Intelligent Low Voltage Series Arc Detection System

by

Ayman M. Othman Mohamed

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2015

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Protection from electric power hazardous has been used ever since applications of electricity were discovered. Hazards in the electric system can be in the form of over voltage or over current leading to catastrophic system and equipment failure, resulting in physical damage or even to human fatalities. Electrical protection is achieved by installing protection devices throughout distribution system to isolate faulty branches and mitigate fault development.

Fire is a principal cause of buildings damages and related personal injuries. A major contributor to buildings' fire originates from electrical arc faults caused by electric distribution equipment and appliances failures. To remedy this problem, regulatory bodies required electric arc faults protection. Over the years this requirement was enforced by different electric codes and expanded to cover most of residential building areas and all living spaces.

Arc fault circuit interrupters (AFCI) are devised to complement existing protection methodologies and devices, focusing on electric arc detection and preventions of subsequent risks, mainly fire ignition. Circuit interruption occurs whenever characteristics of arc failure is detected, either from current, voltage or electromagnetic radiation. Detecting the arc faults, and hence increasing the reliability of interruption, is a challenge, given that some household appliances produce arc-like behaviors in normal operating conditions, like electronic light dimmers and solid state controlled variable speed drives.

This research focuses on developing an intelligent low voltage series arcing detection scheme based on pattern recognition, with immunity to false tripping. This point is the main drawback of most published work and issued patents on arc detection to date, mainly due to the difficulty of modelling such a transient behavior, especially on low

current arc cases. Real data is generated in lab simulating series arc conditions at different combinations of linear and non-linear loads. Appliances current are recorded as well. Two disjoint datasets are used for training and testing of the proposed system with no components shared between the two datasets to verify classifier generality. The proposed pattern recognition method proved to be highly immune to false tripping in line with benchmark regulatory standard, and can be adapted to similar hard to model non-stationary problems.

Acknowledgements

All thanks and praise to Allah almighty for gifting me with the strength and abilities to succeed in completing this research.

I would like to express my gratitude to the thesis advisor, Dr. Ramadan El-Shatshat, for giving me the opportunity to work with him, his comments and advises were of great value during the graduate program.

I would like also to thank research group members and graduate students community for valuable help and friendship. Many thanks goes to the family of electrical and computer engineering department, lab technicians and staff for their assistance and support with experimental part of this research.

I am indebted to my parents, wife and family for their endless support, patience and prayers.

Dedication

To who inspired me the most …

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AC	Alternating Current
ADC	Analog to Digital Converter
AFCI	Arc Fault Circuit Interrupter
CFL	Compact Fluorescent Lamp
CSA	Canadian Standards Association
CWT	Continuous Wavelet Transform
DC	Direct Current
DSP	Digital Signal Processing
DWT	Discreet Wavelet Transform
DWPT	Discrete Wavelet Packet Transform
Ecul-m	Minimum of Squared Euclidean Distances
Ecul-s	Sum of Squared Euclidean Distances
FT	Fourier Transform
FFT	Fast Fourier Transform
FPGA	Field Programmable Gate Array
IIR	Infinite Impulse Response
KKT	Karush-Kuhn-Tucker Condition
LED	Light Emitting Diode
LDA	Linear Discriminant Analysis
MDS	Multidimensional Scaling
MRA	Multiresolution Analysis
Maha-m	Minimum of Estimated Mahalanobis Distances
Maha-s	Sum of Estimated Mahalanobis Distances
ML	Machine Learning
NEC	National Electric Code
PR	Pattern Recognition
PCA	Principal Component Analysis
SBS	Sequential Backward Selection
SBFS	Sequential Backward Floating Selection
SFS	Sequential Forward Selection
SFFS	Sequential Forward Floating Selection
STFT	Short Time Fourier Transform
SVM	Support Vector Machines
WT	Wavelet Transform
WPT	Wavelet Packet Transform

Chapter 1

Introduction

1.1 Introduction

Today, more than ever, we rely on electric power for every aspect in our daily living, from commercial and industrial to households' applications, and its use in transportation is on the rise with increased penetration of electric vehicles.

Electric power is delivered to users at different voltage levels to best suite consumer applications. This power delivery is through insulated conductor wiring with relevant voltage and current rating specifications. Electric power comes with inherent risks associated with delivery and usage, including shocks, and fires started by faulty wiring and appliances, and so various electric standards, best practices, and protective devices have been developed to reduce these risks.

An electric current poses a real hazard when it exceeds predetermined ratings or flows through an unintended path, such as in short circuits and current leakage from devices and cables, and may lead to equipment and property damage. Thus, a range of devices have been introduced over the years, from thermal, differential circuit protection, to ground and arc fault circuit interrupters.

1.2 Motivation

Fire hazard can produce substantial property and financial losses, not to mention loss of lives. There is a diverse range of fire ignition sources, from misuse of flammable materials and gases, to exposure to hot surfaces and electrical sources. Electrical appliances and electrical distribution systems contributed to 10% of fires in Canada during 2007, resulting in 11% of total fire related death cases and 7% of registered fire related injuries [1]. It is estimated that around 47,700 fires were reported in the US during 2011 involving building's electrical distribution systems and devices, leading to 418 civilian deaths and around $1.4 billion in property damage [2].

The US national electric code (NEC) revision introduced in 2002, required the protection of specific living space in new buildings against arc faults, and the stricter requirements of NEC 2011 mandated protection of almost all areas in new as well as existing buildings [3]. The different AFCI products in the market are being limited to high current arc detection, shadowed with unintended tripping for low current arc faults, and requiring operational reliability enhancements to reduce unintended interruption.

1.3 Scope of Work

This research aims to develop and validate a pattern recognition system for low voltage series arc fault detection, with special attention to false tripping immunity. The designed system should be capable to discriminate between real arc and arc-like behaviors produced from different electric appliances to eliminate spurious tripping, while possessing a non-intrusive nature. Practical deployment for online circuit protection is emphasized throughout the research. Canadian standards association (CSA) standards are used as the benchmark for data generation and requirements validation [4].

1.4 Thesis Outline

This thesis consists of six chapters organized as follows, Chapter 2 is a literature review on arc inception process, arc characteristics and current detection techniques, along with a detailed review of arc protection issued patents, with limitations compared to current electrical protection standards. Chapter 3 provides a brief background of signal processing techniques for stationary and non-stationary signal analysis. These techniques constitute the corner stone of this work, by capturing arc discriminant signals and forming the link between the machine and physical system layer. Chapter 4 presents an overview of pattern recognition systems and different types of machine learning typologies and dimensionality reduction. Particular attention is given to support vector machines, as it is used for classification purposes in the proposed system. In Chapter 5, the proposed system is introduced from system level down to component level; different stages and components are detailed, including current sensing, feature extraction and selection, normalization and classifier training, a benchmark arc generator and dataset generation. Concluded with system performance measures evaluation and analysis. Chapter 6 presents conclusions and findings of the work, and list possible further improvements considerations.

Chapter 2

Literature Review

2.1 Introduction

This chapter provides background and reviews the literature on the low voltage electrical arc problem. The discussion includes how an arc is created, what causes it, different types of arcs, factors governing arc initiation and extinguishing, specific characteristics, detection techniques, as well as a review of some patents issued on this problem.

2.2 Electrical Arc

Electrical arcing can be defined as the presence of an unintentional conductive path parallel to an insulation medium between two points of different electric potential, accompanied with luminous discharge, electromagnetic radiation and acoustic waves [4]. This definition can be expanded to include parallel and series arcs. The danger of an electric arc is in its capability to ignite fires, as arc-generated temperatures are very high; high enough to melt conductor metals and ignite insulation and nearby combustible material [5].

2.2.1 How Arcs are Created

An electric arc is a form of electric discharge, resulting from the creation of conductive path between two points of different electrical potential, and serving to equate existing potential difference by moving electrons from a low to a high potential side. Arcs will be created whenever the potential difference is greater than the insulation material breakdown voltage, depending on the insulation medium's specific properties [6].

The main causes of electric arcs are cable insulation failure and loose or broken wire connections [5, 7]. Such failures can be due to material degradation or mechanical damage. Insulation material can deteriorate by aging, high ambient temperature and humidity, or use out of its design limits, like over voltage or exposure to high heat as in the case of conductor over load. Mechanical damage can result from exercising stress or strain force on the conductor, with effects on the conductor or insolation medium characteristics. Loose wiring may result from exposure to vibration, corrosion or incorrectly fastened screws, while broken conductors can be the result of repeated misuse by forcefully pulling cable cords, resulting in electric arcs by the intermittent make and break of connection cycles.

Arc generation and characteristics are controlled by different factors, ranging from electrode materials, physical dimensions, voltage difference between arc terminals, gap separating contacts, conductive path electrical properties, load current and voltage, local cooling and generated electric field in between [8, 9]. Whenever these requirement are present and favorable, arc inception will result in arc initiation; otherwise, arc fault would be suppressed and extinguished.

2.2.2 Types of Arc

Arc faults can occur between any combinations of wires within an electric circuit, in parallel or in series with the connected load.

(a) (b) (c)

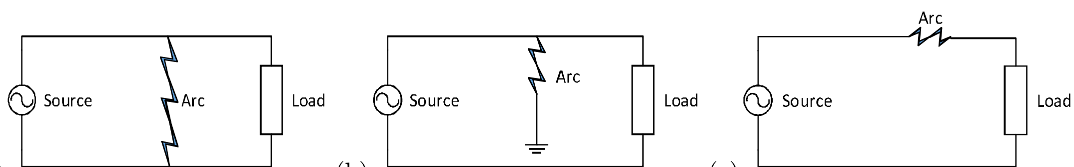


Figure 1. Arc fault types, (a) Line to neutral, (b) line to ground and (c) series arc faults

Parallel arcing is the presence of an arc between two parallel conductors in the circuit, such as in line-to-neutral or line-to-ground shown in Figure 1.a and 1.b. A parallel arc can result from external mechanical damage in insulation materials, causing a current path from one conductor to the other, or from thermal stress caused by conductor over loading or external heat exposure resulting in degraded insulators leading to insulator materials hardening and developing cracks with subsequent arcing [10, 11].

In contrast, series arcing is an unintended arc present in a series with a current-limiting load, as in Figure 1.c. It is caused by improperly fastened wiring or a break within conductor strands with repetitive connections making and breaking under loading condition. The result is generation of copper oxide film with higher resistance and contact point heating to high temperatures, resulting in glowing contact, with the possibility of igniting wire insulation and surrounding materials. Insulation materials' decomposition by arc-produced heat produces flammable gases and pressure waves, which can be ignited by arc currents, which have more than enough energy to ignite the generated gases [10, 11].

A parallel arc's current is higher than the rated current and it is limited by the source supply current and all connected components in the current path, usually to around 75 amperes [8]. A series arc's current on the other hand is within normal current range, with a less than rated *rms* value [12, 13]. Figure 2 shows the difference between normal and series arc currents for a purely resistive load.

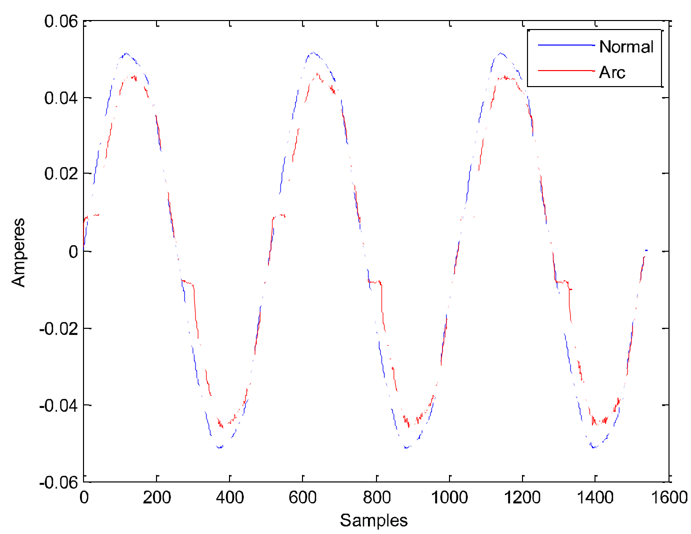


Figure 2. Normal and series arc current of purely resistive load

2.2.3 Arc Fault Characteristics

Electric arc is characterized by luminous discharge, high signal distortion of current and voltage waveforms, negative V-I relationship [14], and an electromagnetic radiations [15, 8, 14]. Also accompanied with generation of radio frequency signal noise, in the megahertz to gigahertz band, which is governed by arc specific variables; like resistance, inductance, capacitance and arc current [15, 8]. The generated noise is present during arc current conduction and suppressed during zero arc current.

2.2.4 Arc Fault Detection Techniques

Arc current waveform has high randomness and unpredictable behavior, this nature is different from one load to another, making it difficult to reliably discriminate arc from non-arc conditions. Distinct characteristics in frequency domain, time domain and frequency-time domains were investigated in the literature [16, 11, 8].

Time domain analysis relies on the self-extinguishing characteristics and broadband radio frequency noise of arc current during zero crossing, half cycle asymmetry and high di/dt in the first half of the half-cycle. Fault current exceeding 75 ampere is considered a distinct feature of parallel arc fault [8]. The induced current gap is governed by the contact separating distance and the ratio between arc resistance and inductive component of the load, bigger current gap is associated with parallel arc. Arc current high rise rate after zero gap is another aspect; which is effected by the current gap duration and arc type, being slower for parallel arc compared to series arc [16, 11].

Frequency domain analysis uses fast Fourier transform (FFT), revealed high third harmonic component which can be effected by loading conditions, elevated spectral energy in the frequency band of 100 Hz to 1.5 KHz, with high frequency component in the range of $2 - 5$ KHz, resulting from the sharp current rise after zero crossing with distinct presence in the case of parallel arcing [11]. Another method proposed in [17] utilizes high pass filter with low cut-off frequency for eliminating power frequency component from voltage signal.

Time-frequency domain methods based on Wavelet transform detection technique is presented in [18, 19], using wavelet coefficients mean-difference algorithm on detailed signal coefficient extracted through dyadic multiresolution wavelet decomposition, and high energy band detection at reference frequencies respectively.

Other methods proposed in [20, 15] were based on acoustic, infrared and radio frequencies generated by arc fault, and electromagnetic radiation in KHz-to-MHz range respectively.

2.3 Masking Loads

Arc fault detectors need to reliably distinguish real arcing conditions from normal arcing generated in some household appliances. Typical loads generating arc-like behavior are electronic light dimmers, air compressors, vacuum cleaners, power tools and power line communication modules. Light dimmers has periods of zero current, air compressors and vacuum cleaners generate high noise and high inrush current during start-ups, while power tools produce current intermittency as a function of rotational speed, power line communication generates noise. All previously mentioned appliances share some characteristics that could be viewed as indication of an arc failure.

2.4 Previous Work and Limitations

The arc detection problem has received a great deal of attention from researchers interested in electrical safety, resulting in different patents and commercially available products. NEC 2002 requires the protection by AFCI for all single phase branch circuits with 125 V supply voltage and current rating from 15 to 20 A [21]. With more attention is given to the problem after the introduction of NEC 2011, which extends the required protection to all distribution circuits in new and existing buildings [3].

U.S. Pat. No. 4,658,322 issued to Rivera, describes a method of arc detection for electrical equipment with vented enclosures. It was devised for use with single equipment in a marine environment with immunity to electromagnetic interference [22]. This method is based on thermodynamic principles through internal and external differential pressure measurement, internal temperature, and a photic sensor to detect the occurrence of arc flash.

U.S. Pat. No. 5,185,684 issued to Beihoff et al, based on the detection of established distinct electromagnetic fields around current carrying conductors during arc conditions; through monitoring specific frequencies using a combination of passband filters and amplifiers to detect arc presence and generating a trip signal. This method uses a special transducer for measuring electric and magnetic fields separately [23].

U.S. Pat. No. 5,223,795 and 5,432,455 issued to Blades; discloses a method based on monitoring radiated energy, as well as high frequency noise in power line voltage and current in 10 KHz – 1 GHz range. This method employs digital signal processing techniques for detecting noise gap present every half-cycle providing immunity to high frequency noise sources [24, 25]. Being not capable of detecting series arc with inductive load or contact arcing is considered the main drawback for this method.

U.S. Pat. No. 4,376,243 and 5,280,404 issued to Ragsdale, designed for DC powered electric rod furnace arc protection. This is achieved by amplifying high frequency signal and counting the number of arc pulses within a predefined sampling window. This method provides good noise immunity and based on analog circuitry and using cheap microprocessor are considered good advantages [26, 27].

U.S. Pat. No. 5,805,398 issued to Rae et al, operates on the principle of time decay accumulation of clipped pulses present during arc faults [28], aims to overcome false tripping due to tungsten bulb burnout and inrush current of cold tungsten filament energization by a solid state dimmer.

U.S. Pat. No. 5,818,237 issued to Zuercher et al, presents a method immune to solid state dimmers and inrush currents. Monitoring signal step increases with two signal envelops; by the addition of detected pulses with attenuation over time, a second envelop uses a lower time constant then the first envelop time constant.

Arc signal is considered present when a random step increase from half-cycle to the other is detected [29].

U.S. Pat. No. 6,128,169 issued to Neiger et al, utilizes high frequency signal content and AC signal energy, with high peaks in either component is characterized for arc fault. Measurement of the line impedance is made using a dedicated circuit and a permanently connected testing load. Arc condition is detected whenever the line impedance increases leading to arc fault [30].

U.S. Pat. No. 8,089,737 B2 issued to Parker et al, presents a method based on sensing and compressing broadband noise present in power circuits and detecting noise signal minimum values [31]. Measuring the noise minimum and maximum values from different half-cycles is achieved through measuring the range between maximum and minimum values of the sensed noise signal; then by counting the number of samples exceeding predetermined first level to determine a condition characterized by:

1. Range of higher than a predefined level.
2. Minimum value occurring at either beginning or end of the half-cycles.
3. Counter is increased if the number of samples higher than first level is greater than the predetermined third level, and decreased otherwise.

Fault condition is detected whenever the counter reaches a set value. Issued patents are capable of identifying parallel arc faults with good accuracy, mainly due to the clear and easily recognizable associated characteristic as described earlier, although they suffer from the inconsistency and low reliability for series arc detection, especially when compared with masking loads.

2.5 Summary and Discussion

This chapter presented the arcing phenomenon, causes behind it, potential consequences, and different types of arcing in low voltage electric circuits. Distinct arc fault features in time, frequency and time-frequency domains investigated in literatures are reviewed, plus existing detection techniques and their main principles, as well as patents issued for arc detection.

Although an arc current possesses distinct features in time, frequency and time-frequency domains, it is not feasible to rely on one feature alone to reliably detect the presence of transient arcing conditions, especially when considering masking loads like lamp dimmers and vacuum cleaners etc. Previous work has focused on the arc detection problem from different perspectives, resulting in a combination of successes and cases of low detection reliability, mainly due to the techniques used and the extreme random behavior of the arcing phenomenon. All patents on arc detection issued to date have been based on a combination of individual aspects, making their methods prone to false tripping rate when applied to all masking loads existent in ordinary households. No method has until now considered analyzing the complete signal waveform and utilizing machine learning principles and techniques to detect arc conditions in low voltage.

Chapter 3

Wavelet Transform

3.1 Introduction

Wavelet transform (WT) has been employed in different fields, ranging from mathematics and engineering to physics. The wavelet theory was formalized and found applications in seismic signal analysis, digital signal processing and computer vision [32]. WT is regarded as an extension to short time Fourier transform (STFT) for non-stationary signals; unlike fixed window used in STFT, WT employs variable window allowing longer window for lower frequencies and shorter window at higher frequencies resulting in better time-frequency resolution. This chapter will present brief background for short time Fourier transform, wavelet transform in continuous and discreet forms, wavelet multiresolution analysis, and wavelet packet transform (WPT).

3.2 Short Time Fourier Transform

Fourier transform (FT) is a well-established spectral analysis technique, where it decomposes original signal to its frequency spectrum contents. By definition, FT is limited to stationary signals only due to the used infinite basis. FT for signal $x(t)$ is given by equation 1.

$$X(f) = \int_{-\infty}^{\infty} x(t) e^{-j2\pi ft} dt \tag{1}$$

Short-Time Fourier Transform was introduced to estimate timely sinusoidal signal components in frequency and phase [32]. STFT for signal $f(t)$ and moving window $g^*(t-\tau)$ is given by equation 2.

$$STFT(\tau, f) = \int_{-\infty}^{\infty} f(t) g^*(t-\tau) e^{-2j\pi ft} dt \tag{2}$$

The fixed window size is a decisive factor for transient signal analysis of the same resolution in time and frequency [33], as smaller window size than transient time would result in loosing vital frequency information, in contrary, longer window will result in better frequency and poor time resolution [34].

3.3 Wavelet Transform

WT provides a compromise over STFT by having a variable time and frequency resolutions, which is achieved using a predefined mother wavelet function instead of Gaussian window as in STFT [32].

WT decomposes the signal by using high-pass and low-pass orthogonal filters, G and H respectively, producing detailed and approximate signals followed by down sampling operation.

Continuous wavelet transform (CWT) is given by equation 3, where $\psi(t)$ is the basis function, a and b are scaling and translation parameters respectively [32].

$$W_{\psi}f(a,b)=\frac{1}{\sqrt{a}}\int_{-\infty}^{\infty}f(t)\overline{\psi(\frac{t-b}{a})}dt \tag{3}$$

Discreet wavelet transform (DWT) can be obtained by discretizing scaling and translation parameters to the form $a=a_0^m, b=nb_0a_0^m, m,n\in Z$ [32], as in equation 4.

$$W(m,n)=\frac{1}{\sqrt{a_0{}^m}}\int_{-\infty}^{\infty}f(t)\psi(a_0{}^{-m}t-nb_0)dt \tag{4}$$

Figure 3 shows wavelet decomposition tree, where *f(n)* is the original discreet signal, a_1 and d_1 are approximate and detailed signal coefficients respectively.

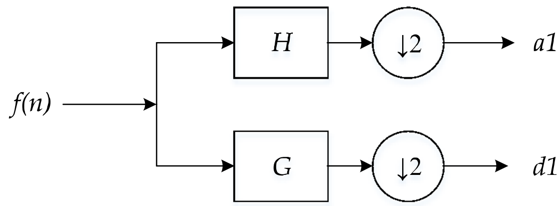


Figure 3. Wavelet transform decomposition

3.4 Multiresolution Analysis

Multiresolution analysis (MRA) is the successive decomposition of signal using specific filter banks, producing a multiresolution decomposition with coarser resolution at higher decomposition level. Pyramidal decomposition is achieved using wavelet filter banks that satisfies containment, decrease, increase, dilation and generator conditions [35].

In wavelet multiresolution framework, approximate signal is further decomposed depending on the decomposition levels required with MRA.

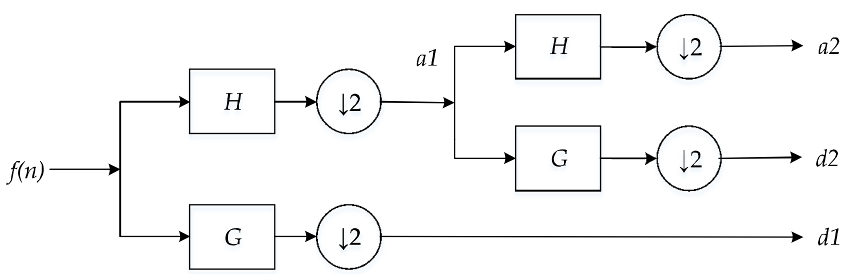


Figure 4. Two level multiresolution wavelet decomposition tree

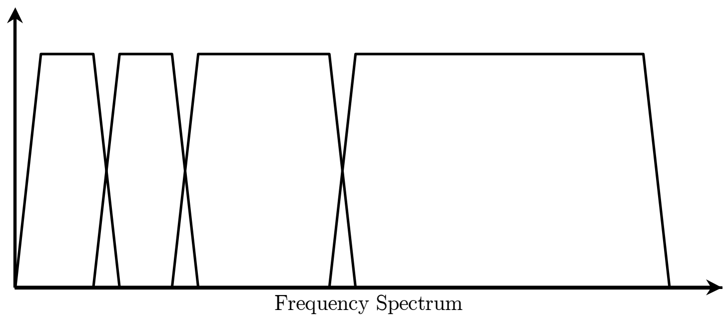


Figure 5. Multiresolution wavelet decomposition frequency spectrum

Figure 4 depicts a two levels multiresolution wavelet decomposition tree, where *f(n)* is the original discreet signal, a_1 and d_1 are approximate and detailed signal coefficients at level 1 respectively, a_2 and d_2 are resulting coefficients at decomposition level 2. Figure 5 shows frequency distribution of DWT decomposed signal at three levels.

3.5 Wavelet Packet Transform

Discrete wavelet packet transform (DWPT) is similar to multiresolution DWT, with the difference of detailed signal coefficients being expanded with every decomposition level using the same wavelet filters, as with multiresolution analysis wavelet transform.

Figure 6 depicts a pyramidal decomposition scheme for signal *f(n)* with resulting approximate and detailed components, while Figure 7 shows the frequency spectrum of decomposed signal components at three levels.

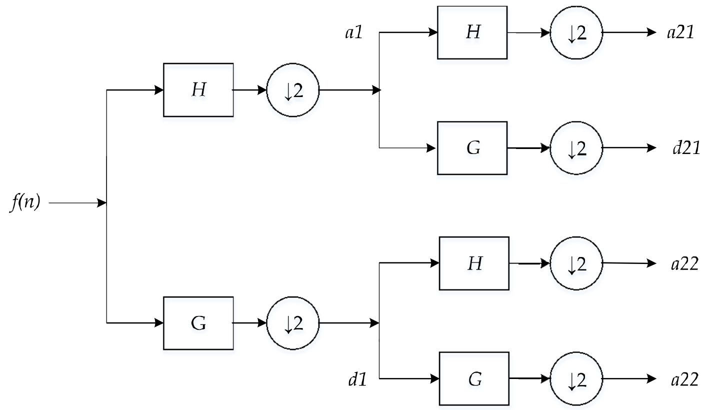


Figure 6. Two level multiresolution wavelet packet decomposition tree

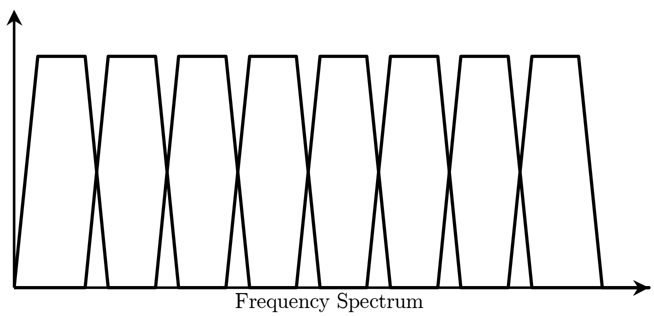


Figure 7. Multiresolution wavelet packet decomposition frequency spectrum

Chapter 4

Pattern Recognition Systems

4.1 Introduction

This chapter introduces the basic concepts of pattern recognition and machine learning principles; presenting benefits, drawbacks and limitations, with focus on supervised learning and support vector machines. Background of the different algorithms for data dimensionality reduction and features selection is reviewed, as well as pattern recognition classification performance measures.

4.2 Pattern recognition

Pattern recognition (PR) is one of the artificial intelligence areas that has been gaining momentum over the past couple of decades. PR applications are found in many different sectors ranging from medical, financial, security and engineering to name a few. PR aims to identify existing patterns in nature and data. The two main tasks of PR are classification and clustering, where the former targets data point assignments between two or more groups, while the latter groups all data points of similar characteristics in one group.

Machine learning will prove useful in any of the following cases [36]:

1. Difficulty of task definition unless by the use of examples; where input and output examples are presented to the machine in a training format in order to model the underlying relationship between input and output data.
2. The possibility of design environment changes during deployment of the intelligent system, which requires system adaptation and no redesign activity, as well as the lack of knowledge of some working environment definition during design phase. Machine learning approach can keep track of the dynamic environment and continually update application related knowledge.
3. The discovery of new and interesting relationships in large data sets.

4.2.1 Types of Machine Learning

Machine learning is typically categorized into supervised learning and unsupervised learning; supervised learning intern is subdivided into classification and regression, where unsupervised learning is also known as clustering.

Classification is the process of classifying data points into different classes or categories based on common features, this process is known as supervised learning. Classification relies on the fact that training data being used is labelled, where label is the information in which class each data point belongs to, these labels are used to train the classifier or classification model by example to distinguish between the existing different classes based on distinct class features. Classes can take a categorical or nominal values [37]; categorical labels as in the case of binary or multi-class problems, while nominal value labels are for regression and forecasting problems. The main goal is estimating the underlying model in the training set between input instances and class labels, and then predicting class labels for novel input instances which has not been seen by the learning algorithm before. In other words, the inferred training model should possess enough generalization to be applied on novel data points, and avoid closely overfitting the training data. Many different classification algorithms exist with varying performance and computational complexities [38], like Bayesian classification, artificial neural networks and support vector machines etc. Some of the current applications of classification are electrocardiogram analysis (ECG), email spam filtering and game playing are. Regression is similar to classification except it uses a real value labels, where the resulting model would be used to forecast the real time series and predict future instances [37].

Clustering on the other hand, tackles unlabeled data to identify what interesting patterns do exist. Data clustering could be viewed as grouping data points with similar features and attributes together, where the number of groups is not defined beforehand. The selection of best clustering algorithm heavily depends on the nature and size of the target data.

4.3 Support Vector Machines

Support vector machines (SVM) is one of the most influential algorithms in the area of machine learning, it is used for classification and regression purposes for linear and non-linear data. SVMs' unique capabilities relies on reaching the global optimum and achieving a good generalization from training data.

SVM works by projecting the training data instance features into a high dimensional features space using a non-linear mapping, then solving for maximum separating marginal hyperplane in the new space [39]. This approach faces two difficulties:

1. How the algorithm will generalize from the training data, especially when most of the separating hyperplanes will not provide good generalization.
2. How to handle and reduce the computational cost of the produced higher dimensional space.

The optimal separating hyperplane is characterized by having the maximum margins between different class vectors. This plane will take into account a subset of the training data called support vectors, where the lower the number of support vectors the higher the generalization ability will be [40].

Considering linearly separable training dataset (X_i, y_i), where training instance X_i is associated with class label y_i, $y_i \in \{+1,-1\}$ and $i = [1,2,...,n]$, where n is the number of training set instances. Separating hyperplanes is given by equation 5.

$$W.x_i + b = 0 \qquad (5)$$

Where W is a weight vector, $W = \{w_1, w_2, ..., w_n\}$, and b is scaler vector. Figure 8 shows SVM optimal plane for two dimensional feature space.

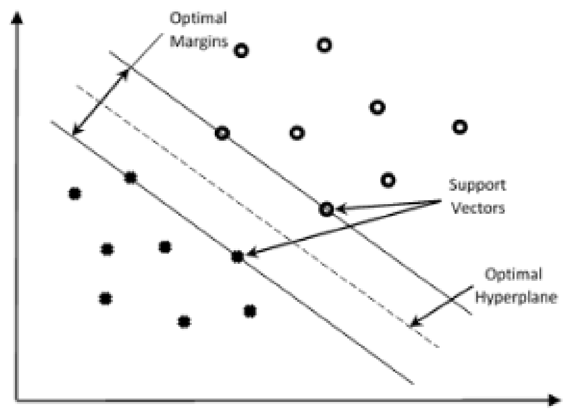


Figure 8. Two dimensional separation problem showing optimum hyperplane, support vectors and optimal margins

Training dataset set is considered linearly separable when there is a vector W and a set of constants b such that

$$W.x_i + b \geq 1, \qquad \text{if } y_i = 1 \qquad (6)$$

$$W.x_i + b \leq -1, \qquad \text{if } y_i = -1 \qquad (7)$$

By combining the above two equations

$$y_i(W.x_i + b) \geq 1 \qquad i = 1, ..., n \qquad (8)$$

The resulting optimal hyperplane equation is given by

$$w_0.x + b_0 = 0 \quad (9)$$

Yielding optimum separation and maximum margins, with maximum separation margin of $\frac{2}{\|W_0\|}$.

Optimal hyperplane is the unique plane minimizing the dot product $W.W$, under constrain (8), which is in the form of quadratic programming problem; such problem solution involves problem formulation using Lagrangian multiplier and solving using Karush-Kuhn-Tucker (KKT) conditions [40]. Obtained support vectors are the bases for classifying data points and directly controls produced classifier complexity [39].

While in the case of linearly inseparable data, the goal will be modified to achieve optimal separation with the error, extension to linear SVM can be achieved using other non-linear mappings and solving in the new higher dimensional space.

Reduction of associated computational cost with dot product calculation in the new feature space is attained by applying specific kernel functions to the input data, which is found to be equivalent to computing dot product in the feature space. This procedure requires the addition of Lagrange multiplier upper bound to the optimization formulation which is best obtained experimentally [39].

SVM differentiate between classes by maximizing the separation distance between classes; being capable of reaching the global optimum regardless of the training set. Other SVM attractive features is the insensitivity to overtraining, where overtraining sensitivity can pose a serious concern during training phase, as which data samples should be used for classifier training, providing good generalization performance and having limited number of parameters to be tuned. Although training computational complexity is high, due to the quadratic optimization problem, and the experimental nature of penalty term setting are the main drawbacks of this technique [41].

4.4 Dimensionality Reduction and Features Selection

Classifier performance is tied to the training sample size, the number of features available and classification algorithm. The curse of dimensionality is where the association between sample space and feature space is of high proportions, with exponential relationship as in the case for naïve table-lookup [42]. Peaking phenomenon is where the probability of sample misclassification does not increase with increased number of features, while the addition of new features will practically reduce the classifier performance in case of low training sample to feature size [41].

Dimensionality reduction of high dimensional features set will benefit pattern recognition algorithms by simplified classifier model, as well as yielding higher accuracy, faster computational time and reduced computational cost. This dimensional reduction should be tackled with care as it might remove some of the discriminant features resulting in classification errors. Dimensionality reduction can be achieved by transforming the existing feature set to a lower dimensional space using linear or non-linear transformations, linear transforms as principal component analysis (PCA) or linear discriminant analysis (LDA), while non-linear transforms as kernel PCA or multidimensional scaling (MDS). The generated features from the new projection may not contain any physical meaning compared with the original features set [41].

Feature selection is the selection of a subset of m features from feature set d that results in the lowest classification error. The obvious method is to test all possible combination in the feature set via exhaustive search, while this approach guarantees optimality on the expense of excessive computational burden, exhaustive search becomes infeasible as d becomes large. The only selection algorithm guaranteeing optimality except exhaustive search is based on branch and bound algorithm [43], other search strategies exists in literature ranging from individual feature ranking, sequential forward selection (SFS), sequential backward selection (SBS), plus l take away r and sequential forward and backward floating selection (SFFS/SBFS). The best group of features may not contain the best individual features and thus optimality is not guaranteed. Sequential search methods suffer from subset nesting, where SFFS and SFBS has the advantage of back tracking over SBS and SFS, which is considered a generalization of plus l take away r providing a close to optimal solution with an affordable computational complexity [41].

4.5 Summary

This chapter has reviewed the basic theory of pattern recognition, listing conditions where pattern recognition techniques is better suited, main types of machine learning and expected benefits, an overview of supervised learning, theory of support vector machines, dimensionality reduction and feature selection.

Machine learning techniques application to problems of high difficulty, or practically impossible to solve using conventional methods, resulting in good performance which can be attributed to machine learning unique capabilities of solving complex formulated problems. SVM classifier unique capabilities places it as one of the top choices for classification problems, this is due to the mix of its favorable attributes and acceptable drawbacks; since training phase is made offline and can use as much data as possible resulting in a better generalization.

Computational complexity reduction can be achieved by reducing the number of features and eliminating any inconclusive features, translating into a reduced feature space and reduced overall complexity and increased efficiency.

Chapter 5

Proposed Solution

5.1 Introduction

This chapter presents the proposed solution for the problem of series arc fault detection in low voltage systems. Firstly, the problem is broadly defined, followed by governing regulatory standards. The proposed solution is then introduced, and its different stages of operation; with elaboration of each process stage. Finally, the complete system's practical evaluation and performance measures are described.

5.2 Problem definition

Arc faults may exist in low voltage distribution wires with the presence of specific preconditions, and may lead fire ignition. The danger of LV arcs is in the difficulty associated with detecting the presence of arc conditions, and reliably discriminating them under different masking loads. Operating on a system frequency of 60 Hz poses a great challenge for computational complexity, especially with the added requirements for AFCI to ”trip whenever 8 arcing half cycles occur within 0.5 s interval” [4]. This requirement dictates that the detection algorithm monitor the current signal, or voltage signal, one half-cycle at a time and detect arc conditions whenever they occurs.

5.3 Proposed Technique

The proposed solution is a pattern recognition system based on wavelet packet transform and classification algorithm to classify arc from non-arc conditions. The PR approach is chosen due to the great difficulty in modeling transient nature of arc fault, and the superior performance of supervised PR methods.

Throughout the design of the system, the emphasis is on keeping computational cost to the minimum, and making the algorithm practically attractive and more realistically implementable for real time monitoring. Using non-intrusive detection technique favors current signal over voltage signal as the basis of arc detection, as it can be measured without altering circuit wires' connection and routing.

5.3.1 Over all system

The Proposed system is based on pattern recognition principles for arc condition detection, and examines current cycle waveform through specific arc features and classification technique. It consists of three main stages, current sensing stage, feature generation and extraction stage, and arc detection stage. The current sensing stage is composed of current transducer and analog-to-digital signal converter; the feature extraction stage uses WPT with special filter bank for signal decomposition and feature extraction from selected WPT node coefficients; and the arc detection stage consists of a PR classifier and instances counter that initiates a trip signal whenever a preset value is reached. Figure 9 shows the overall system components, and Figure 10 shows process flow chart for arc detection.

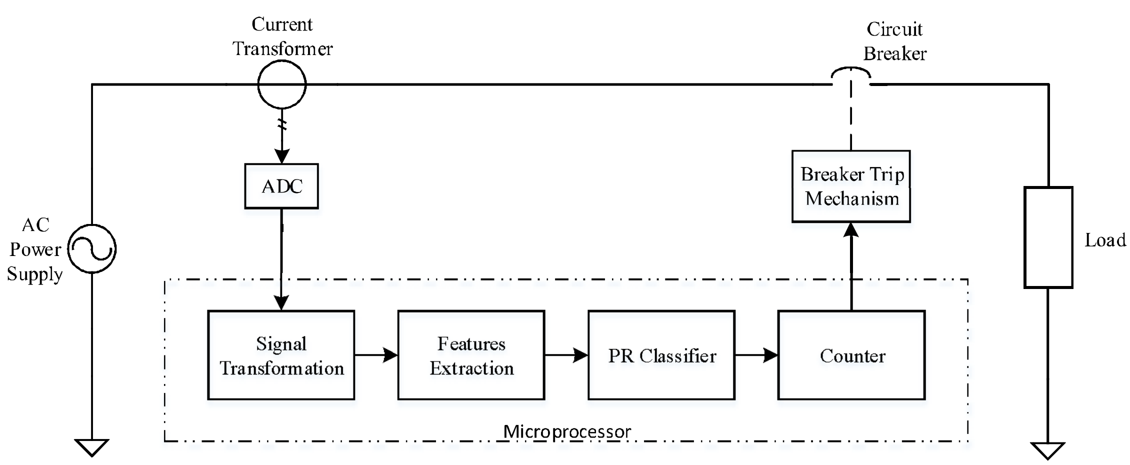


Figure 9. Overall arc fault interruption system

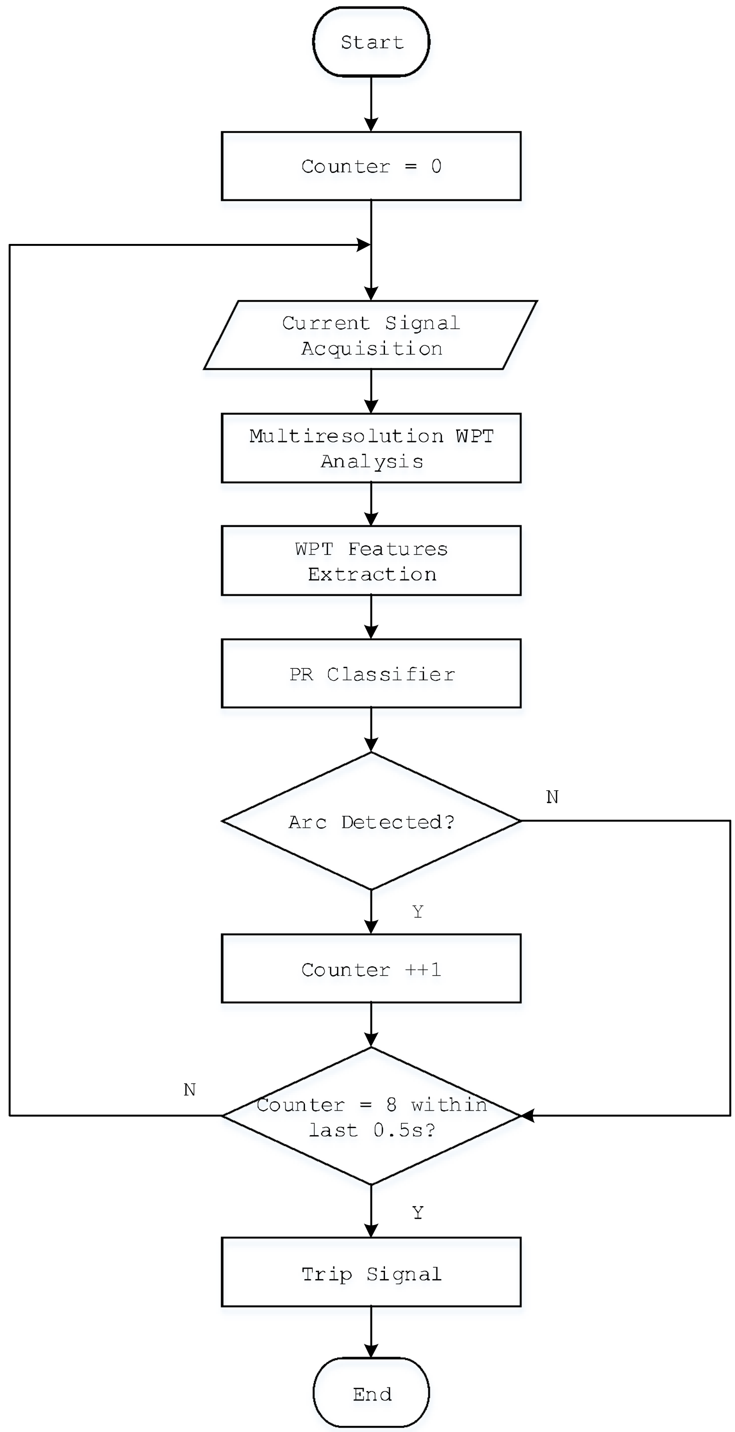


Figure 10. Arc detection process flow chart.

5.3.2 Current Sensing Stage

Current signal monitoring is the first step in the proposed method, and aims to capture the conductor current waveform using a sensitive current transformer with a high frequency response. A high frequency content is expected with transient behavior such as electric arcing. Current sensing is practically achieved using a wideband split core current transformer with internal sampling resistance, as shown in Figure 11.

Figure 11. Split core current transducer

Table 1. Current Transformer Specification

Rated Input Current	10A
Rated Output Voltage	1V
Core Material	Ferrite
Type	Split Core

A split core type is selected for practicality and ease of use with existing circuits for the purposes of real data collection and in later steps complete system validation.

Current signal is sampled at a rate of 30,720 sample per second for a 60 Hz system frequency, corresponding to 512 samples per current cycle. Current signal data is practically acquired using MATLAB Data Acquisition Toolbox.

5.3.3 Feature Generation and Extraction Stage

The sampled current signal does not represent any discriminatory features in its raw form in the time domain. Thus, signal projection into another domain is required to produce a feature format usable in machine learning algorithms. This stage consists of two steps; features generation and feature extraction.

Features are generated by decomposing current signals through multiresolution wavelet packet transform into five decomposition levels, as shown in Figure 12. A special case of infinite impulse response filter (IIR) is used as a WPT filter bank. Two half-band all-pass poly phase filters are used to design low-pass and high-pass filters. The specialized elliptic filter design (L=3) provided excellent frequency roll-off characteristics and lower computational complexity [44, 45, 46], hence very good frequency band separation for the WPT. The output coefficients at each WPT tree node represent signal features in their raw format. This procedure decomposes original signals into smaller bands within the signal frequency spectrum.

The feature extraction step retrieves information contained in the WPT coefficients. WPT tree nodes contain a number of coefficients. Depending on their position within the tree, the higher the decomposition level, the lower the number of coefficients and the smaller the frequency band represented. Various feature extraction methods exist in literature, including coefficient energy, root mean square, or entropy, all of which are suitable for feature extraction from wavelet transform coefficients, but the Euclidean norm method is found to be the most suitable for this problem. Discriminant information is extracted from tree coefficients using the coefficients' Euclidean norm. The Euclidean norm of a coefficient vector $p = (p_n)$ is given by equation 10.

$$\|p\|_2 = \sqrt{\sum_{n=-\infty}^{\infty} p_n^{\;2}} \tag{10}$$

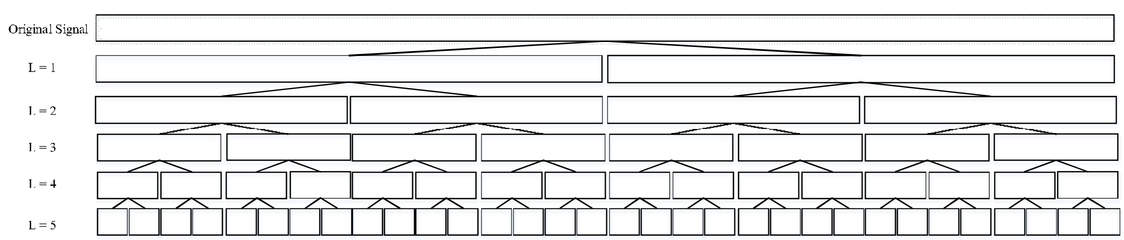


Figure 12. Wavelet packet decomposition tree

The Euclidean norm proves to provide a compromise between classification accuracy and reduced computational cost. Table 2 compares different feature extraction methods, using the same training and testing datasets and a training scheme with a full feature set.

Table 2. Comparison of different feature extraction methods for series arc detection

Feature Extraction Method	Accuracy	
	Training	Testing
RMS	97.4251	97.3730
Euclidean norm	97.9439	98.8730
Entropy “Shannon”	96.6701	97.8495
Entropy “Log-Energy”	92.5895	89.6641

5.3.4 Arc Detection Stage

In this stage, arc detection is carried out using features extracted in the previous stage to classify a current signal as normal or generated from an arc fault, followed by arc instances counter controlling the trip signal generation. Investigation of different machine learning algorithms suggests using SVM due to its various positive traits. This stage consists of several steps, ranging from features normalization and selection, to classifier training.

The generated features preprocessing step eliminates feature dominance and any existing bias due to a higher feature magnitude compared with more representative features of lower magnitude on classification accuracy. Features are scaled using min-max normalization to the range $[0,1]$. By finding the minimum and maximum values, $f_{\min}$ and $f_{\max}$ of the given feature vector f, the normalized feature instance $f_i^{'}$ can be found using equation 11.

$$f_i^{'} = \frac{f_i - f_{\min}}{f_{\max} - f_{\min}} \tag{11}$$

Feature selection aims to identify the features best correlated to the class label and to eliminate any redundant or insignificant features. Using only a subset of features would have a direct effect on classification accuracy and overall performance time, due to reduced feature generation and extraction time and minimized classifier feature space. A floating search was used to investigate feature selection using four selection criteria: the sum and minimum of estimated Mahalanobis distances and the sum and minimum of squared Euclidean distances. Table 3 shows that the minimum estimated Mahalanobis distance provides the most discriminate feature set, with a total of 50 features.

The selected feature set is [*2,31,62,6,61,60,30,59,63,14,13,4,9,26,29,58,12,24, 52,25,28,57,56,54,27,55,51,49,1,15,10,5,48,50,19,7,53,11,23,8,17,32,18,21,40,22,36,35,46,47*] out of 63 feature. This feature vector is not globally optimal, and thus some loss in system accuracy is anticipated. Branch and bound search cannot be used due to its computational complexity and the limited computing power available. Figure 13 shows the selected features decomposition tree.

Table 3. Floating search feature selection methods comparison

Total Number of Selected Features	Accuracy			
	Maha-s	Maha-m	Ecul-s	Ecul-m
10	96.1037	96.7043	81.5933	81.5933
20	96.9177	96.2380	97.4077	97.1864
30	97.2181	97.7634	97.6843	98.1348
40	98.5142	98.5300	98.6406	98.7276
50	98.7908	99.0042	98.8856	98.8619
63	98.8777	98.8777	98.8777	98.8777

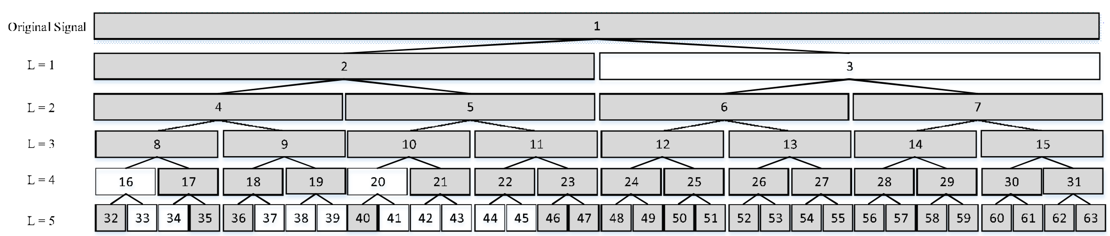


Figure 13. Selected features decomposition tree, selected features are highlighted in grey

SVM classification algorithm is used for arc detection, using a labelled dataset based on current signal extracted features using WPT. The LIBSVM library [47] is used in this research for its performance and portability to other environments, that is, its ease of deployment to an embedded environment. Linear kernel is found to be the most suitable for the generated features, as suggested by principal component analysis (PCA) projection of the training set. SVM is capable of reaching global optima, and above all immune to over training, which would prove helpful during the training phase features-normalization step, since minimum and maximum feature values need to be fixed in the training phase leading to the use of training set with high instance count. The SVM classifier is trained using a 10-fold cross validation training scheme, by randomly splitting the training dataset into 10 parts and using nine parts for training and one part rest for testing. By repeating the process for all other cases and taking their average accuracy, a linear SVM kernel with a cost parameter of 1 is found to produce the highest classification accuracy. PCA also shows high separation margins between feature clusters.

Table 4 shows the training and testing accuracies for the 10-fold partitions.

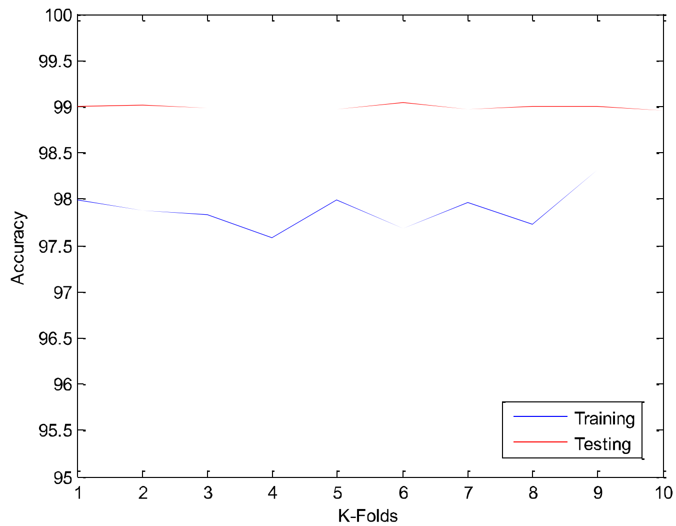


Figure 14. 10-fold cross validation classifier training and testing accuracies

Table 4. 10-fold cross validation classifier training and testing accuracies

Training Cases	Accuracy	
	Training	Testing
1	97.9966	99.0041
2	97.8700	99.0120
3	97.8274	98.9804
4	97.5748	98.9646
5	97.9966	98.9725
6	97.6803	99.0437
7	97.9544	98.9646
8	97.7224	98.9962
9	98.3129	99.0041
10	98.1864	98.9567
Average	97.9122	98.9900

Classifier output is fed to an arc-fault counter initiating a trip signal whenever counter exceeds a preset threshold. The arc counter is similar to a shift-register where each bit is equivalent to one current signal cycle, overall number of bits spans for half a second.

5.3.5 Integrated System

Enabling an intelligent system to gain problem-specific knowledge and achieve benchmark performance requires various steps from training to deployment. Some steps can be carried out offline and the selected parameters used online. Feature scaling, selection and classifier training and tuning are the most important offline steps.

Features are scaled using min-max normalization, where the minimum and maximum value of each feature is calculated during the training phase and used to scale new sample points. Training-feature selection reduces the online feature extraction step, depending on the selected subset and the relative WPT pruned tree sections. Classifier kernel selection and tuning will search for a mapping with the least support vectors. Figure 15 shows the different steps involved in the different system development phases.

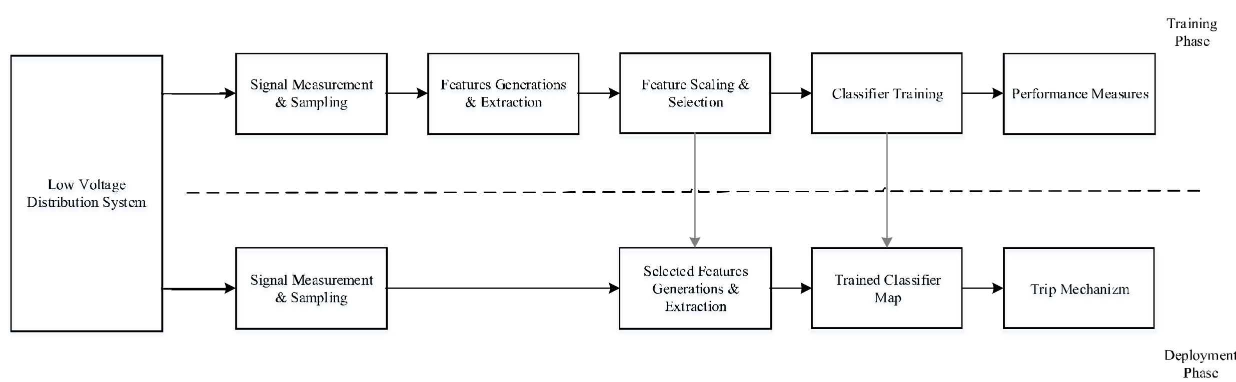


Figure 15. Overall system stages and steps for training and deployment phases

5.4 Dataset

Lab-generated data is used in this research to mimic real life conditions so as not to rely on synthetic arc data. Two separate datasets are used for this purpose: one for training the other for validation. The SVM classifier is trained using a subset of the training dataset and tested using another subset, while the validation dataset is used to validate the classifiers' generalization ability to detect arc conditions present in unseen loads and load combinations.

5.4.1 Dataset Generation

Arc generation was performed in a lab environment in accordance with a benchmark setup [4]. This includes generating arc conditions using different loads ranging from linear and non-linear load combinations; using different kinds of light bulbs, including incandescent, florescent, CFL and LED; as well as, employing other household appliances such as vacuum cleaners, power tools and electronic light dimmers, etc.

An arc generator consists of two electrodes, one of copper and another of carbon graphite placed in the contact path within Plexiglas tubing, with blade connectors providing the arc-generator's external connection. Both electrodes have a diameter of 6.4 mm, with the copper electrode having a pointed sharp end of around 20 mm length for arc initiation, and the carbon electrode having a flat end. The Plexiglas tubing serves as a physical motion guide for contact making and breaking and as an insulator to the surrounding environment. Figure 16 shows the arc generator used in this research.

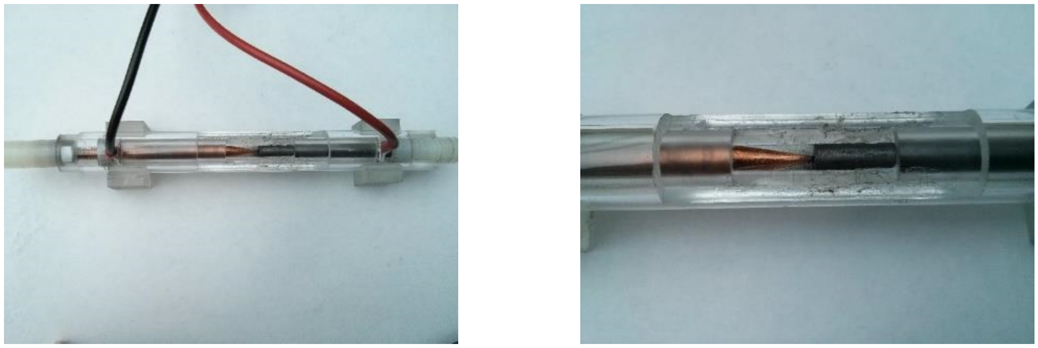


Figure 16. Benchmark arc generator

5.4.2 Training and Validation dataset

The training dataset consisted of linear and non-linear load combinations of varying proportions. The linear load is represented by a purely resistive load, where the non-linear load is in the form of a compact florescent lamp (CFL) of 13W, combined in parallel with variable linear loads resulting in twelve different load combinations. Table 5 presents training set linear and nonlinear conditions with their corresponding wattage. Figure 17 shows some training set load combinations' current waveforms.

Other load cases are added to the training set to widen diversity and increase classifier generality when applied to the validation set. These cases included appliances containing AC motors, variable speed drives, and electronic dimmers. Figure 17 shows some training load combinations' current waveforms for normal and arc conditions. Figure 18 shows the current waveforms for different speeds of a variable-speed controlled AC motor such as in a stand mixer. Figure 19 and Figure 20 show an electronic controlled light dimmer current at different firing angles and three kinds of light bulbs, respectively. Figure 21 shows the through current of different masking loads present in common household appliances. The training set consist of 47,419 sampled current cycles, 18,719 arcing cases and 28,700 non-arcing cases.

Table 5. Training dataset load combination

Combination %		Wattage (W)	
Linear	Non-Linear	Linear	Non-Linear
100	0	100.00	0
95	5	247.00	13
90	10	117.00	13
80	20	52.00	13
70	30	30.33	13
60	40	19.50	13
50	50	13.00	13
40	60	8.67	13
30	70	5.57	13
20	80	3.25	13
10	90	1.44	13
0	100	0	13

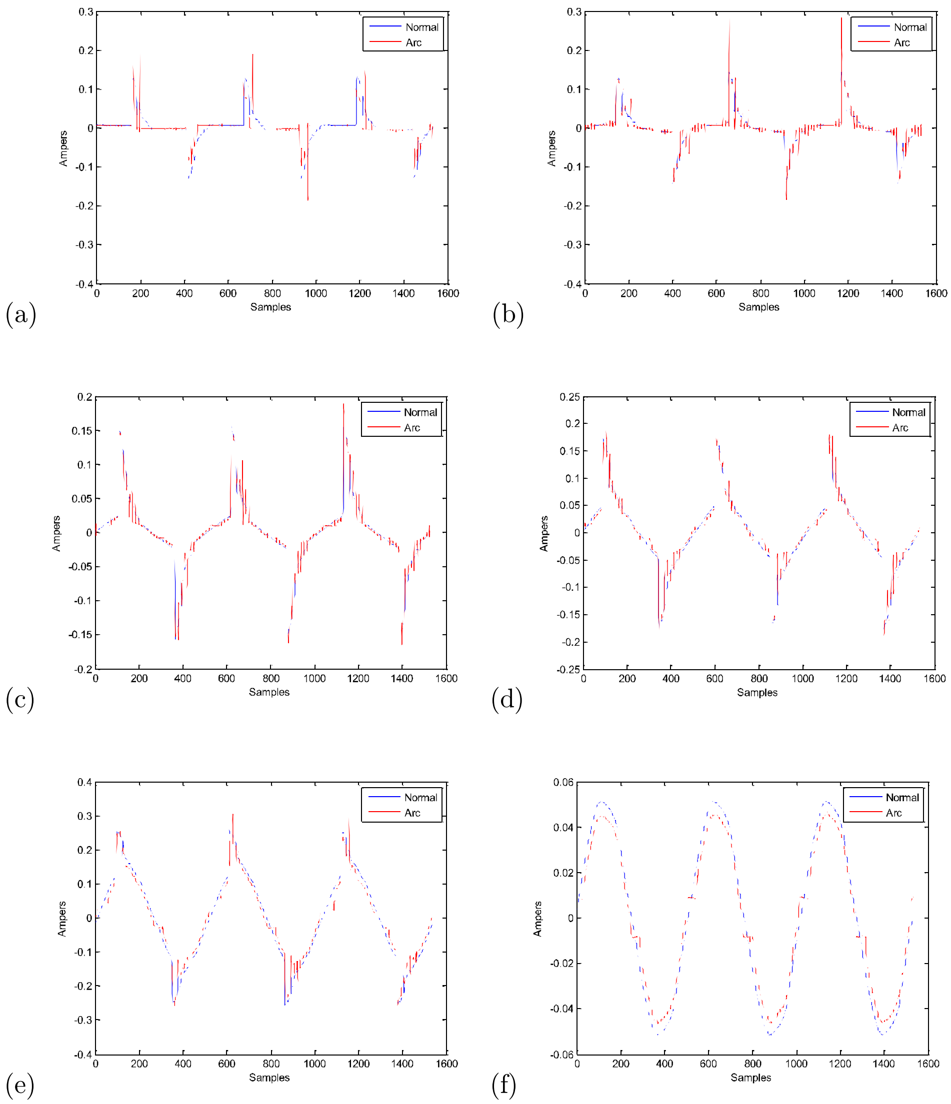


Figure 17. Training set load combinations' waveforms, (a) 0% linear and 100% nonlinear load (000/100), (b) 020/080, (c) 040/060, (d) 060/040, (e) 080/020, (f) 100/000

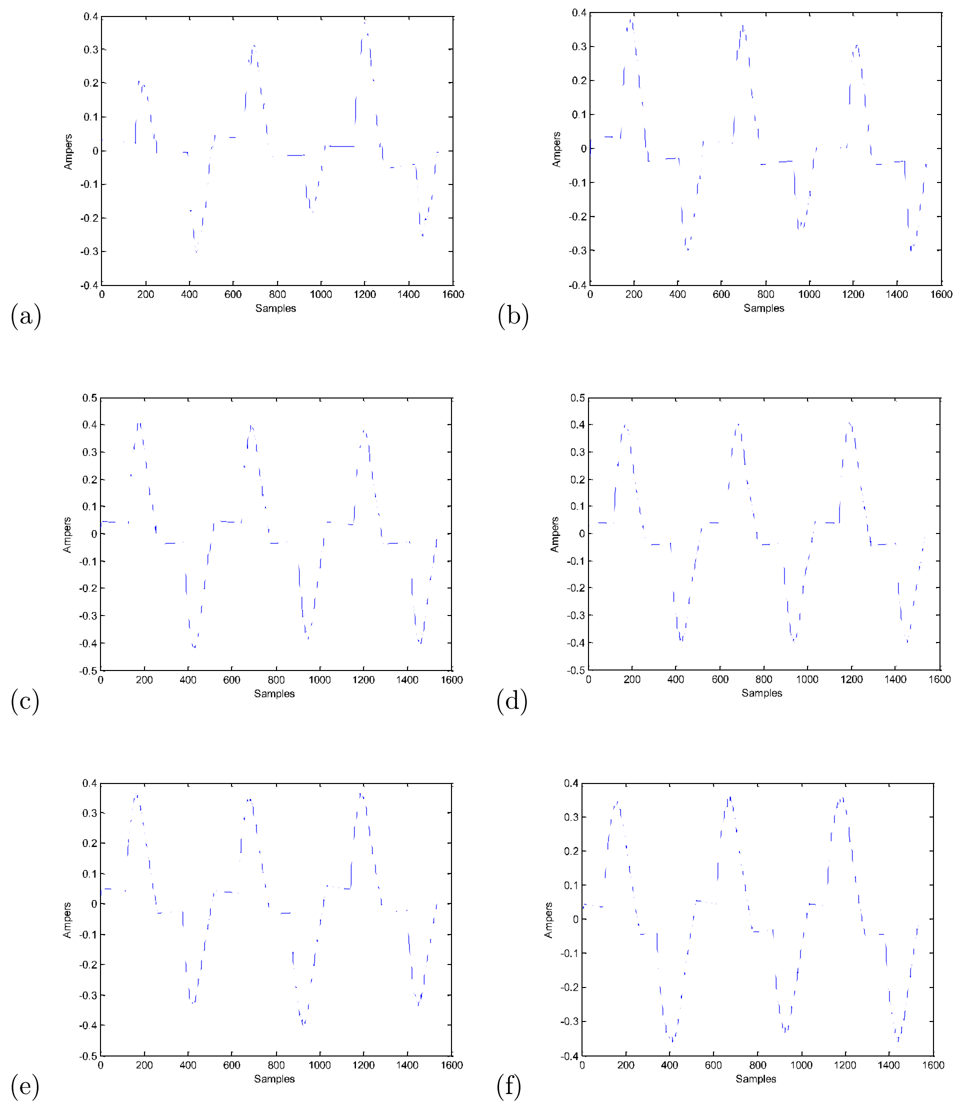
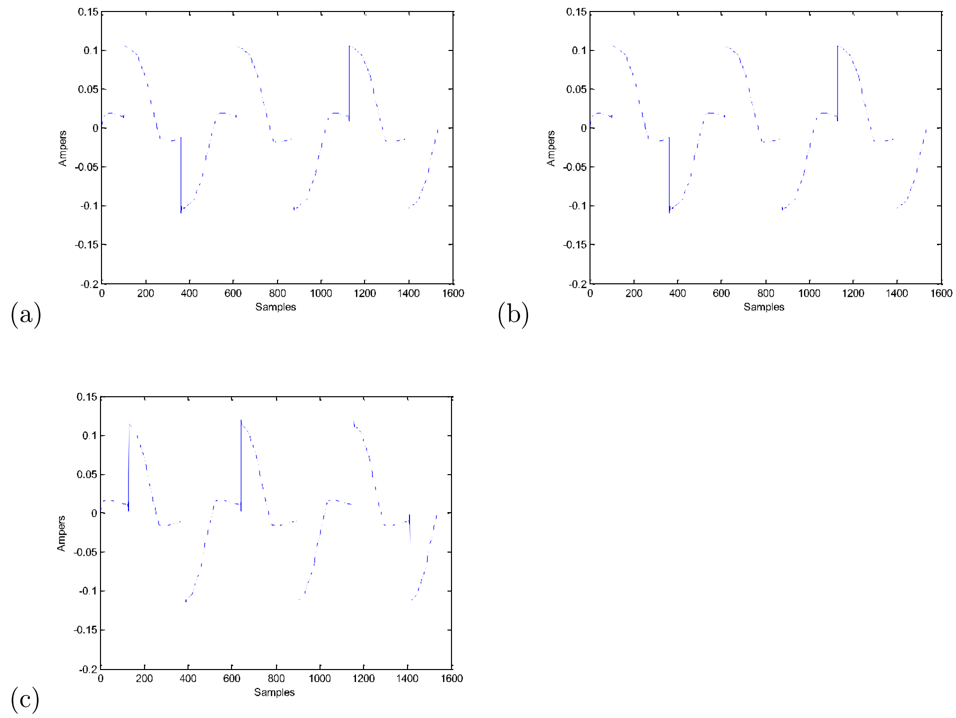


Figure 18. Stand mixer current waveform at different speeds



(a) (b) (c)

Figure 19. Light dimmer current waveform at different firing angles, (a) 60 deg, (b) 90 deg, (c) 120 deg

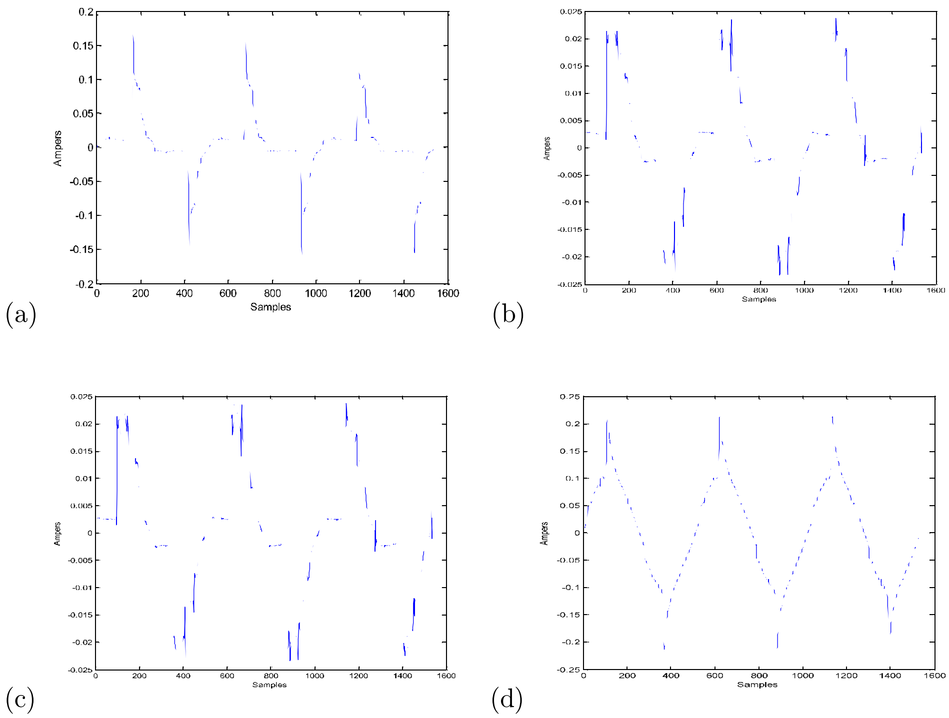


Figure 20. Different light bulbs current waveforms, (a) fluorescent lamp, (b) CFL lamp, (c) LED lamp, (d) all three in parallel

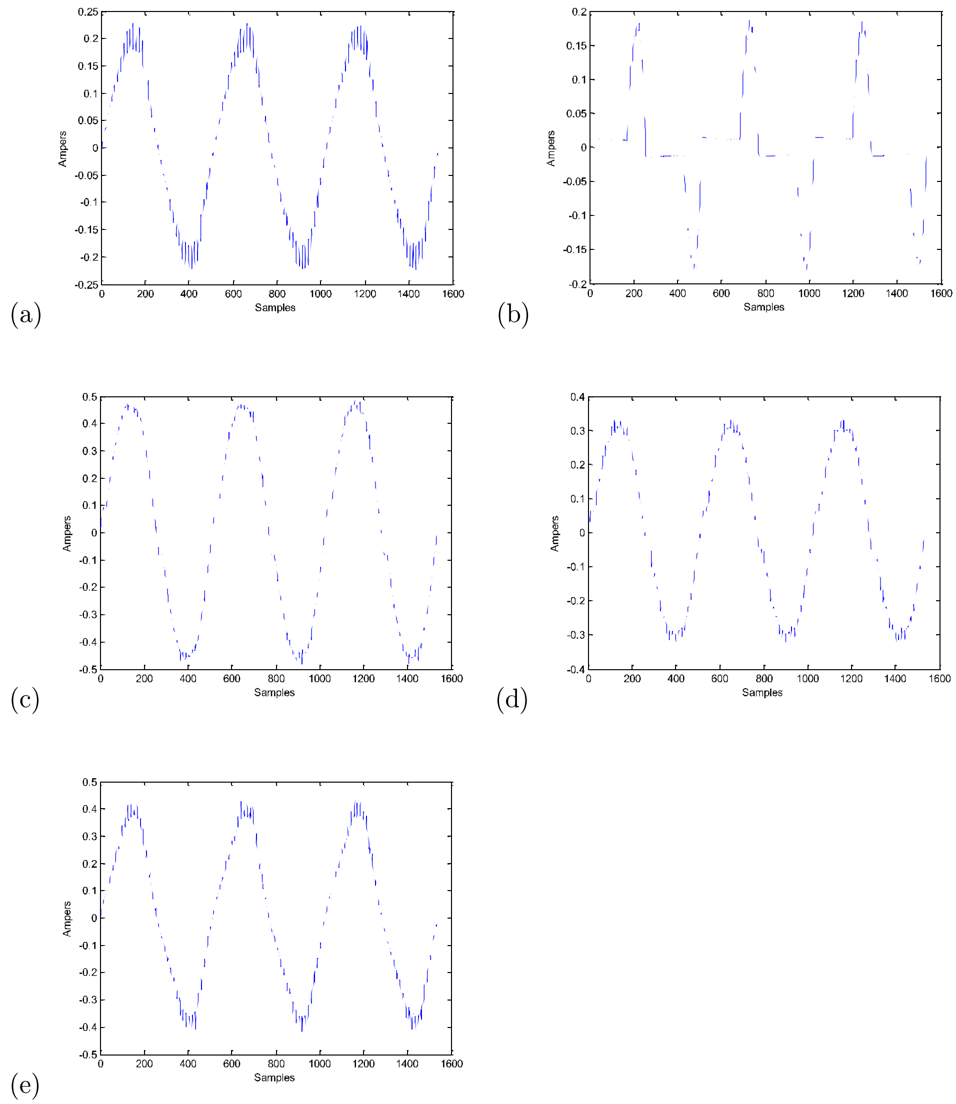


Figure 21. Rotating equipment current waveforms, (a) coffee grinder, (b) food processor, (c) hair dryer at low speed, (d) hair dryer at high speed, (e) power tool at high speed

The validation dataset consists of nine linear and non-linear load combinations, with different proportions than the training set. Table 6 presents the validation set linear and nonlinear combinations with their corresponding wattage. Other load cases are added to the validation set to examine classifier performance and generality, and to comply with the benchmark test in [4]. Specific loads are incorporated in the validation set, namely light dimmers and an electronically controlled variable-speed drive. The validation set consists of 12,653 data points, 1,853 arcing cases and 10,800 non-arcing cases. Figure 22 shows some validation set current waveforms.

Table 6. Validation dataset load combinations

Combination %		Wattage	
Linear	Non-Linear	Linear	Non-Linear
97	3	420.33	13
85	15	73.67	13
75	25	39.00	13
65	35	24.14	13
55	45	15.89	13
45	55	10.64	13
35	65	7.00	13
25	75	4.33	13
15	85	2.29	13

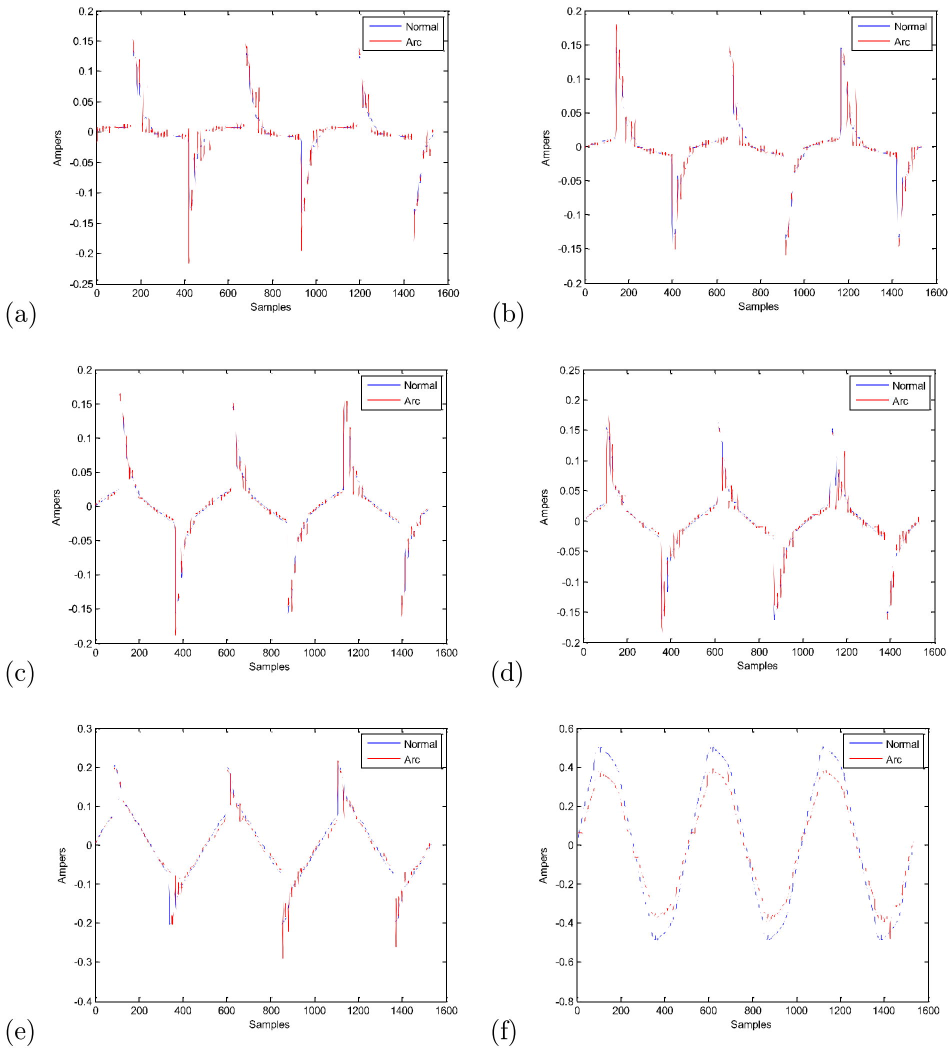


Figure 22. Validation Dataset Load Combinations, (a) 15/85, (b) 25/75, (c) 45/55, (d) 55/45, (e) 75/25, (f) 97/03

5.5 Performance evaluation

The overall performance of the proposed system is analyzed using the overall accuracy, error rate and precision to gives extra confidence, especially with the disjoint training and testing sets. A comparison of one case of training compared with testing is shown in Table 7. Where accuracy is the overall detection rate for positive and negative labels, the error rate is 1-*Accuracy*, precision is the ratio between a true positive and true positive and false positives, and F-measure is the harmonic mean of precision and recall.

Table 8 show a confusion matrix for training and testing as per Table 7. The receiver operating characteristic shows very good classification behavior as shown in Figure 23.

Table 7. Classification and training performance measures

Measure (%)	Training	Testing
Accuracy	98.1864	98.9567
Error	0.01814	0.01043
Precision	99.3370	95.9423
True Positive Rate (Recall)	96.0470	96.9700
False positive rate (*fpr*)	0.4180	0.7030
True Negative Rate (Specifity)	99.5810	99.2900
False negative rate (*fnr*)	3.9520	3.0220
F-measure	0.976643	0.964534

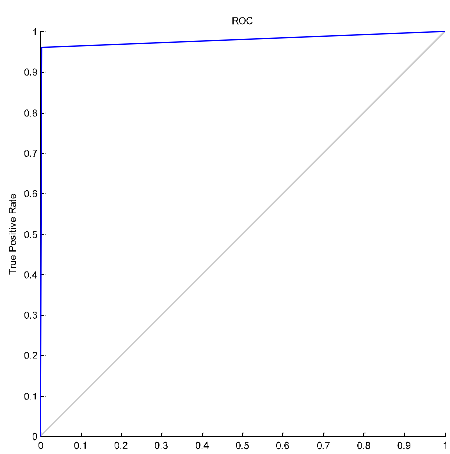
Table 8. (a) Classifier training confusion matrix, (b) classifier testing confusion matrix

(a)

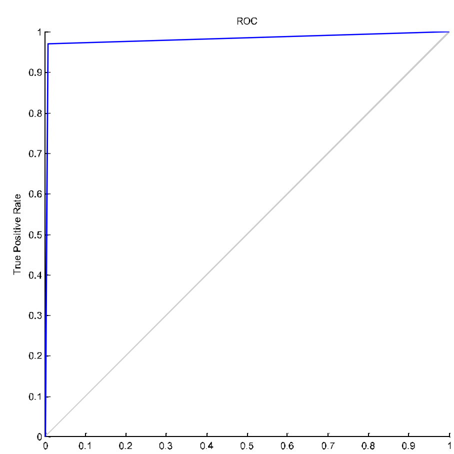
Predicted label	True label 0	True label 1
0	2858	74
1	12	1798

(b)

Predicted label	True label 0	True label 1
0	10,724	56
1	76	1797



(a)



(b)

Figure 23. Receiver operating characteristic, (a) training, (b) testing

5.6 Discussion

The system proposed in this chapter aims to solve series arc-detection problem using signal processing and machine learning techniques. This approach is chosen to cope with the transient nature of the problem and the extreme difficulty of formulating an exact problem model. The proposed system is designed to meet standard requirement performance [4], and to be deployable in an embedded environment by having a manageable computational cost. Investigation of different signal processing techniques combined with machine learning algorithms is found to produce the best performance with the least complexity. From feature generation and extraction to classification technique, computational cost reduction was pursued while performance-accuracy is maintained. Practical fault simulations are carried out in a lab environment mimicking real life loads, generating training and validation dataset using a benchmark arc generator [4].

The classifier performance measure indicates high accuracy with a low error rate. The high true positive rate of 96.79% implies the high capability of arc detection, with the very low false positives rate of 0.703% correlating to high immunity to masking loads and nuisance tripping, coupled with the very high true negative rate of 99.29% supported with high F-measure. These performance figures can be improved further with refinements in feature selection and classifier tuning.

Chapter 6

Conclusion and Future Work

6.1 Conclusion

This research has proposed a practical solution to low voltage series arc detection problem using pattern recognition techniques, with immunity to false tripping due to masking loads. Although other research on this problem has been conducted over the years, most solutions have some weaknesses due to their limited application scope. The proposed method is based on multiresolution wavelet packet signal decomposition and support vector machines. The method delivers satisfactory performance within benchmark standard's requirements and a reduced computational cost. It is also inherently versatile enough to be ported and with minor adjustments used on other difficult-to-model electric protection problems.

Although the proposed system produced very specific and measured performance indices, it still faces some hard and soft limitations. A hardware limitation was encountered in the data acquisition stage, as PC sound card is used for current data acquisition. The actual sampling rate is found to be lower than theoretically calculated to produce the required results. As a result, each captured sampled current cycle is representative of more than one current cycle, with tangible effects when logging current signals in continuous time.

The proposed method is tested using the test dataset containing equipment listed on CSA certification standard, exclusive from the training set, not including any electronically controlled variable speed switch power tools, or switch changeover under load. Nonetheless, a variable speed stand mixer under varying operational speeds is included in the set.

6.2 Future work and system improvements

Optimum methods for deploying the proposed system to an enclosed embedded system need to be investigated, taking into consideration the overall system complexity, plus the individual components level. The architecture used may rely on FPGA or DSP chips, not to mention ADC chip and the embedded system clocks.

Another important aspect for further investigation is the relation between the current signal sampling rate and system accuracy. In other words, what is the lowest sampling rate that achieves the desired performance? Another possibility is the relation between the sampling rate and the required computing power from the system side, since a higher sampling rate directly correlates to high captured definition and system accuracy.

The most important aspect of improvement exists within the classification algorithm itself. Support vector machines produces class labels using decision margins based on support vectors. Reducing the total number of classifier support vectors would reduce computational costs, either by special selection of training set components or by feature engineering within the classification algorithm.

Bibliography

[1] M. Wijayasinghe, "Fire Losses in Canada Year 2007 and Selected Years," Council of Canadian Fire Marshals and Fire Commissioners, Calgary, Alberta, Canada, 2011.

[2] J. John R. Hall, "Electrical," National Fire Protection Association, Apr 2013. [Online]. Available: http://www.nfpa.org/research/reports-and-statistics/fire-causes/electrical. [Accessed 20 Oct 2015].

[3] National Electric Code 2011, NFPA, 2011.

[4] CSA, "Arc Fault Protective Devices, C22.2 No.270-14," Canadian Standard Association, 2014.

[5] G. Gregory and G. Scott, "The arc-fault circuit interrupter: an emerging product," *IEEE Transactions on Industry Applications,* vol. 34, no. 5, pp. 928-933, Sep/Oct 1998.

[6] J. B. Calvert, "Electrical Discharges How the spark, glow and arc work," [Online]. Available: http://mysite.du.edu/~jcalvert/phys/dischg.htm. [Accessed 15 Sep 2015].

[7] J. J. Shea, "Conditions for Series Arcing Phenomena in PVC Wiring," *IEEE transactions on components and packaging technologies,* vol. 30, no. NO. 3, pp. 532 - 539, 2007.

[8] C. E. Restrepo, "Arc Fault Detection and Discrimination Methods," in *the 53rd IEEE holm conference on Electrical contacts*, 2007.

[9] J. A. Wafer, "The Evolution of Arc Fault Circuit Interrupters," in *Proceedings of the Fifty-First IEEE Holm Conference on Electrical Contacts*, 2005.

[10] J. J. Shea, "Identifying Causes for Certain Types of Electrically Initiated Fires in Residential Circuits," in *Aged Electrical Systems*, Chicago, 2006.

[11] S. T. M. M. A. Peter Müller, "Characteristics of Series and Parallel Low Current Arc Faults in the Time and Frequency Domain," in *2010 Proceedings of the 56th IEEE Holm Conference on Electrical Contacts (HOLM)*, 2010.

[12] K. W. R. F. D. George D. Gregory, "More About Arc-Fault Circuit Interrupters," *IEEE Transactions on Industry Applications,* vol. 40, no. 4, 2004.

[13] G. Gregory and G. Scott, "The arc-fault circuit interrupter: an emerging product," *IEEE Transactions on Industry Applications,* vol. 34, no. 5, pp. 928-933, Sep/Oct 1998.

[14] R. Kaufmann and J. Page, "Arcing Fault Protection for Low-Voltage Power Distribution System - Nature of the Problem," *Power Apparatus and Systems, Part III. Transactions of the American Institute of Electrical Engineers ,* vol. 79, no. 3, 1960.

[15] C. J. Kim, "Electromagnetic Radiation Behavior of Low-Voltage Arcing Fault," *IEEE Transactions on Power Delivery,* vol. 24, no. 1, 2009.

[16] W.-J. L. I. M.-S. C. Wiktor Charytoniuk, "Arcing Fault Detection in Underground Distribution Networks—Feasibility Study," *IEEE Transactions on Industry Application,* vol. 36, no. 6, 2000.

[17] D.-w. Park., I.-k. Kim., S.-y. Choi. and G.-s. Kil., "Detection Algorithm of Series Arc for Electrical Fire Prediction," in *2008 International Conference on Condition Monitoring and Diagnosis*, Beijin China, 2008.

[18] L. X. X. D. C. N. C. D. Peiyong DUAN, "An Arc Fault Diagnostic Method for Low Voltage Lines Using the Difference of Wavelet Coefficients," in *IEEE 9th Conference on Industrial Electronics and Applications (ICIEA)*, 2014 .

[19] S. Z. Zhang Ruixiang, "Arc Fault Detection Method Based on Signal Energy Distribution in Frequency Band," in *2012 Asia-Pacific Power and Energy Engineering Conference (APPEEC)*, 2012.

[20] G. S. M. S. T.S. Sidhu, "Arcing fault detection using artificial neural networks," *Neurocomputing,* no. 23, 1998.

[21] NEC, "Natioanl Elcetric Code 2002," *2002.*

[22] N. Rivera, "Arcing Fault Detector". US Patent 4,658,322, 14 April 1987.

[23] B. C. Beihoff., C. J. Tennies., J. Francis S. Richards and W. K.O'Neil, "Frequency Selective Arc Detection". US Patent 5,185,684, 9 Feb 1993.

[24] F. K. Blades, "Method and Apparatus for Detecting Arcing in Alternating Curent Power Systems by Monitoring High-Frequency Noise". US Patent 5,223,795, 29 Jun 1993.

[25] F. K. Blades, "Method and Apparatus for Detecting Arcing in Alternating Curent Power Systems by Monitoring High-Frequency Noise". US Patent 5,432,455, 11 Jul 1995.

[26] C. W. Ragsdale, "Arc Detection System". US Patent 5,280,404, Jan 18 1994.

[27] C. F. Renn. and W. J.Koeller, "Arc detector for electric rod furnace". US Patent 4,376,243, 26 Jan 1981.

[28] T. C. Rae, "Arc Fault Detector with Immunity to Tungesten Bulb Burnout and Circuit Breaker Incorporating same". US Patent 5,805,398, Sep 8 1998.

[29] R. W. M. S. C. S. Joseph C. Zuercher, "Apparatus for Envelope Detection of Low Current Arcs". US Patent 5,818,237, 28 Oct 1998.

[30] B. B.Neiger., R. N.Bradley., J. N.Pearse. and W. J.Rose, "Arc Fault Detector with Circuit Interrupter and Early Arc Fault Detection". US Patent 6,128,169, 3 Oct 2000.

[31] T. J. M. Kevin L. Parker, "Arc Fault Circuit Interrupter and Method Providing Improved Nuisance Trip Rejection". US Patent 8,089,737 B2, 3 Jan 2012.

[32] I. Daubechies, Ten Lectures on Wavelet, Philadelphia, Pennsylvania: Society for Industrial and Applied Mathematics, 1992.

[33] a. C. H. Martin Vetterli, "Wavelets and Filter Banks: Theory and Design," *IEEE Transaction on Singla Processing,* vol. 40, no. 9, pp. 2207, 2232, SEPTEMBER Sep 1992.

[34] O. R. a. M. VETIERLI, "Wavelets and Signal Processing," *IEEE SP Magazine,* pp. 14, 38, October 1991.

[35] M. V. Wickerhauser, Adapted Wavelet Analysis from Theory to Software, Wellesley, Massachusetts: A K Peters, Ltd., 1994.

[36] N. J. Nilsson, Introduction to Machine Learning, Stanford, 1998.

[37] K. P. Murphy, Machine Learning, A Probabilistic Perspective, Cambridge, Massachusetts: MIT Press, 2012.

[38] M. H. Dunham, Data Mining, Introductory and Advanced Topics, Upper Saddel River, NJ: Prentice Hall, 2002.

[39] M. K. Jiawei Han, Data Mining Concepts and Techniques, Elsevier, 2006.

[40] C. CORTES and V. VAPNIK, "Support-Vector Networks," *Machine Learning,* vol. 20, no. 3, pp. 273-297, 1995.

[41] A. Jain, R. Duin and J. Mao, "Statistical pattern recognition: a review," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 22, no. 1, pp. 4-37, Jan 2000.

[42] C. Bishop, Neural Networks for Pattern Recognition, Oxford: Clarendon Press, 1995.

[43] P. Somol, P. Pudil and J. Kittler, "Fast branch & bound algorithms for optimal feature selection," *IEEE Transactions Pattern Analysis and Machine Intelligence,* vol. 26, no. 7, pp. 900-912, July 2004.

[44] A. N. Akansu. and M. J. Smith., "IIR Filter Banks and Wavelets," in *Subband and Wavelet transforms Design and Application*, Dordrecht, The Netherlands, Kulwer Acadmic Publishers, 1996.

[45] L. Eren, M. Unal and M. Devaney, "Harmonic analysis via wavelet packet decomposition using special elliptic half-band filters,," in *21st IEEE Instrumentation and Measurement Technology Conference*, 18-20 May 2004.

[46] L. Eren and M. Devaney, "Bearing damage detection via wavelet packet decomposition of the stator current,," *IEEE Transactions on Instrumentation and Measurement,* vol. 53, no. 2, pp. 431-436, April 2004.

[47] C.-C. Chang and C.-J. Lin, "LIBSVM : a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology,* 2011.