current incentives for more distributed measurements throughout the feeder and new sampling technologies, such as

sensor capacitive sampling, might help the task of obtaining these signals. Further investigations on the sampled signals are to be continued for more relevant features, especially ones that could facilitate real-time implementation, such as a simple current based approach. This means also improving the understanding of the decision nodes in the classifier that lead to such high accuracy to propose a highly adaptable and fine-tuned method to simplify real implementation.

V. Conclusion

Aiming at attesting the classification of small HIF currents with real data, an investigation has been presented concerning the analysis of the HF content using voltage signals. The fault signals were collected, compared to non-fault state recordings and validated using a real large dataset. To enhance the presented classification accuracy, feature extractors and classification methods were evaluated and selected. Regardless the presence of real noise, good accuracy was achieved for all analyzed types of fault and for heavily limited current values. The high security, as a priority parameter, was probably the most relevant result and a pertinent indicator of the method's potential. While future investigations will tackle hardware implementation possibilities, the presented results are promising regarding the mitigation of bushfires ignited by vegetation HIFs.

Appendix A
Feeder and Test Rig Characteristics

The experiments were performed in a dedicated feeder inside a substation in Melbourne, Australia. The substation's feeders are set-up in the three wire configuration with 22 kV nominal voltage. Fig. 5 illustrates the simplified single-line diagram with the CB (Circuit Breaker), ACRs (Automatic Circuit Recloser), RCGSs (Remote Control Gas Switch), and HV Resitors. The CB and ACRs have overcurrent, earth fault, and earth fault sensitivity protection. Represented in the diagram is also the 1.1 nF coupling capacitor, and HV resistors that were put in place to ensure the non-occurrence of internal flashovers.

Not much information concerning the adjacent feeders was made public by the project, neither the energy company which collaborated in the tests. However, it is known that the substation is supplied by a 66 kV sub-transmission system and have two transformers connected in floating Y-Y configuration with secondaries connected to neutral earth resistors. Besides the test feeder, the substation supplies at least ten more consumer distribution feeders, including industrial loads.

To effectively capture the wide-band, low noise signals from the tests, two Omicron 24kV, 1.1 nF coupling capacitor were combined with bottom-end capacitors to form the dual channel capacitive voltage divider (LF and HF channels). Two layers of over-voltage protection were included in each channel, given by a 125 V bidirectional voltage limiting diode, embedded in the coupling capacitor itself and a 350 V spark gap contained in the bottom-end termination box. For the HF channel, a 110 nF bottom-end capacitor with a 220 ohms shunt

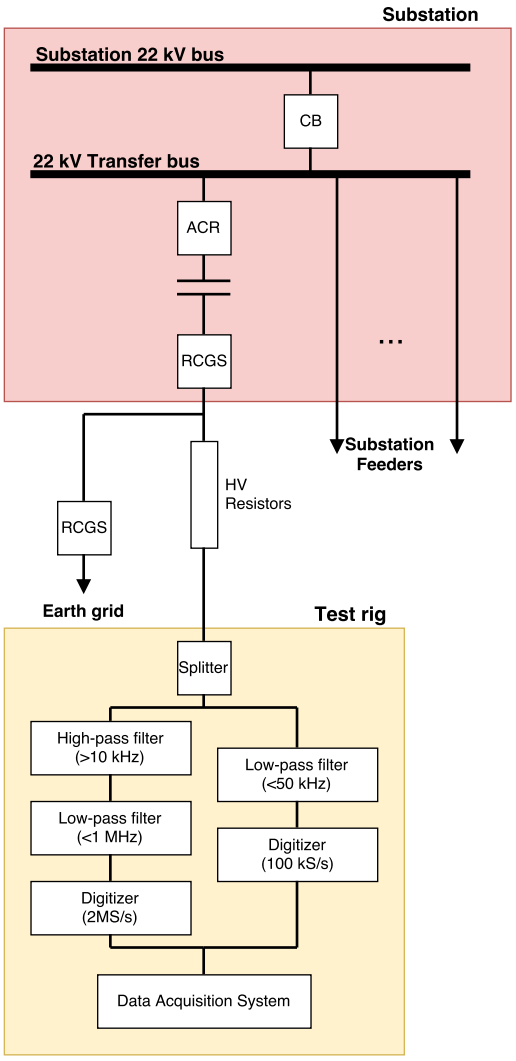

Fig. 5. Unifilar diagram of the feeder and test rig.

resistor provided a ratio of 100:1 at high frequencies and high-pass characteristic with 10 kHz corner frequency. The output signal was processed by a Frequency Device, active 4-pole Butterworth filter to eliminate the 50 Hz signals and low-order harmonics. Moreover, when sampling the signals at 2 MS/s, proper Anti-Aliasing filters were used, represented in the diagram as the Low-pass ($<$1 Mhz) filter.

The data acquisition system used the Gen3i HBM as mainframe, fitted with a HBM GN401 four channel optical input card fed by four HBM GN110-2 opto-isolated digitisers. The digitisers operated at a constant 100 MS/s sampling rate and incorporated anti-alias filters, analog and digital, plus data decimation to reduce the effective sampling rate to the chosen setting. All the information presented in the Appendix regarding the LF channel, test rig, and more, can be found in [2].

Acknowledgment

The authors would like to thank the Powerline Bushfire Safety Program and the Department of Economic Development, Jobs, Transport and Resources from Victoria Australia

for funding such massive program and making data from the tests available. Also, the authors acknowledge Dr. Tony Marxsen, author of the ignition tests' final report, which inspired this work.

REFERENCES

[1] A. Ghaderi, H. L. Ginn Iii, and H. A. Mohammadpour, "High impedance fault detection: A review," *Electric Power Systems Research*, vol. 143, pp. 376–388, 2017.

[2] T. Marxsen, "Vegetation conduction ignition test report - final," Marxsen Consulting Pty Ltd., Department of Economic Development Jobs Transport and Resources, 2015.

[3] N. R. Bahador, F. Namdari, and H. R. Matinfar, "Feature extraction of tree-related high impedance faults as a source of electromagnetic interference around medium voltage power lines' corridors," *Progress In Electromagnetics Research B*, vol. 75, pp. 13–26, 2017.

[4] I. Baqui, I. Zamora, J. Mazn, and G. Buigues, "High impedance fault detection methodology using wavelet transform and artificial neural networks," *Electric Power Systems Research*, vol. 81, no. 7, pp. 1325–1333, 2011.

[5] J. Chen, E. Ambikairajah, D. Zhang, T. Phung, and T. Blackburn, "Detection of high impedance faults using current transformers for sensing and identification based on features extracted using wavelet transform," *IET Generation, Transmission & Distribution*, vol. 10, no. 12, pp. 2990–2998, 2016.

[6] F. B. Costa, B. A. Souza, N. S. D. Brito, J. A. C. B. Silva, and W. C. Santos, "Real-time detection of transients induced by high-impedance faults based on the boundary wavelet transform," *IEEE Transactions on Industry Applications*, vol. 51, no. 6, pp. 5312–5323, 2015.

[7] W. C. Santos, F. V. Lopes, N. S. D. Brito, and B. A. Souza, "High-impedance fault identification on distribution networks," *IEEE Transactions on Power Delivery*, vol. 32, no. 1, pp. 23–32, 2017.

[8] M. Sarlak and S. M. Shahrtash, "High impedance fault detection using combination of multi-layer perceptron neural networks based on multi-resolution morphological gradient features of current waveform," *IET Generation, Transmission & Distribution*, vol. 5, no. 5, pp. 588–595, 2011.

[9] S. Gautam and Brahma, "Detection of high impedance fault in power distribution systems using mathematical morphology," *IEEE Transactions on Power Systems*, vol. 28, no. 2, pp. 1226–1234, 2013.

[10] M. Sarlak, S. M. Shahrtash, and D. Arab Khaburi, "Design and implementation of a systematically tunable high impedance fault relay," *ISA Transactions*, vol. 49, no. 3, pp. 358–368, 2010.

[11] A. Ghaderi, H. A. Mohammadpour, H. L. Ginn, and Y. J. Shin, "High-impedance fault detection in the distribution network using the time-frequency-based algorithm," *IEEE Transactions on Power Delivery*, vol. 30, no. 3, pp. 1260–1268, 2015.

[12] A. Ghaderi, H. A. Mohammadpour, and H. Ginn, "High impedance fault detection method efficiency: simulation vs. real-world data acquisition," in *Power and Energy Conference at Illinois (PECI), 2015 IEEE*. IEEE, Conference Proceedings, pp. 1–5.

[13] 2009 Victorian Bushfires Royal Commission, "Final report," Report, 2010. [Online]. Available: http://www.royalcommission.vic.gov.au/finaldocuments/summary/PF/VBRC_Summary_PF.pdf

[14] *MATLAB and Signal Processing Toolbox Release 2016b*, The MathWorks, Inc., Natick, Massachusetts, United States.

[15] *MATLAB and Statistics and Machine Learning Toolbox Release 2016b*, The MathWorks, Inc., Natick, Massachusetts, United States.

[16] *MATLAB and Wavelet Toolbox Release 2012b*, The MathWorks, Inc., Natick, Massachusetts, United States.

[17] M. J. Reddy and D. K. Mohanta, "A wavelet-neuro-fuzzy combined approach for digital relaying of transmission line faults," *Electric Power Components and Systems*, vol. 35, no. 12, pp. 1385–1407, 2007.

[18] A. H. A. Bakar, M. S. Ali, C. Tan, H. Mokhlis, H. Arof, and H. A. Illias, "High impedance fault location in 11kv underground distribution systems using wavelet transforms," *International Journal of Electrical Power & Energy Systems*, vol. 55, pp. 723–730, 2014.

[19] S. K. Pandey and L. Satish, "Multiresolution signal decomposition: a new tool for fault detection in power transformers during impulse tests," *IEEE Transactions on Power Delivery*, vol. 13, no. 4, pp. 1194–1200, 1998.

[20] M. H. Dhend and R. H. Chile, "Fault diagnosis of smart grid distribution system by using smart sensors and symlet wavelet function," *Journal of Electronic Testing*, pp. 1–10, 2017.

[21] P. Welch, "The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms," *IEEE Transactions on audio and electroacoustics*, vol. 15, no. 2, pp. 70–73, 1967.

[22] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, "Classification and regression trees. wadsworth & brooks," *Monterey, CA*, 1984.

[23] L. Breiman, "Random forests," *UC Berkeley TR567*, 1999.

[24] L. Breiman, "Bagging predictors," *Machine learning*, vol. 24, no. 2, pp. 123–140, 1996.

[25] Y. Freund and R. E. Schapire, "Experiments with a new boosting algorithm," in *icml*, vol. 96, Conference Proceedings, pp. 148–156.

[26] IEEE Power and Energy Society, "Distribution test feeders." [Online]. Available: http://ewh.ieee.org/soc/pes/dsacom/testfeeders.html

[27] M. Sedighizadeh, A. Rezazadeh, and N. I. Elkalashy, "Approaches in high impedance fault detection - a chronological review," *Advances in Electrical and Computer Engineering*, vol. 10, no. 3, pp. 114–128, 2010.

Douglas P. S. Gomes received the B.E. degree in electrical engineering from Federal University of Mato Grosso, Brazil, in 2013 and the M.Sc. degree in Power Systems from University of Sao Paulo, Brazil, in 2016. He is currently working towards his Ph.D. degree at Victoria University, Melbourne, Australia. His research interests are power systems, protection, power quality, and artificial intelligent systems.

Cagil Ozansoy received his B.Eng. degree in electrical and electronic engineering (Hons.) and the Ph.D. research degree in power system communications from Victoria University, Melbourne, Australia, in 2002 and 2006, respectively. He is now working as a Senior Lecturer and Researcher in the College of Engineering and Science, Victoria University. His major teaching and research focus is in electrical engineering, renewable energy technologies, energy storage, and distributed generation. He has successfully carried out and supervised many sustainability-related studies in collaboration with local governments in the past. He has over 25 publications detailing his work and contributions to knowledge.

Anwaar Ulhaq received the B.S.(Hons) in computer software Engineering from University of Engineering and Technology, Lahore, Pakistan. He received M.S. System Engineering degree from G.I.K institute, Topi, Pakistan and Ph.D. (Intelligent Systems) from Monash University, Australia. He is working as a lecturer, information technology in College of Engineering and Science, Victoria University. His research interest includes pattern recognition, signal and image processing and human action recognition.