Multi criteria series arc fault detection based on supervised feature selection

Hien Duc Vu, Edwin Calderon, Patrick Schweitzer, Serge Weber, Nicolas Britsch

Multi criteria series arc fault detection based on supervised feature selection

Hien Duc VU, Edwin CALDERON, Patrick SCHWEITZER, Serge WEBER, Nicolas BRITSCH

***Abstract*— As a field of research arc fault detection in domestic appliances has existed for a long time and many detection algorithms have been published, patterned or implemented on commercial products. None of them, however, guarantees perfect discrimination and all are susceptible to false negatives or false positives (i.e. indicating the absence of arcing fault, when in reality it is present, or recognizing normal functioning as an arcing condition). This phenomenon can be explained by the fact that all methods have been based on some features of arc fault which can be shared with load and network conditions such as noisy loads, the plugging-in or unplugging of appliances, the change of functioning mode of an appliance on a network and so on. A solution for limiting this phenomenon is multi arc-fault feature recognition. This research presents a method for finding and combining arc fault features in order to obtain better performance than using a single arc fault feature. The choice of arc-fault features and the algorithm for combining them are based on machine learning techniques. The method proposed here can be used for different network conditions and loads. The effectiveness of this method has been verified by a number of experimental tests including not only the requirements of standard arc fault detection, but also the most difficult situations such as multiple loads masking and transient loads.**

***Index Terms*— Arc fault detection device; feature selection; series arc fault; multi features arc fault detection; machine learning**

I. Introduction

ARC fault is a harmful condition which may lead to electrical fires. Arc Fault Circuit Interrupters (AFCI) have been introduced in the USA and the National Electrical Code requires their installation in all living areas. A standard IEC already exists in Europe for arc fault detection devices. Line to ground and line to line arc faults can be easily detected due to their effect on current level [1]. However, it is more complicated to detect series arc fault and this is an active area of research. No solution has been found so far which guarantees the detection of all arc faults without ever producing undesirable tripping.

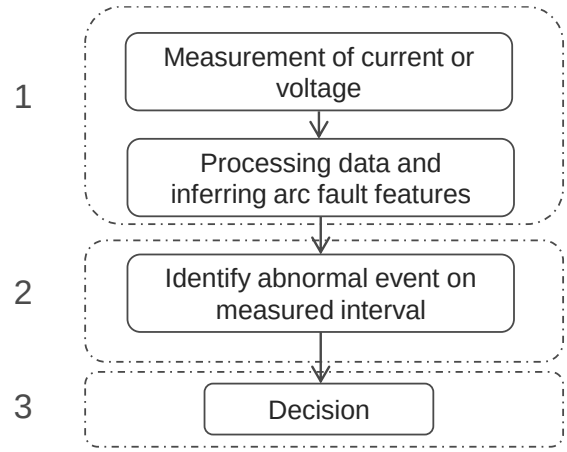

Figure.1 Arc fault detection algorithm.

In general an arc fault detection algorithm (figure 1) can be divided to three main parts: Measurement - feature extraction (1), classification between normal and abnormal situation (2) and decision (3). The purpose of feature extraction part is: inferring arc fault feature (AFF) from acquired signals. To achieve a good performance at detection, appropriate AFFs are mandatory.

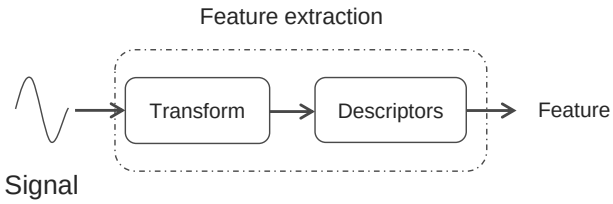

Figure.2 Arc fault features extraction.

Some detection algorithms have a direct feature extraction part such as fractal, current integral and current variation [2] [3] [4], thus AFF can be inferred without any additional steps. For the others, the feature extraction part can be divided into two sub-parts - transformation and descriptors (Figure 2). A number of transformations have been used on arc fault detection fields, including Fourier transform [5] [6], Wavelet transform [5] [7] [8], filtering [6] and correlation [9]. A variety of descriptors have also been mentioned with either frequency or temporal analysis [5] [10] [11] [12] [13]. These include a variation of sub-spectrum energy between two adjacent power cycles as a descriptor for discrete Fourier transform (DFT) [5], harmonic ratio - DFT [10], mean value of the differences [11] on low-frequency spectra of the current measured in two subsequent observations with chirp zeta transform [12] as well as eigenvalue – Kalman filtering [13].

The classification part can be simple, with a fixed threshold for AFF [14], or more complicated with an artificial neural network (ANN) [5] [15] [16] and a support vector machine (SVM) [2] [17]Counting techniques or fuzzy logic are used as a decision strategy [12] [18] [19].

The main difficulty of series arc fault detection lies in discriminating between arcing and normal situations. In general, in order to detect a dangerous arcing condition one or more AFFs can be used [4] [11]. Some of the most commonly used AFFs include current zero crossing, broadband noise and randomness of current variation [9] [20] [21]. However, these features can also be found in normal functioning networks. A vacuum cleaner, for example, may produce a number of arc features at zero crossing. The randomness of current variation can also be mimicked by changing the functioning mode of an appliance, e.g. by switching the power supply to an appliance on and off within a short space of time. Broadband noise may also be generated by electromagnetic interference or noisy loads.

As a result of this problem, every detection algorithm proposes a threshold for discriminating between dangerous arcing and normal situations, thus maintaining a compromise between an unwanted trip (a false negative) and failure to detect an arc fault (a false positive). Each of these situations has a cost associated with it. An unwanted trip may lead to a loss in work time or an interrupted service whilst failure to detect an arc fault is more problematic and may lead to a loss of equipment and even possibly life. It is therefore essential that these errors should be minimized are minimized as far as possible. It can be seen that one AFF may give better detection performance than another AFF in a number of situations and vice versa for the rest [22].

TABLE I
REPRESENTATIVE METHODS AND CORRESPONDING ACCURACIES

Method	Feature type		Lowest prediction accuracy	Problematic al loads
	Direct feature	Transform		
Sparse representation and neural network [16]	250 sparse coefficient from every current half cycles	-	88%	Computer
Chirp Zeta Transform and current difference [12]	The mean value of the difference between two subsequent observation windows	CZT transform with low frequency	96.7%	Electrical drill
Fractal theory and SVM [2]	Box-Counting Dimension and Information dimension	-	98%	Micro oven and induction cooker
Autoregressive Bispectrum Analysis [17]	-	Two-dimensional Fourier transform of third-order cumulants	97%	Vacuum cleaner
High-frequency energy and current variation [3]	Current integral of one period	Short-time Fourier transform	96%	Electrical drill

Table I presents several detection methods with different features and prediction accuracies and the most problematic type of load for each method. In the case of a simple load, such as resistive or inductive, the prediction accuracy of some methods can reach 100%. Even with nonlinear loads (computers, electric drills etc.) some algorithms still stay very close to 100% when the detection task becomes more complicated. Each method struggles only with one or several types of load and therefore a multi-criteria approach may lead to a state-of-the-art result, if all available methods can be correctly combined.

The idea of using more than one several AFF to increase detection performance has been mentioned in some publications and patents, such as verifying the presence of different AFFs [11], or using time characteristics together with the analysis of frequency [3]. However, some very important elements for multi AFF detection are still missing. Amongst these are the choice of AFFs which should be used together and the combination algorithm used to make an efficient detection algorithm from chosen AFFs. This paper presents a methodology for achieving multi-feature arc fault detection and the aim of our research is to take advantage of the numerous arc fault features and to create the most efficient detection algorithm from them.

In this paper we propose a method that involves two successive stages of selection and combination which

The proposed method involves two successive stages of selection and combination which are briefly described in Section II. Section III presents the first step for selecting the best set of descriptors for a given transformation and expected performance. Section IV presents the second step, which aims to select the features and the best combinations between them. The experimental results based on transforms and descriptors found in the literature are presented in section V .and demonstrate that superior performances can be obtained using the proposed method.

II. GENERAL DESCRIPTION

The proposed method consists of two main steps: building a pool of arc fault features, selecting the relevant features then using them together in an efficient way (Figure 3).

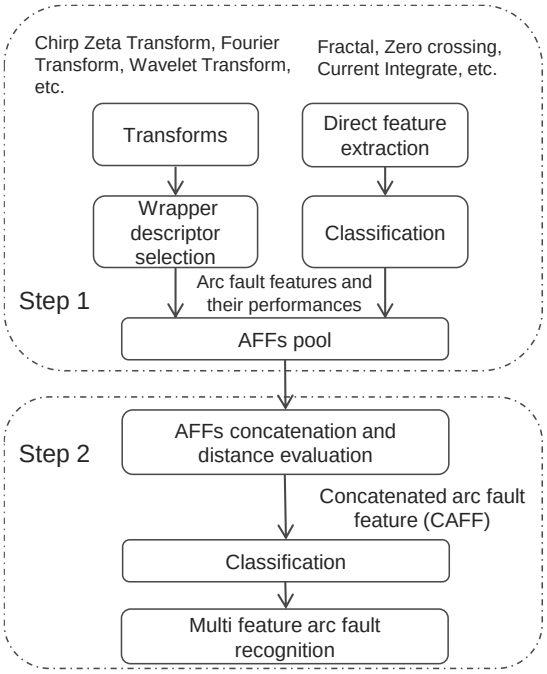

Figure.3 Multi AFFs detection methodology

A. *Arc fault feature pool construction*

1. Transform and descriptor selection

For AFFs which have separate transforms and descriptors, the wrapper selection method was applied in order to guarantee the best set of descriptors for a given transformation. Selection was based on a classification algorithm, which may be any machine learning technique such as an artificial neural network, a support vector machine etc. Prediction mean squared error was used as the criterion for evaluating the chosen classification method. This is the most commonly used cost function for evaluating classification machine learning based techniques [23].

2. Direct feature extraction

Some AFFs can be obtained by the direct feature extraction method [2] [3]. The only information required before adding them into the AFF pool is their respective accuracy in the detection task. Each AFF should be evaluated using the chosen classification algorithm at the descriptor selection stage in order to determine the information required.

B. *Arc fault feature combination*

The second step consists in ranking all AFFs, removing the irrelevant ones and keeping only those performing well. The combined arc fault feature (CAFF) is then made by combining the best AFF with the other AFFs which have passed the ranking stage. Euclidean distance on feature space has been used as the criterion for selecting which AFF should be combined with the best performing AFF. To be more precise, the ratios of the distance between the false positive (FP) and the false negative (FN) elements and the correct classified elements of the best performing AFF after concatenation were evaluated. The CAFF was considered the best if it had the highest ratio of distance. At the end, the CAFF was used as input for arc fault recognition.

III. WRAPPER DESCRIPTOR SELECTION

This section explains the necessity for descriptor selection for each transform and the development approach with the wrapper method.

A. *Transform and optimal descriptor sets*

Many possible descriptors can be used for a given transform. All feature extraction methods employ one or several descriptors. For example Artale et al used chirp zeta as transform (CZT) and the mean value of the differences between the two low-frequency spectra of the current measured in two subsequent observation windows and the differences between the maximum values of spectra in specified frequency intervals [12]. Hadziefendic et al presented an algorithm which used Fourier transform (FT) and the fifth current harmonic [10]. These descriptors are interchangeable between different transforms. The fifth current harmonic may also be worth using with CZT and the other descriptors with FT.

Descriptors are often chosen based on the observation of experimental data. A review of the literature shows that descriptor sets are frequently chosen with the help of a relatively small (generally less than 10) number of appliances [2] [9] [12] [14] [22]. Consequently, the performance of a descriptor set can only be justified by the relevant experimental test.

Further, an arc fault detection method should be able to work on a number of different installations. The step for determining the optimal set of descriptors is very important for obtaining the desired AFF. It should be noted that the optimal descriptor set will be strictly related to the transform.

If only a few descriptors from all the available descriptors are used, a large amount of information provided by the transform will be wasted. Conversely, using a large number of descriptors also has disadvantages, requiring sophisticated detection rules which are hard to achieve with the classical heuristic method of observation and multi-thresholds. Detection rules based on machine learning can simplify the task. However, too many descriptors may induce a loss of generalization and an over fitting problem due to noise and redundancy [24]. Therefore only an optimal number of descriptors should be used.

B. *Wrapper descriptor selection*

Technically, both heuristic and machine learning based rules may help to establish the optimal descriptor set. The classic observation and thresholds technique can perform well if all possible situations are taken into account. However, in domestic electrical networks, there are a large number of frequently used appliances and network conditions and it is unrealistic to perform a sufficient number of tests to cover all cases. Thus, the generalization property of a descriptor must be given priority. For this reason, the machine learning technique was chosen instead of hand-crafted heuristic rules. The feature selection approach on machine learning was used to determine the best set of descriptors for each transformation.

Machine learning based selection method can be divided into three categories: supervised, unsupervised and semi-unsupervised. Supervised selection can be categorized as the filter method, wrapper method and embedded method. The

filter method separates descriptor selection from the classification algorithm and relies on the measurement of general characteristics from data, namely distance, dependence, mutual information and correlation. The wrapper method uses the predictive accuracy of a given classification algorithm to determine the quality of the selected subset. The embedded method is a hybrid of the wrapper and filter methods and uses both statistical analysis and model fitting with a classification algorithm [23] [24]. In this case, wrapper selection is the most appropriate method as it measures the usefulness of descriptors rather than the relevance of their correlation with training data such as principal component analysis and the filter method. This property of the wrapper method is very important because a descriptor which is inefficient by itself can provide a significant improvement in performance when taken with others descriptors [24] [25] . Computation costs and the risk of over fitting can be ignored since the number of descriptors considered for each transform is less than twenty for arc fault application.

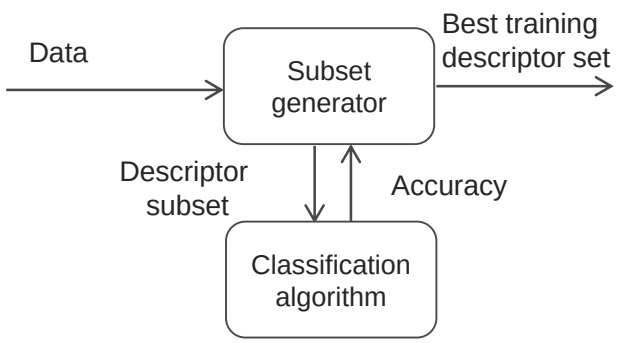

Figure.4 Wrapper descriptor selection

Figure 4 presents the wrapper descriptor selection principle which works by generating candidate subsets from all available descriptors and then evaluating each subset with a classification algorithm. Several strategies for generating subsets can be listed, each with their own advantages and disadvantages, including exhaustive search (brute force [26], branch and bound [27]), heuristic (hill climbing, best first search) and meta-heuristic (genetic algorithm and particle swarm) [28] [29]. Compared to the other methods, the exhaustive search method guarantees that the best subset of features will be found. The drawback of exhaustive search, however, is computation cost, which may increase exponentially with the number of descriptors. This method therefore becomes impractical as the number of descriptors increases. In arc fault application this same drawback can be ignored and in our work exhaustive search has been retained. Each generated subset was evaluated with the same classification algorithm. The best subset can be obtained when every generated subset has been examined.

C. Development approach

The method proposed requires a database of electrical signals collected from experimental domestic networks. This database is divided into two datasets - training and test. Ideally the training dataset contains a fixed number of appliances and network situations and the test dataset contains more situations and appliances than the training dataset.

Every feature extraction method which is composed of transform and descriptor steps is evaluated. For each transform all possible descriptors are used to create the set of all descriptors. Figure 5 illustrates the selection method.

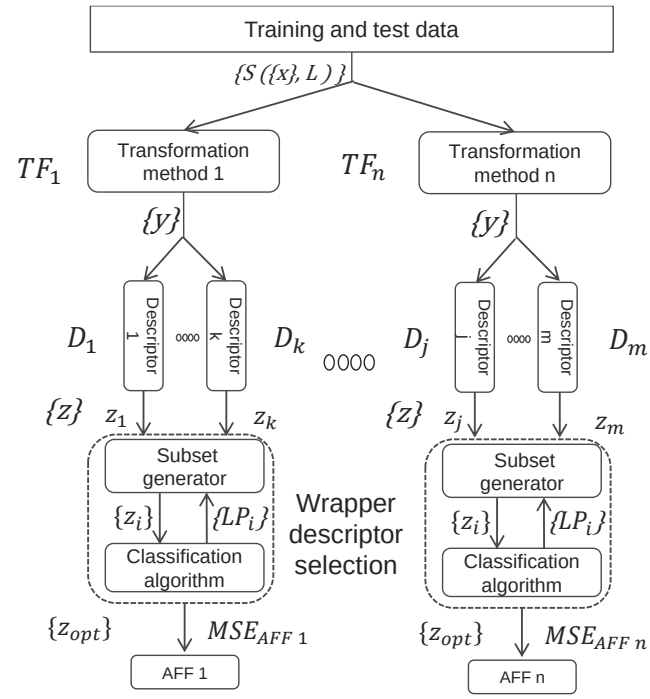

Figure.5 Wrapper method for descriptor selection

The training dataset contains a very large number of labeled samples S_i, each sample containing a discrete time sequence $\{x\}$ of current or voltage measured, and the associated label L_{Si}:

$$\begin{cases} L_{Si} = 1 \ \ in\ case\ of\ arc\ fault \\ L_{Si} = 0\ otherwise \end{cases}$$

n transformation methods will be considered:

$$TF_1, TF_2\ , \dots\ TF_n$$

The result obtained after a transformation can be noted as:

$$\{y\}\ \ with\ \{x\} \xrightarrow{TF} \{y\}$$

For a transform under evaluation k descriptor D can be used, for example:

$$D_1\ , \dots D_k : \{y\} \xrightarrow{D_1\ ,\dots D_k} \{z_k\} := \ z_1, z_2, \dots\ z_k$$

For a given subset $\{z_i\}$, $\{z_i\} \subseteq \{z_k\}$ a classification algorithm can be deployed on the training data to discriminate between arc and no arc samples. Mean square error (MSE) is the objective function used for evaluating trained classifier on test data.

$$MSE(\{z_i\}) = \frac{1}{NT}\sum_{m=1}^{NT}(L_m - \ LP_m)^2$$

NT is number of samples on the test dataset.

LP_m is the prediction of label given by the trained classifier of a sample m .

$\{z_i\}$ is the chosen subset .

The subset$\{z_{opt}\}$ is the optimal subset $\leftrightarrow \{z_{opt}\} \subseteq \{z_k\}$, MSE $(\{z_{opt}\}) \leq$ MSE $(\{z_i\})\ \forall\ \{z_i\} \subseteq \{z_k\}$.

The descriptor selection problem consists in finding the $\{z_{opt}\}$ and MSE $(\{z_{opt}\})$ for every transformation.

One simple solution for obtaining a result consists in evaluating every possible subset of $\{z_k\}$. However, this solution has the highest computational cost and many unnecessary subsets would be evaluated. For example, if one of the most relevant descriptors is removed, the classification performance is affected. It is no use evaluating all the subsets without this relevant descriptor. In order to avoid this problem, the search solution, the branch and bound elimination algorithm, can be deployed.

First the $\mathrm{MSE}(\{z_k\})$ *is* calculated, the bound value θ= MSE $(\{z_k\})$ is set to the child nodes (subset derived from $\{z_k\}$) on the first level of depth. On this level, all descriptors will be removed one by one from z_1 to z_k and the corresponding subset: $\{z_{k1}\}, \ldots, \{z_{kk}\}$ $with$ $\{z_{ki}\} \backslash \{z_k\} = z_i$ $(1 < i < k)$ is evaluated. z_i can be defined as a significant descriptor if $\mathrm{MSE}(\{z_i\}) \geq \theta$ or less-significant descriptor if $\mathrm{MSE}(\{z_i\}) < \theta$.

After this first level, either depth-first search or breadth-first search algorithms [30] can be used. The depth-first search algorithm gives priority to investigating on branch before backtracking, and the breadth-first search to exploring neighbor nodes first before going deeper. If the number of descriptors is too high the breadth-first algorithm may help to limit the calculation time by fixing a depth level. Conversely, if a performance has been defined, the depth-first algorithm may converge faster to optimal solution with the lowest number of descriptors. In this paper the depth-first algorithm has been used.

The child nodes are created by removing one less-significant descriptor from the parent nodes. This newly created child node must be different from any node on the left in order to avoid any unnecessary computation. The upper bound for the any child node can be defined as follows: $\theta_{Child} = MSE\ (\{z_{Parent}\})$ where $\{z_{Parent}\}$ is the subset which generates the child node.

For a given transformation the algorithm stops when every node has been evaluated. As a result both optimal subsets of descriptor - $\{z_{opt}\}$ and MSE $(\{z_{opt}\})$ are determined. Figure 6 shows an example using a depth-first search.

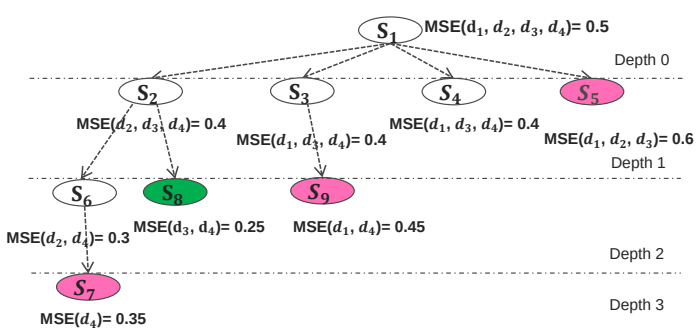

Figure.6. Example of depth-first branch and bound elimination for descriptor selection. MSE should be minimized. The descriptor subset was evaluated from S1 to S9. After the first depth level evaluation, in the step 5(S5), descriptor 4 (d4) is considered as a significant descriptor and therefore all subsets without d4 are not evaluated. Red nodes violate the upper bound (MSE of parent node), therefore no more child nodes are generated from these nodes. Subset (d3, d4) is the optimal subset in this example.

Pseudo code for descriptor subset selection based on backward elimination:

```
1.  Descriptor set {z_k} := z_1 , z_2, ... z_k
2.  MSE{z_k} = Compute (MSE of classifier trained from {z_k})
3.  Global variable min_MSE = MSE{z_k} ,best_set = {z_k} ,
4.  List of evaluated descriptor set : EVA_list
5.  List of MSE value corresponds for each removed descriptor :
    MSE_list
6.  List of less significant descriptor : LSD_list
7.  for each z_i of {z_k}
8.   {z_ki}  =  remove z_i from {z_k}
9.   MSE{z_ki} = Compute (MSE of classifier trained from {z_ki})
10. MSE_list[i] = MSE{z_ki}
11.     if MSE{z_ki} < min_MSE
12.         LSD_list = LSD_list append z_i
13.     end if
14. end for
15. for each z_i of LSD_list
16.     {z_n} =  remove z_i from {z_k}
17.     B_ELI( {z_n} , MSE_list[i ] )
18. end for
19. end of program
20.
21. Function B_ELI is
22. Input: Descriptor set {z_n} and MSE
23. for each z_i of LSD_list
24.  {z_ni} = remove z_i from {z_n}
25.     if {z_ni} ∈ EVA_list
26.         continue
27.     end if
28.     EVA_list = EVA_list append {z_ni}
29.     MSE{z_ni} = Compute (MSE of classifier trained from {z_ni})
30.     if MSE {z_ni} < MSE
31.         if MSE {z_ni} < min_MSE
32.             min_MSE = MSE{z_ni}
33.             best_set = {z_ni}
34.         end if
35.         B_ELI( {z_ni} , MSE{z_ni} )
36.     end if
37. end for
38. return
```

Descriptor selection for AFF pool construction is completed when all optimized descriptor sets for every transformation have been found. The pool of available arc fault features is noted as P.

IV. Arc Fault Feature Combination

The combination method consists of 3 steps: the ranking step on P, the distance evaluation of different generated CAFFs and the classification step which uses the CAFFs with the highest ratio of distance. The ranking step helps to find the most relevant arc fault feature (AFF_1) and to remove the irrelevant (poor detection performance) AFFs. $\{AFF_1\}$ has the lowest MSE (MSE $(\{AFF_1\}) \leq$ MSE $(\{AFF_i\})$ $\forall$ i≠1, $\{AFF_i\} \subseteq P$).

If MSE $(\{AFF_1\})$ is higher than the desired error (noted d_mse), a solution for obtaining the performance we want might be to combine AFF_1 with the other AFFs. The simplest method for combining features without losing any information is concatenation of features to create a new feature vector. For example in the case of two AFFs, from $\{AFF_1\} =: z_1, z_2, \ldots z_i$ and $\{AFF_2\} =: z_j, z_{j+1}, \ldots z_{j+k}$ concatenated feature $\{CAFF\} =: z_1, z_2, \ldots z_i, z_j, z_{j+1} \ldots z_{j+k}$ can be formed.

By using the result from AFF_1 and its trained classifier, the data set $\{S\}$ can be divided into four groups:

$$\forall S_i \in \{S\} \begin{cases} S_i \in \{TP_{AFF1}\} \leftrightarrow L_{Si} = 1, LP_{Si} > ThS \\ S_i \in \{TN_{AFF1}\} \leftrightarrow L_{Si} = 0, LP_{Si} \leq ThS \\ S_i \in \{FN_{AFF1}\} \leftrightarrow L_{Si} = 1, LP_{Si} \leq ThS \\ S_i \in \{FP_{AFF1}\} \leftrightarrow L_{Si} = 0, LP_{Si} > ThS \end{cases}$$

TP, FP, FN and TN stand for true positive, false positive, false negative and true negative groups. The calculated label LP_{Si} is a real number between 0 and 1, thus a threshold ThS is needed to define these groups. Since the sensitivity and specificity (proportion of false positives and false negatives) of the model are not of concern in this paper, the threshold ThS is fixed at 0.5 (which is analog to the balanced value).

Since AFF1 is insufficient to completely resolve the detection problem, the four groups FP_{AFF1}, FN_{AFF1}, TP_{AFF1} and TN_{AFF1} cannot be easily distinguished. The groups FP_{AFF1} and FN_{AFF1} always overlap with groups TN_{AFF1} and TP_{AFF1}. An illustration of group distribution is shown in Figure 7.

If any other AFF is more efficient than AFF_1 for sensitive elements (FP_{AFF1} or FN_{AFF1}), this AFF may help to better separate the groups FP_{AFF1}, FN_{AFF1} from TP_{AFF1} and TN_{AFF1}. As the distance between these groups and their distribution on feature space are related to prediction performance [31], higher detection performance can be expected after this combination. An illustration of group distribution on CAFF feature space is show in Figure 7.

If it is assumed that the group TP_{AFF1} has k element, the centroid of this group after concatenation (CAFF feature space) can be determined with the following relation:

$$Centroid\ (TP_{AFF1}) = \frac{{TP_{AFF1}}^1 + {TP_{AFF1}}^2 + \cdots + {TP_{AFF1}}^k}{k}$$

The centroids of the other groups can be determined with the same formula. Let $\vec{u}, \vec{v} \in R^n$ be the centroids of group TP_{AFF1} and TN_{AFF1}, the inter-group distance d_1 between TP_{AFF1} and TN_{AFF1} can be calculated:

$$d_1 = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2 \ldots (u_n - v_n)^2}$$

Similarly for the other inter-group distances:

$$d_2: distance\ (TP_{AFF1}, FP_{AFF1})$$
$$d_3: distance\ (TN_{AFF1}, FN_{AFF1})$$

The most efficient CAFF is that which has the best separation between groups $TP_{AFF1} - FP_{AFF1}$ and $TN_{AFF1} - FN_{AFF1}$. In other words, the one with highest sum of ratio distances:

$$\frac{d_2 + d_3}{d_1}$$

The higher the inter-group distances d_2, d_3, the better we can expect separation. However, it is also important to take into consideration the dimensions of different CAFFs feature space and it is for this reason that distance d_1 has been used as a normalized coefficient.

The next step consists in discriminating between arc fault and non-arc situations with the help of a classification method and the most efficient CAFF (created from AFF1 and another AFF).

If the mse of CAFF is still higher than the desired mse, the number of AFFs used for combination should be increased. More specifically, after each stage of evaluation if the desired performance is still out of reach, the number of complementary AFFs is increased by one. The algorithm stops when the desired error has been achieved or all possible CAFFs have been examined. In the second case, the pool of arc fault features should be revised (add new transforms and use more suitable descriptors).

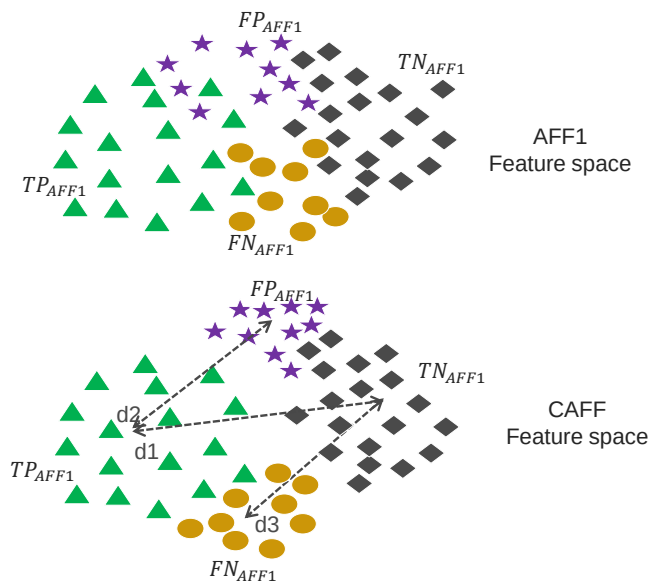

Figure.7 Illustrative of group distribution when CAFF provides a better discrimination than AFF

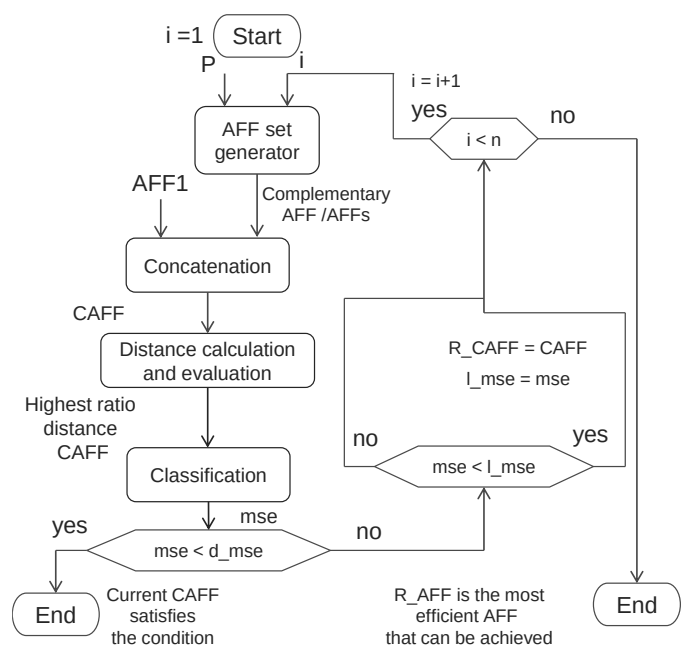

Figure.8 Arc fault feature combination based on inter-group Euclidean distance. L mse is the lowest mse up to the current stage, R_CAFF is the best CAFF that can be achieved up to the current stage, i: the number of complementary AFFs used for combination at the current stage, n: the number of complementary AFFs available.

Figure 8 describes the combination method in detail. At the beginning i is equal to 1 (stage 1). Every possible CAFF (AFF1 concatenate with another AFF) is evaluated. The

highest distance CAFF is used to train a classifier and then the classifier's performance is compared to the desired mse. If the current CAFF satisfies the condition, the combination algorithm reaches its end. Otherwise, CAFF will be noted as R_CAFF and its mse as l_mse. The variable i will be incremented by 1 (stage 2) and every CAFF (concatenated from AFF1 and two other AFFs) is evaluated. The process is repeated until a satisfied CAFF has been found or i equals n. In the second scenario, R_CAFF is the most efficient CAFF and l_mse is the lowest error that can be achieved.

In comparison to the wrapper selection method used in the previous section, the distance evaluation method greatly reduces the computation cost because there is only one classifier which requires training at each stage. The computation of ratio distance is negligible compared to the cost of the training classifier. If the AFF pool has n elements, in the worst case scenario the wrapper method needs to train $\sum_{k=1}^{n} \frac{n!}{k!(n-k)!}$ classifiers and only n classifiers for the distance evaluation method.

V. Experimental Result

This section describes in detail how the method presented in this paper has been applied. The results obtained can be found at the end of the section.

A. *Arc fault detection database*

A database is essential for the proposed method and the construction of the database is very important as it directly affects how efficient the obtained detection method can be. The number of appliances, combinations and disturbances presented in an electrical network are very large and it is impractical to build a universal database that covers all possible situations. There is no need to make a universal detection method for all installations and it is always possible to construct a useful database for any installation when the number of situations is narrow.

In this series of experiments the European household network was studied. For this installation, the standard IEC 62606 –"General requirements for arc fault detection devices" was used as the main reference for database construction. The database contained the sample with arc fault and non-arc fault situations.

In order to generate arc fault samples, the following configurations were used:

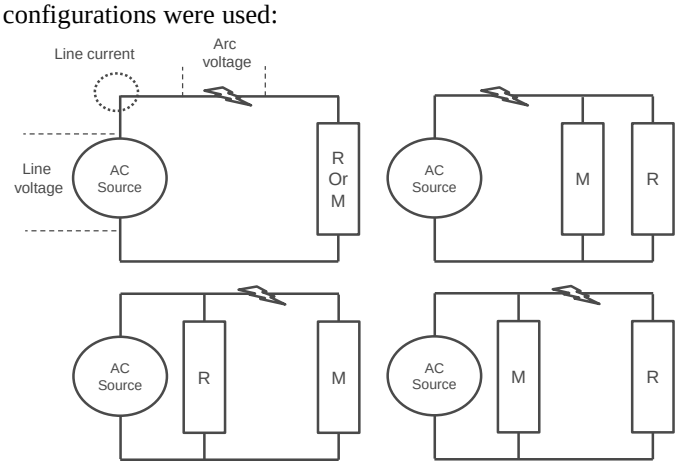

Figure.9 R: Resistive load, M: Masking load – list of masking loads can be found below. Several network disturbances are also added to each configuration. The AC source (230V – 50 Hz) generated in accordance with the standard for distribution networks in Europe

The arc fault is created by a carbonized cable specimen according to standard IEC 62606.

The non-arc samples are also generated with the same configuration. In order to guarantee good performance against false positive errors, many samples contain the transient state of appliances. In addition, the standard cross-talk test has been also taken into account in the database. The following types of appliances were used for constructing the database: masking loads (air conditioner, air compressor, computer, dimming lamp, electric drill, vacuum cleaner, halogen lamp, fluorescent lamp and hair dryer) and resistance. At least two or three different brands were used for each type of appliance.

Each sample contains a measurement of network voltage, current and arc voltage for a fixed duration. It is essential to measure the current for the method presented.

Measuring arc voltage helps to accurately label the samples and every period of each sample is labelled. If arc voltage stays at noise level for one period, the period is labelled as normal. If the arc voltage wave form corresponds to an arcing situation, the period is labelled as arc influenced. The period which is not completely affected by arc fault is not labelled (example shown in Figure 10). There are a total of 16,231 samples – 3,405 samples with arc fault, and 12,826 without arc fault. The signals were acquired with the sampling frequency of 1 MHz, because all feature extraction methods in this study operated at a frequency lower than 1 MHz

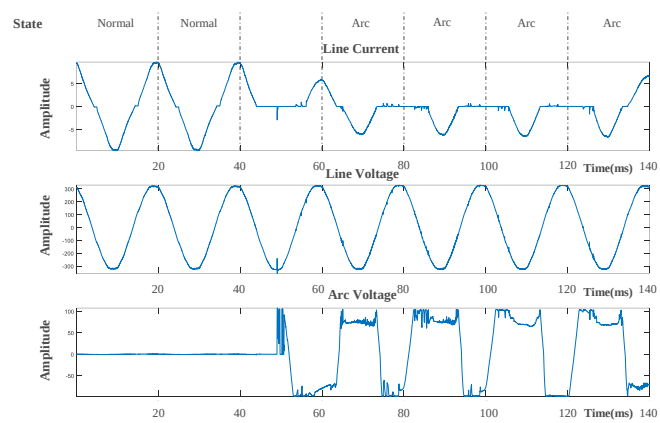

Figure.10 Example of signal

B. *AFFs pool construction*

After the database was established, the next step consisted in creating an arc fault features pool. This step was accomplished by applying different feature extraction methods to the current signature of the database. In this series of experiments only those feature extraction methods which are composed of transforms and descriptors were used.

1. *Transformations*

Five different transforms were chosen in this study because they are used in arc fault detection literature [12] [18] [22] [32].

All transforms are defined below; $\{x_n\}$ represents the current time series input and $\{y_k\}$ is the result obtained with the respected transform, N is the number of elements in a current series:

- Current finite difference:
$$y_k = |x_{k+1} - x_k|;\ 1 \le k \le N\text{-}1$$
- Discrete Fourier transform (FFT):
$$y_k = \left|\sum_{m=1}^{N} x_m . e^{-i\frac{2\pi}{N}km}\right|;\ 1 \le k \le N$$
- Chirp Zeta transform (CZT):
$$y_k = \left|\sum_{m=1}^{N} x_m . e^{-i\frac{2f\pi}{f_s N}km}\right|;\ 1 \le k \le N$$

Four different frequency bands were considered for FFT and CZT:

1 kHz - 10 kHz
10 kHz - 20 kHz
50 kHz - 80 kHz
80 kHz - 100 kHz
100 kHz - 150 kHz

- Wavelet

$$y(m,k) = \frac{1}{a^m}\sum_{m=0}^{N-1} x_n . g\left(\frac{k-b}{a^m}\right)$$

$g(.)$ is the mother wavelet
m is the decomposition level, in this paper a=2 .

For mother wavelet: Daubechies 4 (DB4) and Meyer (Dmey) were selected with decomposition level 2, 3, 4 and 5 (LVL2, 3, 4, 5) for each wavelet respectively.

Figure 11 shows an example of transform with arc fault and non-arc signal.

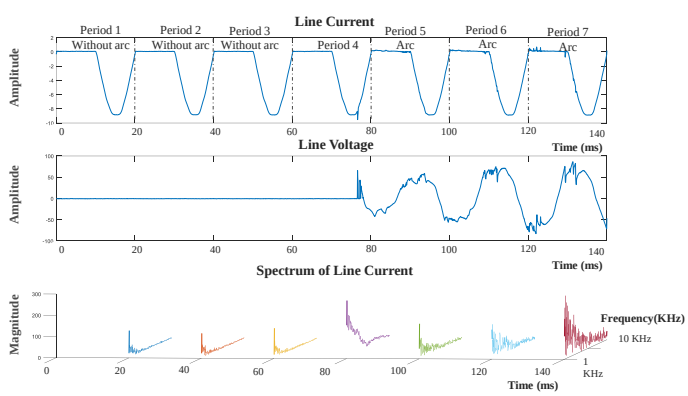

Figure.11 Fourier transform with non-arc and arc fault signal

2. Descriptors

The results after each transformation are discrete series and they can be noted as $\{y_n\} := y_1, y_2, \dots, y_n$. In order to finish the feature extraction step, every transform needs to be associated with several descriptors. Based on the literature of arcing detection [32], the groups of descriptors were chosen as follows.

The first group of descriptors is based on statistical analysis. In this paper, the first, second, third and fourth order moment were chosen with the aim of measuring the shape of $\{y_n\}$.

- Mean value: $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_n$
- Variance: $s^2 = \frac{1}{n}\sum_{i=0}^{n}(y_n - \bar{y})^2$
- Skewness: $\frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_n-\bar{y}}{s}\right)^3$
- Kurtosis : $\frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_n-\bar{y}}{s}\right)^4$

The second group of descriptors relates to the analysis of the first peak of $\{y_n\}$ (the peak's value, location and duration). Finding all the points of $\{y_n\}$ around the first peak, which are higher than the average value, allows the duration of the signal to be found. An illustration of these descriptors can be found in Figure 12.

- The max value: $y_m \in \{y_n\}$, $y_m \ge y_i\ \forall\, y_i \in \{y_n\}$
- Normalized index of the max: $\frac{m}{n}$
- Normalized highest pulse duration:

$$\frac{k-j}{n},\begin{cases} j \le m,\ y_i \ge \bar{y}\ \forall\ j \le i \le m \\ y_{j-1} < \bar{y}\ or\ j = 1 \\ k \ge m, y_i \ge \bar{y}\ \forall\ m \le i \le k \\ y_{k+1} < \bar{y}\ or\ k = n \end{cases}$$

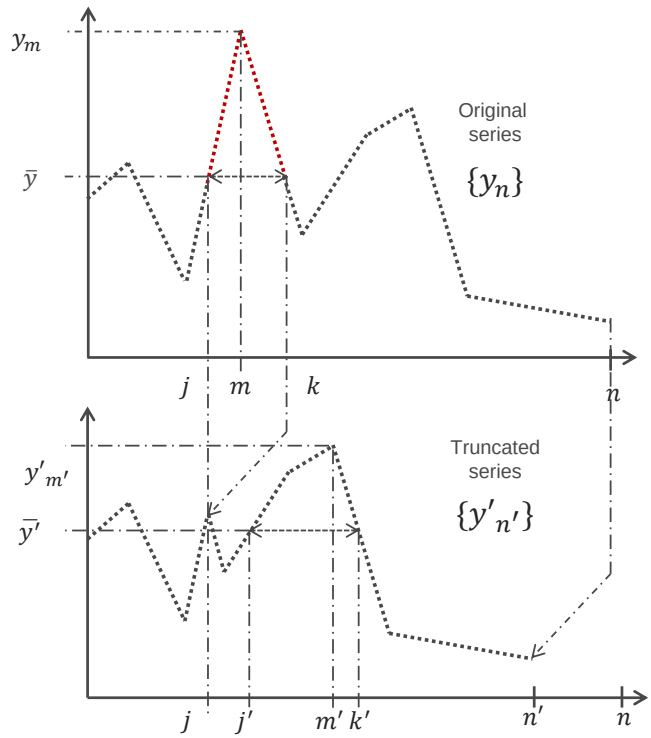

Figure.12 Illustrative of customize descriptors 2 - 7

The last group of descriptors gives information about the second peak of data series $\{y_n\}$. The highest pulse is removed (all points between j and k of $\{y_n\}$) in order to find the second peak. The truncated series $\{y'_{n'}\}$ can be defined as:

$$\{y'_{n'}\} := y'_{1'}, y'_{2'}, \dots, y'_{n'} = y_1, y_2, \dots y_j, y_k, \dots, y_n$$

The first peak of $\{y'_{n'}\}$ is now equivalent to the second peak of the data series.

The second peak value is:

$$y'_{m'} \in \{y'_{n'}\},\ y'_{m'} \ge y'_i\ \forall\, y'_i \in \{y'_{n'}\}$$

Normalized index of the second peak: $\frac{m'}{n'}$

Normalized second pulse duration:

Note that $\bar{y'} = \frac{1}{n'}\sum_{i=1}^{n'} y_{n'}$

$$\frac{k'-j'}{n'}, \begin{cases} j' \le m',\ y'_i \ge \bar{y}'\ \forall\ j' \le i \le m' \\ y'_{j-1} < \bar{y}'\ or\ j' = 1 \\ k' \ge m',\ y_i \ge \bar{y}'\ \forall\ m' \le i \le k' \\ y'_{k+1} < \bar{y}'\ or\ k' = n' \end{cases}$$

The figure 13 shows an example of 10 descriptors for a given transform and their numbering.

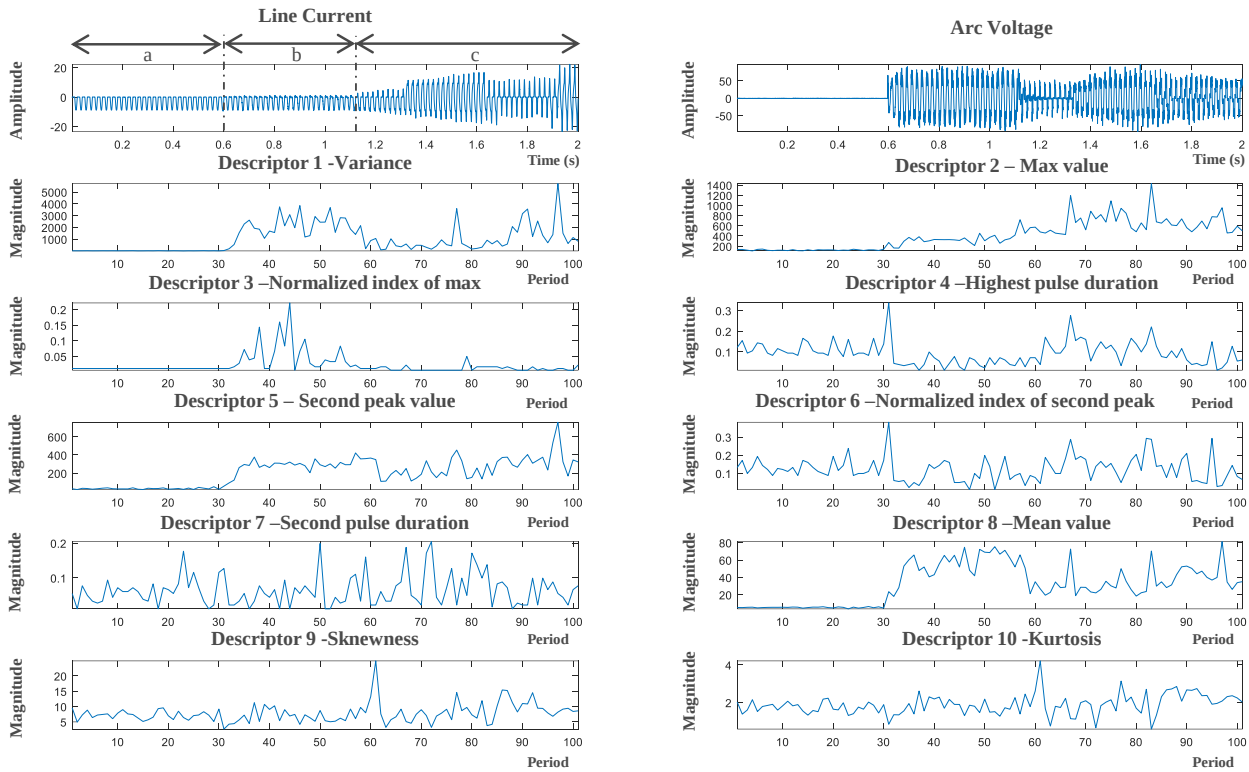

Figure.13 Example of ten descriptors, DFT frequency band 1-10 KHz with non-arc and arc fault signal
a-Current without arc fault at steady state; b-Current with arc fault at steady state; c-Current with arc fault and transient (change of functioning mode).

3. Neural network wrapper descriptor selection

It was our objective to select the set of descriptors that provides the best performance for each transform selected in this article; irrelevant descriptors for each transform were eliminated. In order to accomplish this, a fully connected feed forward artificial neural network (ANN comprising four layers) was used as the classifier. The first hidden layer is composed of 20 neurons and the second one of 8 neurons (Figure 15).

The inputs of the neural network are composed of the arc features obtained from the different descriptors whose calculation procedure is explained in Figure 14.

The analysis is based on the line current without arcing (labeled 0) or with arc (labeled 1). The calculation is performed in a window of 20 ms for each descriptor. Three successive periods (60 ms) with the same label provide three different sets of values (z_i, z_{i+1}, z_{i+2}). Based on the ten descriptors selected, the neural network therefore has 30 entries. One of the 10 descriptors was subsequently excluded and this was done for each of the ten descriptors. In each case, the new generated subset is composed of 27 values for the ANN input. Two or more descriptors can then be removed from the list for analysis.

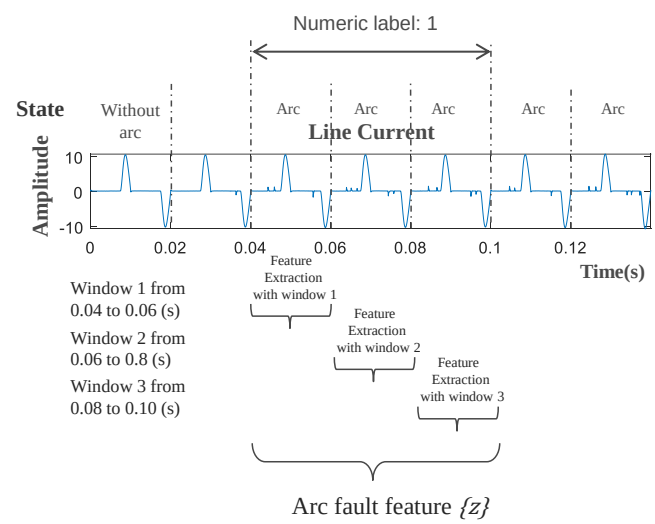

Figure.14 Label and arc fault feature construction

For each generated subset, the mean squared error (when the value of the ANN output equals LP) is estimated according to the following equation:

$$MSE(\{z_i\}) = \frac{1}{NT}\sum_{m=1}^{NT}(L_m - LP_m)^2$$

In this experiment, NT = 16231 and for any given sample m of the database, the squared error can be calculated as follows

$$Squared\ Error(\{z_i\}) = \begin{cases} (1-LP)^2 & \text{When an arc occurs} \\ (0-LP)^2 & \text{When no arc occurs} \end{cases}$$

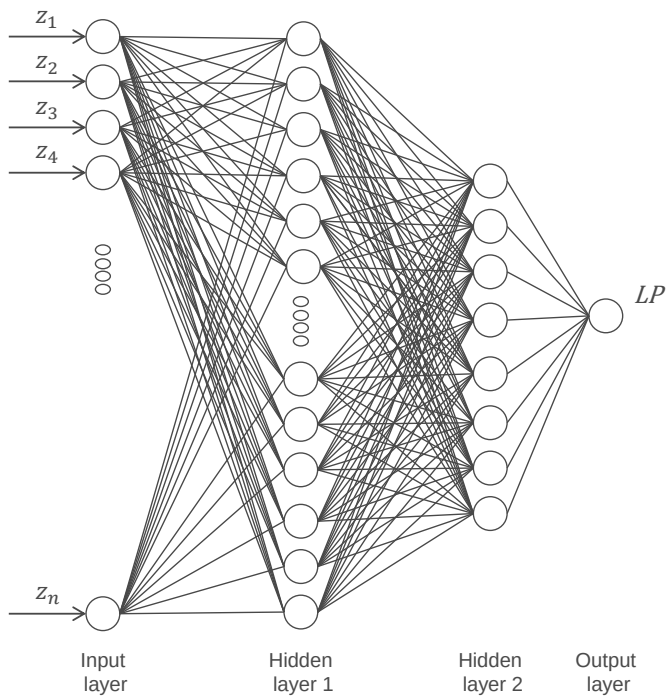

Figure.15 Neural network for arc fault detection

Figure 16 shows the results obtained from the error depending on the descriptors considered. A « removed descriptor » of 0 indicates that all the 10 descriptors have been considered, 1 refers to the fact that the first descriptor has been removed and so on. In some cases two or more descriptors can be removed with the backward elimination algorithm mentioned above.

The results are presented with respect to the frequency band rather than by the type of transform.

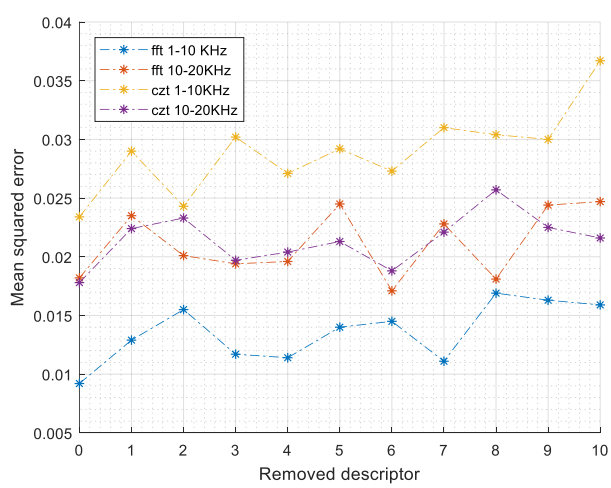

Figure.16 Descriptor selection with low frequency FFT and CZT (1- 20 KHz)

At the frequency band 1-10 kHz the FFT has lower MSE than CZT. For the frequency band 10-20 kHz, FFT and CZT give similar results (the red and magenta lines respectively), their respected minimal MSE are 0.0171 and 0.0178. The optimized descriptor sets are almost the same (all descriptors are important) except for FFT at 10-20 KHz. Descriptor 6 should be removed in order to achieve the lowest MSE. (Figure 16)

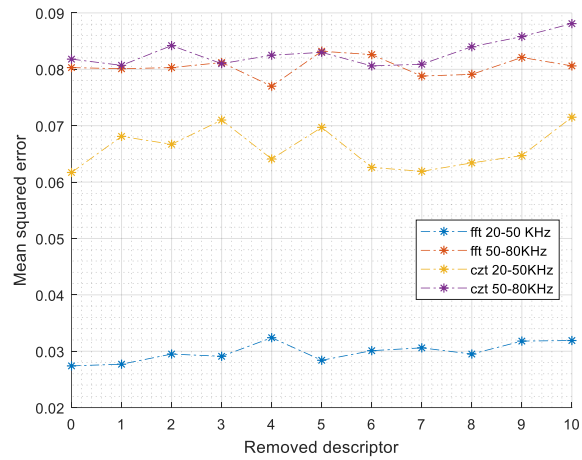

Figure.17 Descriptor selection with middle frequency FFT and CZT (20 - 80 KHz)

The frequency bands 20-50, 50-80 kHz show poorer performance compare to the 1-10 kHz and 10-20 kHz band. All descriptor are necessary for FFT and CZT at the frequency band 20-50 kHz. The descriptor 6 needs to be removed for CZT 50- 80 kHz. The analysis of the results obtained from FFT in the band 50-80 kHz show that descriptors 2,3,5,6,9 and 10 are relevant for the detection (Figure 17).

It remains to be decided whether the other descriptors should be retained or not. The process is illustrated in Figure 18. By considering the set that contains all descriptors (z), the mean square error is equal to 0.0803. When one of the descriptors 1, 4, 7 or 8 is removed, the error decreases (z_1 , z_4 , z_7, z_8). Three subsets, derived from subset z_1, can be generated, namely $z_{1,4}$, $z_{1,7}$, $z_{1,8}$ (descriptor {1,4}; {1,7} or {1,8} removed). Only subset $z_{1,7}$ reduces the error (MSE = 0.077 < 0.081) and subsets $z_{1,4}$ and $z_{1,8}$ show higher errors in comparison to the parent subset; therefore descriptor 4 or 8 should not be removed with descriptor 1. As a result, subsets $z_{1,7,4}$ and $z_{1,7,8}$ were not evaluated. The subsets derived from subset z_4, that is subsets $z_{4,1}$, $z_{4,7}$, $z_{4,8}$, can be created. Subset $z_{4,1}$ was evaluated before and the error of subset $z_{4,7}$ and $z_{4,8}$ is yet to be found. Since their errors are higher than the error obtained with z_4, there is no need to evaluate further. Similarly, the last subset $z_{7,8}$ has a higher error than the error obtained with z_8. In conclusion, descriptors 1 and 7 should be removed in order to achieve the best performance.

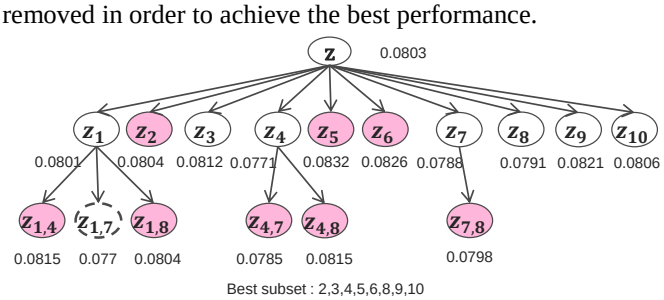

Figure.18 Descriptor selection process

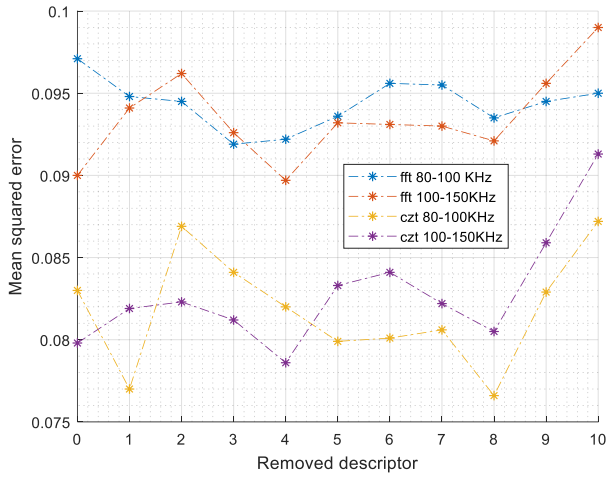

Figure.19 Descriptor selection for high frequency FFT and CZT (80 - 150 KHz)

Figure 19 shows that the detection performances of frequency bands 80-100 and 100-150 kHz are mostly identical to the results obtained with frequency bands 20 -50 and 50-80 kHz. Some descriptors, such as 8, 6, and 4, are less efficient for this band.

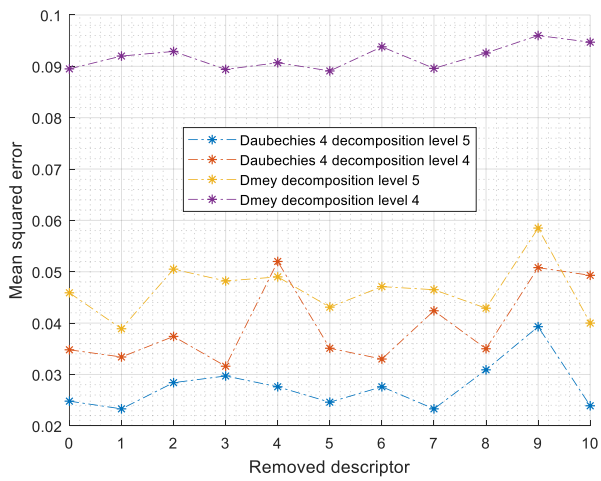

Figure.20 Descriptor selection for wavelet transforms-decomposition level 4 and 5

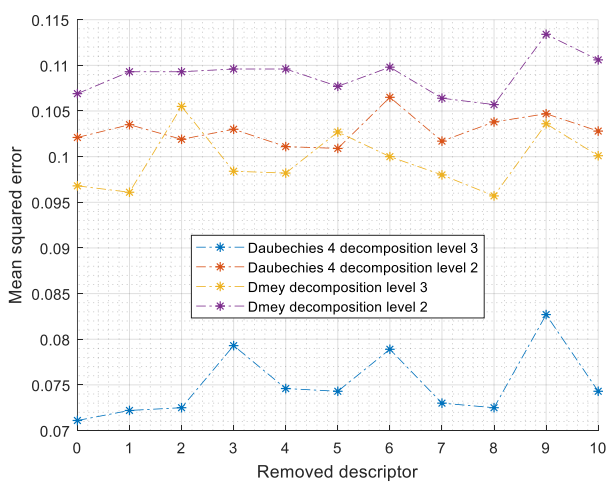

Fig.21.Descriptor selection for wavelet transforms- decomposition level 2 and 3

Same as FFT and CZT, wavelet transform shows better performance at lower frequency band (high level of decomposition).According to the results, the Daubechies 4 wavelet gives better results than the Dmey wavelet but overall performance lower than FFT and CZT. For the decomposition level 5 of Daubechies 4 and Dmey wavelets the respected descriptor 7 and 1 should be removed. For the decomposition level 2, 3 and 4 the descriptor 5, 8 are most the irrelevant. (Figure 20 and 21)

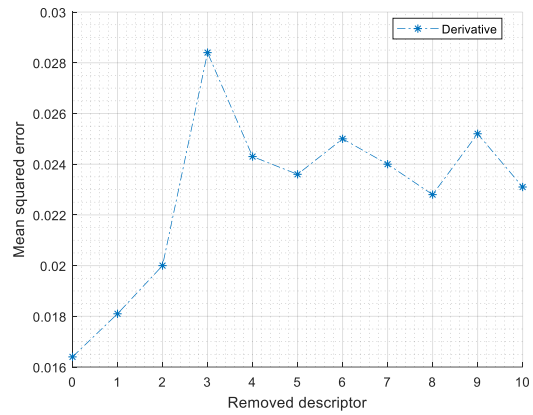

Figure.22 Descriptor selection for derivative

Figure 22 shows the results from the transform based on derivative analysis. The minimum value of mse (0.016) is obtained when all descriptors are used together. The performance of derivative transform is good and only slightly lower than that obtained with FFT at frequency band 1-10 kHz.

The small variations in error show that some arc fault features should be fine-tuned in order to achieve the best performance. As fine-tuned elements, their associated descriptors need to be carefully evaluated for the aimed specific requirements or installation with the corresponding database. For a given transform, the optimal subset of descriptor is not absolute it depends on different situations considered. Table II summarizes the results obtained for all the analyzed transforms.

TABLE II
ARC FAULTS FEATURES EXTRACTION

Transform	Optimized descriptor list	MSE	Accuracy
FFT 1-10 kHz	1,2,3,4,5,6,7,8,9,10	0.0092	99.11%
CZT 1-10 kHz	1,2,3,4,5,6,7,8,9,10	0.0234	97.6%
FFT 10-20 kHz	1,2,3,4,5,7,8,9,10	0.0171	98.13%
CZT 10-20 kHz	1,2,3,4,5,6,7,8,9,10	0.0178	98.12%
FFT 20-50 kHz	1,2,3,4,5,6,7,8,9,10	0.0274	96.87%
CZT 20-50 kHz	1,2,3,4,5,6,7,8,9,10	0.0617	92.07%
FFT 50-80 kHz	2,3,4,5,6,8,9,10	0.077	91.54%
CZT 50-80 kHz	1,2,3,4,5,7,8,9,10	0.0806	89.68%
FFT 80-100 kHz	1,2,3,4,5,6,7,8,9,10	0.0912	89.68%
CZT 80-100 kHz	1,2,3,4,5,6,7,9,10	0.0766	90.04%
FFT 100-150 kHz	1,2,3,4,5,7,8,9,10	0.0897	89.80%
CZT 100-150 kHz	1,2,3,5,6,7,8,9,10	0.0786	89.56%
DB4 LVL5	1,2,3,4,5,6,8,9,10	0.0233	97.58%
Dmey LVL5	2,3,4,5,6,7,8,9,10	0.0389	95.57%

DB4 LVL4	1,2,4,5,6,7,8,9,10	0.0316	96.34%
Dmey LVL4	1,2,3,4,6,7,8,9,10	0.0891	88.59%
DB4 LVL3	1,2,3,4,5,6,7,8,9,10	0.0711	91.59%
Dmey LVL3	1,2,3,4,5,6,7,9,10	0.0957	87.90%
DB4 LVL2	1,2,3,4,6,7,8,9,10	0.1009	87.75%
Dmey LVL2	1,2,3,4,5,6,7,9,10	0.1057	87.20%
Derivative	1,2,3,4,5,6,7,8,9,10	0.0164	98.26%

Many transforms keep all descriptors in order to achieve a superior performance. A redundant descriptor was added to demonstrate the efficiency of the selection process. Descriptor 11 is equal to (1 – descriptor 2); the value 1 was chosen because the value of descriptor 2 is normalized between [0; 1] and the result is shown in Table III. The MSE of each transform is increased when the redundant descriptor is added. Removing descriptor 2 may decrease the error in some cases. This may be explained by the fact that descriptor 11 contains information on descriptor 2. Removing any other descriptor other than descriptor 11 (for example descriptor 9) increases the error. Using redundant descriptor may affect the detection performance. This phenomenon has been studied with theoretical analysis and empirical evidence on several researches [33].Redundant descriptor does not provide addition information and it may confuse the learning algorithm. During the learning process, interesting relations between relevant descriptors can be ignored because the presence of redundant descriptor (there is a chance that the learning algorithm focus on redundant relations) therefore a lower performance can be expected.

TABLE III
DESCRIPTOR SELECTION WITH REDUNDANT DESCRIPTOR

Transform	MSE		
	All descriptor	Descriptor 2 removed	Descriptor 9 removed
FFT 1-10 kHz	0.0118	0.0117	0.0145
CZT 1-10 kHz	0.0256	0.0262	0.0347
CZT 10-20 kHz	0.0212	0.0239	0.0280
FFT 20-50 kHz	0.0282	0.0289	0.0320
CZT 20-50 kHz	0.0662	0.0636	0.0689
FFT 80-100 kHz	0.0954	0.0928	0.0955
DB4 LVL3	0.0725	0.0718	0.0827
Derivative	0.0176	0.0175	0.0232

C. Features combination

Descriptor selection shows that the most efficient AFF is discrete Fourier transform with all descriptors at frequency band 1-10 kHz (noted as AFF1). One way to achieve a better performance on the detection task consists in combining arc fault features, including AFF1. As mentioned above, the method to find complementary AFF is based on the sum of ratio distances between true and false positives, false negatives and true negative groups of AFF1 after concatenation. There are 42 FP and 101 FN samples when only AFF1 is used for the classification.

The results with one complementary AFF are shown in Table IV. The number of shared FN and FP samples between AFF1 and the other AFFs are much lower than the number of samples in AFF1 (FN and FP) groups which means that complementary AFF can provide additional information and increase detection performance.

The sums of ratio distance are listed in the fourth column; in this case AFF1 with the CZT transform 10-20 kHz has the highest sum of ratio distance. As expected, this combination has the lowest mean squared error and the highest accuracy.

Combination with the derivate method also gives an exceptional result. Overall, CAFFs always give better results than single AFFs. The best accuracy can be achieved with a single AFF of 99.11% and an MSE equal to 0.0092. With CAFF the accuracy can go higher, up to 99.81%, and lower the MSE to 0.00173.

The combination of AFF1 and derivative leads to a higher MSE but is more accurate compared to Debauchie's wavelet. This can be explained by the fact that all classifiers have been trained with MSE as objective functions and therefore accuracy is not optimized.

TABLE IV
ARC FAULTS FEATURES AND RATIO DISTANCE

Complement AFF	Shared FP With AFF1	Shared FN With AFF1	Sum of ratio distance	MSE	Accuracy
CZT 10-20 kHz	21	11	1.51	0.00173	99.81%
CZT 1-10 kHz	16	13	1.44	0.00293	99.7%
DB4 LVL5	35	13	1.43	0.00386	99.61%
Derivative	13	6	1.36	0.00396	99.7%
FFT 10-20 kHz	30	10	1.34	0.0041	99.58%
FFT 20-50 kHz	21	27	1.34	0.00452	99.56%

The results with two complementary AFFs are shown in Table V. The best performance combination is composed of CZT 10 – 20 kHz, DB4 LVL5 and FFT 1-10 kHz. As expected, this combination has the highest sum of ratio distance. The overall results are slightly better compared to the combination of two AFFs. Variation in performance gain across different combinations of AFFs can be observed. In several cases, using two complementary AFF perform worse than the best performance combination at the previous stage.

TABLE V
ARC FAULTS FEATURES AND RATIO DISTANCE

Complement AFFs	Sum of ratio distance	MSE	Accuracy
CZT 10-20 kHz & DB4 LVL5	1.52	0.00049	99.85%
CZT 10-20 kHz & CZT 1-10 kHz	1.49	0.00066	99.84%
CZT 1-10 kHz & DB4 LVL5	1.43	0.00100	99.83%
CZT 10-20 kHz & Derivative	1.41	0.00112	99.84%
DB4 LVL5 & FFT 10 - 20 kHz	1.39	0.00145	99.82%
CZT 10-20 kHz & FFT10-20 kHz	1.38	0.00151	99.81%
CZT 1-10 kHz & Derivative	1.37	0.00158	99.8%
DB4 LVL5 & Derivative	1.37	0.00160	99.8%
Derivative & FFT 10 - 20 kHz	1.37	0.00160	99.8%
CZT 10-20 kHz & FFT20-50 kHz	1.36	0.00163	99.81%
CZT 1-10 kHz & FFT 20-50 kHz	1.36	0.00185	99.73%
CZT 1-10 kHz & FFT 10-20 kHz	1.34	0.00200	99.75%
DB4 LVL5 & FFT 20 - 50 kHz	1.34	0.00230	99.71%

Derivative & FFT 20 - 50 kHz	1.33	0.00300	99.7%
FFT 10-20 kHz & FFT 20-50 kHz	1.24	0.00402	99.6%

In this experiment, the number of AFFs used for combination steps is limited to 3. Using more than three AFFs for combination can give better results in the training process but the ANN tends to lead to over-fitting (accuracy on the training set is higher than accuracy on the testing set). This phenomenon can be explained by the fact that adding more AFFs consequently increases the size of input vector and the ANN becomes more complex. More data is therefore needed to correctly train the ANN. Approximately 10 minutes training time is required for each CAFF made from 2 AFFs and 30 minutes for CAFF when composed of 3 AFFs. Eight and a half hours are required to find the optimal CAFF with the exhaustive selection method whilst the distance evaluation method requires only 42 minutes. This demonstrates the usefulness of the proposed AFF combination method.

D. Computational costs and hardware platform

All the calculations in this paper were performed with an average CPU (Core i7-2600) and the total time needed to obtain the final results from raw data is about one day. It varies, depending on the complexity and the number of transform-descriptor used. The proposed method can be run in any typical desktop PC; it can also be accelerated with the use of GPU. The final proposed detection solution can be implemented in an embedded platform for example Xilinx-Zynq7010 without real-time issues. The idea of integrate directly the time and resources constraints in the selection and combination process may be interesting for the future work.

VI. Conclusion

In this paper, a methodology for optimizing arc fault detection performance with plural arc fault features has been presented. The main originality of the proposed method is the use of supervised feature selection. The method consists in creating an arc fault feature pool and finding a combination of those features which satisfy the desired performance. The wrapper selection method was first used on every transform to find the most efficient descriptor set. This selection step which examines all possible descriptors and removes the irrelevant or redundant descriptors was necessary in order to make a reliable arc fault feature pool. Secondly, a supervised selection method based on Euclidean distance was used to find an appropriate combination in the pool of arc fault features. The combination of several arc fault features helps to reach a detection performance that cannot be achieved by using only one arc fault feature. Experimental results with basic (standard IEC 62606) and complicated situations (transient, multiple masking loads or disturbance on power network etc.) have demonstrated the efficiency of the proposed methods. Twenty-one specific transforms associated with 10 different descriptors were evaluated and in terms of accuracy, the combination of FFT, CZT and DB4 can reach 99.85%.

ACKNOWLEDGEMENT

This work was supported by Hager Group.

References

[1] P.Muller, S.Tenbohlen, R.Maier and M.Anheuser, "Characteristics of series and parallel low current arc faults in the time and frequency domain," *Proceedings of the 56th IEEE Holm Conference on Electrical Contacts*, pp. 1-7, 2010.

[2] B. Jieqiu, Z. Yi, D. Zhiqiang and Z. Hongqiang, "Arc fault identification method based on fractal theory and SVM," *2014 International Conference on Power System Technology*, pp. 1182-1187, 2014.

[3] K. Yang, R. Zhang, J. Yang, C. Liu, S. Chen and F. Zhang, "A Novel Arc Fault Detector for Early Detection of Electrical Fires," *Sensors*, 2016.

[4] R. Zhang, K. Yang, Q. Wu and J. Yang, "Research on Low-voltage Arc Fault Detection Based on BP Neural Network," *Applied Mechanics and Materials*, pp. 499-502, 2014.

[5] L. Yu-Wei, W. Chi-Jui and W. Yi-Chieh, "Detection of serial arc fault on low-voltage indoor power lines by using radial basis function neural network," *Electrical Power and Energy Systems*, pp. 149-157, 2016.

[6] S. Jovanovic, A. Chahid, J. Lezama and P. Schweitzer, "Shunt active power filter-based approach for arc fault detection," *Electric Power Systems Research*, vol. 141, pp. 11-21, 2016.

[7] C.-J. Wu, Y.-W. Liu and C.-S. Hung, "Intelligent Detection of Serial Arc Fault On Low Voltage Power Lines," *Journal of Marine Science and Technology*, vol. 25, pp. 43-53, 2017.

[8] A. T. Renjini Raveeendran, "Series Arc Fault Detection Using Discrete Wavelet Transform," *International Journal of Science and Research*, 2014.

[9] J. Lezama, P. Schweitzer, E. Tisserand, J.-B. Humbert, S. Weber and P. Joyeux, "An embedded system for AC series arc detection by inter-period correlations of current," *Electric Power Systems Research*, vol. 129, pp. 227-234, 2015.

[10] N. Hadziefendic and Z. Radakovic, "Detection of series arcing in low-voltage electrical installations," *European Transactions on Electrical Power*, pp. 423-432, 2009.

[11] C. Restrepo and J. Henson, "Systems, devices, and methods for detecting arcs". US Patent US7110864B2, 8 3 2004.

[12] G. Artale, A. Cataliotti, V. Cosentino, D. D. Cara, S. Nuccio and G. Tine, "Arc Fault Detection Method Based on CZT Low-Frequency Harmonic Current Analysis," *IEEE Transactions on Instrumentation and Measurement*, vol. 66, pp. 888-896, 2017.

[13] S. Zhang, F. Zhang, P. Liu and Z. Han, "Identification of Low Voltage AC Series Arc Faults by using Kalman Filtering Algorithm," *Elektronika ir Elektrotechnika , ISSN*, vol. 5, 2014.

[14] H. Guan, B. Wang, Z. Zhao, S. Bimenyimana and Q.

Wang, "Arc fault Current's Power Spectrum Characteristics and Diagnosis Based on Welch Algorithm," *International Journal of Engineering Science and Computing,* vol. 6, 2016.

[15] T. A.Kawady, N. I.Elkalashy, A. E.Ibrahim and A.-M. I.Taalab, "Arcing Fault Identification using combined Gabor Transform-neural network for transmission lines," *Electrical Power and Energy System,* pp. 248 - 258, 2014.

[16] Y. Wang, F. Zhang and S. Zhang, "A New Methodology for Identifying Arc Fault by Sparse Representation and Neural Network," *IEEE Transactions on Instrumentation and Measurement,* vol. 67, pp. 2526-2537, 2018.

[17] K. Yang, R. Zhang, S. Chen, F. Zhang, J. Yang and X. Zhang, "Series Arc Fault Detection Algorithm Based on Autoregressive Bispectrum Analysis," *Algorithms,* vol. 8, pp. 929-950, 2015.

[18] S. Lesecq and A. Barraud, "Arcing Fault Detection Using Wavelet Transform," *IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes,* vol. 36, pp. 345-350, 2003.

[19] C. R. Jr., "Arc fault detection using fuzzy logic". US Patent 8 054 592 B2, 8 11 2011.

[20] Y. Liu, F. Guo, Z. Ren, P. Wang, T. N. Nguyen, JiaZheng and X. Zhang, "Feature Analysis in Time-domain and Fault Diagnosis of Series Arc Fault," *IEEE Holm Conference on Electrical Contacts,* pp. 306-311, 2017.

[21] C. E.Restrepo, "Arc Fault Detection and Discrimination Methods," *Electrical Contacts - 2007 Proceedings of the 53rd IEEE Holm Conference on Electrical Contacts,* pp. 115-122, 2007.

[22] P. Qi, S. Jovanovic, J. Lezama and P. Schweitzer, "Discrete Wavelet Transform Optimal Parameters Estimation for Arc Fault Detection in Low-voltage Residential Power Networks," *Electrical Power Systems Research,* pp. 130-139, 2017.

[23] J. Tang, S. Alelyani and H. Liu, "Feature Selection for Classification: A Review," *Data Classification: Algorithms and Applications,* 2014.

[24] G. Isabelle and E. Andre, "An Introduction to Variable and Feature Selection," *Journal of Machine Learning Research,* pp. 1157-1182, 2003.

[25] J. C. Ang, A. Mirzal, H. Haron and H. N. A. Hamed, "Supervised, Unsupervised, and Semi-Supervised Feature Selection: A Review on Gene Selection," *IEEE Transactions on Computational Biology and Bioinformatics,* vol. 13, pp. 971-989, 2016.

[26] M. M. Christiansen, K. R. Duffy, F. d. P. Calmon and M. Medard, "Brute force searching, the typical set and Guesswork," *IEEE International Symposium on Information Theory,* pp. 1257-1261, 2013.

[27] P. Somol, P. Pudi and J. Kittler, "Fast Branch & Bound Algorithms for Optimal Feature Selection," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 26, pp. 900-912, 2004.

[28] A. A. Naeini, M. Babadi, S. M. J. Mirzadeh and S. Amini, "Particle Swarm Optimization for Object-Based Feature Selection of VHSR Satellite Images," *IEEE Geoscience and Remote Sensing Letters,* vol. 15, pp. 379 - 383, 2018.

[29] X.-Y. Liu, Y. Liang, S. Wang, Z.-Y. Yang and H.-S. Ye, "A Hybrid Genetic Algorithm with Wrapper-Embedded Approaches for Feature Selection," *IEEE Access,* vol. 6, pp. 22863-22874, 2018.

[30] R. Webster, "Useful AI tools-a review of heuristic search methods," *IEEE Potentials* , pp. 51-54, 1991.

[31] L. J.-H. and O. S.-Y., "Feature selection based on geometric distance for high-dimensional," *Electronics Letters,* vol. 52, pp. 473-475, 2016.

[32] J. Lezama, P. Schweitzer, S. Weber, E. Tisserand and P. Joyeux, "Arc Fault Detection Based on Temporal Analysis," *IEEE 60th Holm Conference on Electrical Contacts,* pp. 1-5, 2014.

[33] L. Yu and H. Liu, "Efficient Feature Selection via Analysis of Relevance and Redundancy," *Journal of Machine Learning Research,* pp. 1205-1224, 2004.