Czech Technical University in Prague
Faculty of Electrical Engineering
Department of Measurement

AC ARC FAULT DETECTION
DETEKCE OBLOUKU VE STŘÍDAVÝCH OBVODECH

by

KAZIM YİĞİT BAŞER

Master's Thesis
Supervisor: doc. Ing. Radislav Šmíd, Ph.D.
Field of Study: Cybernetics and Robotics

Prague, May 2019

Supervisor:
doc. Ing. Radislav Šmíd, Ph.D.
Department of Measurement
Faculty of Electrical Engineering
Czech Technical University in Prague
Technická 1902/2
166 27 Prague 6
Czech Republic

MASTER'S THESIS ASSIGNMENT

I. Personal and study details

Student's name: **Baser Kazim Yigit** Personal ID number: **471698**

Faculty / Institute: **Faculty of Electrical Engineering**

Department / Institute: **Department of Measurement**

Study program: **Cybernetics and Robotics**

Branch of study: **Cybernetics and Robotics**

II. Master's thesis details

Master's thesis title in English:

AC Arc Fault Detection

Master's thesis title in Czech:

Detekce oblouku ve střídavých obvodech

Guidelines:

Based on extensive bibliographic research, develop methods for detection of AC arc faults in residential and commercial buildings using analysis of load current. Compare different approaches to signal features computation, e.g., frequency distribution, correlation, higher-order spectral analysis, filter banks, time-frequency analysis, fractal dimension, etc. Evaluate the performance of the detection for standard household appliances.

Bibliography / sources:

[1] Jinmi Lezama et al.: An embedded system for AC series arc detection by inter-period correlations of current, Electric Power Systems Research, Volume 129, 2015, Pages 227-234,

[2] Amna Farooq Husain: Series Arc Fault Detection in the Presence of Household Electrical Loads, diploma thesis, CTU FEE, 2017.

[3] K. Zeng, L. Xing, Y. Zhang and L. Wang: Characteristics analysis of AC arc fault in time and frequency domain. 2017 Prognostics and System Health Management Conference (PHM-Harbin), Harbin, 2017, pp. 1-5.

Name and workplace of master's thesis supervisor:

doc. Ing. Radislav Šmíd, Ph.D., Department of Measurement, FEE

Name and workplace of second master's thesis supervisor or consultant:

Date of master's thesis assignment: **14.02.2019** Deadline for master's thesis submission: ______

Assignment valid until:
by the end of summer semester 2019/2020

doc. Ing. Radislav Šmíd, Ph.D.
Supervisor's signature

Head of department's signature

prof. Ing. Pavel Ripka, CSc.
Dean's signature

III. Assignment receipt

The student acknowledges that the master's thesis is an individual work. The student must produce his thesis without the assistance of others, with the exception of provided consultations. Within the master's thesis, the author must state the names of consultants and include a list of references.

18.02.2019
Date of assignment receipt

Student's signature

I, KAZIM YİĞİT BAŞER declare that I carried out the research, complete the assigned thesis,'AC Arc Fault Detection' by myself and I referenced materials I used.

Date:

Signature:

Abstract

This thesis aims to propose a detection algorithm for AC Arc Fault Detection Devices(AFDD) using direct digitization and supervised machine learning algorithms.

The price of high frequency, direct digitization devices have been steadily decreasing in recent years and it is expected that this trend will continue in the future. Therefore, data processing using direct digitization is becoming a feasible method to use in AC arc fault detection devices.

The main purpose of this thesis is to analyze arc faults in suitable domains and to propose a method for detection based on supervised machine learning algorithms. In order to analyze arc faults, measurements were gathered according to IEC62606 standards. After suitable analysis, this data was then used to design supervised machine learning algorithms to achieve successful detection performance.

In particular, the main contributions of the thesis to arc fault detection research are as follows:

1. An optimized algorithm for detection that can be utilized both in existing solutions and in the proposed solution in this thesis

2. An automated feature selection tool from current waveforms that is done heuristically in existing solutions.

Keywords:

Supervised Machine Learning, AC Arc Fault Detection, Arc Fault Detection Device(AFDD) Support Vector machine, Feature extraction, Fourier transform.

Abstract

Cílem této práce je vyvinout detekční algoritmus pro detekci střídavého oblouku v zařízeních AFDD s vyuitím přímé digitalizace a strojového učeni.

Cena obvodů pro rychlou přímou digitalizaci stále v posledních letech klesá a očekává se pokračování tohoto trendu. Proto bude moné pouití přímé digitalizace v zařízeních pro detekci oblouku AFDD.

Hlavním cílem práce je analýza oblouku ve vhodných oblastech a návrh metody detekce zaloené na strojovém učení s učitelem. Pro analýzu byla provedena měření podle standardu IEC62606. Po analýze byla tato data pouita pro návrh algoritmu strojového učení s učitelem k dosaení spěšné detekční schopnosti. Hlavním přínosem práce k výzkumu

detekce oblouku je:

1. optimalizovaný algoritmus pro detekci, který mue být pouit v existujících řešeních a v řešení navreném v této práci

2. nástroj pro automatizovaný výběr příznaků, který je v existujících řešeních prováděn heuristicky

Klíčová slova:

Strojové učení s učitelem, detekce stř ídavého oblouku, zařízení pro detekci střídavého oblouku, SVM, extrakce příznaků, Fourierova transformace

Acknowledgements

I would like to express my appreciation to my supervisor doc. Ing. Radislav Šmíd, Ph.D. for being a constant source of support and leading me whenever I needed assitance throughout my thesis. Without his guidance throughout the duration of research, this thesis could not have been written.

I would like to thank Dr.Özlem Şen for helping me whenever I needed support. Together with her guidance, she inspired me with her virtues that I will keep as a guide in my life.

I am grateful to my colleagues at Eaton EEIC with whom it was a pleasure to work. I express my special thanks to Adam Gabriel for trying to show his support when needed.

Additionally, I would like express my gratitude to the proof readers who helped me edit this thesis, namely Sara Loucille Steil, Deepak Koranga, Jeremy Kramer, Arturo Montes de Oca Zapiain, Nicola Zaru and, Anna West.

I need to thank my friends Mert Daloğlu and Ulas Ozipek for being there for me.

The author thanks the International Electrotechnical Commission (IEC) for permission to reproduce Information from its International Standards. All such extracts are copyright of IEC, Geneva, Switzerland. All rights reserved. Further information on the IEC is available from www.iec.ch. IEC has no responsibility for the placement and context in which the extracts and contents are reproduced by the author, nor is IEC in any way responsible for the other content or accuracy therein.

Last, but not least, I deeply appreciate the efforts of my family, my father Adnan and my mother Ayten. Without their constant support, infinite patience, understanding, care and love, I would not have accomplished to write this thesis.

Contents

List of Figures

List of Tables

Abbreviations

AFDD	Arc Fault Detection Device
ALCI	Appliance Leakage Current Interrupter
a-PFC	Active Power Factor Correlation
ANN	Artificial Neural Network
CFL	Compact Fluorescent Lamp
DFT	Discrete Fourier Transform
GFCI	Ground Fault Circuit Interrupter
GFI	Ground Fault Interrupter
GIL	General Incandescent Bulb
Ktoe	Thousand Tonnes of Oil Equivalent
LED	Light Emitting Diode
MCB	Miniature Circuit Breaker
MCCB	Moulded Case Circuit Breaker
NCA	Neighborhood Component Analysis
n-PFC	No Power Factor Correlation
p-PFC	Passive Power Factor Correlation
RBF	Radial Basis Function
RCBO	Residual Current Circuit Breaker with Overcurrent Protection
RCCB	Residual Current Circuit Breaker
RCD	Residual Current Device
SMPS	Switched Mode Power Supply
SPIM	Single Phase Induction Motor
SVM	Support Vector Machine

Chapter **1**

Introduction

In this chapter, the motivation for thesis subject " AC arc fault detection" is specified, challenges in design of arc fault detection devices are addressed and outcomes and structure of thesis are stated.

Electrical fires are one of deadliest and common hazards of the 21stcentury. According to Fire Safe Europe which is an European association for fire safety in buildings, *200000 fires are reported in Europe each year and 90% of fires in the European Union happen in buildings. 4000 people are killed by fire in Europe every year which is 11 deaths per day. 7000 people are hospitalized in Europe each year due to severe injuries caused by fire. 126 billion Euro which is equivalent to 1% of European GDP is eaten up by fire damage each year.* [1] From 2014 to 2016, an estimated 24,000 residential building electrical fires were reported to United States fire departments each year and these fires caused an estimated 310 deaths, 850 injuries and $871 million in property loss. [3][4]

According to the Geneva Association, 25% of fires are ignited by electrical failure in Europe.[2]. In only 17% of residential building electrical fires, the fire spread was limited to the object where the fire started.[3] Residential building electrical fires occurred most often in the winter month of January (12%) which is considered because of an increase in demand of heating.[3]

Together with deaths and injuries, there are also economical losses caused by fires. In the U.S, residential building electrical fires cost $27.500 loss per fire.[3]

Different protective devices are used to prevent deaths, injuries, and economic loss. Most widely used devices are circuit breakers, introduced in more detail in Section 2.3.1. Circuit breakers provide protection against overloads and short circuits. Residual current devices explained in Section 2.3.2 are used to detect the imbalance between live and neutral wires, namely, leakages. However, these devices cannot provide protection against arc faults. Arc fault detection devices, introduced in more detail in Section 2.3.3, are necessary to employ in order to provide more protection.

Since unspecified short-circuit arc (23%), and short-circuit arc from defective, worn insulation (11%) are a big portion of the factors that contribute to the ignition of residential electrical fires, together with the malfunction (43%) and other electrical failure[3], arc fault detection devices are important with respect to saving lives, preventing injuries and economical loss.

Lastly, according to Transparency Market Research (TMR), the arc fault detection device market revenue has reached the $3.769 billion in 2017 and it is expected to grow with a solid 5.3% compound annual growth rate until 2025 where revenue is projected to increase 5.596$ billion by 2025. [5] This presents a promising opportunity for this technology to be adopted by the market.

1.1 Motivation

Arcing is defined as "luminous discharge of electricity across an insulating medium, usually accompanied by the partial volatilization of the electrodes"[6]

When arc fault occurs, it generates broadband noise up to 1GHz. Arc fault detection is a challenging task because of the complexity of phenomena explained in Section 2.2. However, introduction of new loads with different noise characteristics, power line communication devices, or different type of distribution systems such as ring circuitry, make detection even more challenging for existing detection methods. In light of this information, it is needed to investigate advancements in arc fault detection technology and to propose more robust solutions.

A number of parameters make it difficult to identify arc fault events. These include wide spectrum noise, complex cause-effect relationships, and load dependant signal properties. As such, it is important to employ a variety of signal processing approaches, including:

1. Time domain signal processing

2. Frequency domain signal processing

3. Time-frequency domain signal processing

Generally, the approaches above are combined in many products on the market to increase robustness and minimize nuisance tripping. Robustness issues decrease the protection level while nuisance tripping has economical impacts due to unnecessary de-energization of circuitry. As an answer to this, there are numerous algorithms being developed to increase robustness of arc fault detection.

Arc fault detection is a binary classification task, a task of classifying the elements of a given set into two groups (predicting which group each one belongs to) on the basis of a classification rule.[24] Considering the advancements in data acquisition devices and classification algorithms rapidly developed in recent past, it becomes feasible to propose a new detection algorithm using direct digitization for data acquisition and supervised machine learning algorithms for detection.

1.2 Goals of the Thesis

This thesis is intended to analyze arc faults that occur in low voltage level residential areas using suitable signal processing techniques and to propose a novel detection algorithm for arc fault detection using supervised machine learning algorithms.

In order to investigate arc faults in detail, measurements in compliance with the IEC:62606 standard [6] that contains general requirements for arc fault detection devices are used because using real measurements is more reliable and the result is guaranteed compared to other methods such as modelling arc faults.

To design an arc fault detection algorithm, data acquired by measurements need to be analyzed in suitable domains. In this thesis, main goal is to achieve best detection performance by applying suitable signal processing techniques.

Another goal of the thesis is to design supervised machine learning algorithms for arc fault detection and then to design an algorithm to achieve successful and robust detection performance.

In particular, the main contributions of the thesis to arc fault detection research are as follows:

1. A successful detection algorithm that is suitable to use with a proposed data acquisition method, namely, direct digitization. The algorithm can also be utilized by existing detection solutions to increase efficiency and robustness.

2. An automated feature selection tool, for analyzing current waveforms, that is used for pinpointing the frequency bands that provide sufficiently enough data to classify an arc fault event successfully. This is done heuristically in existing solutions.

1.3 Structure of the Thesis

The thesis is organized into six chapters as follows:

1. *Introduction*: Describes the motivation behind arc fault detection and gives statistical data. Main contributions of the thesis are introduced.

2. *AC Arc Fault Detection*: Introduces AC arc fault types, arc characteristics and protection devices. Load types and characteristics are introduced.

3. *AFDD Tests and Measurements*: Tests and measurements according to IEC:62606 are introduced. Measurement results and test setups used in measurements are explained.

4. *Direct Digitization Approach*: Proposed algorithm for arc fault detection is explained in detail. Basics of support vector machine are introduced. Designed feature extraction algorithms are presented. Proposed detection algorithm from signal processing to detection are explained in detail. Performance metrics and criteria are introduced.

5. *Main Results*: Results of research are summarized according to decided metrics.

6. *Conclusion*: Findings are discussed and possible topics for further research is presented.

Chapter 2

AC Arc fault detection

In this chapter, AC arc fault types and corresponding protection devices with a focus on arc fault detection device is explained.

2.1 AC Arc Faults

According to IEC62606:2013 which is the international standard published by International Electrotechnical Commission for general requirements for arc fault detection devices, an arc is defined as a luminous discharge of electricity across an insulating medium, which usually results in the partial volatilisation of the electrodes. The definition of an arc fault is given as a hazardous unintentional arc between two conductors.[6] An AFDD is described as a device able to detect and mitigate the effects of arc faults by disconnecting the circuit when such a fault is detected.[7]

Although arcing or sparks may occur under normal operation in loads such as electrical drills or air compressors, these are not classified as hazardous. As such the standard for a harmful arc, in arc fault detection research is much higher. Since an AFDD is intended to de-energize the circuit when a hazardous arc occurs, these devices must be able to ignore non-hazardous arcing. When distinguishing between hazardous arc faults and arcing due to normal operation, some loads can still present a large challenge in arc fault detection research

According to [8] , typical causes of an arc faults are as follows:

1. Cords or wires having a loose connection
2. Crushed wires
3. Damaged or misapplied electrical equipment

4. Ageing installations

5. Pets and rodent bites

Illustration of common arc fault causes are shown in Figure 2.1.

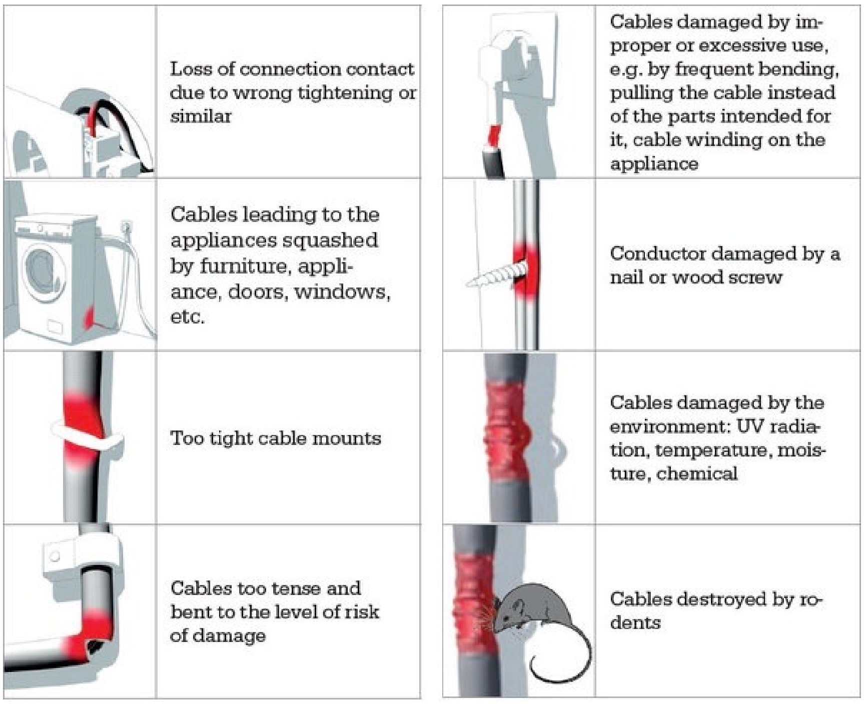

Figure 2.1: Illustration of fault causes[63]

According to the U.S. Fire Administration, *the leading specific items most often first ignited in residential building electrical fires are electrical wire, cable insulation (31%) and structural member or framing (18%). The leading factors that contribute to the ignition of residential building electrical fires are other electrical failure, malfunction (43%), unspecified short-circuit arc (23%), and short-circuit arc from defective, worn insulation (11%).*[3]

Arc faults are typically at the nominal current or just below, therefore it is difficult to detect using traditional protection devices. Small arcs may grow in time as damage ton insulation worsens. Arc faults are detected by using the fact that high frequency noise appears in the waveform as a result of arcing and breakdown of the fault current close to zero-crossing of the driving voltage.[8]

According to IEC62606, there are three types of arc faults which are earth arc faults, parallel arc faults and series arc faults. These are introduced in more detail in the following sections.

2.1.1 Parallel Arc Fault

Parallel arc fault is defined as an arc fault where the arc current is flowing between active conductors in parallel with the load of the circuit.[6] The total current in the circuit increases depending on load impedance and arc impedance. There is no current flowing through the earth conductor. Residual current devices(RCDs), explained in more detail in section 2.3.2, do not provide protection for parallel arc faults, whereas, miniature circuit breakers(MCBs) might provide protection depending on load and arc impedance whether tripping conditions such as current/time are presented or not. Parallel arc is illustrated in Figure 2.2.

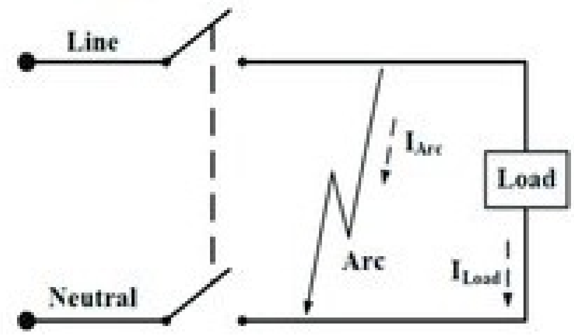

Figure 2.2: Parallel Arc[10]

2.1.2 Ground Arc Fault

Ground arc fault is defined as an arc fault where the current is flowing from an active conductor to earth. [6] Ground faults are sometimes called residual current faults or earth leakage faults. Typically, they have much lower current compared to short circuit. A ground fault can happen in permanent wiring, in an appliance itself, or in the cord of appliance. For some ground faults AFDD can trip if other protection is not installed. For instance, if an RCD(see section 2.3.2) is not installed, an AFDD trips for higher nominal currents than RCD, or some cases where overcurrent protection devices do not provide protection because the impedance of the faulty circuit can be too high.[61] More detailed information about device characteristics of AFDD are given in the section 2.3.3. Ground arc is illustrated in Figure 2.3.

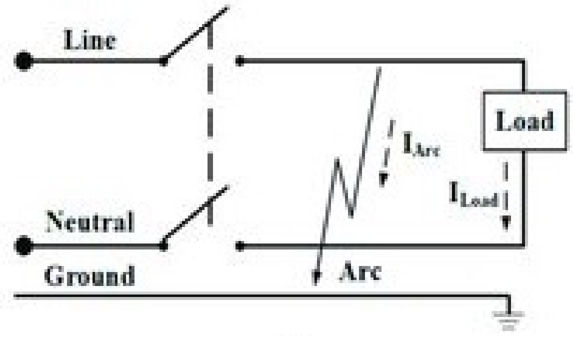

Figure 2.3: Ground Arc[10]

2.1.3 Series Arc Fault

IEC62606 standards define series arc faults as arc faults where the current is flowing through the load(s) of the final circuit protected by an AFDD.[6] Arc voltage is in series with the load and it behaves like an electrical component serially connected to acircuit. Therefore, load current is different than expected current in normal operation of load.

Additionally, it behaves as a voltage divider and causes a decrease on the load voltage. Loads that provide a constant power output, such as power supplies, compensate this with an increase in current consumption. Total power dissipation in the circuit increases. RCDs and MCBs cannot provide protection for such a fault[9]. Therefore, arc fault detection device(AFDD) is required to fill this protection gap in addition to MCBs and RCDs. Series arc is illustrated in Figure 2.4.

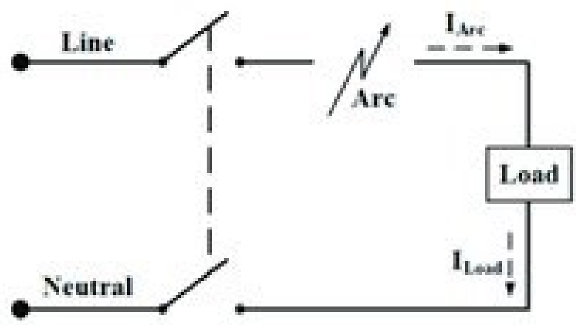

Figure 2.4: Series Arc[10]

2.2 Arc Fault Characteristics

According to a US patent filed by Siemens[12], during the time that the arc is conducting current, it produces wide-band, high-frequency noise ranging from about 10kHz to 1GHz. The inventor of [12], also stated that the resulting characteristic pattern of high-frequency noise with synchronous gaps is unique to arcing and therefore an algorithm for analyzing repetitive patterns in the amplitude of the noise can be used to detect arcing.

Beacuse the noise generated by arcing is wide-band and reaches frequencies up to 1GHz, any informative combination of frequency spectrum can be used to detect arcing. However, there are some advantages of using a bandwidth of 1 to 50MHz, which are clearly stated in [12] and the advantages are listed as follows:

1. Household appliances are intentionally designed to minimize the noise above 1MHz since it can interfere with radio broadcasts. Therefore, noise generated by load itself is minimal above 1MHz so that it is suitable to use in the detection of an arc fault.[12]

2. Loading effects of devices connected to the line can be presented. Therefore arc noise signal is attenuated. Power distribution lines behaves as transmission lines at high frequencies. Other devices that are supplied by the same power distribution line are inductively isolated from the power distribution line by their main cords and internal wires, which limits the amount of attenuation they can produce. Using a frequency band of 1 to 50MHz will eliminate the possible effects of these loading effects.[12]

Noise generated by the arc fault appears on both the line voltage and load current solely when arcing conducts. Amplitude of the noise is exactly 0 as the arc extinguishes and reignites at zero-cross section of line voltage,i.e every half-cycle of the line frequency. This is the reason that synchronous gaps are observed on the noise generated by an arcing. For resistive loads, arc voltage is in phase with the line voltage. Therefore, these gaps occurs simultaneously with the zero-crossing of line voltage. For reactive loads, arc voltage and the gaps may shift in phase up to plus or minus 90° depending on the line voltage.[12] Reactance of the load in series with the arc is determinative if the gaps occur simultaneously with the zero-crossings of line voltage. However, regardless of phase shifts, gaps in the noise generated by arcing are equal in to half of the line frequency cycle.

2.3 Protection Devices

Protection devices in residential power systems are introduced in following sections. There are three different subcategories. Devices with same functionality have different naming in different places and they are often combined.

2.3.1 Miniature Circuit Breaker(MCB) and Moulded Case Circuit Breaker (MCCB)

Miniature circuit breakers(MCB) and moulded case circuit breakers(MCCB) protect residential power distribution lines against overloads and short circuits. Additional to fuses that are used for protection as early as 1864 [13], MCBs bring a switching function. MCBs often work with thermal or thermal-magnetic principles whereas recent MCCBs offer higher rated currents and employ electronic tripping units and adjustable breaker characteristics. MCBs and MCCBs also do not require replacement after tripping if their rated capabilities are not exceeded. They are the most common protection devices employed on the world. There are three types of MCBs: Type B, C and D. Type B trips at 3 to 5 times of the rated current and are used for domestic and commercial installations having little or no switching surges, whereas type C trips at 5 to 10 times the rated current and is designed for general use in commercial or industrial application with a greater use of fluorescent lighting and motors which may cause nuisance tripping of type B breakers. Type D breakers trip 10 to 20 times the rated current and are suitable for industrial applications where transformers, large motors, welding cause high inrush switching surges.

Figure 2.5: Miniature Circuit Breaker[17]

2.3.2 Residual Current Device(RCD) or Residual Current Circuit Breaker (RCCB)

While the exact origins of the RCD are unknown, the technology appeared on the market around the 1950s and was used by utility companies against 'energy theft' in order to prevent usage of current from phase to earth instead of phase to neutral.[14] Operating principle of RCDs is given in Figure 2.6.

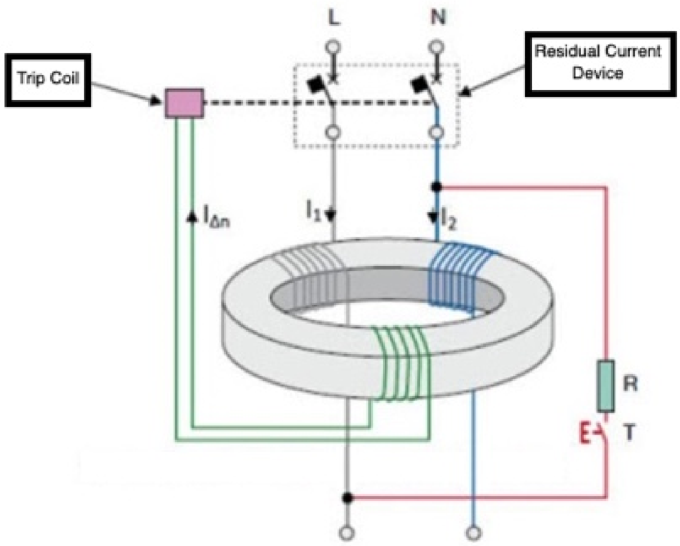

Figure 2.6: RCD Working Principle[16]

Their working principle is based on monitoring the wiring continuously to detect leakages. When there is no fault, vectorial sum of the currents on live wire and neutral wire($I_1 + I_2$ shown in Figure 2.6) is equal to 0. In case of a fault, if $I_1 + I_2$ exceeds the rated residual operating current $I_{\delta n}$, regarding circuit at the secondary side of toroid sends a trip signal to residual current device.[16] Typically, 5mA (30mA in some places) is considered as an imbalance which can cause fatal outcomes in case of an electrocution.

An RCD detects the imbalance in the live wire and neutral wire, however, it does not provide protection against overload and short circuit as an MCB does, except in the special case of a short circuit from live to ground. There are available products on the market that often integrate these two together. These are called:

1. In Europe: Residual current circuit breaker with overcurrent protection(RCBO)
2. In the United States and Canada: Ground fault circuit interrupter(GFCI), or ground fault interrupter (GFI), or an appliance leakage current interrupter (ALCI).
3. In Australia: Safety switch or RCD

There are also outlet versions of GFCIs in US market, which is simply a GFCI included in a dual socket. Unlike a GFCI, GFCI outlets only provide protection for devices plugged into it or other outlets downstream from the outlet. [18]

2.3.3 Arc Fault Detection Device (AFDD)

The arc fault detection device shown in Figure 2.7, also known as an arc fault circuit interrupter is a circuit breaker that trips when an electric arc is detected in the circuit in order to prevent electrical fires.[20]

Figure 2.7: Arc fault detection device[19]

IEC62606:2013 clearly states that, during a series arc fault, there is no leakage to ground. Therefore, RCDs cannot detect such a fault. Furthermore, load current is reduced because of the impedance of a series arc; therefore, the current level is also lower than the threshold of the circuit breakers and fuses. When arcing occurs between a phase and a neutral conductor, the only limitation to the current is the impedance of installation. Conventional circuit breakers are not intended to trip for such situations.

Additionally, AFDDs have to distinguish between hazardous and non-hazardous arcs that occur under the normal operation of some electrical loads such as power drills or brushed motors. This is a critical requirement because over-sensitive design of an AFDD can be prone to nuisance tripping. Nuisance tripping has economical impacts whereas under sensitive design comes with lack of protection. This trade-off has to be taken into consideration while designing an arc fault detection algorithm.

IEC62606:2013 contains general requirements for arc fault detection devices including its performance, testing and characteristics. According to [6], limit values of break time for $U_n = 230V$ AFDDs at low arc currents up to 63A are given as follows.

Test arc current (r.m.s values)	2.5A	5A	10A	16A	32A	63A
Maximum break time	1s	0.5s	0.25s	0.15s	0.12s	0.12s

Table 2.1: Limit values of break time for Un = 230 V AFDDs

State of the art devices generally employ a high frequency analog front-end monitoring circuit combined with a microprocessor. However, introduction of new kinds of loads and new, cheap and noisy components cause robustness issues, thereby causing either economic loss or lack of protection. Although there are already many different products offered in different markets, advancements in signal processing devices and classification algorithms makes new detection algorithms feasible.

This thesis is intended to propose a novel detection algorithm chain using high frequency direct digitization to obtain current waveform of loads and then to propose a detection algorithm using machine learning algorithms such as support vector machine, k-nearest neighbor and others, which are investigated in more detail in Chapter 4. Such an approach has some advantages compared to existing algorithms, namely:

1. Although machine learning algorithms can be computationally expensive in training phase, obtained classifiers do not require high processing power and obtained classifiers are very simple to implement on a microprocessor. Therefore, obtained classifiers can be even used together with the existing solutions in order to increase robustness and efficiency.

2. Existing analog front ends are designed using expert knowledge, and therefore they are designed based on heuristic data. Implemented methodologies in the proposed arc fault detection technique for extracting features in Section 4.2.1 offer a tool to automatically extract informative subsets of a given dataset which is simply frequency domain representation of current waveforms.

3. Since all the processing, training and testing is implemented on both MATLAB and Python environments, it is simple to modify the algorithms whereas designing an analog front-end takes much more time.

2.4 Residential Loads and Characteristics

A systematic approach needs to followed in order to investigate the maximum number of load categories and their effects, because the performance of the algorithm designed in Chapter 4 is proportional to the coverage of the feature space to real life situation. Namely, if feature space constructed from measurements covers a small portion of real life situation, algorithm would have satisfactory performance metrics in the design phase, however, it would not perform as expected after implementation. In [21] and [22], categorization of household loads was established in a systematic manner, and a similar methodology with minor modifications is also applied in this thesis. There are numerous home appliances from different manufacturers on the market which have different characteristics even if they are same types of appliances. Selection of loads and test methodology is introduced in more detail in chapter 3 . Load types given in Figure 2.8 are investigated in more detail in the following sections.

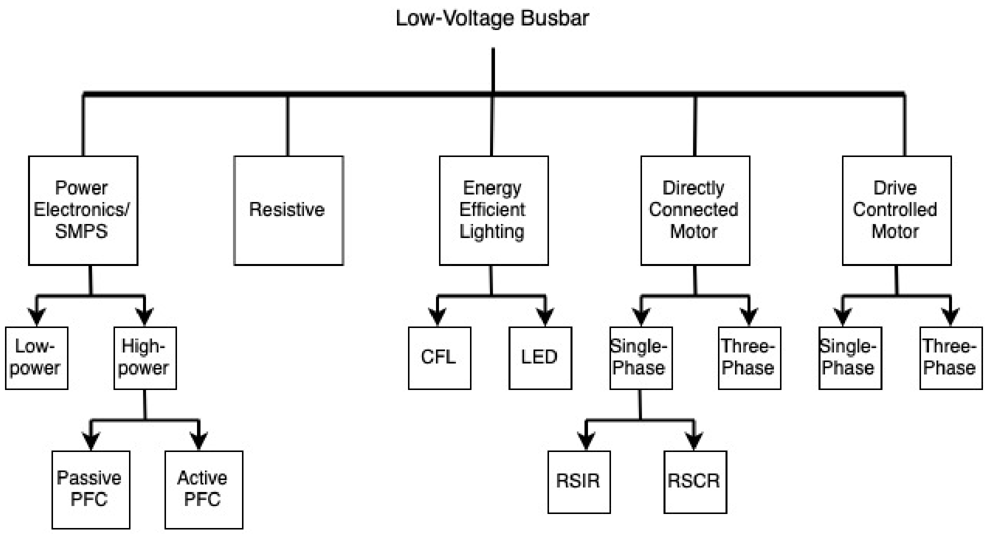

Figure 2.8: Residential Load Types

2.4.1 Resistive Loads

Devices such as general incandescent lamps, space and water heaters, electrical cookers, and similar loads are classified as resistive loads. These devices are assumed to act as ideal resistors.[22] Voltage and current waveforms are in phase. It is known that inrush current of resistive loads is small because current reaches steady-state instantaneously and does not exceed it.

2.4.2 Power Electronics/Switched Mode Power Supply

Numerous number of devices are susceptible to variations in voltage levels, and therefore, switched mode power supplies are used to supply a regulated dc voltage. Personal computers, televisions and monitors are widely used examples that utilize SMPSs. According to [25], the total energy consumption by household appliances grew 1.7% every year on average from 1970 to 2013 in the United Kingdom and consumer electronics that generally utilize SMPSs have the biggest residential consumption with an estimation of 1868 ktoe.

According to related standards[23], switched mode power supplies are divided into three subcategories, as follows:

1. No Power Factor Correlation(n-PFC): SMPS with a rated power smaller than 75 W do not have to meet harmonic legislation. Therefore, they do not need to include a power factor correlation circuit.

2. Active Power Factor Correlation(a-PFC): Some SMPS with rated power larger than 75 W include an additional dc-dc converter circuit in order to shape the input current to have appropriate sinusoidal waveform in-phase with the supply voltage. [22]

3. Passive Power Factor Correlation(p-PFC): p-PFC utilizes a large inductor in the path of current transmission since inductors resist change in current, therefore resulting a waveform that is smoother and has less harmonic content.

P-PFC has a wider usage than a-PFC because of its easier implementation and significantly lower cost. However, usage of a-PFC increases as the cost of components decreases steadily.

2.4.3 Energy Efficient Lighting

Light emitting diodes(LEDs) and compact fluorescent lamps (CFLs) are the most widely used loads that falls into this category. Due to the economic benefits of LEDs compared to general incandescent bulbs(GILs), energy efficient lighting has been constantly gaining popularity.

According to the International Energy Agency, LED lighting sales expeienced terrific growth in recent years. In 2011, LEDs and fluorescent lamps were formed 40% of total sales with 1% and 39% of the market share respectively. However, in 2018, these two formed 82% of total sales, with LEDs at 40% and fluorescent lamps at 42% of total sales. It is expected that this trend will continue and LEDs will capture the 80% of total sales and fluorescent lighting will capture 20% of total sales by the end of 2030. [26]

Figure 2.9 shows the North America industrial and commercial LED lighting market share in USD millions from 2012 to 2022.

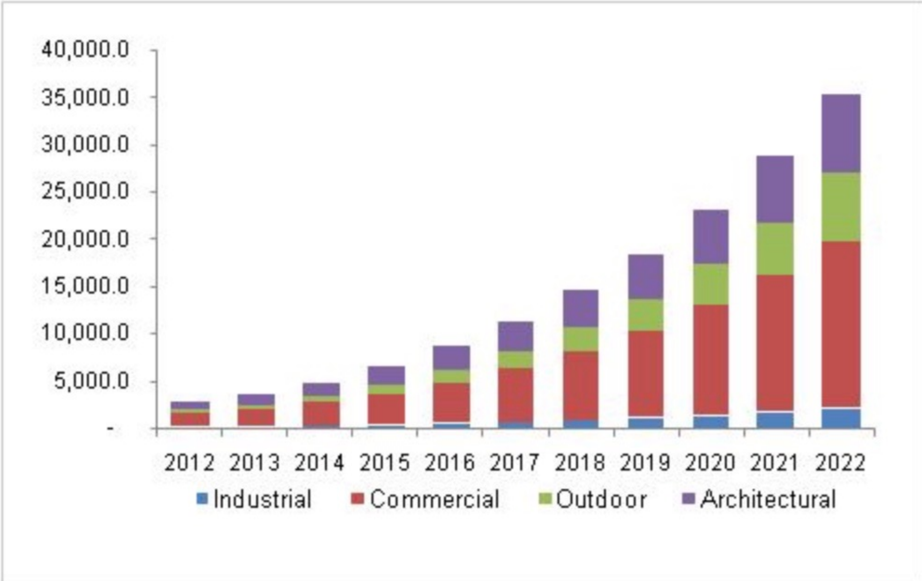

Figure 2.9: North America Industrial and Commercial LED Lighting Market Share, 2012-2022 (USD millions)[27]

CFLs of 8, 11 and 18 W are widely replaced with 40, 60 , 100 W GILs accordingly because of similar output lumens together with much less power consumption.[22] Because of nonlinearity, harmonics are introduced to the supply system by CFLs.

2.4.4 Single Phase Induction Motor

Figure 2.10 shows a widely used single phase induction motor.

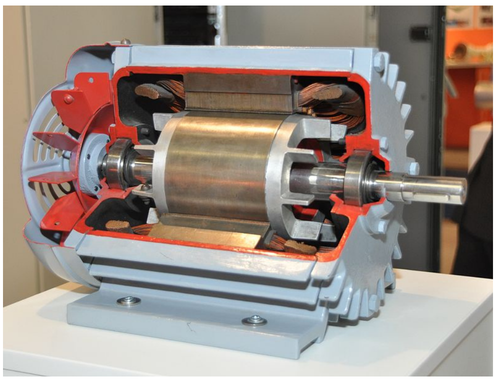

Figure 2.10: Single Phase Induction Motor[28]

Single phase induction motors are one the most common types of loads found in devices such as refrigerators, dishwashers, freezers, and air conditioners, which all contains SPIMs. The reason they are called 'single phase' is because they work with single phase residential AC supply. Three phase motors are not included in this thesis since they are outside the scope of this research. Single phase induction motors are used when a fixed speed is needed. SPIM has a simple design.

2.4.5 Universal Motor

The universal motor is one of the most widely used components found in numerous appliances, such as blenders, vacuum cleaners, hair dryers, drill machine, sanders, and others. Another use is for applications that require speed control. Universal motors can operate with AC and DC. The size of universal motors is relatively small compared to AC motors that operate on the same frequency. The main features of universal motors are its capability of working at high speeds, its relatively high start torque, and its small size and compact design.[21] The disadvantages posed by the universal motors is the acoustic and electromagnetic noise generated by them.

Figure 2.11: Universal Motor

Chapter 3

AFDD Tests and Measurements

In this chapter, European test procedure and equipment used in the measurements are presented. Tested residential loads are introduced. Measured loads are analyzed in time domain and frequency domain. In order to analyze the arc fault events, different approaches can be employed. In this thesis, real measurements are employed because of the reliability of measurement procedures and guaranteed results.

3.1 European Standards Test Procedure

IEC 62606:2013 are the standards that cover the general requirements of arc fault detection devices including performance, functionality, and testing. To generate an arc fault, either a carbonized cable specimen or an arc generator can be used. These two are used for the same purposes, however, there are some minor differences between them in terms of testing methodology and specifications which are explained in more detail in subsequent sections. For instance, energy generated by a cable specimen is 2.5 times the energy generated by an arc generator.Therefore the maximum break time values in Table 2.1 need to be also 2.5 times faster which is shown in Table 3.1.

Test arc current (r.m.s values)	2.5A	5A	10A	16A	32A	63A
Maximum break time	0.4s	0.2s	0.1s	0.06s	0.048s	0.048s

Table 3.1: Limit values of break time for Un = 230 V AFDDs if arc generator is used

Cable specimen is used in the experiments carried out in this thesis and they are prepared according to IEC:62606:2013 standards. Preparation of carbonated cable specimen is covered in Section 3.1.1.

3.1.1 Cable Specimen

Illustration of cable specimen is shown in Figure 3.1.

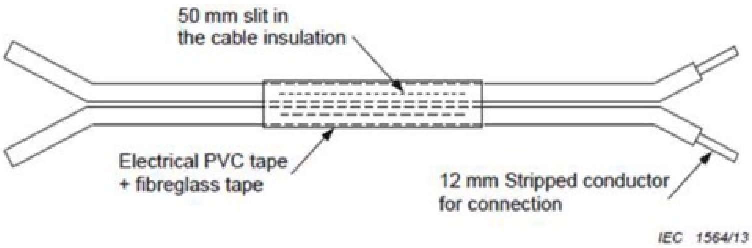

Figure 3.1: Cable Specimen[6]

A cable that contains two separate conductors with a cross-sectional area of 1.5mm^2, that could be tied together with a suitable tape or a cable that contains parallel conductors can be used for experiments. There are three specific types of cables stated in the standard. OMY 2x1, 50mm^2 cables were used for experiments in this thesis.

The methodology is stated in the standards as follows[6]:

1. The material and geometry of a cable has to be taken into consideration to perform a satisfactory carbonization between the conductors and to initiate arcing by applying the rated voltage.
2. The cable specimen needs to be prepared longer than 200mm and inner wires need to each be separated each end of the specimen for 25mm.
3. Outer insulation needs to be slit with a depth of 50mm to expose the inner conductors. Strands in the inner conductors need to remain undamaged.
4. Two layers of tape have to be wrapped around the slit. The first layer is an electrical grade black PVC tape. The second layer is a fiberglass tape.
5. The cable specimen has to be connected to the test bench to the further end to the slit. Inner insulation of conductors has to be stripped approximately 12 mm in this end.

After the cable is prepared, it will be conditioned to form a carbonized conductive path through insulation between two inner conductors as follows:

1. The cable specimen needs to be connected to a circuit that provides 30mA short circuit current and an open circuit voltage at least 7kV. This connection needs to be sustained either for 10 seconds or until the smoking stops.
2. As the last step, it is needed to connect the cable specimen to a circuit that provides 300mA short circuit current at a voltage of at least 2 kV or adequate to current flows through it. This step needs to be applied at least one minute or until the smoke stops.

The carbonized cable specimen is approved to be complete, if it has a resistance value equal to the resistance of 100W/230V incandescent lamp or an equivalent resistance value of path draws 0.3A at 120V. Completed cable specimen according procedure above is shown in Figure 3.2. These are used in measurements.

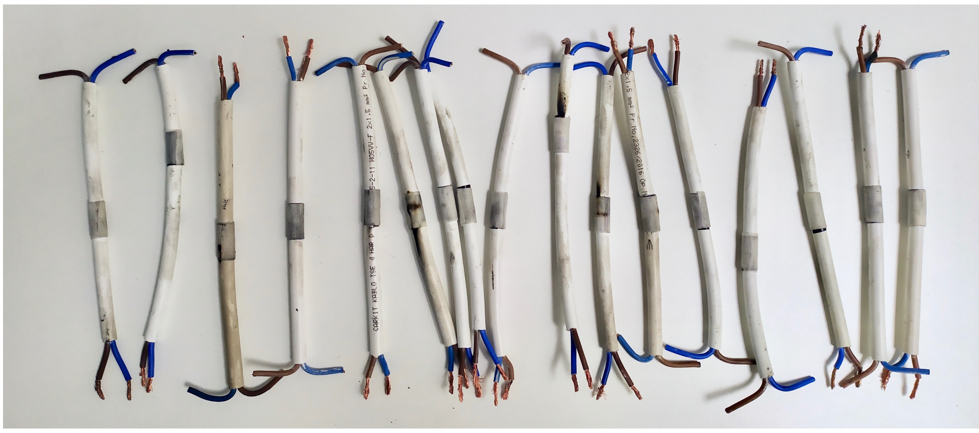

Figure 3.2: Prepared Cable Specimen

3.1.2 Arc generator

An arc generator is an apparatus that generates an arc by moving an electrode to another stationary one. One electrode is made of carbon-graphite with a diameter of 6mm $\pm$ 0.5mm and the other one is made of copper. With an appropriate distance between them, it generates a consistent arc between two electrodes. Figure 3.3 demonstrates an arc generator.

Conditions of electrodes are also important. It can be necessary to replace or to sharpen them eventually.

Since the energy provided by an arc generator is 2.5 times less than the carbonized cable specimen, an AFDD shall de-energize the circuit 2.5 times faster if arc generator is used instead of cable specimen.[6]

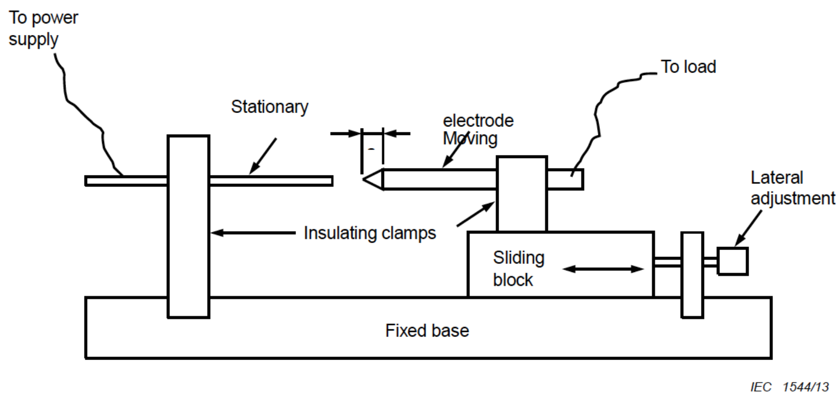

Figure 3.3: Arc generator[6]

3.1.3 Experiments Described in Standards

Experiments to verify the functionality of an AFDD are described in [6]. Following sections address regarding experiments in context of this thesis.

3.1.3.1 Test for series arc fault

Circuit diagram for series arc faults using a carbonized cable specimen prepared according to methodology in Section 3.1.1 is presented in Figure 3.4. All the measurements are performed at the rated voltage of an AFDD which is 220V in Europe.

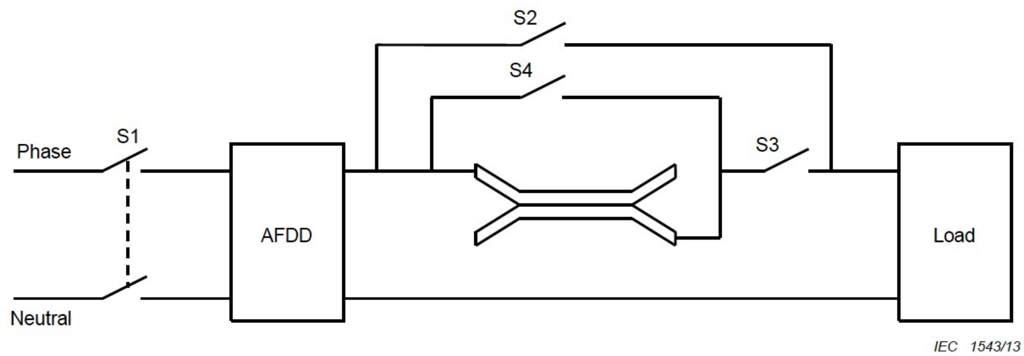

Figure 3.4: Test circuit for series arc fault test[6]

Tests cover the possible situations that can be faced in case of a series arc fault, i.e, sudden appearance of series arc fault, inserting a load with series arc fault and closing on series arc fault. For all the cases, AFDD shall not exceed the tripping values given in Table

2.1. Following tests are intended for verifying correct operation of AFDD in case of a series arc fault. The same methodology can be used without an AFDD in order to observe the current waveform.

3.1.3.1.1 Inserting A Load with Series Arc Test switches S3 and S4 need to be in open position whereas the switch S1 and AFDD need to be in closed position. A resistive load should be employed to adjust the test arc current to lowest arc current value that is provided in table 2.1. Then, test switch S2 is opened. While test switch S1 and AFDD are in closed position and test switches S3 and S4 are in open positions, switch S3 is abruptly closed to feed the load with a series arc fault. It is needed to repeat the test for the rated value of AFDD and break time needs to be in the given values in table 2.1 to verify the correct operation. [6]

3.1.3.1.2 Sudden Appearance of Series Arc in the Circuit All the switches S1, S2, S4 and the AFDD needs to be in the closed position. Test arc current needs to be set from lowest arc current value up to the rated current of AFDD by a resistive load. Then, the test switch S2 is opened. As a last step, test switch S3 is abruptly closed in order to insert the load in series with the carbonized cable specimen. Break time needs to be in the given values in Table 2.1 to verify the correct operation.[6]

3.1.3.1.3 Closing on Series Arc Test switches S1, S3 and AFDD need to be in closed position. Test arc current needs to be set to the lowest arc current in table 2.1 by a resistive load. Test switch S1 and S2 need to be opened one after another respectively. While switches S1 and S4 are in open position, test switch S1 is abruptly closed in order to insert the load with a series arc fault. Break time needs to be in the given values in Table 2.1 to verify the correct operation. It is needed to repeat the test for the rated value of AFDD.[6]

3.1.3.2 Tests for Parallel Arc

There are two types of test that are tests for parallel arc with limited current and parallel arc cable cutting test which explained in detail in following sections.

3.1.3.2.1 Parallel Arc with Limited Current An AFDD needs to clear an arcing fault if number of half-cycles of arcing shown in Table 3.2 occurs within a period of 0.5s.[6]

Test arc current (r.m.s values)	75A	100A	150A	200A	300A	500A
N	12	10	8	8	8	8
Notes: 1) Test arc current values are prospective currents before arcing in the circuit 2) N is the number of half cycles at the rated frequency.						

Table 3.2: Maximum allowed number of arcing half-cycles within 0.5 s for AFDDs

In order to accomplish a test according to these requirements, an arcing half-cycle is considered to be all the current signals occuring in a period of 10ms for a device rated 50Hz(8.3ms for a device rated at 60Hz). In that period of time, some current can be observed but it shouldn't be continuous. In explanation, before and after each period of current flow, it should be an instance with very reduced current or no current. Very reduced current is defined as the current with amplitude less than 5% of available current or current that continues for no more than 5% of the duration of half cycle[6]. This can last for several half cycles or a portion of a half cycle. Fault current level needs to be 75A and 100A. Figure 3.5 shows the test circuit for parallel arc with limited current.

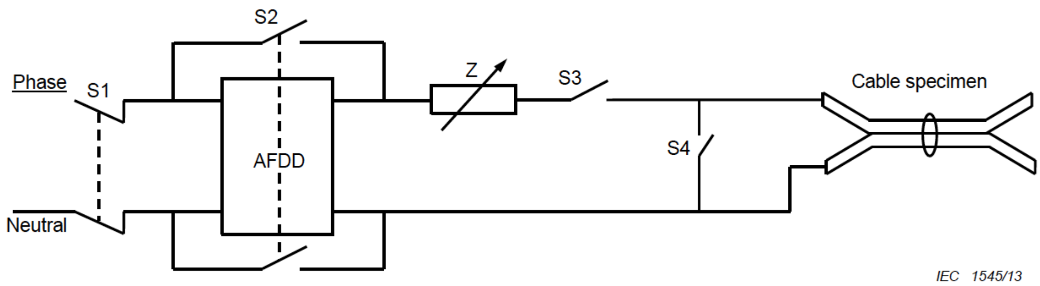

Figure 3.5: Test circuit for parallel arc fault test[6]

Line impedance Z, shown in figure 3.5, is used to adjust the test current to 75A when all switches are in closed position. Then, switches S2, S3 and S4 will be opened, AFDD and the switch S1 will be closed.

Test needs to be repeated with a current adjusted to 100A using the same methodology above.

Correct operation of AFDD is verified if it clears an arc fault and if the number of half-cycles of arcing mentioned in table 3.2 occurs within 0.5s. 0.5s period starts with the first arcing half-cycle.

3.1.3.2.2 Parallel Arc Cable Cutting There are minor differences in methodology between the parallel arc cable cutting test and parallel arc with limited current test. The number of half-cycles, definition of half-cycle and requirements about currents are the same. However, test circuit has slight changes as it can be seen from Figure 3.6.

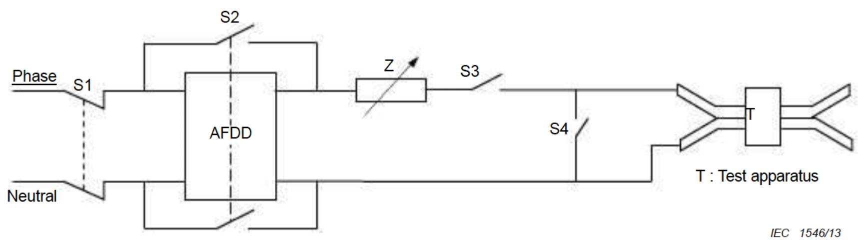

Figure 3.6: Test circuit for parallel arc cable cutting test[6]

Another difference is the test apparatus, T, which is shown in Figure 3.7.

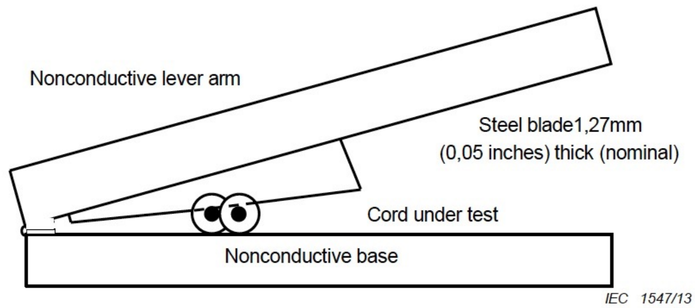

Figure 3.7: Test Apparatus[6]

The steel blade should be 3 mm thick with approximate dimensions of 32 mm by 140 mm for a 230 V AFDD. Using the apparatus shown above, the blade needs to be positioned in order to make a solid contact with one conductor and arcing contact is made with the second conductor.[6]

Line impedance Z, shown in Figure 3.6, is used to adjust the test current when all switches are in closed position. While test switches S1 and S3 are in the closed position, the blade is forced to cut through the insulation therefore, blade makes a solid contact with one conductor and then point contact with the other conductor.[6]

AFDD needs to clear an arc fault according to values given in Table 3.2 which occur in a period of 0.5s.

3.2 Measurement Setup

The test circuit is adopted from section 3.1.3.1 with minor differences to capture the current waveform. Figure 3.8 shows the test circuit used in measurements. Carbonized cable specimen is prepared according to Section 3.1.1.

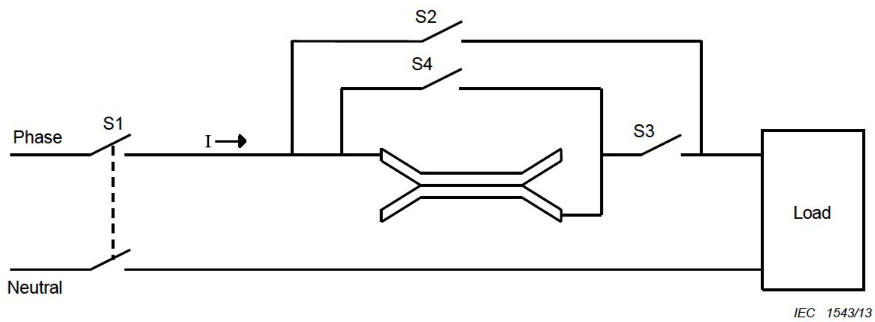

Figure 3.8: Test Circuit

High voltage power sourced used for preparing the cable specimen and the box for easy replacement of cable specimen and safety are shown in Figure 3.9.

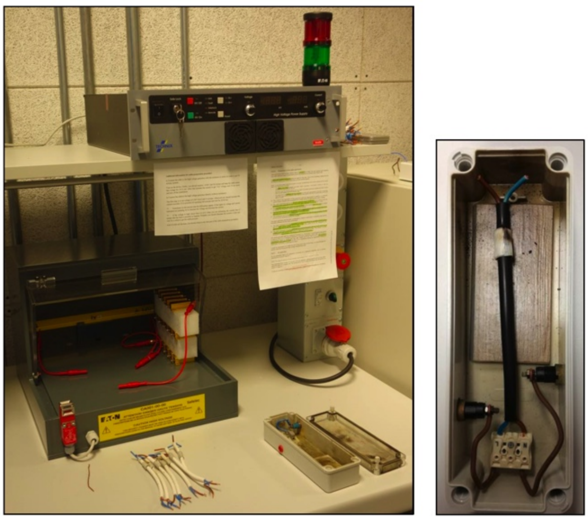

Figure 3.9: High voltage power source and safety equipment

Two different set of measurements were used to design the algorithm proposed in Chapter 4. The main difference is the sampling frequency of high speed data acquisition devices. In order to make a preliminary design and to investigate the feasibility of machine learning classifiers for AC arc fault detection, an oscilloscope with 5MHz sampling frequency was used to gather the data. Therefore, it covers a bandwidth up to 2.5MHz.

After preliminary work is completed, an oscilloscope with a higher sampling frequency is employed. Sampling frequency of oscilloscope is 62.5MHz, therefore, measurements covered a bandwidth of 31.25MHz which is sufficient to cover the required bandwidth to design a detection algorithm. This setup is more suitable for arc fault detection research since noise generated by arcing appears also in higher frequencies than 2.5MHz. Equipment list used in these experiments are given in Table 3.3.

Test Equipment	Specifications
Oscilloscope	Tektronix 5 Series Mixed Signal Oscilloscope (62.5 Msps)
Current Probe	Tektronix TCP0020 50 MHz 20A Current Probe
High Voltage Power Source	Technix - 10 kV, 300 mA
High Voltage Test Cage	Sefelec
Fume Extractor	Weller - WFE2ESKIT1

Table 3.3: Equipment list used in measurements

This measurement setup enhanced measurement capabilities since it covers a bandwidth up to 31.25MHz and gives more freedom to observe and manipulate data onboard. Also, oscilloscope used for the second dataset calculates fast fourier transform that can be either used to inspect the data onboard or to save it and use after. FFT computed by oscilloscope is only used for inspection purposes. All the signal processing work is completed in MATLAB which is explained in more detail in Chapter 4.

3.3 Measured Loads and Analysis

Table 3.4 shows the loads tested in the first dataset that measured with 5MHz sampling rate.

Measured Load Type	Load Name
SMPS(p-pfc)	Power Drill
SPIM(Directly Connected Motor)	Desk Fan
Resistive	Heater
Lighting	Fleurescent
Resistive	General Incandescent Lamp

Table 3.4: Measurements for first dataset

Since the measurements for first dataset were only used for preliminary work and investigation of feasibility, they are not explained in the following sections. Only measurements in the second dataset which are shown in Table 3.5 are explained.

Due to time and resource limitations, it was not possible to measure all different categories stated in Figure 2.8. However, most challenging loads in terms of arc fault detection were used in experiments. Loads are selected according to their usage statistics. Most sold loads in Czech market were chosen. Measured loads that are used in design of detection algorithm are given in Table 3.5.

Measured Load	Load Type
Heaters	Resistive Load
LED	Energy Efficient Lighting
Fluorescent Lamp	Energy Efficient Lighting
Cultivator	Universal Motor
Power Supply	Switched Mode Power Supply with a-PFC
Heaters with Dimmer	Resistive Load

Table 3.5: Measured Loads by Group

The same number of measurements were planned to be taken for arcing and normal working conditions of different loads. However, it is simpler to capture normal operation of a load since there is no need for preparation of the cable specimen. Therefore for some loads, number of measurements with an arc fault are less than normal operation.

3.3.1 Heaters

Resistive loads are the easiest load types for arc fault detection because of their consistent and unique characteristics. Additionally, they do not contribute to nuisance tripping. In the following table, measured resistive loads(heaters) and their combinations are given.

Measured Load (by name)	Power (W)	Current (A)	Number of Measurements
Sahara	400	1.739	42
Sahara+Eurom	900(400+500)	3,913	42
Sahara+Concept	1150(400+750)	5	42
Sahara+Tristar	1600(400+1200)	6.957	42
Tristar+Concept	2050(1200+850)	8.913	42
Eurom+Tristar+Concept	2550(500+800+1250)	11.087	42
Eurom+Sahara+Tristar +Concept	3350(500+400 +1200+1250)	14.565	42

Table 3.6: Resistive Loads used in measurements in details(Heaters)

Two different types of measurements are done for every combination of heaters. One is to observe current waveform in normal working condition and the other is to observe current waveform of the load with a series arc. Figure 3.10 and Figure 3.11 show the regarding waveforms of heater with the lowest power consumption(400W) and the combination of heaters with the highest power consumption(3350W). For the case with series arc fault, shoulders at the zero crossings are visible without further analysis. Also, it is observed that characteristics of a resistive load for arcing and non-arcing do not show remarkable changes based on the power consumption.

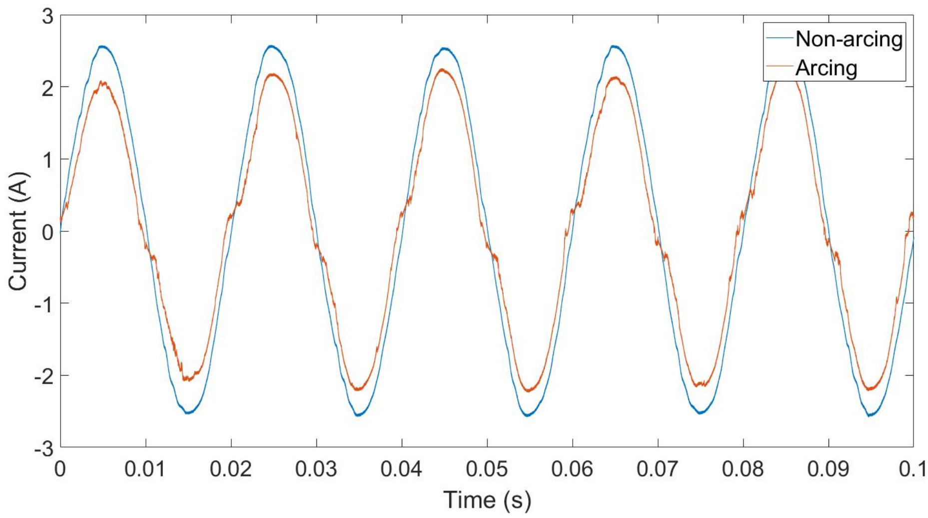

Figure 3.10: Current waveform of electrical heater(400W)(Red: Arcing, Blue: Non-arcing)

Figure 3.10 is the measurements of the heater with lowest power consumption, 400W. Figure 3.11 shows the measurement of combination of heaters with a power consumption of 3350W in total.

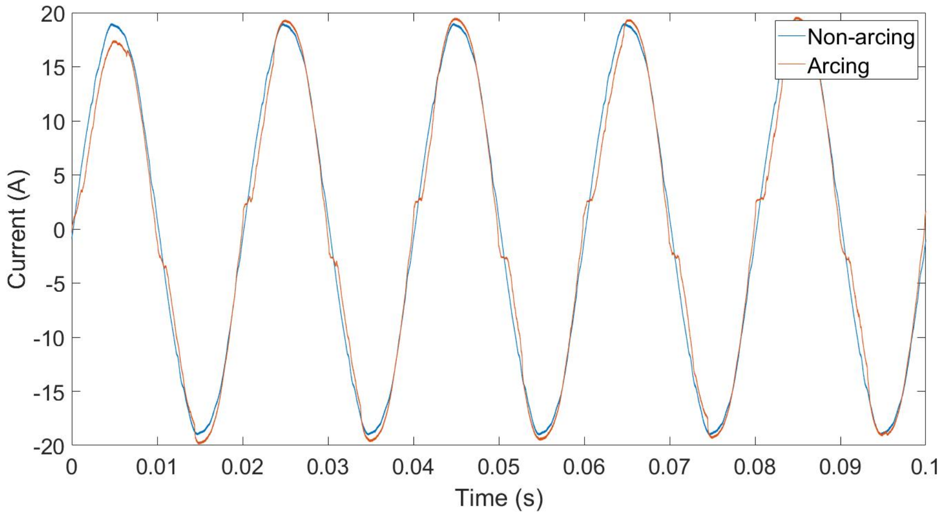

Figure 3.11: Current waveform of electrical heaters(3350W)(Red: Arcing, Blue: Non-arcing)

As it can be seen from Figure 3.10 and 3.11, when arc distinguishes and re-ignites around zero-crossings, it causes shoulders to appear in the waveform and these are visible by eye inspection without a need of further analysis.

3.3.2 Switched Mode Power Supply(SMPS)

Two different type of measurements are done for two different combination of SMPSs. One is to observe current waveform in normal working condition and the other is to observe current waveform of the load with a series arc. Loads are selected based on their availability on the market. Two different coolers that are widely used in computers are used in measurements. One type of measurements is done with one load which is Cooler Master(600W) and second type is combination of two SMPSs with a-PFC which are Cooler Master(600W) and Platimax(600W), therefore, total power is 1200W in total. If a series arc occurs, current waveform of a power supply with active power factor correlation differs remarkably compared to current waveform under normal working conditions. Figure 3.12 shows the current waveform of the 600W switched mode power supply with active power correlation with and without arcing whereas Figure 3.13 shows the current waveform of the 1200W switched mode power supply with active power correlation with and without arcing.

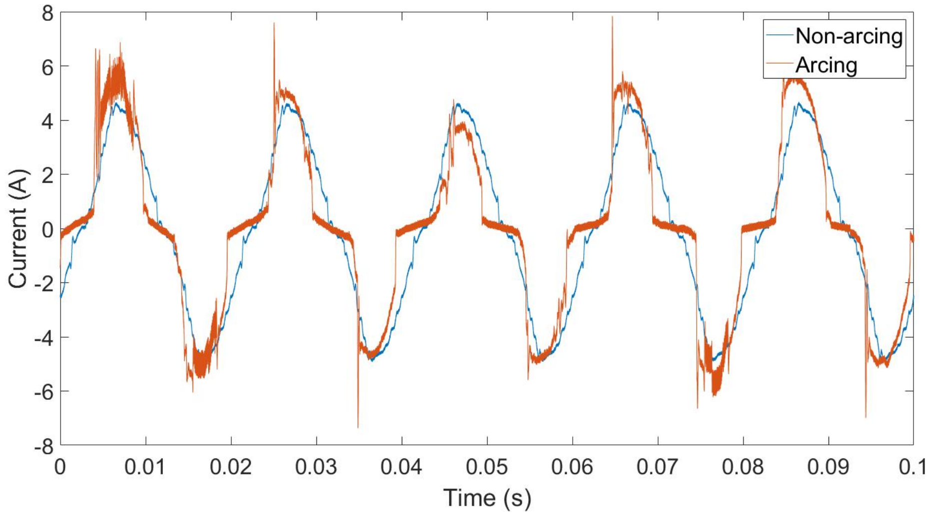

Figure 3.12: Current waveform of SMPS(Cooler Master) with A-pfc(600W) (Red: Arcing, Blue: Non-arcing)

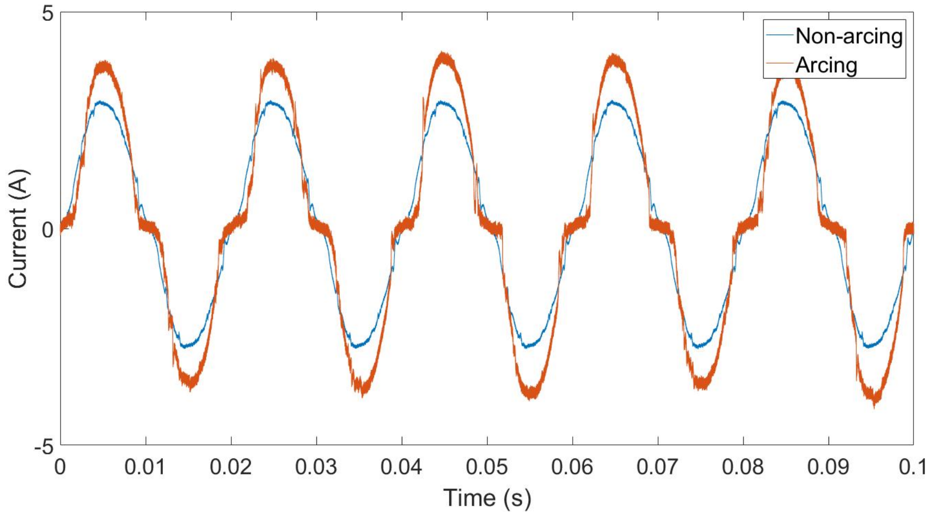

Figure 3.13: Current waveform of SMPSs(Cooler Master+Platimax) with A-pfc(1200W) (Red: Arcing, Blue: Non-arcing)

3.3.3 Universal Motor

Current waveform of a load(Cultivator) that contains an universal motor is given in Figure 3.14. Shoulders at zero-crossings are visible. Two different combination of loads with universal motors were tested. Loads are chosen by their usage statistics. These are widely used loads in Czech Republic and Europe.

First load is a single 1400W cultivator, and the second load is a combination of cultivator and vacuum cleaner(2600W in total).

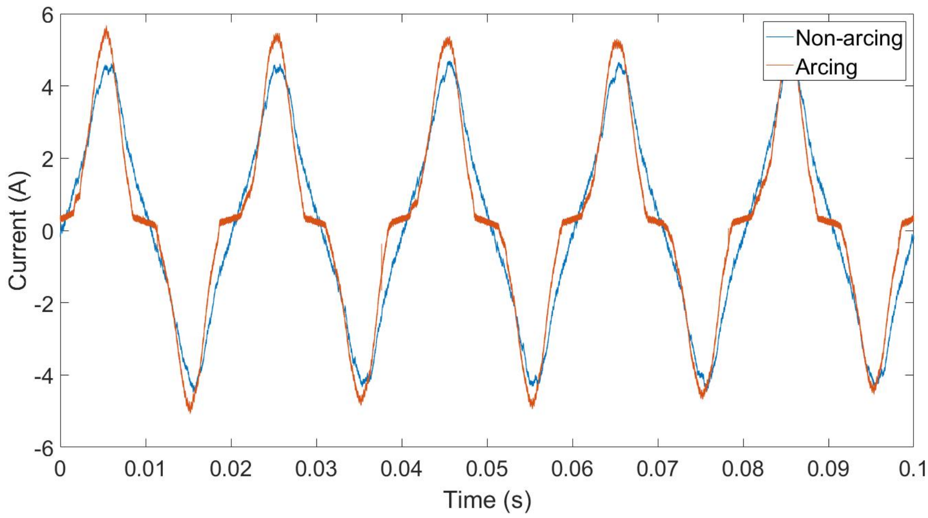

Figure 3.14: Current waveform of Universal Motor(Cultivator)(1400W)(Red: Arcing, Blue: Non-arcing)

In Figure 3.15, two loads which are cultivator and vacuum cleaner are combined. Shoulders near zero-crossing are more visible in Figure 3.16 compared to Figure 3.15. Therefore, it is concluded that increasing power demand of loads makes shoulders around zero-crossing more visible which is one of the very important characteristics of an arc fault. Shoulders near zero-crossings appear on most the loads, however they are much visible in universal motors, especially, when power consumption is relatively higher.

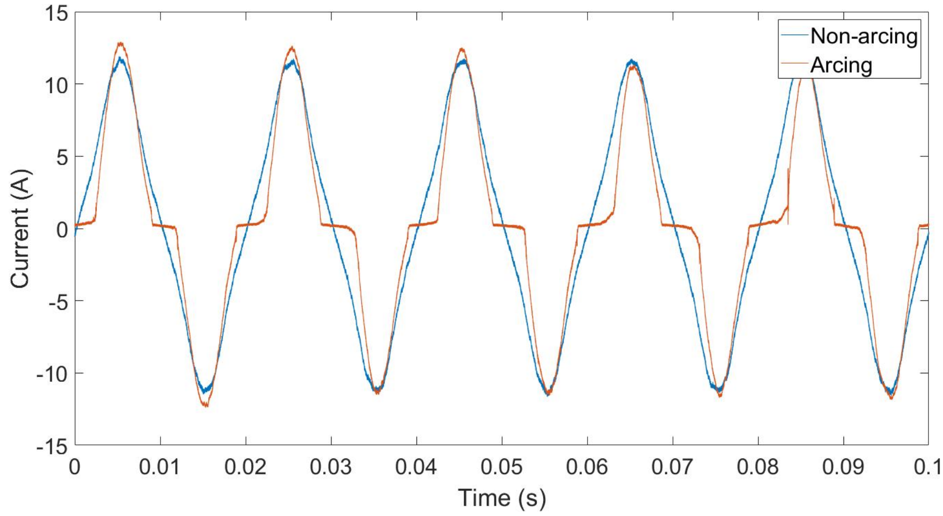

Figure 3.15: Current waveform of Universal Motor(Cultivator+Vacuum Cleaner)(2600W)(Red: Arcing, Blue: Non-arcing)

3.3.4 Light Emitting Diode(LED)

Measured combinations of LEDs and their specifications are given in Table 3.7.

Measured Load (by name)	Power (W)	Current (A)	Number of Measurements
1xLED MAX	44	0.19	39
3xLED MAX	132	0.57	39
5xLED MAX	220	0.95	39
10xLED MAX	440	1.9	39
10xLED MAX+Tristar Heater	1240	4	39

Table 3.7: Energy efficient lighting loads used in measurements in details(LEDs)

Measured LEDs has different cable lengths in the measurement circuit to inspect the effect of cable lengths. However, no difference is observed. Figure 3.16 shows the current waveform of LED under normal operation and with series arcing. Differences around zero-crossings are visible. When arcing distinguishes and reignites, current waveform is remarkably different. Also, there are peaks which do not happen under normal working conditions.

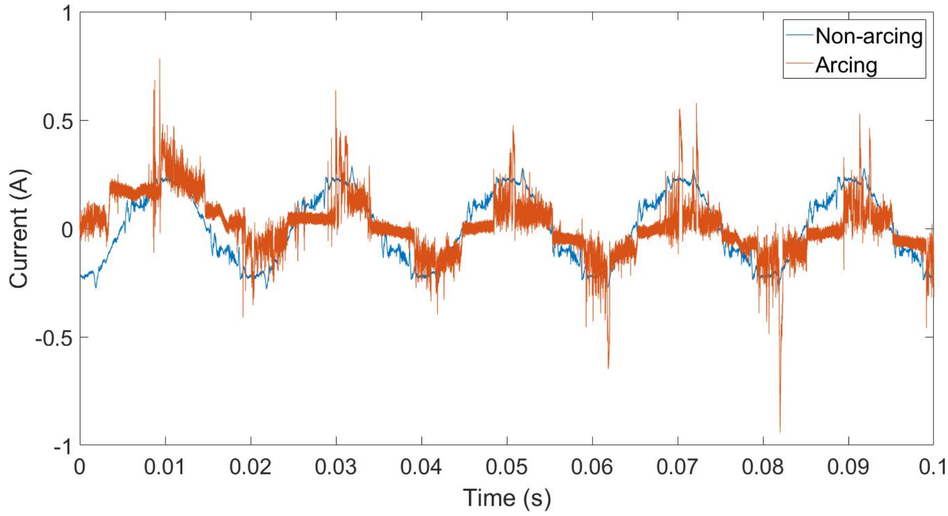

Figure 3.16: Current waveform of LED(44W)(Red: Arcing, Blue: Non-arcing)

The current waveform of 10 LED MAX(440W) and Tristar(800W) is given in Figure 3.17. When compared to Figure 3.16, current consumption is much higher.

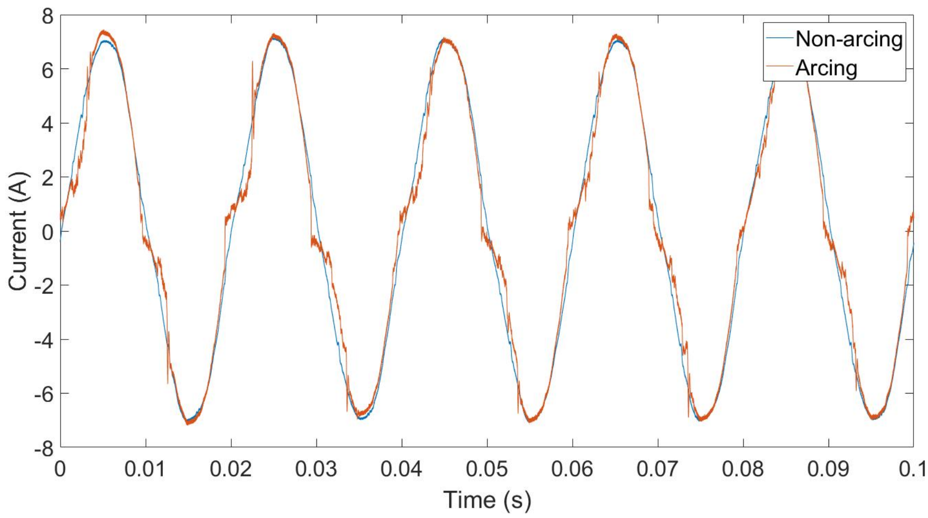

Figure 3.17: Current waveform of 10 LEDs+Tristar(1240W)(Red: Arcing, Blue: Non-arcing)

In Figure 3.16, current values are much lower, therefore, spikes and noise on the current waveform is more visible. It is concluded that arcing causes remarkable changes in current waveforms. That is also why LEDs have good classification rates that explained in Section 5 in detail.

3.3.5 Fluorescent Lamps

Fluorescent lamps are already noisy loads under normal working conditions. However, current waveform is remarkably different while series arcing fault.

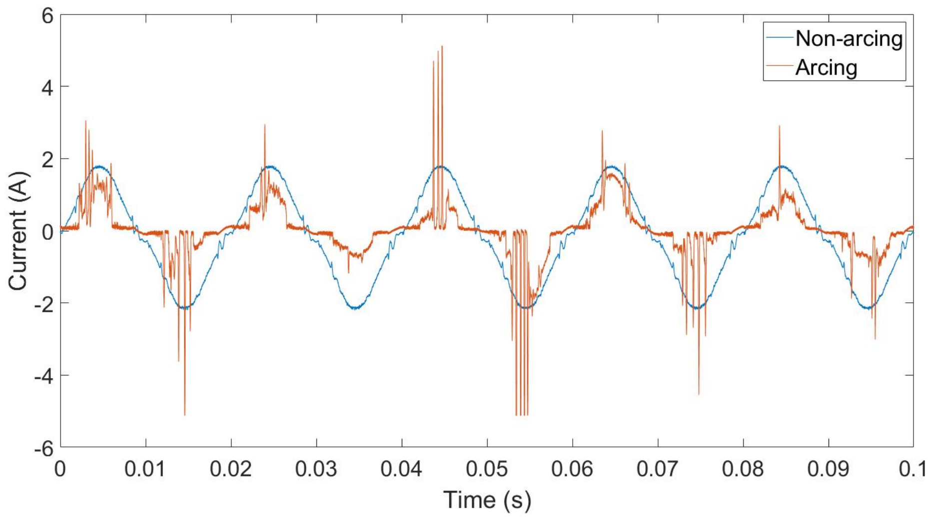

Figure 3.18: Current waveform of 10 Fluorescent Lamps(360W in total)(Red: Arcing, Blue: Non-arcing)

As it can be seen from Figure 3.18, there are visible spikes in the current waveform while arcing. This characteristic of fluorescent lamps makes them easier to classify when an arc fault occurs. These differences while arcing and non-arcing, can be observed in frequency domain too. Because of unique characteristics of fluorescent lamps, they are not problematic loads for classification algorithm.

3.3.6 Dimmer

Dimmers are important because they manipulate the voltage waveform applied to a load. A dimmer with 3 different settings is measured connected to a resistive load. Resistive load(heater) is dimmed with 0, 90 and 180 degree and current waveforms are measured for both arcing and under normal operation. Measurement results are as follows:

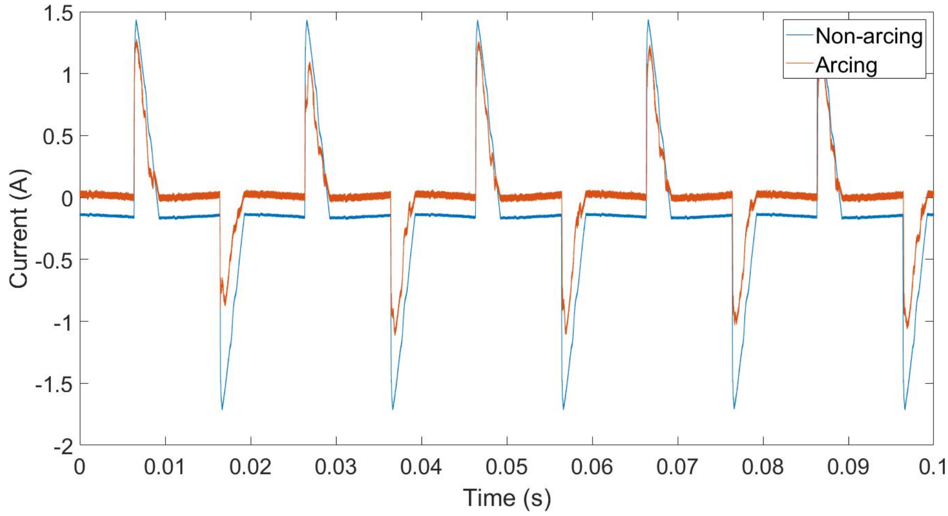

Figure 3.19: Current waveform of Dimmer(Minimum setting(0 degree)) with 400W resistive load(Heater)

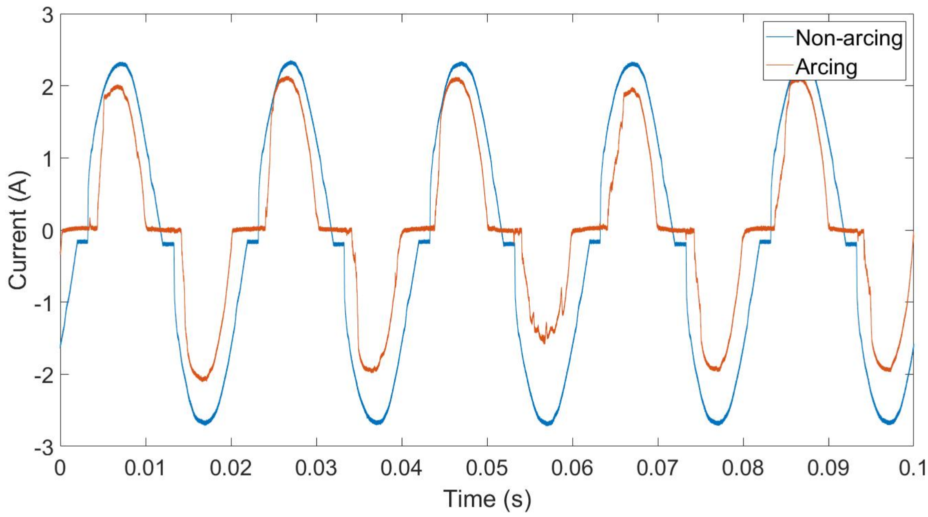

Figure 3.20: Current waveform of Dimmer(Maximum setting(180 degree)) with 400W resistive load(Heater)

Dimmers are one of the hardest loads to classify as arcing or non-arcing because dimmer is used to change current waveform of connected load to manipulate the energy consumed on the unit. However, difficulty of detection for dimmers is the way it manipulates the current waveform. Dimmed resistive load's current waveform and current waveform under normal operation of resistive loads are very similar as it can be seen from Figure 3.19 and Figure 3.20. Although there are differences while arcing and normal operation, general characteristics are similar and it is another reason that makes dimmers hard to classify while arcing.

Chapter **4**

Direct Digitization Approach

In this chapter, novel detection methodology that consists of direct digitization and supervised machine learning algorithms is introduced. Fundamentals of support vector machine is presented. Generation of feature space and feature selection algorithms are explained. Performance metrics are specified.

Existing arc fault detection methods are based on high frequency analog front end monitoring/conditioning circuit combined with a microprocessor. Analog front end is designed to monitor and to detect specific noise characteristics in different bands in frequency spectrum. The information used in analog front end is heuristically found and then implemented on the analog circuit.

Existing algorithms used in detection are well established and are optimized to detect most of the arc faults successfully. Also, they are designed in a way to distinguish between hazardous and non-hazardous sparks in the circuit. However, introduction of new kinds of loads, cheap and noisy components or power line communication devices cause robustness issues. Robustness issues result in either nuisance tripping that causes economical loss because the circuit is de-energized unnecessarily, whereas not tripping when it is necessary causes lack of protection.

Price of high speed acquisition devices are decreasing constantly in recent years and they are expected to be cheaper in the future. In the light of this knowledge, it becomes feasible to employ direct digitization instead of analog front end since the same information used in analog circuits is also achievable by direct digitization. Using direct digitization has advantages compared to analog front end. It is easier to modify detection algorithms and it gives more freedom on signal processing. Since more types of information can be gathered using direct digitization, it allows a various number of algorithms to implement for detection. However, analog circuit based solutions are much cheaper to produce therefore, analog front end is more feasible as of today for industrial usage.

To analyze and to design an algorithm for arc fault detection, arc faults needs to be investigated in detail. There are two possible ways to achieve deep understanding of arcing phenomena. The first way is to construct an arc model which requires deep understanding of arc phenomena, therefore, highly skilled human resource and time. The second way is to use real measurements explained in chapter 3. Using real measurements is more suitable methodology for the context of this thesis because arc model is already a complex task itself to complete in limited time whereas using measurements are easier, much faster and more accurate for the measured load types.

General system diagram for proposed methodology for AC arc fault detection is given Figure 4.1.

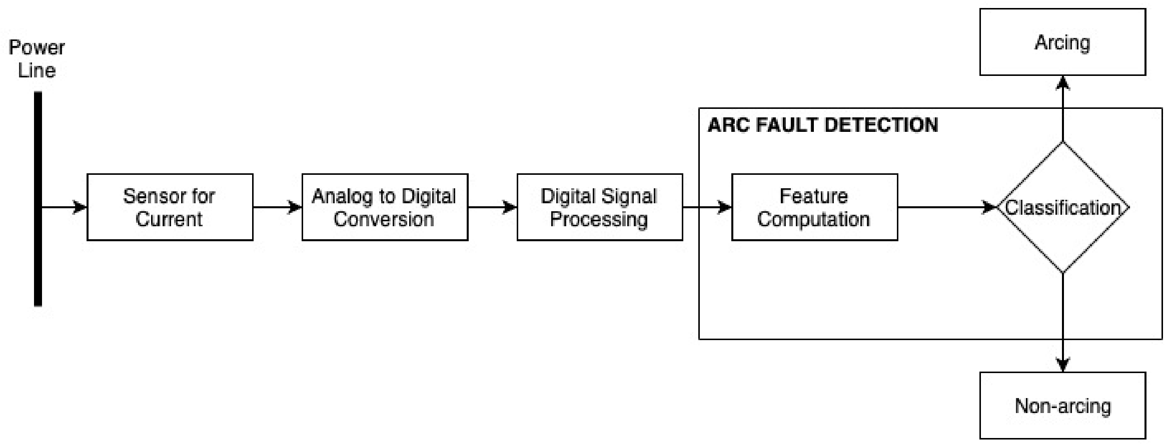

Figure 4.1: General System Diagram of Designed Method

Arc fault detection is a binary classification task which is a task of classifying the elements of a given set into two groups (predicting which group each one belongs to) on the basis of a classification rule.[24] For the context of this thesis, these groups are features extracted from captured current waveforms with series arc fault that explained in Section 2.1.3 and normal working conditions of different loads.

Informative features extracted from measurements with suitable methods allows us to employ so-called *learning from examples* paradigm. The crucial point about the learning from examples paradigm is that the number of measurements needs to be sufficiently high to cover all real life scenarios. Otherwise, algorithm trained using a small portion of real life scenarios can achieve successful performance metrics in the design phase however, they will not work sufficiently after implementation on the hardware. This is because, they are designed based on the measurements taken but not for the real situation, i.e *feature space* is not sufficiently rich. Therefore, quality and number of measurements have a crucial importance for success rates of designed algorithms for the implementation phase.

Feature extraction is another crucial task to design a successful classifier. It is also crucial in the phase of implementation on hardware . It is because if suitable features are not selected from feature space, namely all the bandwidth used for detection, detection algorithm works with both informative and uninformative data, therefore it is needed to process more data then necessary. It is not feasible in terms of computational complexity and it has other drawbacks such as:

1. It decreases accuracy of machine learning algorithm
2. It increases training time
3. It can cause overfitting

These drawbacks and benefits are investigated in more detail in section 4.2.1.

For the context of arc fault detection algorithm, selection of features simply means selection of proper bands that can be informative enough for classification. This is one of the main purposes of this thesis because an algorithm that is capable of choosing informative bands automatically for classification will ease the design of future algorithms, whereas current designs are based on expert knowledge and heuristic data. Proposed metaheuristics are investigated in section 4.2.1.1 and 4.2.1.2.

The goal of binary classification task is to learn a function $F(x)$ that minimizes the misclassification probability $P\{yF(x) < 0\}$, where y is the class label with + 1 for positive and 1 for negative. [29] There are numerous effectual binary classification algorithms such as kernel methods[30] , ensemble methods[31] , and deep learning methods[32] . Support vector machine, SVM,[33] is a powerful kernel method. Boosting, [34],[35] and random forest (RF) [36] fall into ensemble methods. Deep learning methods are based on artificial neural networks (ANNs) [37].

According to [65], k-nearest neighbors algorithm (k-NN) is a non-parametric method used for classification and regression. [64]. The input consists of the k closest training examples in the feature space. The output is a class membership. An object is classified by a plurality vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors where k is a positive integer.[65]

All classification techniques mentioned above have advantages and disadvantages. These advantages and disadvantages have been taken into consideration according to the analyzed data. SVM, which is introduced in more detail in section 4.1, performed well for a wide range applications such as finance, time series prediction[38], biological data processing for medical diagnosis[40], face recognition[41] and so on.Particularly, performance of SVM is proven when analyzed data is irregularly distributed or distribution of data is unknown. Main advantages of SVM algorithm in context of this thesis are stated as follows:

1. SVM assumes that classification data is linearly separable. By using the *kernel trick* which is to transform data into another dimension that has a clear dividing margin between classes of data [42](see section 4.1), SVM gains the capability of nonlinear classification whereas SVM without using *kernel trick* works as a linear classifier. By using kernel function, SVM performs nonlinear classification tasks even for nonlinearly dependent data and for data that has different functional forms.[43]

2. Kernel function implicitly contains nonlinear transformation inside, therefore, it is not required to make any assumptions on functional form of the transformation to make the data linearly separable. Transformation already runs implicitly on a proven robust theoretical background, therefore, a priori work is not needed on that.[44]

3. SVM provides an optimal and unique solution because the optimality problem is convex. SVM differs from neural networks in this context because neural networks can hand in various number of solutions that can possibly lie on a local minima therefore, robustness issues can be faced when different datasets are provided.

4. If parameters are properly set, SVM provides a good generalization error(out-of-sample error) which is a measure of how accurately an algorithm is able to predict outcome values for previously unseen data.[45] In explanation, if generalization grade is appropriate, SVM can be robust even if the training set is biased.

4.1 Fundamentals of SVM

The support vector machine was introduced by Vladimir N. Vapnik and Alexey Ya. Chervonenkis in 1963[33] and it has been developed through the years. In 1992, Bernhard E. Boser, Isabelle M. Guyon and Vladimir N. Vapnik suggested a way to create nonlinear classifiers by applying the kernel trick to maximum-margin hyperplanes. Statistical learning theory developed by Vladimir Vapnik and co-workers at ATT Bell Laboratories in 1995 improved the algorithm and transformed the algorithm to it's final version.

The support vector machine is a linear classifier that assumes training data is linearly separable. In [33], it is stated that SVM chooses the particular classifier which separates the classes (+1,-1) with maximal *margin*. Optimal hyperplane which separates the classes is defined by a weight vector W and a bias term b as:

$$w^T x + b = \begin{cases} \geq 1, & \text{class } +1 \\ \leq -1, & \text{class -1} \end{cases} \tag{4.1}$$

Figure 4.2 shows separating hyperplane in 2-dimensional space including so-called support vectors and margin.

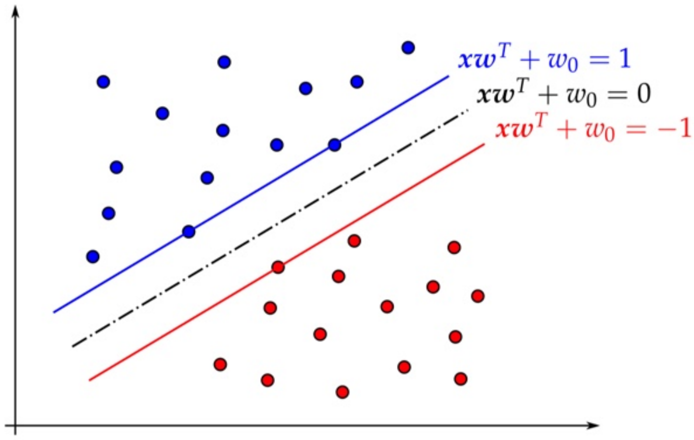

Figure 4.2: Optimal separating hyperplane, support vectors and decision boundary(Separable case)

The margin is defined as the width of the largest tube not containing samples that can be drawn around the decision boundary and it can be represented as follows.

$$Margin = \frac{2}{\|\vec{w}\|} \tag{4.2}$$

To achieve the widest separation, it is needed to maximize margin. In order to maximize the margin,$\|\vec{w}\|$ is minimized. Therefore, the maximum margin can be achieved by solving the quadratic optimization problem below.

$$Min \frac{1}{2} \|\vec{w}\|^2 \tag{4.3}$$

$$s.t \qquad y_i(W^T x_i + b) \geq 1 \tag{4.4}$$

In equation (4.4), y_i is 1 for samples from class (+1) and -1 for samples from class (-1). Since this is a constrained optimization problem, Lagrangian multiplier is used for solution.

$$L = \frac{1}{2} \|\vec{w}\|^2 - \sum \alpha_i [y_i(\vec{w} \cdot \vec{x_i} + b) - 1] \tag{4.5}$$

$$\frac{\partial L}{\partial \vec{w}} = \vec{w} - \sum \alpha_i y_i x_i = 0 \quad \Longrightarrow \quad \vec{w} = \sum_i \alpha_i y_i x_i \tag{4.6}$$

$$\frac{\partial L}{\partial b} = -\sum \alpha_i y_i = 0 \quad \Longrightarrow \quad \sum_i \alpha_i y_i = 0 \tag{4.7}$$

If equation (4.6) and (4.7) are inserted into (4.5). L is as follows where α_i is Lagrangian multiplier.

$$L = \sum_i \alpha_i - \frac{1}{2} \sum \sum \alpha_i \alpha_j y_i y_j x_i x_j \qquad\qquad where \qquad \alpha_i \geq 0 \tag{4.8}$$

Equation (4.9) shows that optimization depends on only dot product of pairs of samples $x_i x_j$ because the other terms in the equation are constant. By inserting the findings in (4.6) and (4.7) into decision rule (4.1), following decision rule is obtained where every term except $\vec{x_i}\vec{u}$ are constant.

$$\sum \alpha_i y_i \vec{x_i} \vec{u_i} + b = \begin{cases} \geq 1, & \text{class +1} \\ \leq -1, & \text{class -1} \end{cases} \tag{4.9}$$

When optimization problem is solved, training points with $\alpha_i^* > 0$ are the support vectors(SVs), and w^* is calculated as follows:

$$w^* = \sum_i \alpha_i^* y_i x_i \tag{4.10}$$

Kernel function is used to add the ability of making nonlinear classification to SVMs. Kernels are used to map the input data to a new multidimensional space. Feature vectors are mapped into the new image space by using mapping K. SVM performs linear classification task in the newly mapped image space, however, this corresponds to nonlinear classification in the original space. Widely used kernel functions are given as follows:

1. Polynomial Kernel: $K(x_i, x_j) = (x_i \cdot x_j + 1)^p$
2. Radial Basis Function Kernel: $K(x_i, x_j) = e^{-\gamma \|x_i - x_j\|^2}$ where $\gamma > 0$
3. Sigmoid Kernel: $K(x_i, x_j) = tanh(\eta x_i \cdot x_j + v)$

4.2 Feature Extraction

In order to process the data and to investigate the noise characteristics of different loads in normal working condition and with an arc fault; as well as to construct a suitable feature space to train machine learning algorithms, it is needed to process the data which is current waveforms of loads. Domain of analysis is important since the quality, namely informativeness of feature space is important. Following domains can be employed to create the feature space for the machine learning algorithms.

1. Time-domain
2. Frequency Domain
3. Time-frequency Domain

There are numerous papers in literature that propose a detection algorithm using each of the three domains above. [47], [48], [49] [50] propose frequency domain based approaches for different applications. [47], uses FFT to obtain signal energy and calculates the change in it for arc faults in photo-voltaic systems. This approach is similar to the approach applied in this thesis with a difference in analysis of band selection and detection methodology. In this thesis, given approach to determine the frequency bands are done using feature selection algorithms in section 4.2.1.

[48], proposes a method for arc fault detection using FFT and wavelet packet decomposition. The drawback of using FFT is stated as an inherent trade-off between time and frequency resolution. Using FFT also gives global frequency components of a signal without time information.[48] In order to have a better resolution of frequency, the time window needs to be extended while the sampling frequency is fixed. The most crucial finding of this paper is the duration of the time window for FFT, which is stated as 200 ms based on extensive analysis. Therefore, it has to be taken into consideration while designing the detection algorithm because longer time used for signal processing means there is less time available for the detection algorithm to run. This trade off has be taken into consideration in the design process of general detection methodology. Machine learning classifiers, especially SVM, are advantageous from this point of view because although SVM can be computationally expensive in the training phase, it does not require high computational power as a classifier, therefore it is fast compared to most other algorithms proposed in literature. Therefore, analysis done in [48] contributes to prove feasibility of SVM classifier to use for AC arc fault detection based on frequency domain analysis.

[49] and [50] also uses frequency domain analysis but for DC arc fault detection. The arc characteristics observed in AC and DC applications has differences however, they also use information extracted from current waveforms by applying frequency domain analysis.

It is more important the quality of feature space for detection algorithm and construction of the feature space according to needs than the number of signal processing techniques applied. As it can be seen from the studies above, all the signal processing methods can be informative enough to design an arc fault detection algorithm if the information extracted using any method is sufficient for classification. However, time domain based solutions are mostly applied for DC arc fault detection algorithms and it is not a common approach employed according to extensive literature research, therefore, this approach is not investigated in this thesis.

In order to construct the feature space using frequency domain analysis, it is first needed to perform the *Fourier Analysis*. Discrete Fourier Transform(DFT) is the most important discrete transform, used to perform Fourier analysis.[46] In this thesis, single-sided spectrum of current waveforms of different loads are computed on MATLAB and single-sided spectrum is obtained for further processing. Then, the main component of single sided spectrum and harmonics are removed since they don't contain any valuable information but mask valuable information to be used in designed algorithm.

The signal processing chain employed in this thesis is given in Figure 4.3. Figure 4.4 shows a single sided spectrum of a heater. In Figure 4.5 harmonics related to line frequency are removed. Finally, in Figure 4.6, one element of feature space with 1000 features that include total energies when frequency spectrum divided to 1000 is shown.

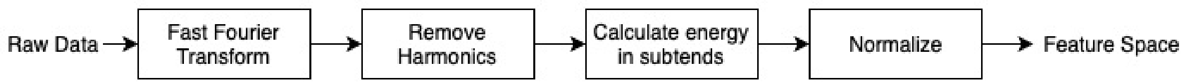

Figure 4.3: General Diagram of Signal Processing Chain

Residential AC supply is 50Hz, 50Hz has the biggest amplitude compared to higher frequencies. There are also harmonics associated with the line frequency, 50Hz. However, frequency spectrum around 50Hz and associated harmonics are not informative since all the loads work with the same line frequency. Therefore, line frequency and associated harmonics need to be removed for further processing. Figure 4.4 shows single sided spectrum of current waveform of heater under normal working conditions from 0Hz to 400Hz which is an interval that is already not used in phase of design but shows an illustration of harmonic removal process.

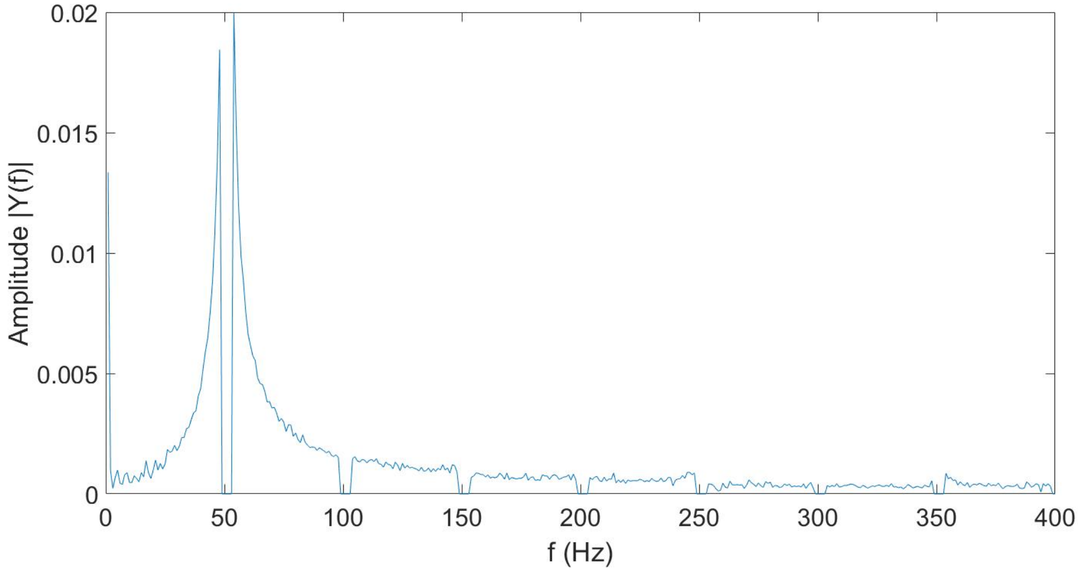

Figure 4.4: Single-sided spectrum of current waveform of heater without harmonics under normal working conditions (0-400Hz)

In order to construct a feature space to train machine learning algorithm, spectrum data obtained by removing the line frequency and harmonics is divided into intervals of different bandwidths. Total energy in these intervals are calculated. If these intervals cover a wide bandwidth, then noise data is eliminated in the intervals because noise levels are expected to be relatively small compared to total energy in a wide frequency range. If intervals are too small, then it is computationally ineffective for feature selection algorithms that will be explained in more detail in section 4.2.1. Therefore, this trade-off has to be taken into consideration while making a decision on the length of intervals. The spectrum is divided into 100 and 1000 intervals and total energy for both intervals are calculated. However, after experiments, it is observed that using 1000 features is more suitable because when 100 features are employed, frequency bands are too wide to be informative, namely, noise levels are not distinguishable in a wide range of spectrum.

Last step of obtaining the feature space is normalization of features, which is simply done by dividing all the total energies in intervals to the amplitude at the line frequency. This is needed to eliminate the differences caused by different powers of different loads. Normalization provides a feature space depending on the relative normalized amplitudes by removing the effects of different power demands of loads.

Figure 4.6 shows the normalized features for one load in an interval of 21MHz.

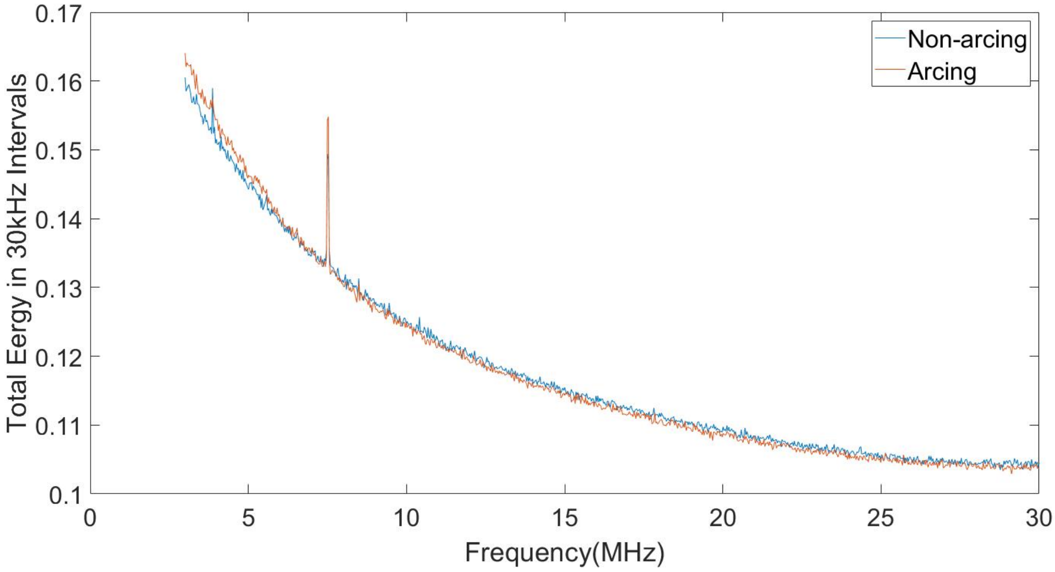

Figure 4.5: Feature vector in 2D for one load(heater) from 3MHz to 30 MHz, Feature Index: 98-1000

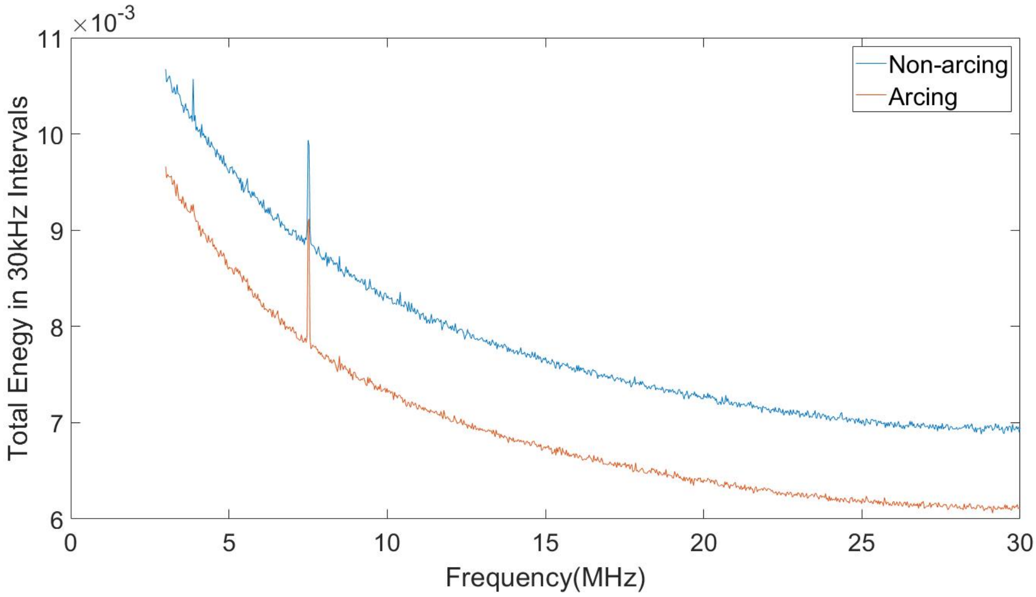

Figure 4.6: Normalized feature vector in 2D for one load(heater) from 3MHz to 30 MHz, Feature Index: 98-1000

4.2.1 Feature Selection

Feature selection has a significant importance in machine learning, especially when handling with high dimensional data.

In order to find the most correlated frequency bands that are informative for classification, there are many feature selection algorithms available such as neighborhood component analysis or sequential feature selection. Extracting relevant features for a given dataset with labels are crucial because feature selection:

1. Improves Accuracy: This is because misleading data can decrease the modelling accuracy of machine learning algorithms therefore the algorithm performs less because of the uninformative data for classification

2. Reduces Training Time: With a smaller feature space, it is faster to train the machine learning algorithm and training is less computationally expensive.

3. Reduces Overfitting: When feature space is smaller, it is less likely to make a decision based on uninformative data

There are two different feature selection metaheuristics that are widely used in literature which are filter approach and wrapper approach. There are also hybrid methods that are designed for different purposes to increase robustness and accuracy. Figure 4.7 shows flow diagrams of these approaches.

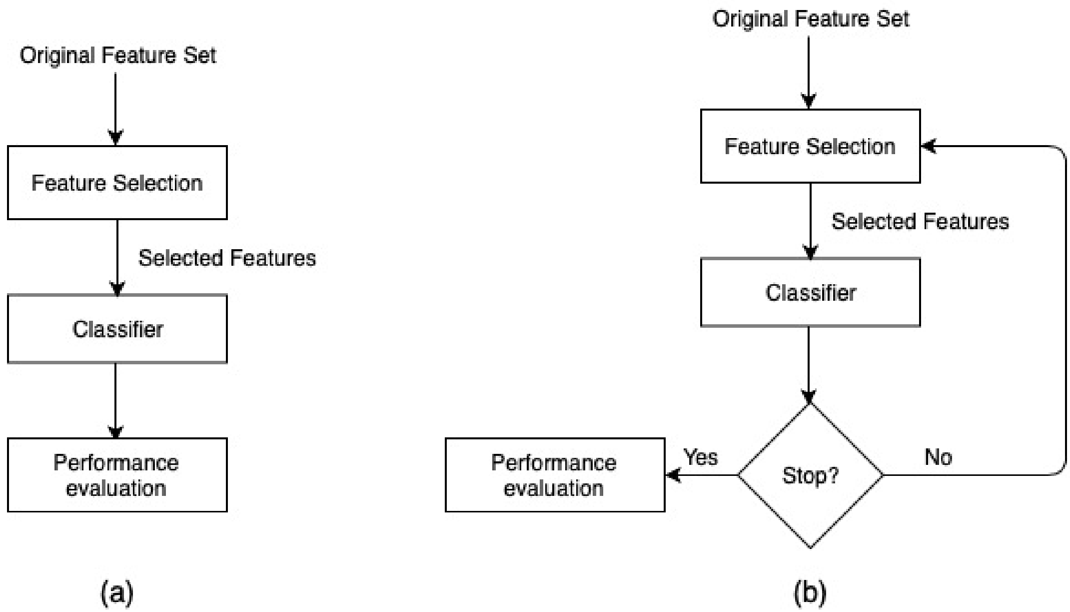

Figure 4.7: Feature Selection Approaches: a) Filter Approach, b) Wrapper Approach

Feature selection approaches employed in thesis are introduced in Section 4.2.1.1 and Section 4.2.1.2.

4.2.1.1 Neighborhood Component Analysis

Neighborhood component analysis (NCA) is a non-parametric and embedded method for selecting features with the goal of maximizing prediction accuracy of regression and classification algorithms.[51] One drawback of neighborhood component analysis is the computational cost, but there are proposed methods such as fast neighborhood component analysis to decrease computational cost by applying statistical tricks and prior calculations. [52]. However, computational cost of NCA is acceptable for this thesis since feature space does not consist of very high number of features for NCA, namely, maximum 1000 for a bandwidth of 31.25Mhz.

Neighborhood component analysis algorithm[53] starts by building a full graph with each feature as its node. Weight of each edge between any nodes are denoted as p_{ij}. p_{ij} is the probability that feature x_i selects x_j as its neighbor and it is calculated as follows:

$$p_{ij} = \frac{exp(-d_A^2{}^{ij})}{\sum_{t \in N_i} exp(-d_A^2{}^{it})} \tag{4.11}$$

N_i is the set of neighbors of x_i and d is the distance between two features. Then, neighborhood component analysis learns a linear transformation A which maximizes the log likelihood so that using transformation, each feature chooses the features with same labels as itself as neighbors.[54] This objective function is represented as follows:

$$maxL(A) = \sum_i log(\sum_{j \in N_i} 1\{y_i = y_j\} \cdot p_{ij}) \tag{4.12}$$

4.2.1.2 Sequential Feature Selection

Sequential feature selection algorithm used in this thesis selects a subset of feature space that has the best classification metrics for a given label of y, which is the class label by sequentially selecting until classification metrics do not improve. Rows of feature space corresponds to frequency spectrum divided into intervals and columns,therefore corresponds to total energy in that specific band for each measurement.

The algorithm starts with an empty feature set, then it creates candidate subsets by sequentially including each feature not selected yet to empty set. With every iteration, the algorithm performs a *10-fold cross-validation* which is introduced in more detail in section 4.4.3.

According to [55], there are four key steps in procedure of sequential feature selection which are:

1. Subset generation
2. Evaluation of Subset
3. Stopping Criteria
4. Result Validation

These key steps can be designed in different ways according to needs however, general procedure is introduced in Figure 4.8 as follows:

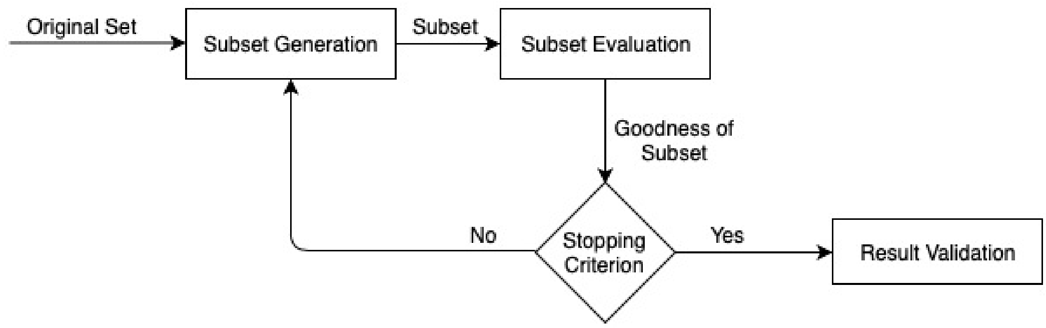

Figure 4.8: Sequential Feature Selection General Procedure

4.3 SVM Training

All the calculations and training of support vector machine and other machine learning algorithms were implemented using MATLAB statistics and machine learning toolbox. General design process is demonstrated in Figure 4.9.

First step of the design process is to define the requirements. Secondly, it is needed to measure different loads in sufficiently enough numbers. Generation of feature space is very critical step because, accuracy of classifiers are heavily depends on feature space. It is needed to consider possible situations that can be faced by a load and then to generate the feature space because if feature space is not properly generated, then even if classification algorithm performs well in its feature space, it will not have same classification rates after implementation on hardware. All the generated features are stored in a database to use after if needed. After selecting the features by using metaheuristics in Section 4.2.1, classifier is trained and optimized. Grid-search is used to find optimal parameters of the classifiers. Grid-search is a way to find to optimal parameters of a classifier but parameters found by grid-search depends on the range of the grid. In order to prevent doing more time-consuming processes as shown in Figure 4.9, firstly parameter optimization is redone because it is much simpler compared to other tasks in design process. The flow diagram is applied until successful classifier is achieved.

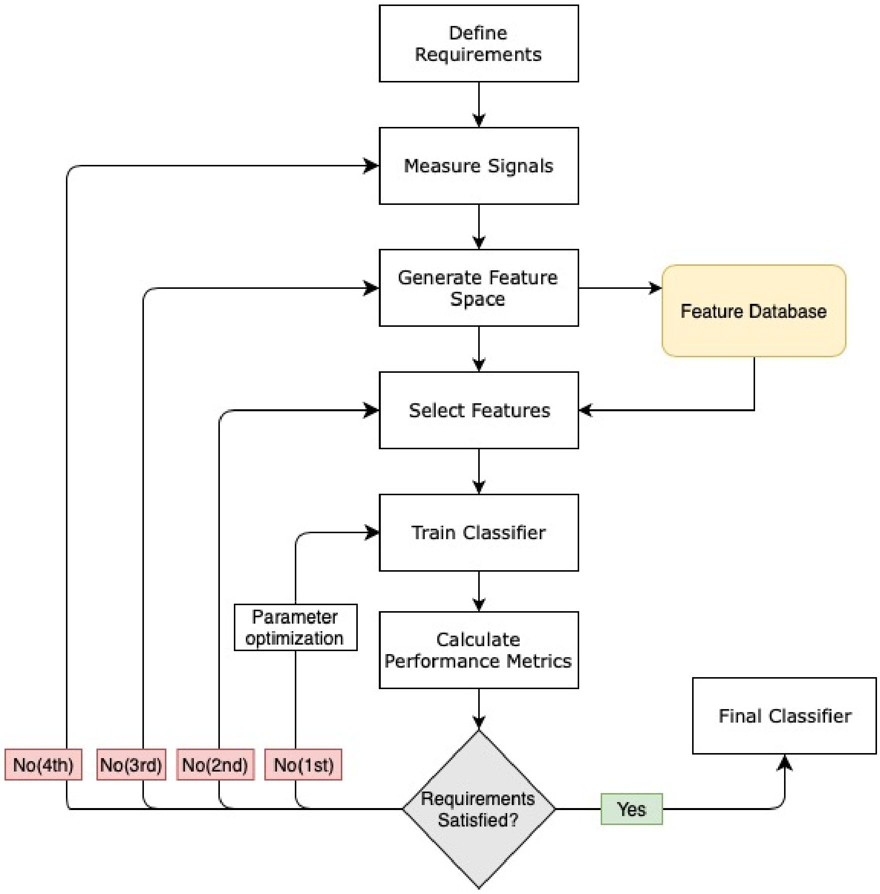

Figure 4.9: Classifier Training Process

4.4 Performance Evaluation

Performance of designed classifiers has to be evaluated using suitable metrics and analysis. There are measures such as misclassification rate and metrics derived from confusion matrix given in Table 4.1. These metrics are introduced in more detail in Section 4.4.1. Also, a basic methodology to evaluate the classifier is to use cross-validation technique which is explained in Section 4.4.3.

It is needed to avoid overfitting or underfitting which decreases the generalization ability of classifier that's explained in chapter 4, in order to obtain a successful design and to set the regarding parameters of the machine learning algorithm. Overfitting or underfitting need to be avoided since the overfitted classifier performs well on training data, however it does not perform the same way for unseen new data due to its low generalization grade.

This phenomena is shown in Figure 4.10.

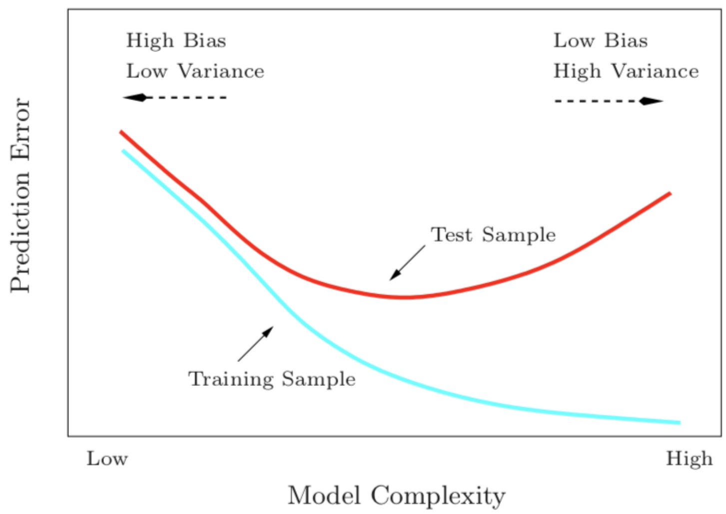

Figure 4.10: Testing and training error as a function of model complexity[56]

As it can be seen in figure 4.10, training error decreases with increasing model complexity but testing error is minimal to a certain degree of model complexity. [57] The way of checking whether the classifier is overfitted or not is to search for another classifier for the same domain such that the training error is higher but testing error is lower. If such a classifier exists, then it means that classifier is overfitted.

4.4.1 Performance Metrics

In order to evaluate performance of classifiers, performance metrics derived from the confusion matrix are employed. The confusion matrix is shown in Table 4.1.

	Condition Positive	Condition Negative
Predicted Condition Positive	True Positive(TP)	False Positive(FP)
Predicted Condition Negative	False Negative(FN)	True Negative(TN)

Table 4.1: Confusion Matrix

Metrics derived from confusion matrix are as follows:

1. Positive Predictive Value: $PPV = \frac{TP}{TP+FP}$

   The positive predictive value(PPV) is the proportion of true positive predictions to all positive predictions of a classifier.[59]

2. Negative Predictive Value: $NPV = \frac{TN}{TN+FN}$

   The negative predictive value(NPV) is the proportion of true negative predictions to all negative predictions of a classifier.[59]

3. False discovery rate: $FDR = 1 - PPV = \frac{FP}{TP+FP}$

   The false discovery rate is the proportion of false positive predictions to all predicted condition positives of a classifier.[59]

4. False omission rate: $FOR = 1 - NPV = \frac{FN}{TN+FN}$

   The false omission rate is the ratio of false negative predictions to all predicted condition negatives of a classifier.[**?**]

5. Accuracy: $ACC = \frac{TP+TN}{TP+TN+TN+FN}$

   Accuracy is the ratio of correctly classified predicitons among all predictions.[59]

6. Prevalence: $Prevalence = \frac{TP+FN}{TP+TN+TN+FN}$

   Prevalence is the ratio of condition positive which is sum of true positives and false negatives among all predictions.[59]

7. True Positive Rate: $TPR = \frac{TP}{TP+FN}$

   True positive rate, also called as sensitivity, recall or probability of detection in some fields[60] is the proportion of actual positives that are correctly identified as such.[59]

   –

8. False Positive Rate: $FPR = \frac{FP}{FP+TN}$

   False positive rate, also called as fall-out or probability of false alarm is calculated as the ratio between the number of negative events wrongly classified as positive and the total number of actual negative events.[59]

9. False Negative Rate: $FNR = \frac{FP}{FP+TN}$

   False negative rate, also called as fall-out or probability of false alarm is the proportion of positives which yield negative test outcomes with the test, i.e., the conditional probability of a negative classification result given that the condition being looked for is present.[59]

10. True Negative Rate: $TNR = \frac{FP}{FP+TN}$

    True negative rate also called as specificity measures the proportion of actual negatives that are correctly classified as such (e.g., the percentage of non-arcing waveforms that are correctly classified as non-arcing).[59]

4.4.2 Cross-validation

According to [11], cross-validation, also called rotation estimation is one of numerous model validation methodologies used to identify how a classifier will generalize to an independent dataset. It is employed in the design process of machine learning algorithms to investigate the ability of the classifier on unseen data. The way it works is to use a limited subset of dataset for training, and to test the performance of it on the data reserved for test.

Cross-validation has a single parameter, k, that is the number of folds that a dataset is to be separated into. It is a very popular methodology because it is very simple to employ and it gives a better understanding of the classifiers performance compared to simpler methods such as train/test split. [66]

Procedure is as follows:

1. Seperate dataset into k random subsets with same amount of data
2. For each group, take one out as a hold to use as a test set, then use all others as a training set
3. Train a model using the training set and test it with the test set
4. Record the classifier performance
5. Repeat this procedure from 2nd step for each fold
6. Calculate average loss by averaging k different performance metrics

Each data in the dataset should be assigned to only one group and the group that particular data point belongs to should not change through the procedure. This means that each data point is used once in testing and k-1 times for training the classifier.

In this thesis, separation of groups are done in a way that all different load types have an equal number of data in the test and training sets. Namely, data is randomly separated into groups but subgroups are also taken into consideration.

Chapter 5

Results

In this chapter, achieved results by using different supervised machine learning classifiers are presented. Selected features for SVM are explained. Achieved classification results by using selected features are compared and best classifier is decided based on defined performance criteria in Chapter 4. SVM with RBF kernel outperformed all classifiers with 93% classification rate .

As initial work, support vector machine classifiers were trained using all the features(1000 features for a bandwidth of 31MHz). Achieved bandwidth is 31.25MHz, however, bandwidth up to 31MHz is used. Every element of feature space contains the total energy in an interval of 30kHz. Data used for testing purposes were not included in any phase of training of classifiers. Therefore, it is unseen data which is 10% of whole data that are reserved for testing purposes for obtained classifiers. 10-fold cross validation is applied and 10-fold loss were calculated according to methodology introduced in Section 4.4.2.

Table 5.1 shows the confusion matrix using linear kernel.

	Condition Positive	Condition Negative
Predicted Condition Positive	21	10
Predicted Condition Negative	6	18

Table 5.1: Confusion Matrix-SVM(Linear Kernel)

10-fold loss of support vector machine with linear kernel is 29%. Therefore, classification rate is 71%. However, it is also inspected that most of the misclassified data are composed of dimmers and fluorescent lamps whereas, classification was successfully done most of the

LEDs, resistive loads, universal motors and power source with active power correlation. 29% misclassification rate is an expected result since data is not linearly separable. Also, it is observed that while arcing, current waveform of dimmers with resistive load is similar to current waveform of resistive loads in normal operation without arcing. This is the reason that dimmers with resistive loads are hard to classify, especially using SVM with linear kernel.

Table 5.2 shows classification results of SVM trained using second order polynomial kernel. Polynomial kernel allows support vector machine to make nonlinear classification to some degree. Benefit of employing nonlinear classifier can be observed from classification results.

	Condition Positive	Condition Negative
Predicted Condition Positive	21	9
Predicted Condition Negative	6	19

Table 5.2: Confusion Matrix-SVM(Polynomial Kernel(degree 2))

10-fold loss using second polynomial kernel is 27%. It is also inspected that most of the misclassified data are composed of dimmers and fluorescent lamps whereas, classification was successfully done most of the LEDs, resistive loads, universal motors and power source with active power correlation. Therefore, in terms of accuracy for different types of loads, this classifier shows same characteristics with SVM with linear kernel. Classification rate is increased by 2% compared to classifier with linear kernel. This can be further increased by increasing the polynomial order however, it is needed to consider to avoid overfitting.

Table 5.3 shows classification results of SVM trained using third order polynomial kernel.

	Condition Positive	Condition Negative
Predicted Condition Positive	22	9
Predicted Condition Negative	5	19

Table 5.3: Confusion Matrix-SVM(Polynomial Kernel(degree 3))

10-fold loss using third order polynomial kernel is 25%. It is also inspected that most of the misclassified data are composed of dimmers and fluorescent lamps whereas, classification was successfully done most of the LEDs, resistive loads, universal motors and power source with active power correlation. In terms of accuracy for different types of loads, this classifier shows slightly better performance than SVMs with second order polynomial kernel and linear kernel by 2% and 4% respectively. Polynomial kernel is not widely used in literature. Increasing the polynomial order causes overfitting. Therefore, higher orders are not implemented and not discussed in this thesis. The most important parameter of polynomial kernel is order. Since, there is a fact that higher orders lead to overfitting, it is not needed to apply grid-search or any other optimization algorithm on polynomial order.

Table 5.4 shows confusion matrix obtained by SVM classifier with radial basis function kernel. Radial basis function has two parameters to tune which are C and γ. Parameter γ can be considered as the inverse of the radius of influence of samples selected by the model as support vectors.[62] The C parameter trades off correct classification of training examples against maximization of the decision functions margin.[62] If C value is larger, accepted margin is smaller if all training points are classified correctly by the decision function. Therefore, smaller C implies a larger margin but with a trade-off with training accuracy. Namely, when C is larger, classifier is more penalized for misclassified data and hyperplane bends over around misclassified data points in order to obtain a better classification result. However, it is not desired because it lowers the generalization ability of classifier and it decreases the performance on unseen data.

It is observed that for very small values of γ, classifier cannot capture the complexity of data because it is inverse of standard deviation of radial basis function. Standard deviation is used as a similarity measure between two different data. Very small γ value means an RBF with a large variance. Therefore, two data points that are far from each other are classified in the same class. This is verified by decreasing γ and getting similar result with linear kernel because when γ is very small, it behaves like a model with hyperplane that separate center of pairs with high density in two different classes.[62]

	Condition Positive	Condition Negative
Predicted Condition Positive	26	5
Predicted Condition Negative	1	23

Table 5.4: Confusion Matrix-SVM(RBF Kernel)

As it can be seen from Table 5.4, classifier using RBF kernel outperforms classifiers with linear and polynomial kernel. 10-fold loss is 11%. In order to optimize the parameters, a logarithmic grid from 10^{-3} to 10^{3} is applied for grid search. Best classifier is achieved when $C = 10$ and $\gamma = 1$. It is also observed that higher values of both C and γ cause overfitting. Performance of classifier can be slightly increased using a more detailed grid around achieved optimal parameters. Therefore, it can be said that designed classifier is suboptimal in terms of the grid used. However, SVM has better generalization performance compared to most of the classifiers as stated in Chapter 4, therefore, even if it is optimized again with a grid around $C = 10$ and $\gamma = 1$, the increase in performance will be minimal.

K-nearest neighbor classifier has one parameter to set which is k. 3 different classifiers are trained for different k values(2,3,5). Table 5.5 shows the confusion matrix of 2-nearest neighbor classifier. Classification rate is 75% which is similar to SVM with linear kernel. Also, it is observed that wrongly classified data is composed of mostly resistive loads and dimmers. The reason is the similarity between normal dimmer behaviour and resistive load's behaviour while arcing.

	Condition Positive	Condition Negative
Predicted Condition Positive	22	9
Predicted Condition Negative	5	19

Table 5.5: Confusion Matrix-k-nearest neighbors(k=2)

Table 5.6 shows the confusion matrix of 3-nearest neighbors classifier. Classification rate is 71%. It has considerably lower performance compared to SVMs and the 2-nearest neighbor. Wrongly classified data are composed of all loads except LEDs.

	Condition Positive	Condition Negative
Predicted Condition Positive	21	10
Predicted Condition Negative	6	18

Table 5.6: Confusion Matrix-k-nearest neighbors(k=3)

Table 5.7 shows the confusion matrix of 5-nearest neighbors classifier. Classification rate is 69%.

	Condition Positive	Condition Negative
Predicted Condition Positive	20	10
Predicted Condition Negative	7	18

Table 5.7: Confusion Matrix-k-nearest neighbors(k=5)

Decision tree is another classifier that are widely used in classification tasks. Different decision trees are designed with different maximum number of allowed splits, namely, first decision tree is designed without a limitation in maximum number of allowed splits. Second decision tree has 7 allowed branches. Last decision tree has 11 maximum number of allowed splits. However, it is very important to state that training sets for decision trees and Adaboost are slightly different than training sets used for all other classifiers. In SVM and k-nearest neighbors, algorithms are trained using all the bandwidth achieved, therefore, 1000 features that correspond to 31MHz bandwidth were used whereas frequencies lower than 1MHz are excluded because of the reasons that are stated in Section 2.2 while training decision trees and Adaboost algorithm.

Table 5.8 shows the confusion matrix of decision tree without a limitation in number of splits. Classification rate is 80%. Number of splits are 32.

	Condition Positive	Condition Negative
Predicted Condition Positive	22	6
Predicted Condition Negative	5	22

Table 5.8: Confusion Matrix-Decision Tree

Table 5.9 shows the confusion matrix of decision tree with maximum allowed number of splits is 7. Classification rate is 65%.

	Condition Positive	Condition Negative
Predicted Condition Positive	18	10
Predicted Condition Negative	9	18

Table 5.9: Confusion Matrix-Decision Tree-7 Splits

Table 5.10 shows the confusion matrix of decision tree with maximum allowed number of splits is 11. Classification rate is 71%.

	Condition Positive	Condition Negative
Predicted Condition Positive	21	10
Predicted Condition Negative	6	18

Table 5.10: Confusion Matrix-Decision Tree-11 Splits

Desion tree has a better performance than trained k-nearest neighbor algorithms. When there is no limitation on number of splits, it has 80% classification rate which makes its performance better than SVMs with linear and polynomial kernel.

Table 5.11 shows the confusion matrix of Adaboost algorithm. Classification rate is 82%.

	Condition Positive	Condition Negative
Predicted Condition Positive	21	4
Predicted Condition Negative	6	24

Table 5.11: Confusion Matrix-Adaboost

Adaboost performs better than k-nearest neighbors, decision trees and SVMs with linear and polynomial kernel. Adaboost also has best false positive rate which is 3% better than SVM with RBF kernel. This is an important performance metric because of the consequences of nuisance tripping that explained in Section 2.3.3 in more detail. Adaboost and SVM are two classifiers that have better generalization ability compared to other designed classifiers in this thesis. In literature they showed different generalization characteristics on different datasets. Therefore, 10-fold loss is an important indicator of their generalization ability.

In order to evaluate the performance of classifiers and to compare, performance metrics presented in Section 4.4.1 are calculated for all designed classifiers. Table 5.12 shows performance metrics of different classifiers using all bandwidth.

Table 5.12 shows that, SVM with RBF kernel outperforms other classifiers with a classification rate of 89% when all bandwidth is used for training the classifiers(Lower than 1MHz is excluded in decision tree and adaboost). These performance metrics form a basis for comparison for following section, where frequency bands are selected because the objective of feature selection is to achieve the same or better classification performance compared to using full bandwidth. It also makes it more feasible to employ general methodology proposed in thesis for industrial applications because smaller bandwidth is more achievable compared to using 31MHz bandwidth .

Classifier	ACC	TPR	SPC	PPV	NPV	FPR	FDR	FNR
SVM(Linear)	0.71	0.78	0.64	0.68	0.75	0.36	0.32	0.22
SVM(Polynomial-2nd order)	0.73	0.78	0.68	0.70	0.76	0.32	0.30	0.22
SVM(Polynomial-3rd order)	0.75	0.81	0.68	0.71	0.79	0.32	0.29	0.18
SVM(RBF)	0.89	0.96	0.82	0.83	0.95	0.17	0.16	0.03
2-nearest neighbors	0.75	0.81	0.68	0.71	0.79	0.32	0.29	0.18
3-nearest neighbors	0.71	0.78	0.64	0.68	0.75	0.36	0.32	0.22
5-nearest neighbors	0.69	0.74	0.64	0.67	0.72	0.36	0.33	0.26
Decision Tree (Maximum number of split)	0.80	0.81	0.79	0.79	0.81	0.21	0.21	0.18
Decision Tree (Number of split:7)	0.65	0.67	0.64	0.64	0.66	0.36	0.36	0.33
Decision Tree (Number of split:11)	0.71	0.78	0.64	0.68	0.75	0.36	0.32	0.22
Adaboost	0.82	0.78	0.86	0.84	0.80	0.14	0.16	0.22

Table 5.12: Performance Metrics of Different Classifiers Using All Bandwidth

As it can be seen from Table 5.12, SVM with RBF kernel has the best classification rate with 89%. Another important metric is false positive rate because of the consequences of nuisance tripping that explained in Section 2.3.3. False positive rate of SVM with RBF kernel is 0.17 which is the second lowest among all designed classifiers. 2-nearest neighbor also performed similar to SVM with linear and polynomial kernels. However, performance of k-nearest neighbors algorithms decrease while k increases. Therefore, it is concluded that classification of data points on the boundaries of different classes are problematic and generalization ability of k-nearest neighbors algorithm is not sufficient for such a dataset where variance in data is high. Decision tree performs similar to k-nearest neighbors. Best classification performance was achieved when there was no limitation on number of splits. Using 32 splits, classification rate is 80% which makes decision tree the third best designed classifier in terms of performance. Adaboost has 82% classification rate which is second best performance among all classifiers. Impressive result achieved by adaboost is its false positive rate which is 14% that makes adaboost the best classifier in terms of FPR.

SVMs have better performance metrics in general. Also, better generalization ability of SVMs makes SVM with RBF kernel is the best choise among all designed classifier. Therefore, in the following sections, features are selected in order to use with SVM and results are given.

5.1 Feature Selection

Different frequency bands are selected according to methodologies presented in Section 4.2.1. Combinations of these bands are used to train the classifiers. Quality of selected bands is evaluated according to the performance of classifiers. Filter approach that introduced in Section 4.2.1 employs a methodology that features are selected based on decision rules in feature selection methodology. Therefore, performance of classifier is not a parameter for filter approach in the phase of feature selection whereas wrapper approach already takes performance into account while selecting the features and it gives more robust and guaranteed results in terms of resulting classifier's performance.

In this thesis, forward sequential feature selection is applied. In forward sequential feature selection, selected data is an empty matrix in the beginning. While algorithm runs, it checks every combination of features by adding one by one and it selects the features that give the biggest increase in performance until performance converges or until all combinations are tried. Therefore, it does not lie on a local minima. Algorithm trains a classifier for every combination of features, therefore, it is computationally expensive. Backward sequential feature selection is another sequential feature selection method where selected features set includes all features in the beginning, then they are removed one by one for every iteration of algorithm. Therefore, it is less computationally expensive. Backward sequential feature selection is not employed in this thesis but it stays a future work to do.

Selected frequency bands by different algorithms(neighborhood component analysis and sequential feature selection) for best classification performance, methodology of selection and performance metrics of trained classifiers are given in the following sections.

5.1.1 Neighborhood Component Analysis

Figure 5.1 shows feature weights which are their importance between input data and their class labels around 5MHz(4.5MHz to 5.5MHz) for an interval of 1MHz divided in 30kHz intervals.

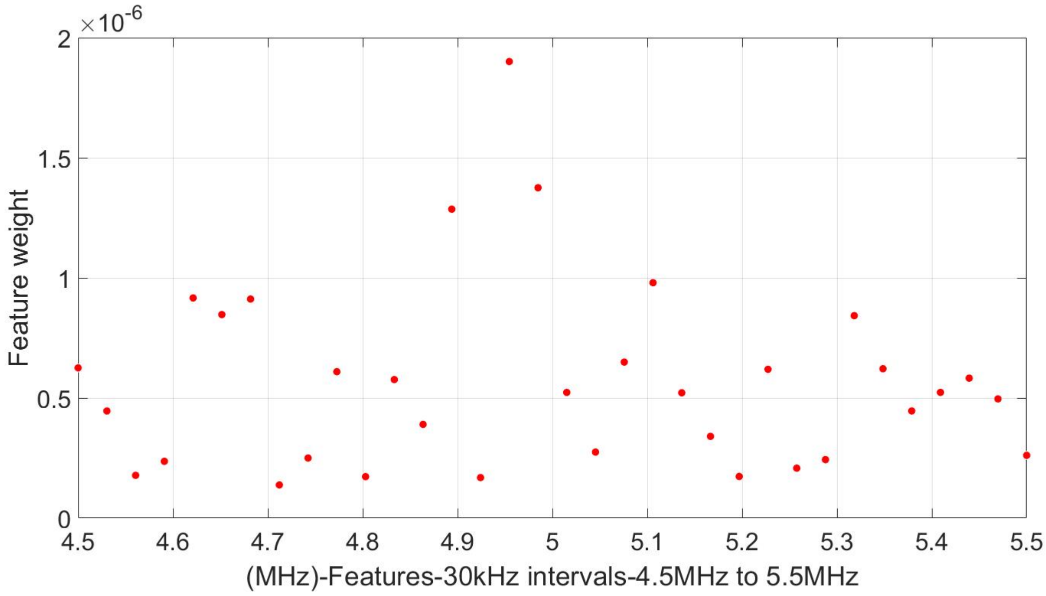

Figure 5.1: Feature selection using NCA in 4.5MHz-5.5Mhz

In order to select the intervals, a threshold value for feature weights has to be decided. When feature weight threshold is chosen $0.5x10^{-6}$, there are 19 intervals to select and total bandwidth used is 570kHz between 4.5MHz and 5.5MHz. When feature weight threshold is chosen $0.25x10^{-6}$, there are 25 intervals to select and total bandwidth used is 750kHz between 4.5MHz and 5.5MHz.

Figure 5.2 shows feature weights which are their importance between input data and their class labels around 17MHz(16.5MHz to 17.5MHz) for an interval of 1MHz divided in 30kHz intervals.

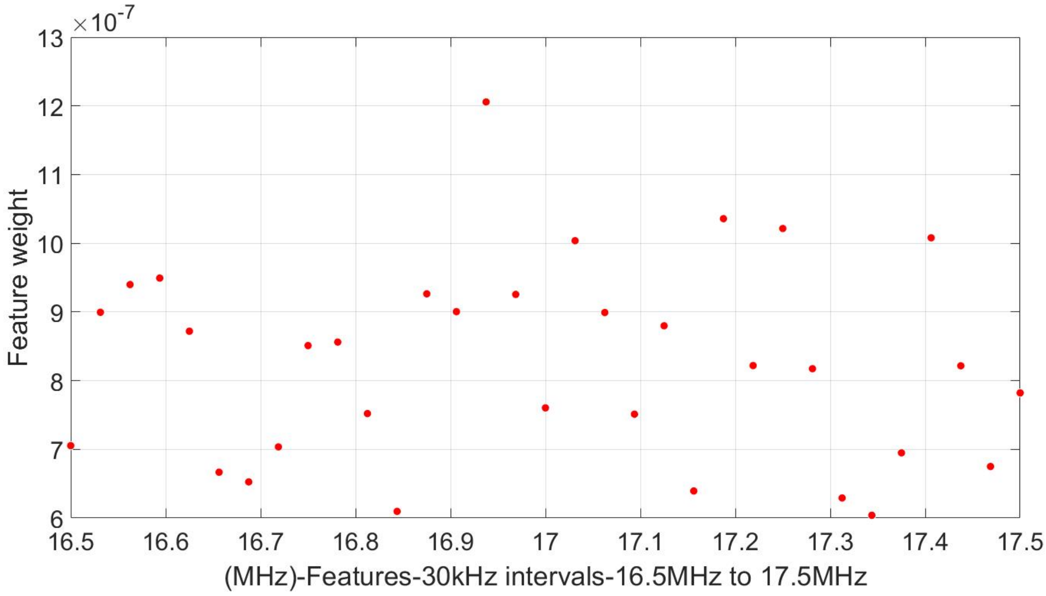

Figure 5.2: Feature selection using NCA in 16.5MHz-17.5Mhz

When threshold value is decided as $8x10^{-7}$, there are 19 features to select which correspond to 570kHZ in total. When threshold value is decided as $7x10^{-7}$, there are 25 features to select which correspond to 750kHZ in total.

Neighborhood component analysis is suitable to use for small bandwidths where variance of input data is relatively small. Total energy decreases proportionally while frequency increases. Therefore, NCA is not suitable for selecting bands in the whole bandwidth. Figure 5.3 and Figure 5.4 prove that NCA is not suitable for selecting features from a wide bandwidth since feature weights decrease proportionally while frequency increases and NCA employs similarity measures to differentiate between input data. Therefore, it is not possible to set a threshold value and select different features. Figure 5.3 shows feature weights obtained by NCA from 15MHz to 23MHz.

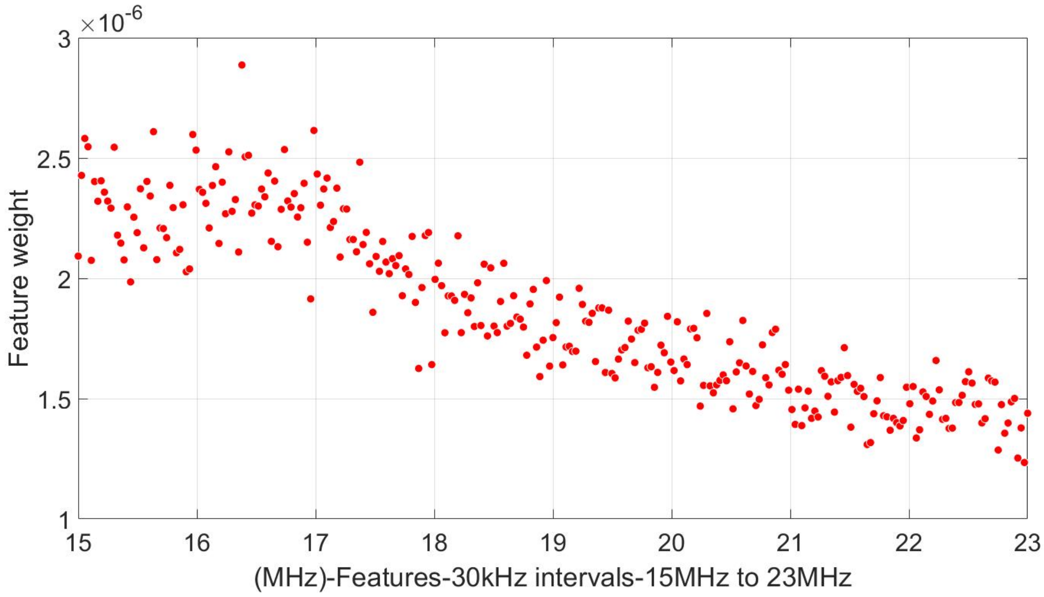

Figure 5.3: Feature selection using NCA in 15MHz-23Mhz

Figure 5.4 shows feature weights from 7MHz to 30MHz obtained using NCA.

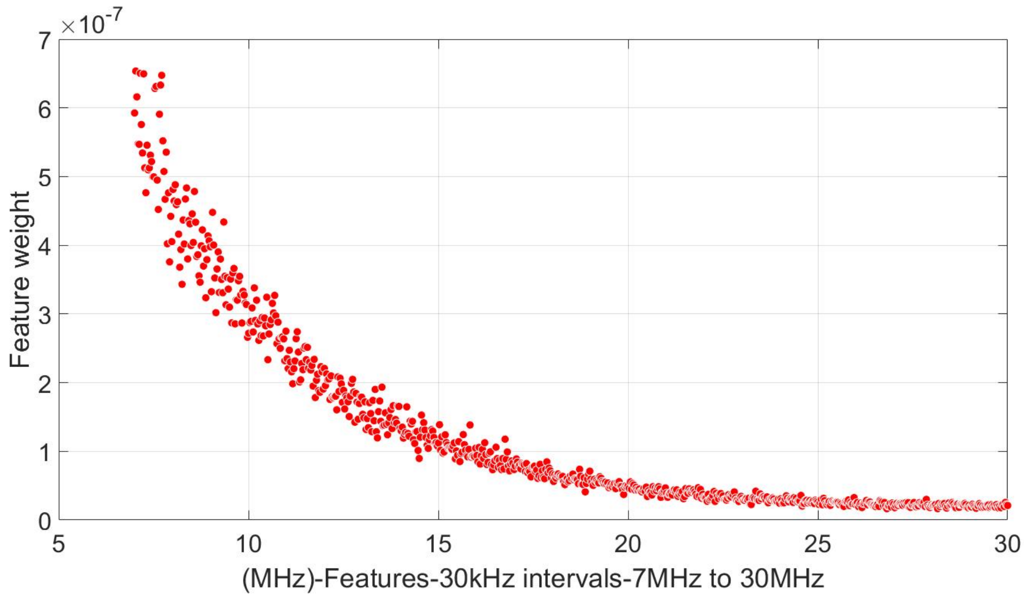

Figure 5.4: Feature selection using NCA in 7MHz-30Mhz

It is concluded that NCA is not suitable for selecting between all the features because of the nature of used dataset where variance in the data is high.

5.1.2 Sequential Feature Selection

Sequential feature selection has a high computational cost because in every iteration of algorithm, it runs a classification training process until given performance criteria is met or performance converges. Also, increasing number of features increase computational cost exponentially. Using sequential feature selection for 1000 features is computationally demanding, therefore, some frequency intervals needs to be excluded. For instance, it is not needed to investigate lower frequencies such as lower than 1MHz. Figure 5.5 shows the most important two features if frequencies lower than 1MHz were not excluded.

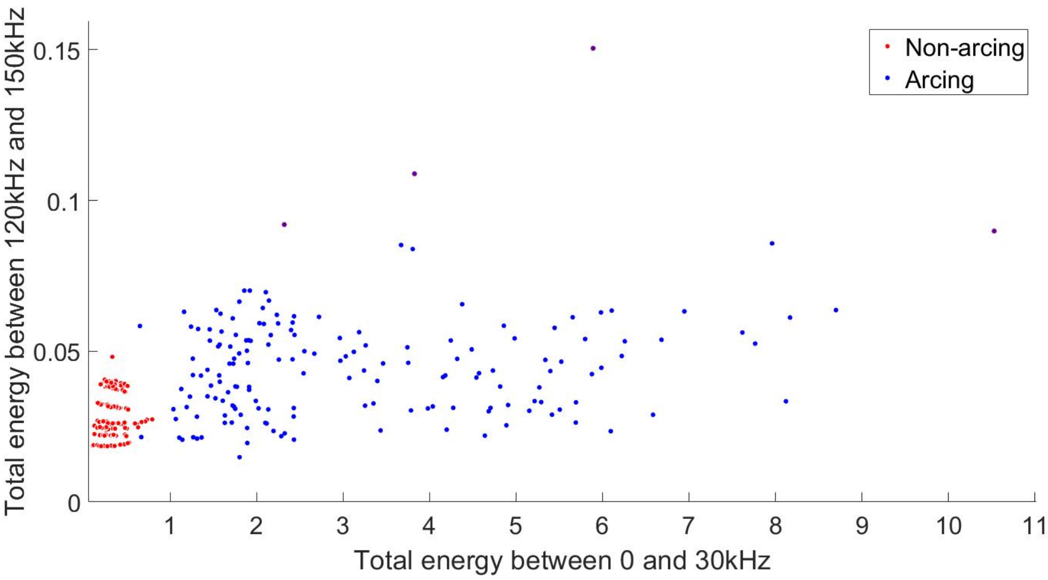

Figure 5.5: Illustration of 2 features(X Axis:Total energy in 0-30kHz, Y Axis:Total energy in 120-150kHz)(Red:Non-arcing, Blue:Arcing)

Figure 5.5 shows the total energy of resistive loads between 0 to 30kHz on the x-axis and total energy between 120kHz to 150kHz on the y-axis. Because of the reasons that are stated in Section 2.2, these intervals are not suitable for detection because of masking conditions and noise generated by other devices, therefore, they have to be excluded for feature selection algorithm. However, these intervals give a good illustration of energy difference in different intervals for arcing and non-arcing situations. If these bands were suitable to use in classification algorithms, they would be very informative and they would be classified with very high classification rates by SVM or other algorithms because difference in energies are even visible by eye inspection in 2D graph.

In order to narrow the scope of bandwidth that feature selection selects the features from, frequency intervals used in literature and know-how from existing devices are helpful to decide initial subsets. In total of 12.4MHz bandwidth was covered using sequential feature selection. 12.4MHz bandwidth corresponds to 413 features in feature space. 5MHz and 17MHz are main frequencies that form the biggest single portions of initial bandwidth with 2.5MHz of bandwidth each(5MHz in total). Other 7.4MHz bandwidth was randomly chosen from 1MHz to 30MHz. When 36 features are selected which correspond to 1.08MHz bandwidth, performance converges with better classification rate than using all bandwidth. Using 26 selected features which correspond to 690kHz, performance is similar to using all bandwidth.

Performance metrics for classifiers that are trained using all bandwidth showed that SVM outperforms all other classifiers. Therefore, sequential feature selection was applied only for support vector machine classifiers with different kernels.

Figure 5.6 shows the frequency interval that most of features are selected using sequential feature selection. There are 66 features shown in Figure 5.6 from 16Mhz to 18MHz. 14 features are selected by sequential feature selection among shown features. All the shown data in Figure 5.6 are the measurements of LEDs while arcing occurs. However, it is important to state that selected features in these interval together with the other selected features in different bands give the best classification results.

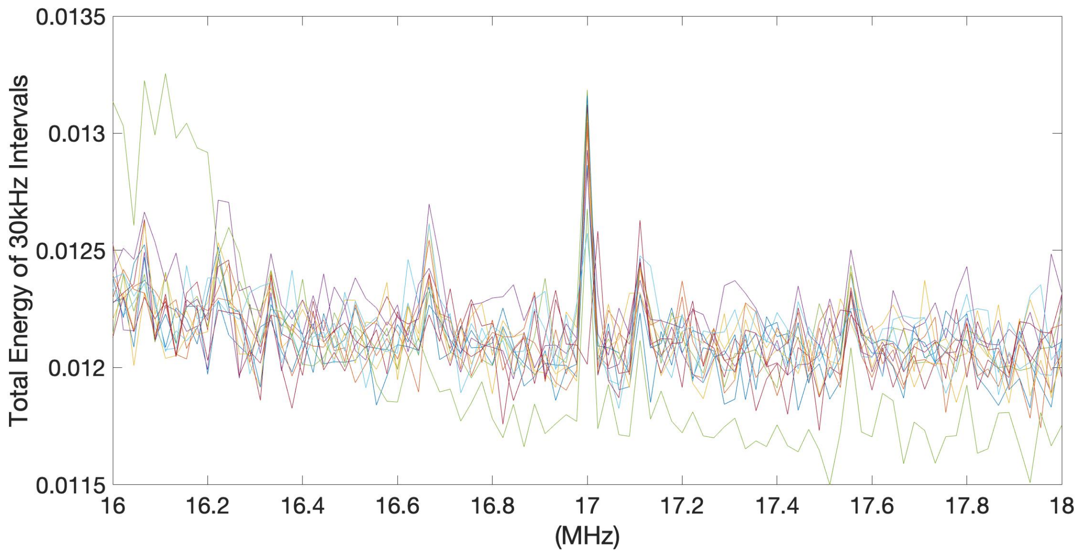

Figure 5.6: Frequency band that most of the features are selected-LEDs while arcing(16Mhz to 18MHz)

5.1.3 Classification Performance Using Selected Features

Trained algorithms using selected features perform better than using all bandwidth because of reduced dimension of feature space. In this way, only informative data for classification are used for training of classifiers. Therefore, performance increased by removing data that cause noise for classifiers. In order to train the algorithms, a total of 1.08MHz bandwidth is used. However, it is important to keep in mind that 1.08MHz bandwidth is not an one-piece band but combination of 36 bands with 30kHz spectrum.

Table 5.13 shows confusion matrix obtained with support vector machine with linear kernel using selected features.

	Condition Positive	Condition Negative
Predicted Condition Positive	22	7
Predicted Condition Negative	5	21

Table 5.13: Confusion Matrix-SVM(Linear Kernel)-Selected Features

Table 5.14 shows confusion matrix obtained with support vector machine with second order polynomial kernel.

	Condition Positive	Condition Negative
Predicted Condition Positive	23	7
Predicted Condition Negative	4	21

Table 5.14: Confusion Matrix-SVM(Polynomial Kernel (degree 2))-Selected Features

Table 5.15 shows confusion matrix obtained with support vector machine with third order polynomial kernel.

	Condition Positive	Condition Negative
Predicted Condition Positive	23	5
Predicted Condition Negative	4	23

Table 5.15: Confusion Matrix-SVM(Polynomial Kernel (degree 3))-Selected Features

Table 5.16 shows confusion matrix obtained with support vector machine with RBF kernel.

	Condition Positive	Condition Negative
Predicted Condition Positive	26	3
Predicted Condition Negative	1	26

Table 5.16: Confusion Matrix-SVM(RBF kernel)-Selected Features

Classifier	ACC	TPR	TNR	PPV	NPV	FPR	FDR	FNR
SVM(Linear)	0.78	0.81	0.75	0.76	0.81	0.25	0.24	0.18
SVM(Polynomial-2nd order)	0.80	0.85	0.75	0.77	0.84	0.25	0.23	0.14
SVM(Polynomial-3rd order)	0.83	0.85	0.82	0.82	0.85	0.18	0.18	0.14
SVM(RBF)	0.93	0.90	0.90	0.96	0.95	0.10	0.10	0.03

Table 5.17: Performance Metrics of Classifiers Using Selected Features

When dimension of feature space is reduced by removing uninformative features and selecting ones for best classification performance by sequential feature selection, performance of the all algorithms increased. Accuracy of SVM with linear kernel increased to 78% from %71. Accuracy of SVM with second order polynomial kernel increased to 80% from %73 percent. Accuracy of SVM with third order polynomial kernel increased to 83% from %75 percent. Accuracy of SVM with radial basis function kernel increased to 93% from %89 percent. Dimension reduction by feature selection caused biggest performance increase in linear and polynomial kernel because feature space is not linearly separable and dimension has an important impact of performance of linear classifiers or SVM with polynomial kernel.

Dimmers used with resistive loads(heaters) are the most wrongly classified load group using all types of classifiers. It is inspected that current waveform of dimmers are similar to resistive loads while arcing occurs. This is the main reason that it is difficult to classify dimmers. LEDs and single phase induction motors are classified correctly in most of the SVMs, especially, SVM with RBF.

SVM with RBF outperformed all other classifiers with a classification of 93% using selected features as input. Figure 5.7 shows receiver operating characteristics of designed support vector machine with RBF kernel using data selected by sequential feature selection.

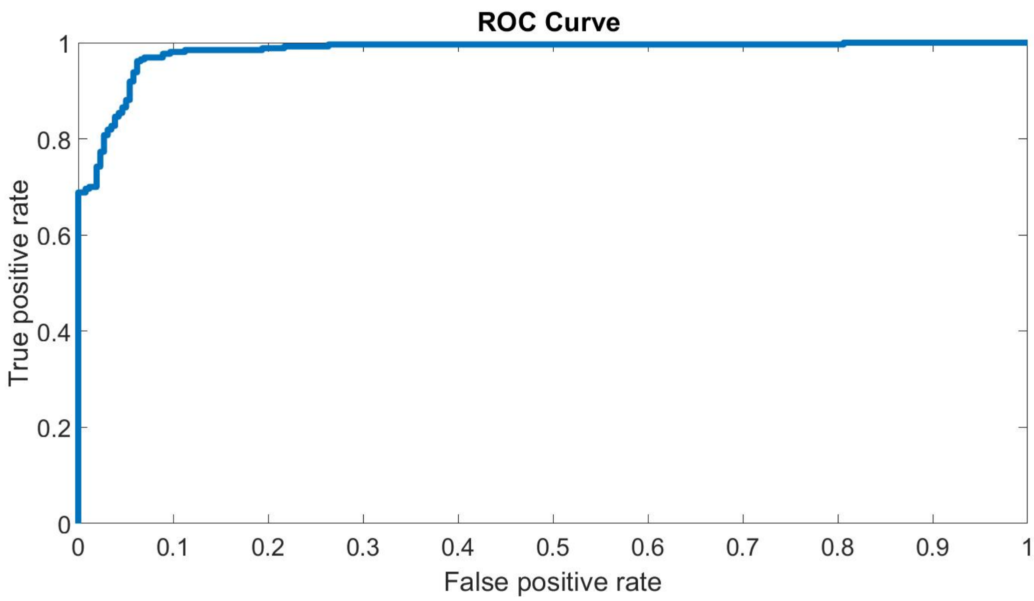

Figure 5.7: Receiver Operating Characteristics of Support Vector Machine with RBF kernel using selected data

Chapter 6

Conclusions

Direct digitization as data acquisition method combined with supervised machine learning algorithms based on features extracted in frequency domain is suitable technique for AC arc fault detection. 5MHz and 17MHz are the best main frequencies to use for detection purposes. Support vector machine with RBF kernel has a classification rate of 93%. Using selected bands to train the algorithms improves classification accuracy considerably. SVM with RBF kernel outperformed SVM with linear kernel by 7%. Classification rates of SVMs with polynomial kernel, for both second order and third order, were 7% and 8% less respectively when RBF kernel was used.

Arc fault detection by using direct digitization was realized in this thesis and feasibility of method is proved by classification rates. Direct digitization is more flexible compared to existing detection methods in terms of possibilities and freedom of signal processing opportunities. Bandwidth is divided into sub-bands of 30kHz and feature selection algorithms were applied. In existing solutions, bandwidths of 30kHz are not feasible to design an analog front end, however, using direct digitization and properly selected 36 features which means a total of 1.08MHz bandwidth, are able to classify arc faults with 93% accuracy.

The advantage of direct digitization is the ability of using relatively small intervals(such as 30kHz). Using such small bands are advantageous because for wider bands, noise characteristics can be masked by another signal and if there is a stochastic or deterministic signal that causes masking, all the information gathered through that wide band becomes unavailable. However, direct digitization prevents this because even if there is a masking situation, it only affects a single band but others still remain informative.

Two different feature selection metaheuristics were applied to select suitable features. It was seen that neighborhood component analysis is not effective when input data has high variance. Since, high frequency components of frequency spectrum are much smaller than lower frequencies, NCA is not able provide effective selection of features for wide bandwidths. However, when variance of input is not high, namely, smaller band is given as input, NCA was able to perform feature selection more efficiently.

Sequential feature selection is effective for choosing suitable bands because algorithm itself takes classifier performance into account in selection process. 1.08MHz of bandwidth(36 features of 30kHz) was chosen by sequential feature selection. By using these features, best classifier's(SVM with RBF) performance was increased to 93%.

Dimmers have the lowest classification rates in the tested loads. It is inspected that current waveforms of dimmers under normal operation are similar to current waveforms of resistive loads while arcing fault occurs. This is the main reason that dimmers has lower classification rates compared to other loads.

LEDs and power sources with active power correlation has the best classification rates and they are classified correctly using SVM with RBF kernel as classification algorithm.

SVM classifiers with different kernels(linear, second order polynomial, radial basis function), k-nearest neighbors classifiers with different k values(2,3,5), decision tree with different limits of splits(maximum, 7,11) and Adaboost algorithm are trained and their performance are compared. Feature space is generated using frequency domain analysis and all feature space was used to train algorithms for comparison purposes. SVM with RBF kernel outperformed all other classifiers with its classification rate(89%) before feature selection. Adaboost also has successful performance metrics, especially, false positive rate. However, accuracy and other performance metrics of SVM with RBF kernel are better than Adaboost. These metrics and generalization ability of SVM make it the most suitable classifier to employ in AC arc fault detection. Therefore, best features are selected to train SVMs by sequential feature selection. In total 1.08MHz that corresponds to 36 features are selected and SVM with RBF kernel has 93% classification rate when it is trained with these selected features.

This thesis proves that using direct digitization as data acquisition method and frequency domain analysis is increasingly feasible method for AC arc fault detection.

Main purposes of this thesis are to design a successful detection algorithm by employing direct digitization and supervised machine learning classifiers. Objective is achieved with a 93% classification rate.

Designed feature selection algorithms form a basis to select features for both designed algorithm in this thesis and for existing analog frontend solutions. However, some modifications are necessary to make for analog frontend solutions. For instance, it would not be feasible to design an analog circuit that has 36 or more band-pass filters whereas it is very simple in direct digitization. Therefore, feature selection algorithm can be modified to choose consecutive bands so that they are usable with analog frontend solutions.

6.1 Future Work

Future work can be summarized as follows:

- Performance of supervised machine learning algorithms is heavily dependant on construction and informativeness of feature space. It can be further increased by increasing the number of measurements, load scenarios, and combinations of loads namely, expanding the feature space with different data.
- Although most of the achieved bandwidth and widely used bands in literature are investigated with sequential feature selection algorithm, covering all the achieved bandwidth can increase the performance of designed algorithm.
- Obtained classifier(SVM with RBF kernel) is optimal by the grid used in grid-search to tune the regarding parameters of algorithm. Therefore, performance can be slightly increased by fine tuning around the achieved parameters.
- Modified accuracy score can be implemented to penalize false positive classifications. Therefore, nuisance tripping can be avoided better by training the algorithms to avoid false positives.
- Feature space for SVM does not have be constructed by the analysis in frequency domain. Analysis and features extracted from other domains such as time-frequency domain(STFT), or wavelet transformation can be used to increase performance and robustness of the designed detection method.
- In order to use the feature selection methodologies offered with existing solutions, it is needed to modify the feature selection algorithm to choose consecutive bands in feature space because approach implemented in this thesis freely chooses different bands without considering their positions. However, such selection of bands would not be feasible for existing solutions.
- Recurrent neural networks, especially long short term memory units which is a special kind of RNN explicitly designed to be capable of learning long-term dependencies[57], are very interesting algorithms to employ for AC arc fault detection. Using deep learning instead of supervised machine learning ease the process of feature selection together with other advantages. Therefore, it should be further investigated to use LSTM networks for AC arc fault detection.

References

[1] Fire Safe Europe, FSEU publications, 2019, https://firesafeeurope.eu/fseu-publications/, online , accessed: 11.05.2019

[2] The Geneva Association World Fire Statistics Bulletin No. 29 Newsletter April 2014

[3] United States Fire Administration Residential Building Electrical Fires (2014-2016),2018, Topical Fire Report Volume 19 Issue 8 Newsletter April 2014

[4] United States Fire Administration, National Estimates Methodology for Building Fires and Losses, August 2012, https://www.usfa.fema.gov/downloads/pdf/statistics/res_bldg_fire_estimates.pdf

[5] Transparency Market Research, Global Arc Fault Circuit Interrupter Market: Booming Construction Industry in Middle East, Asia Pacific to Drive Growth, 2017, https://www.transparencymarketresearch.com/pressrelease/arc-fault-circuit-interrupter-market.htm

[6] International Electrotechnical Commission IEC:62606 General requirements for arc fault detection devices: Edition 1.0. Copyright 2017 IEC Geneva, Switzerland. www.iec.ch Switzerland, 2013-07

[7] Wikipedia, Arc-fault circuit interrupter *https://en.wikipedia.org/wiki/Arc-fault_circuit_interrupter*. accessed: 20.04.2019.

[8] Electrical circuit protection for low voltage residential and light commercial installations *http://promotions.eaton.com/content/dam/beacon/residential/UK/home/for-installers/fire-safety-season/circuit-protection/brochure/AFDD+-range-brochure.pdf*. accessed: 16.02.2019.

[9] Jean-Mary Martel, Series arc faults in low-voltage AC electrical installations, Universittsverlag Ilmenau 2018

[10] Hong-Keun Ji, Guoming Wang, Woo-Hyun Kim, Gyung-Suk Kil, Optimal Design of a Band Pass Filter and an Algorithm for Series Arc Detection,2018, Energies 11(4):992 DOI: 10.3390/en11040992, licence by: CC BY 4.0

[11] Wikipedia, Cross-validation (statistics), https://en.wikipedia.org/wiki/Cross-validation_(statistics), online, accessed:10.01.2019

[12] F. Blades, "Method and apparatus for detecting arcing in AC power systems by monitoring high frequency noise", United States of America Patent US 5729145 A 17 March 1998.

[13] Wikipedia, Fuse(electrical) *https://en.wikipedia.org/wiki/Fuse(electrical. accessed* : 12.04.2019.

[14] Low Voltage Protection-RCD *https://electrical-engineering-portal.com/residual-current-device-rcd*. accessed: 26.04.2019.

[15] Wikipedia, Residual-Current Device, https://en.wikipedia.org/wiki/Residual-current_device accessed: 27.04.2019.

[16] Bob Dean, Residual-current device *https://www.electricalengineeringtoolbox.com /2016/01/how-residual-current-device-rcd-works.html*. accessed: 27.04.2019.

[17] Eaton UK, Miniature Circuit Breaker *http://www.eaton.uk.com/EatonUK/Products Solutions /Electrical/ProductsServices/ResidentialLight-Commercial /CircuitProtection/EatonResidentialRange6A/index.htm*. accessed: 09.04.2019.

[18] Martin M, Understanding the differences between ground fault and arc fault circuit interrupters *http://www.maximumspaces.com/understanding-the-differences-between-ground-fault-and-arc-fault-circuit-interrupters.html*. 12.02.2018 accessed: 12.04.2019.

[19] Eaton s.r.o, Arc fault detection device brochure, http://www.eatonelektrotechnika.cz/cz/obloukova-ochrana-afdd.html. February 2018

[20] Jean-Francois Rey, Do you know an Arc Fault Detection Device (AFDD) can prevent from an electrical fire? *http://https://blog.schneider-electric.com/power-management-metering-monitoring-power-quality/2013/07/03/do-you-know-an-arc-fault-detection-device-afdd-can-prevent-from-an-electrical-fire/*. 03.07.2013 accessed: 11.05.2019.

[21] Amna Farooq Husain, Series Arc Fault Detection in the Presence of Household Electrical Loads 2017 Czech Technical University

[22] A. Collin, J. Acosta, B. Hayes and S. Djokic, Component-based Aggregate Load Models for Combined Power Flow and Harmonic Analysis 7th Mediterranean Conference and Exhibition on Power Generation, Transmission, Distribution and Energy Conversion 2010

[23] IEC 61000-3-2:2014, Electromagnetic compatibility (EMC) - part 3-2: Limits for harmonic current emissions (equipment input current :s 16 A per phase)," 2007.

[24] Wikipedia, Binary Classification, https://en.wikipedia.org/wiki/Binary_classification , *online*, accessed: 01.02.2019

[25] UK Department of Energy and Climate Change, Energy Consumption in the UK (2013), Domestic energy consumption in the UK between 1970 and 2012 25.07.2013

[26] International Energy Agency, Tracking Clean Energy Progress-Lighting, https://www.iea.org/tcep/buildings/lighting/ 13.03.2019

[27] Viribright, Comparing LED vs CFL vs Incandescent Light Bulbs, 2016, https://www.viribright.com/lumen-output-comparing-led-vs-cfl-vs-incandescent-wattage/, online, accessed: 17.05.2019

[28] Gabriel Brindusescu, Better Electric Motors to Arrive? Scientists Discovered New Magnetic Properties in Iron, 2014, https://www.autoevolution.com/news/better-electric-motors-to-arrive-scientists-discovered-new-magnetic-properties-in-iron-80619.html

[29] Science Direct https://www.sciencedirect.com/topics/computer-science/binary-classification accessed: 11.05.2019.

[30] Hoffmann, R.; Fabian,T.; Dr, F., 2008. Comparison of yields and nutritive value of different spring green forage mixtures. Acta Agric. Slov., 92 (suppl 2): 143-148

[31] Polikar, R.: Ensemble based systems in decision making. IEEE Circuit Syst. Mag. 6, 21-45,DOI: 10.1109/MCAS.2006.1688199, October 2006

[32] Yoshua Bengio, Learning Deep Architectures for AI,Foundations and Trends in Machine Learning, 2009, ISBN: 978-1-60198-294-0

[33] Drucker, H., D. Wu and V.N. Vapnik, 1999. Support Vector Machines for Spam Categorization. IEEE Transaction on Neural Networks

[34] Yoav Freund and Robert E. Schapire, A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting, Journal of computer and system sciences 55, 119-139 (1997)

[35] Friedman N, Linial M, Nachman I, Pe'er D, Friedman N, et al. (2000) Using Bayesian networks to analyze expression data. J Comput Biol 7(3-4):601-20

[36] Breiman, L., Random Forests, Machine Learning (2001) 45: 5. https://doi.org/10.1023/A:1010933404324

[37] Bishop DV, North T, Donlan C.,Genetic basis of specific language impairment: evidence from a twin study. Dev Med Child Neurol. 1995 Jan;37(1):56-71.

[38] F Fernndez-Rodrguez, S Sosvilla-Rivero, Testing nonlinear forecastability in time series: theory and evidence from the EMS,1999 Economics Letters 59, 49-63

[39] Van Gestel, Tony, et al., 2001. Financial Time Series Prediction Using Least Squares Support Vector Machines Within the Evidence Framework, IEEE Transactions on Neural Networks, Volume 12, Number 4, July 2001, Pages 809-821

[40] K. Veropoulos, N. Cristianini and C. Campbell, The Application of Support Vector Machines to Medical Decision Support: A Case Study, 2009

[41] Azriel Rosenfeld, Harry Wechsler, Pattern Recognition: Historical Perspective and Future Directions, 2000

[42] Wikipedia, Kernel method, https://en.wikipedia.org/wiki/Kernel_method, online, accessed: 12.01.2019

[43] Li H., Qi F., Wang S. (2005) A Comparison of Model Selection Methods for Multiclass Support Vector Machines. In: Gervasi O. et al. (eds) Computational Science and Its Applications ICCSA 2005. ICCSA 2005. Lecture Notes in Computer Science, vol 3483. Springer, Berlin, Heidelberg

[44] Laura Auria, Rouslan A. Moro, Support Vector Machines (SVM) as a Technique for Solvency Analysis, 2008, German Institute for Economic Research Discussion Papers

[45] Wikipedia, Generalization error, https://en.wikipedia.org/wiki/Generalization_error, accessed: 21.03.2019

[46] Wikipedia, Fourier Transform,https://en.wikipedia.org/wiki/Discrete_Fourier_transform, accessed: 01.05.2019

[47] Grichting, B.; Goette, J.; Jacomet, M.: Cascaded fuzzy logic based arc fault detection in photovoltaic applications. 2015 International Conference on Clean Electrical Power (ICCEP), 2015, (178183), doi:10.1109/ICCEP.2015.7177620.

[48] Cao, Y.; Li, J.; Sumner, M.; aj.: Arc fault generation and detection in DC systems. In 2013 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC), IEEE, 2013, s. 15, doi:10.1109/APPEEC.2013.6837123.

[49] Yuan Gao ; Junjun Zhang ; Yandan Lin ; Yaojie Sun, An Innovative Photovoltaic DC Arc Fault Detection Method through Multiple Criteria Algorithm Based on a New Arc Initiation Method, 2014, IEEE 40th Photovoltaic Specialist Conference (PVSC), IEEE, doi: 10.1109/PVSC.2014.6925613

[50] Gab-Su Seo, Katherine A. Kim, Kyu-Chan Lee, Kyung-Jun Lee, Bo-Hyung Cho, A New DC Arc Fault Detection Method Using DC System Component Modeling and Analysis in Low Frequency Range, 2015,IEEE Applied Power Electronics Conference and Exposition (APEC), doi: 10.1109/APEC.2015.7104690

[51] Mathworks, https://www.mathworks.com/help/stats/neighborhood-component-analysis.html, online, accessed: 04.05.2019

[52] Wei Yang, Kuanquan Wang, Wangmeng Zuo, Fast neighborhood component analysis, 2012 Neurocomputing Volume 83,(31-37), https://www.sciencedirect.com/science/article/pii/S0925231211007120

[53] J. Goldberger, G. E. Hinton, S. T. Roweis,R. Salakhutdinov, Neighbourhood components analysis, in Advances in neural information processing systems, 2004, pp. 513520.

[54] Dong Wang, Xiaoyang Tan, Bayesian Neighbourhood Component Analysis, 2017, IEEE Transactions on Neural Networks and Learning Systems(Volume: 29 , Issue: 7),pp. 3140 - 3151 doi:10.1109/TNNLS.2017.2712823

[55] Vipin Kumar, Sonajharia Minz, Feature Selection: A literature Review, 2014, Smart Computing Review, vol. 4, no. 3, June 2014,pp. 211-229

[56] Trevor Hastie, Robert Tibshirani, Jerome Friedman, The Elements of Statistical Learning Data Mining, Inference, and Prediction, 2001, pp. 38

[57] Petr Pok, BE3M33UI Artificial Intelligence Lecture Slides Non-linear models, Basis expansion, Overfitting, Regularization, 2019, Czech Technical University in Prague Faculty of Electrical Engineering Dept. of Cybernetics

[58] Fletcher, Robert H. Fletcher ; Suzanne W. (2005). Clinical epidemiology : the essentials (4th ed.). Baltimore, Md.: Lippincott Williams Wilkins. p. 45. ISBN: 0-7817-5215-9.

[59] Wikipedia, Confusion Matrix, https://en.wikipedia.org/wiki/Confusion_matrix, online, accessed:10.05.2019

[60] Mathworks, "Detector Performance Analysis Using ROC Curves MATLAB Simulink Example", www.mathworks.com, 2016, online, accessed: 14.05.2019

[61] Siemens , "5SM6 AFD Units - Technology Primer," Siemens AG , 2016.

[62] Scikit-learn, RBF SVM parameters, https://scikit-learn.org/stable/auto_examples/svm/plot_rbf_parameters.html, online, accessed: 19.05.2019

[63] Siemens, Sentron 5SM6 AFD units Technology Primer, https://assets.new.siemens.com/siemens/assets/public.1492678703.d8eedd48f 65bce75615cabcc29ccfa5b19b85cfc.technikfibel-brandschutz-en.pdf, online, accessed: 15.05.2019

[64] Altman, N. S. (1992). "An introduction to kernel and nearest-neighbor nonparametric regression". The American Statistician. 46 (3): 175185. doi:10.1080/00031305.1992.10475879. hdl:1813/31637.

[65] Wikipedia, k-nearest neighbors algorithm, https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm, online, accessed: 21.05.2019

[66] Jason Brownlee, A Gentle Introduction to k-fold Cross-Validation, 2018, Statistical Methods, online, accessed: 11.04.2019