Cracow University of Technology
Department of Electrical and Computer Engineering
Institute of Electromechanical Energy Conversion, E-2

MULTICRITERIA DIAGNOSIS OF SYNCHRONOUS MACHINES.

DOCTORAL DISSERTATION
BY
José Gregorio Ferreira De Sá

Supervisor: dr hab. Adam Warzecha

Kraków, December 2016

DEDICATION

Thanks to my family for supporting every initiative, experiment, new trial and error…
thank you for supporting my insatiable curiosity
and the need to find new ways to solve the next puzzle.

Thanks to science, which gives me the excuse of doing what I love; finding answers.

"What would life be if we had no courage to attempt anything?"
— Vincent van Gogh

"Whatever the mind of man can conceive and believe, it can achieve."
— W. Clement Stone

DECLARATION

This dissertation is the result of my own work and includes nothing, which is the outcome of work done in collaboration except where specifically indicated in the text. It has not been previously submitted, in part or whole, to any university of institution for any degree, diploma, or other qualification.

Signed: _____

Date: _____

José Gregorio Ferreira De Sá

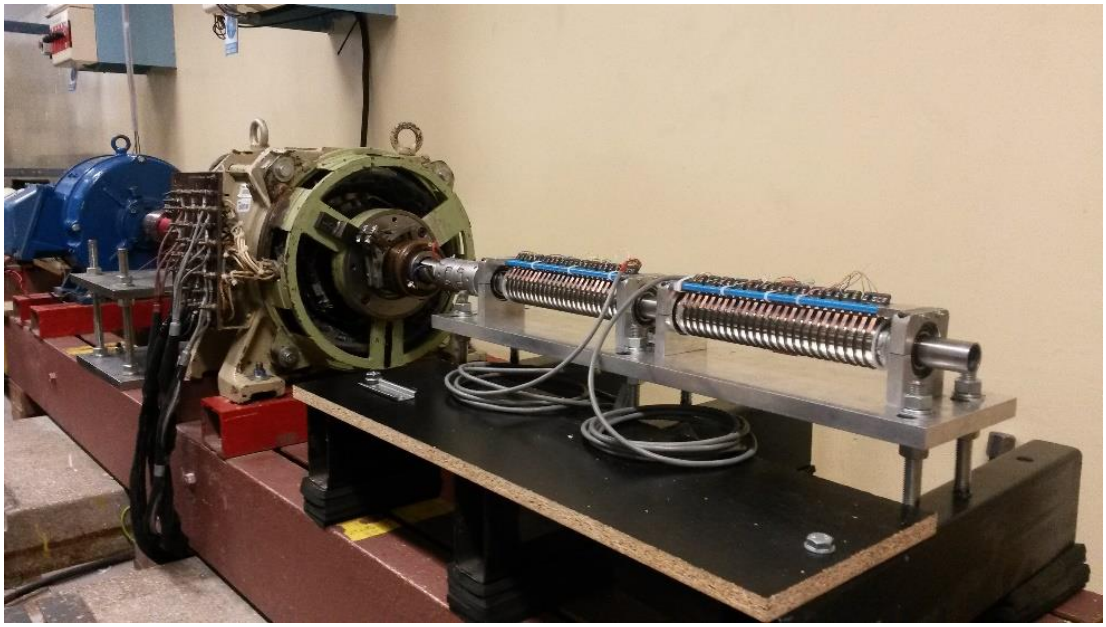
PH.D. THESIS SUMMARY.

Develop a smart multivariate method capable of early fault detection is the aim of this experimental research. It is intended to identify which of the available signals are best suited for monitoring and classification of faults in electrical machines.

It is strongly believed that this method of acquiring data would allow timely and reliable detection of faults, even further when used in combination with an automated monitoring algorithm or process.

This research tends to deal with all the physical phenomena through practical experiments by developing a reliable and safe experimental platform. This platform was specially designed and adapted to carry out various non-invasive tests at different levels of severity. Then, the most optimal group of features is identified and used to assess the condition of the machine through machine learning classification algorithms.

A methodology capable of classifying ten different machine conditions, with an accuracy of 99.5% is defined, and all the steps necessary for its implementation are carefully explained.



Laboratory of Electrical Machines, Institute of Electromechanical Energy Conversion.
Cracow University of Technology

ACKNOWLEDGEMENTS

Financial support from the Marie Curie FP7-ITN project "Energy savings from smart operation of electrical, process and mechanical equipment– ENERGY-SMARTOPS," Contract No: PITN-GA-2010-264940 is gratefully acknowledged.

All the measurements were made in Laboratory of Electrical Machine at the Institute of Electromechanical Energy Conversion, Cracow University of Technology. The synchronous machine used in this research was specially constructed to perform fault diagnosis test and made available by prof. K.Weinreb.

CONTENT

LIST OF TABLES

1 INTRODUCTION

In this first chapter, the motivations driving this research are explained. In addition, direct applications over more efficient electric drives are discussed and used as justification to support the condition based monitoring strategy established in the following chapters.

1.1 RESEARCH STATEMENT

Principally used in high power operations and widely used as alternating current generators, Synchronous Machines (**SM**) can be found where constant speed operation, power factor control or high operating efficiency (or both) are required. They are doubly excited machines, brushless or supplied by an external DC source; two electrical inputs are necessary during normal operations. Its higher maintenance costs and high initial capital are the reasons why synchronous machines are underestimated over induction motors in general low-power requirements industrial applications.

Induction motors, favored for their simplicity and ruggedness, are extensively used in general industrial applications. The number of articles devoted to fault detection and condition monitoring technique **[71], [21]** is more meaningful and specialized than the number of papers dedicated to synchronous machines. The number of publications devoted to fault diagnosis in synchronous machines has led to literature related to permanent magnets wind turbine's generators or motors, synchronous generator, and less often, to synchronous reluctance motors **[3][40]**.

For ratings exceeding 0.75 kW per rpm **[49]**, a rule of thumb is that SMs are less expensive than squirrel-cage motors. In Figure 1-1, and for traditional construction motors, it is possible to appreciate in which broad areas of applications either synchronous are more suitable induction motors.

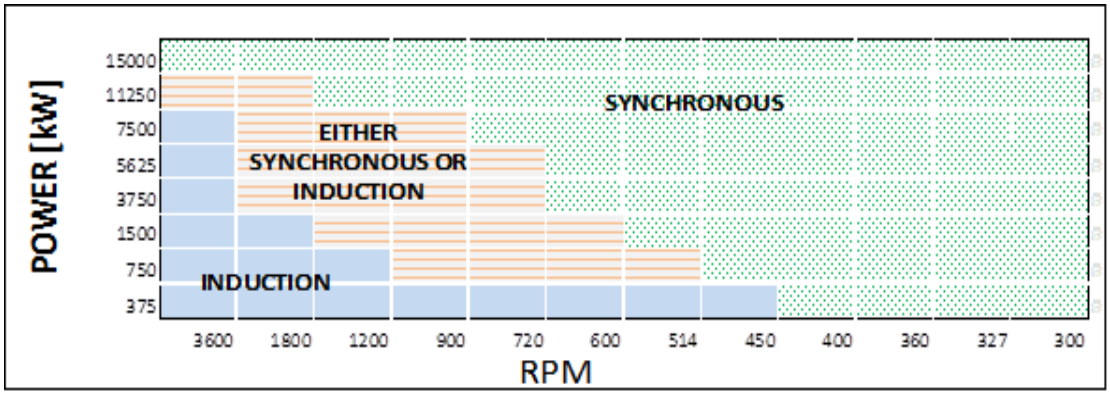


Figure 1-1. General areas of application of synchronous and induction motors

Synchronous machines are used for Power factor correction as well and thus improve energy efficiency. In high power, and heavy load applications synchronous motors are increasingly the motor of choice **[73]**. In addition to this, in applications that require power factor correction, synchronous motors also provide high torques and constant speed under load variation, resulting in low operating and maintenance costs. It is an important characteristic that should be taken into account when selecting it for industrial applications.

All these advantages explain the use of the SM in various applications as large compressors and fans in chemical and petrochemical applications, water and wastewater plants, fans pumps and compressors in steel plants, mills and crushers in cement works and pulp and paper extruders.

Perhaps, all these distinct features have made it more complicated, expensive and less accessible to carry out experimental tests on fault detection in SM. These are the main reasons why for this research a specially designed SM and data acquisition system were built. With this system, it is possible to perform and test different non-intrusive faulty conditions on synchronous motors.

Only a traditional brush/ring SM construction type is considered, focused on researching and developing practical solutions to asset the condition of the motor, and when possible, prevent incipient failures and reduce operational cost.

The need to reduce production costs and expand the return on investment, and to reduce CO_2 emissions is leading to new ways of operating existing processes **[58]**.

Industry and the academia are working extensively finding new opportunities for energy savings.

Rotating machinery and electrical equipment are becoming more integrated giving new possibilities for energy saving through equipment management, automation, and optimization **[61]**. Motors are often part of a complex industrial production system. Energy-efficient electric drives should be well dimensioned, and the control strategy designed according to load requirements **[80]**. The need for more integrated systems is leading to new industrial processes and new ways of operating existing processes. Integrated systems require a continuous flow of information that allows monitoring of their components.

Under these premises, condition monitoring systems are becoming more accurate, less expensive **[14]**, and a fundamental component of modern electric drive systems. Their applications in complex systems allow the definition of the integrity of the system required for long-term scheduling. Loading and maintenance strategy are the two primary tasks that directly link the information provided. Trying to minimize the number of unexpected shutdowns, understand the behavior and/or early detection of faults are desirable in every system. Therefore, it seems clear and reasonable, to develop systems that inform about the condition of the components of a system allowing a better operation and thus, an improvement in efficiency.

1.2 SYNCHRONOUS MACHINE, FAULT DIAGNOSIS

As of January 31st of 2016, the number of publications devoted to electric machines is 17480 in IEEE associated conferences, journals, and magazines. Taking advantage of the online IEEExplore publications search engine, was possible to obtain the list of publications related to fault diagnosis and signal processing, concerning electric machines.

After exporting the results and processing the number of IEEE keywords related to each publication, it was possible to estimate the trends related to the study of electric machines, by listing the top 25 most frequents words per topic. The main physical phenomenons studied are the “torque,” “voltage” and the “magnetic flux,” with superior attention to “permanent magnet motors.” From the Table 1-1, we can

find other topics associated with electrical machines such as "control systems," "hybrid electric motors," "DC motors," etc., out of the scope of this research.

Table 1-1. Electrical machines related publications topics

	Electric Machines					
		17480	Fault Diagnosis	1146	Signal processing	842
	Terms	Freq%	Terms	Freq%	Terms	Freq%
1	rotors	9.51%	circuit faults	13.40%	neural networks	8.04%
2	induction machines	8.96%	neural networks	12.20%	vibrations	6.50%
3	torque	8.17%	signal analysis	8.30%	pulse width modulation	4.96%
4	stators	6.19%	vibrations	7.10%	wavelet analysis	4.85%
5	harmonic analysis	6.18%	wavelet analysis	6.70%	torque control	4.74%
6	voltage	4.84%	bars	6.30%	feature extraction	4.30%
7	synchronous machines	4.64%	feature extraction	4.80%	hardware	4.30%
8	stator windings	4.43%	machinery	3.60%	velocity control	3.85%
9	magnetic flux	3.68%	current measurement	3.30%	fuzzy logic	3.74%
10	permanent magnet motors	3.55%	fuzzy logic	3.30%	machinery	3.74%
11	finite element methods	3.47%	vibration measurement	3.10%	machine vector control	3.63%
12	testing	3.05%	shafts	2.60%	sensorless control	3.63%
13	control systems	3.03%	training	2.10%	current measurement	3.52%
14	fault detection	3.01%	insulation	2.00%	motor drives	3.52%
15	reluctance motors	2.94%	support vector machines	2.00%	voltage control	3.52%
16	mathematical model	2.83%	data mining	1.90%	circuit faults	3.41%
17	equations	2.66%	switches	1.90%	digital signal processors	3.41%
18	windings	2.59%	laboratories	1.90%	costs	3.19%
19	coils	2.57%	rotating machines	1.90%	parameter estimation	3.08%
20	magnetic analysis	2.57%	stator cores	1.80%	shafts	3.08%
21	iron	2.43%	power generation	1.70%	switches	3.08%
22	hybrid electric vehicles	2.26%	transient analysis	1.70%	vibration measurement	2.86%
23	inverters	2.21%	vectors	1.60%	instruments	2.75%
24	dc motors	2.13%	artificial intelligence	1.60%	laboratories	2.75%
25	costs	2.11%	costs	1.60%	power generation	2.75%

There are twice more publications related to "induction machines" than to "synchronous machines," and the principal methodology applied in these publications are "harmonic analysis" and using "finite elements methods." Both methodologies tend to be time and processing cost expensive. Neglecting the keywords "stator" and "rotor," the main machine components analyzed are the "stator winding" and "coils," as well as the "iron."

Analyzing the specific keywords associated with fault diagnosis, “circuit faults,” “vibrations” and rotor “bars” were the most studied type of fault in these publications. Although the applicability of the so-called advanced processing methods, such as “neural networks,” “wavelet analysis” and “fuzzy logic,” became widely available during the last decade, the number of publications devoted to these methods is considerably higher than traditional methods.

When looking at the trends associated with signal processing, apart from the methods discussed before, other signal processing methods used in these publications are focused on control applications such as “pulse width modulation,” “torque control,” “velocity control,” “machine vector control,” etc.

Traditional electric drives monitoring systems are based on three phases voltage, current regulatory systems and mechanical interactions such as vibration, wear-out, and leakages. In more critical systems, torque and velocity measurements are collected as well. Without question, assessing the condition of the machine requires robust and accessible processing capabilities; it is always a challenge to identify all the factors that better describe the state of the system.

1.3 WINDING FAULTS

Overheating, voltage peaks and unbalance are the causes, about eighty percent (80%), of winding damage in electrical motor failures. Either as a motor or generator, salient poles synchronous machines winding failures can appear at any sudden time such as the winding break of a parallel branch or a short circuit of a section of the field winding. Hence, the need for a reliable and fast winding fault detection system. The authors in **[84]** have developed a mathematical model to study the winding failure of a salient pole synchronous machine calculating the spectra of the stator and field current.

Condition monitoring systems are not new and have been studied for decades. Among others, in this research, the contributions in condition monitoring of rotating electrical machines, published in **[70][71][69]** and in the novel introduced data-driven techniques **[92], [93]** were studied and reviewed in detail. As well as the contributions published in the development of accurate mathematical models of electrical machines and diagnosis techniques **[54], [67],[66]**

The detection of stator or rotor winding short-circuit has already been analyzed for several years. In 1996, a novel method for detecting short circuits in both the stator and rotor windings of synchronous generators had been proposed to detect changes in the harmonic content of the rotor and the stator current spectrum. More recently, new diagnostic tools to determine when significant insulating aging has occurred were introduced such as polarization-depolarization current, dielectric spectroscopy and on-line leakage current monitoring [68].

Searching for documentation associated with the stator or rotor short-circuit fault, or both, most of the available literature relate to the monitoring of failures in synchronous generators [31][60] or in permanent magnet SM [96]. In [45] the effects of stator interturn short-circuit are analyzed using the field current. The same author in [46], added a rotor search coil for voltage signature analysis, to identify winding short-circuit in synchronous motors. In [24] the authors have used spectral analysis of stator current using Park's vector to detect an intern-turn fault in the Synchronous motor.

A review of different fault conditions included broken rotor bars and end-rings rotor faults, together with the typical harmonic content can be found in [43]. The authors in [51] have designed a search coil, wounded around the shaft, and analyzed the fault conditions looking at the axially directed fluxes; so-called leakage flux monitoring. A practical implementation of two Neural Networks was introduced in [35], in which the authors have a satisfactory classified rotor, rolling bearing and stator faults in induction motors.

1.4 MULTICRITERIA DIAGNOSIS

The development of a fresh approach to multicriteria analysis derives from rapid developments in computational power and equipment and more accurate and efficient algorithms. Multicriteria analysis aims to compare different features extracted from a group of transducer installed on the machine, according to a variety of criteria.

Mathematical models are an abstraction of physical objects. The use of different mathematical approaches to model electric drives for fault diagnosis was found in an extensive number of publications and books. The authors in [29] have designed a methodology based on parameter estimation and knowledge processing using

static and dynamic models of machines for different operation modes, symptom generation and/or model-based fault detection analytical redundancy. In [84][13][85] the authors have contrasted the theoretically derived formulas of the salient poles synchronous machine with the Fourier spectrum of measured currents and identifying additional frequencies of additional harmonics due to machine windings inherent asymmetries.

Efficient supervision, fault detection, and diagnosis of faults, considering causal fault-symptom relationships together with advanced methods for fault were studied in the book [30]. Another model based publication, for induction motors, considering a fast computational method to perform on-line monitoring can be seen in [63]. The same principal author in [65] provides examples of the possibilities and limitations of using frequency analysis and mathematical models in Faulty machines.

An extended review of MCSA describing the different types of fault and the signatures they generate and their diagnostics schemes is presented in [44]. The authors in [5], described the recent trends in condition monitoring and fault diagnosis of rotating machinery and their interactions with the process they belong.

All these publications share a set of assumptions that hardly appears in complex and composed of many interrelated components real-world systems. Each machine should be identified and classified according to the environment in which they are located. Accurate fault diagnosis systems should consider the historical performance and capable of assessing the current state of the machine.

Apart from specifics of the model, on each machine appears variances related to manufacturing, installation, operation and maintenance. In this sense, this research focuses on experimental procedures to create a multicriteria methodology that leads to a rapid assessment of the machine, and under the steady condition, measure the nominal operation, so any deviation from this might be categorized as abnormal behavior and, when demonstrated, as a fault condition.

The authors in [62] have presented a new concept to diagnose induction motors remotely. The so-called Distributed system for diagnostics of inductions motors takes advantage of the latest sensor technologies and telecommunication

architectures to acquire and process motor current signals, then, transmit the condition of the motor to a remote location.

The process adopted here can be seen in Figure 1-2, in the following chapters it will be described, and its applicability demonstrated.

Figure 1-2. Data-Driven, condition based maintenance.

To establish the base that leads to an original contribution to knowledge, for this experimental research I have designed a new system that allows monitoring of 37 different types of signals, described in Chapter 2. These additional signals and the features extracted from them constitute the multicriteria framework through which the assessment of the system will be obtained and used as a base on fault diagnosis.

2 EXPERIMENTAL PLATFORM

In this second chapter, all the components of the experimental platform are described. The platform is installed in the Laboratory of Electrical Machine, Institute of Electromechanical Energy Conversion, Cracow University of Technology. It is designed to the research of internal faults of synchronous machines.

In experimental studies, it is essential to have an exact number of variables under consideration and control; this is why I meticulously designed and tested the platform. Over two years, several configurations and sensor types were subjected to different tests to guarantee the reproducibility of the experiments and decrease, as much as possible, the variability of all the collected signals.

2.1 THE SYNCHRONOUS MACHINE

The object of study is a synchronous machine specially designed and adapted to carry out various non-invasive tests and under different levels of severity. It is a four Salient Poles Synchronous Machine (SPSM), externally excited designed and built in Saint Petersburg. The constructional details of the machine were taken from the documentation made available by prof. K.Weinreb.

The geometrical dimensions have been extracted from the original documentation and updated with the current stator and rotor winding

configuration, from all the available configurations. It is important to take into account that this machine has suffered several modifications over the years.

This machine has the particular characteristic that its phases are externally accessible from a panel terminal, which it is used to interchange the connections that comprise it. It is possible to test serial or parallel branches winding configuration using one of two layers of coils. The main parameters are:

Table 2-1 Machine parameters

Power [kW]	7.5
Voltage [V]	400
Phase winding voltage	230
Current [A]	13.5
Rotational speed [rpm]	1500
Power factor	0.8
Frequency [Hz]	50
Field current [A]	8.56
Field voltage [V]	50
Efficiency [%]	85.2
Mass [Kg]	263
Protection level	IP23
Cooling	IS01
Assembly	IM100

Main components design schemes of the SM were redrawn in Figure 2-1, turns carrying current out of the page are considered as negative. In the following sections geometrical features of the machine were carefully measured and verified, as well as the description of the rotor, so the reader can repeat and validate the results presented here.

Moreover, when building a Finite Element Model (**FEM**), it is critical to have the definition of all machine dimensions, winding configuration, as well as the type of materials used in its construction.

The range and dynamism of all the variables under considerations, under the faulty and healthy condition, are recognized through the model. Then, for each physical magnitude, a specific transducer is selected and installed, so the signals can

be recorded using the appropriate data acquisition systems, as described in Chapter 3.

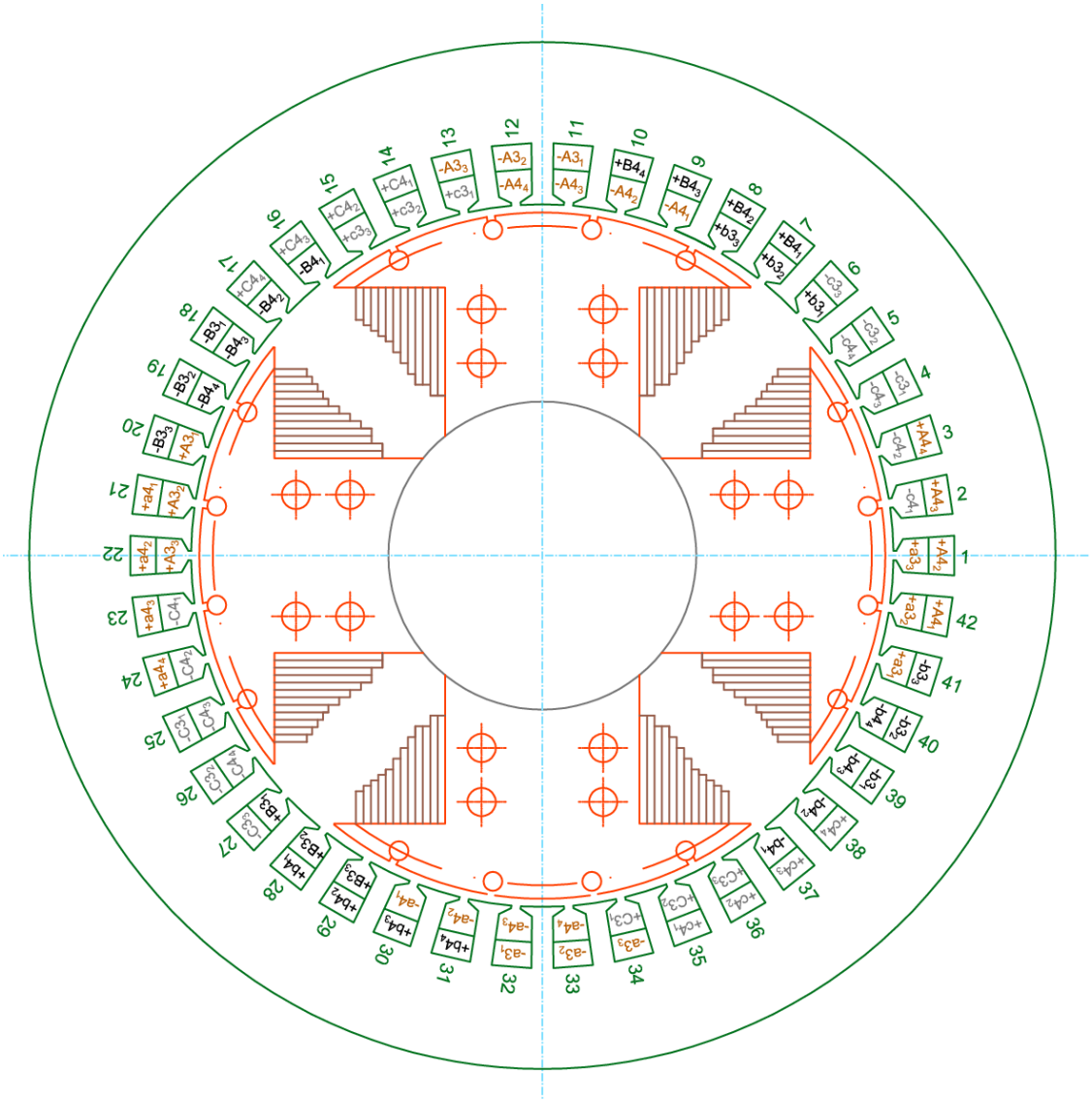


Figure 2-1. Cross section of the magnetic core and stator winding distribution of the tested machine.

2.2 DESCRIPTION OF THE STATOR AND ROTOR

Traditional a.c. machines' stator consists of several coils in each phase distributed in slots around the inner surface of the stator. In Figure 2-1, it is possible to observe the distribution of the stator coils and the connections to each coil allowing different winding configuration.

The setup chosen is parallel branches, giving the possibility of measuring the asymmetries between branches when analyzing the constructional asymmetries or when these asymmetries are seeded as faults for its analysis.

2.2.1 STATOR

Each of the parallel branches is formed by a group of 4-coils in series with another group of 3-coils. These parallel branches, two per phase, create a 3-phase 4-poles nearly sinusoidally distributed winding.

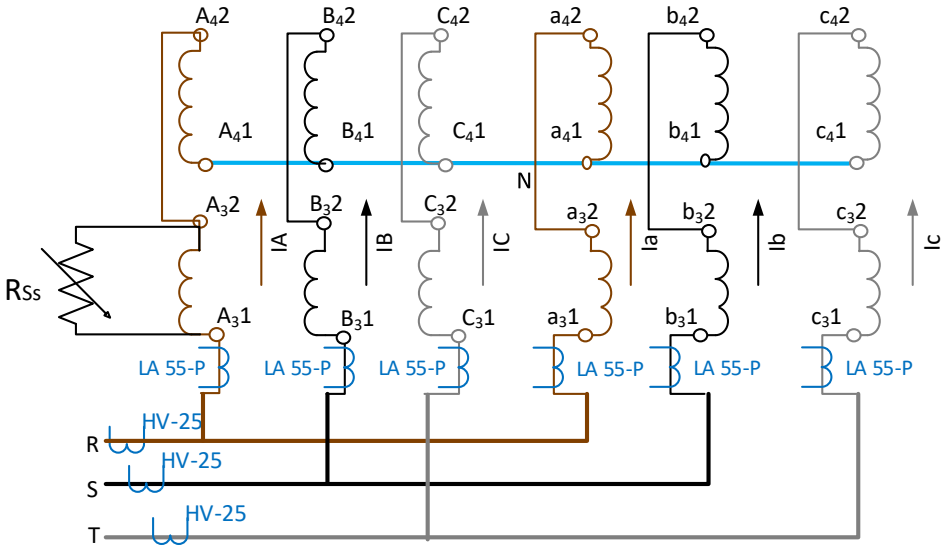


Figure 2-2. Stator winding, parallel configuration.

It is also possible to see in Figure 2-2 a variable external resistance. The resistance can be placed in each coil, and it is used to create winding asymmetries by deviating a portion of the current. Thus, affecting the symmetry of the winding and, therefore, affecting the magnitude of the magnetic field generated by this coil. Three different set-ups are considered as fault type, represented in Figure 2-3.

The current transducers, LEM HV-25, are represented in the figure as well; for the phase current and each of the parallel branches.

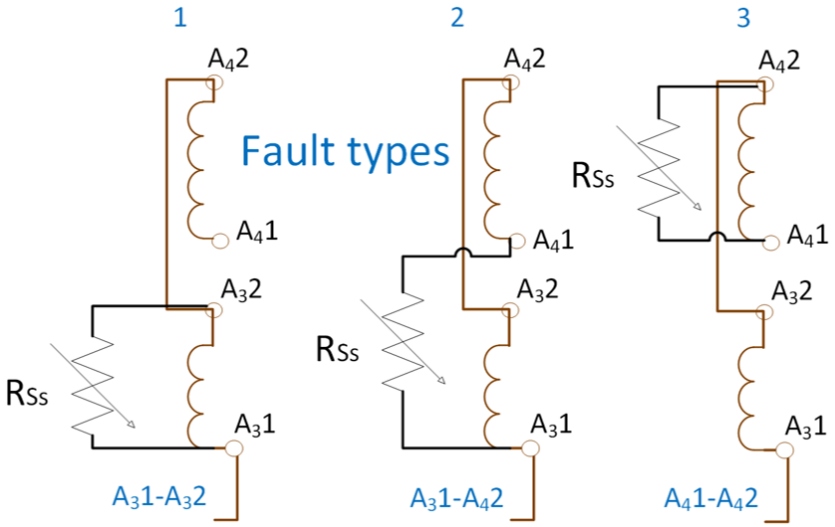


Figure 2-3 Asymmetries created representing the fault types under study.

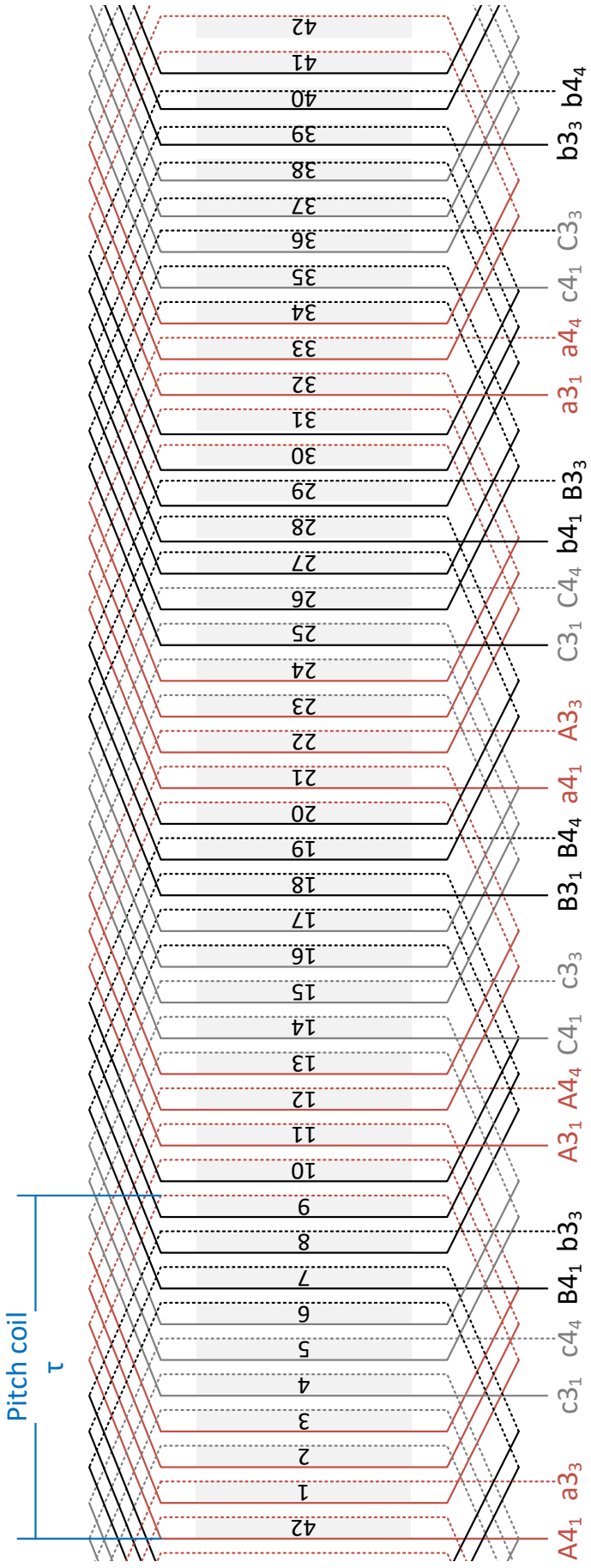


Figure 2-4. Stator winding scheme

Figure 2-4 describes a fractional slot winding. Compared with integer slot winding, the most significant disadvantage of fractional slot windings is the presence of MMF (magnetomotive force). However, the advantage of fractional slot winding is the reducito0n of chosen higher harmonics in the MMF.

Detailed explanations of the winding distribution are provided in section 3.1. The SM in the laboratory has a combination of winding step shortening and coil side shift in a slot **[53]**.

In each slot, there are two sets of conductors, forming the so-called double-layer. Each conductor is made up of 12 copper strands, Figure 2-5.

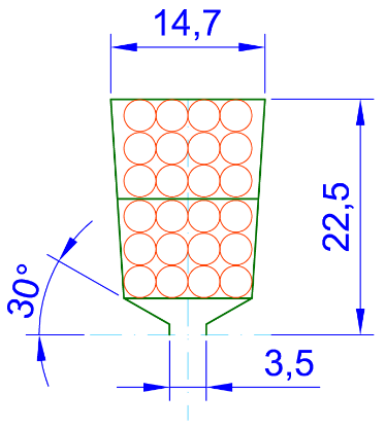


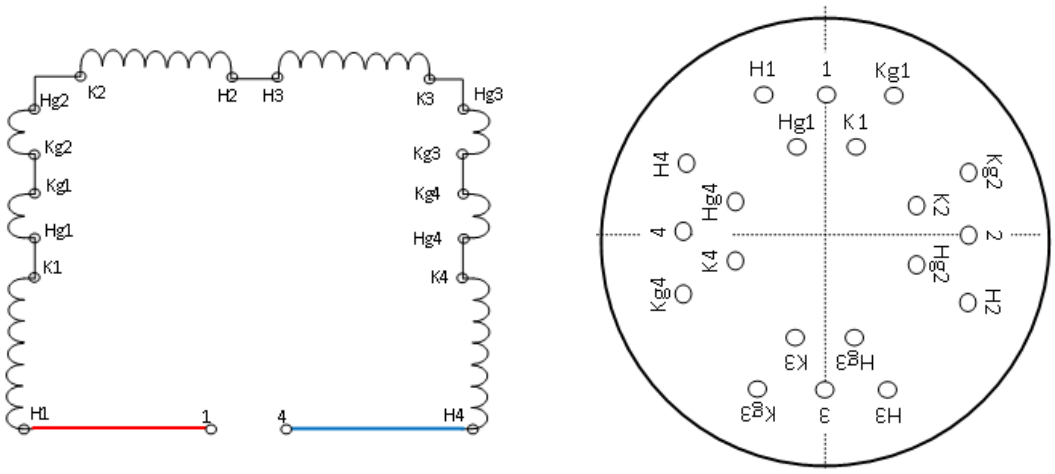
Figure 2-5. Stator slot - cross section

2.2.2 ROTOR

The rotor is a moving component, and its behavior has been traditionally monitored and used for fault detection. The rotor has four poles. In each pole, four round equidistant holes are designed to allocate the damping bars forming the squirrel-cage typical in induction motors

The coils around each pole, are divided into two groups used to create asymmetrical field current or simulate rotor shorten coil. In Figure 2-6, it is possible to see how to build two levels of shortening; 10% or 90% of the total coils per pole.

The uses of rotor shortening fault analysis were published in **[17]** as part of a series of publications created while building the experimental platform. In **[19]** was proved that MCSA analysis is sufficient to detect the asymmetries, produced by shortening of rotor coils, by looking at the variations in the fundamentals harmonics.



(a) (b)

Figure 2-6. (a) Equivalent schema of the excitation winding.(b) Panel used to create different configurations.

2.3 ROTOR-MOUNTED SENSING SYSTEM

Condition monitoring based on measurements inside the rotor is a field that remains under-explored. The objective of building this system is to demonstrate the feasibility of taking measurements from sensors placed over the rotor and illustrating the additional benefits of monitoring those variables. Those signals that in the past were difficult to obtain due to technological limitations and signals obtained from new types of sensor are simultaneously considered.

A rotor-mounted sensing system is not a new idea. The available literature related to the monitoring and diagnostics of sensors installed over the rotor are devoted to large generators. Using scanning technology, by installing infrared sensors, it was possible to elaborate the thermal map in which faulty cases were studied on Hydro generators in 1993 **[8]**. The air gap, the stator core shifting, partial discharge and magnetic flux have also been investigated **[15]**, **[57]**, **[56]**.

Numerous works and publications are focused on the thermal identification of an electric motor, as can be found summarized in **[22]**. To measure the signals from inside the machine, contact or non-contact measurement techniques can be applied. Non-contact techniques are more suitable for large and low-speed machines while contact measurements techniques require more intervention on the motor in the benefit of its higher accuracy.

There are several approaches, such as radio telemetry and light transmission used to transmit the data from the rotating to the stationary part. A recent study has

shown the significances of Doppler Effect on radio telemetry **[22]**. Wireless transmission is less invasive but also requires the installation of batteries and advanced circuitry designs to collect and transmit the signal from the sensors with high accuracy and resolution. Specialized wireless fault detection systems are a challenged that will be only possible to face if it is justified by precise and accurate fault indicators measured only inside the rotor.

A conventional contact measurement technique is slip rings, in which the signal is transmitted to a sliding contact part. Conventionally, a rotational part (ring) mounted on the shaft, is connected to a fixed element (brush) over a carbon brush. One limitation of this system is the degradation and compatibility of the materials used, in consequence, the speed range is limited, and the electrical noise can distort the signal.

The two considered configurations are shown in Figure 2-7, the effects of the machine vibrations over the brushes-rings were the principal factor to take into consideration.

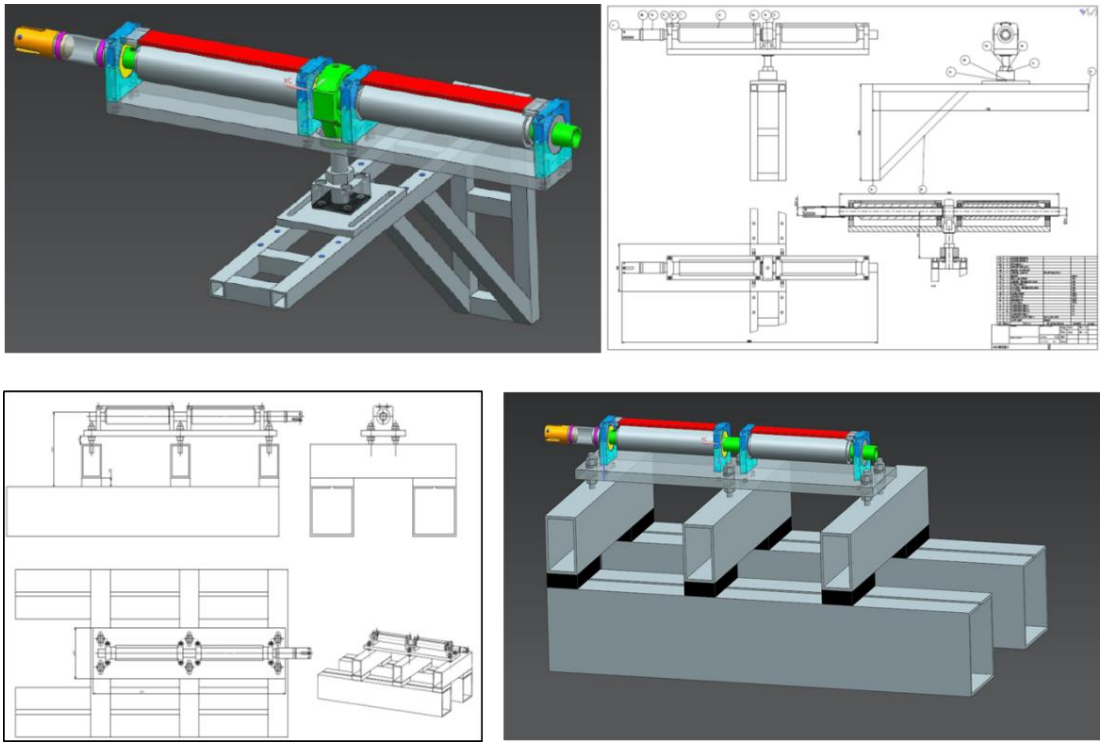


Figure 2-7. Rotor signal collector, prototypes

The designed system, called in this research Rotor Signal Collector (**RSC**), uses the latest materials available on the market, provided by one of the leading companies in the field. The rings are made of silver alloy, and the brushes composition are mainly of copper and graphite. After more than fifteen hours of

experimentation, in cycles no longer than twelve to fifteen minutes there is not significant observed degradation of the ring-brush sets.

By using two 25 slip-ring sets, expressly designed for this purpose, it is possible to obtain up to 50 signals from a group of sensors mounted all over the rotor.

For all the above, a flexible rotor-mounted sensing system was designed and built (see Figure 2-8), with the aim of developing a smart multivariate method capable of early fault detection. The configuration chosen is more robust against vibrations generated in the motor than the other prototype with one pivotal support.

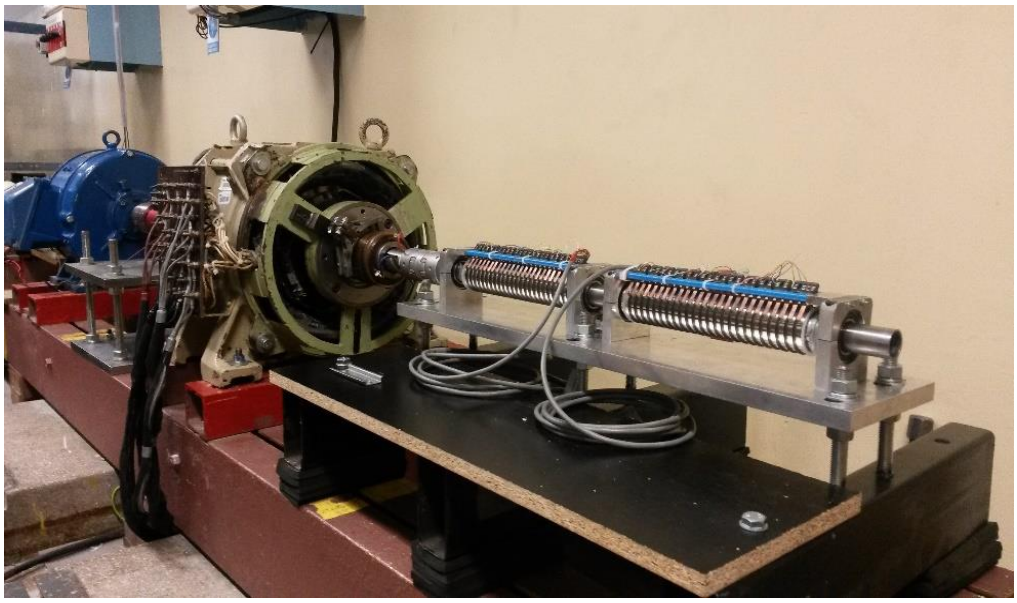


Figure 2-8. Functional and ready RSC

2.4 DATA ACQUISITION SYSTEM

All experiments were performed using the same winding configuration, represented in Figure 2-2 and particular attention was given to guarantee the repeatability of these. The sampling ratio used to collect the data was 19 kSamples/seconds; twelve minutes in total were collected, per each working set point and machine condition, six datasets of 120 seconds.

A National Instrument NI-USB 6255 acquisition card was used to collect 37 signals; this card has 80 analog inputs (16-bit) at 125 MS/s single-channel (750 kS/s aggregated). All the sensors were placed taking into account the intense magnetic fields environment. To reduce external and internal electromagnetic effects, twisted pair shielded wire was used, and when it was possible, placed between poles to connect the sensors to the RSC.

The total number of sensors installed is thirty-seven. Besides of the twenty-one sensors connected through the RSC, there are sixteen most commonly used external transducer for supply voltages, motor phase current, torque, and rotational speed.

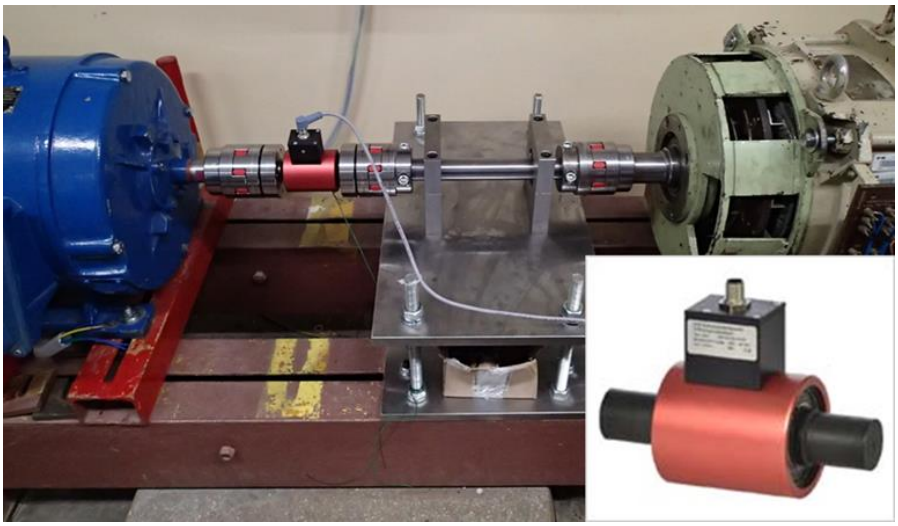


Figure 2-9. Torque-meter assembly

2.4.1 EXTERNAL SENSORS – TRADITIONAL SIGNALS

- *KTR Dataflex 32 torque measuring shaft type:* state of the art torquemeter used to measure the load torque and the rotational speed.
- Machine neutral to ground voltage.
- *LEM current transducer LA 55-P:* used to measure the waveform and magnitudes of the three-phase main motor currents, currents through the parallel branches and the field current.
- *LEM voltage transducer LV 25-P:* used to measure the three-phase to neutral voltage ("Y" configuration).

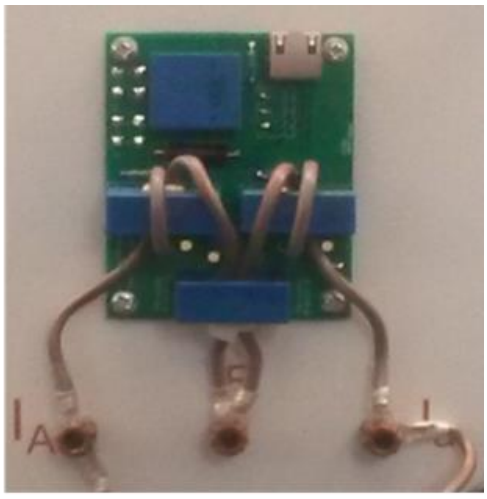


Figure 2-10. Voltage-Current per phase circuitry.

2.4.2 ROTOR-MOUNTED SENSORS

- *Rogowski coil on rotor-cage-bars* **[79][1]**: a Rogowski coil was installed on each bar in two of the poles, so it is possible to measure the induced EMF due to flowing currents as a consequence of higher harmonic of magnetic field distribution into air-gap or asymmetries during synchronous steady-state.
- *Rotor-cage-bars and front-end ring voltage drop:* the damper voltages are said to be zero since the bars are all shorted together **[55]**. However, in the presence of asymmetries or transient states, the induced EMF in each bar is different. Therefore, the potential between bars is expected to be different from zero.
- *Machine sound variations:* two type of microphones were installed; noise canceling, placed parallel to the shaft axis capable of measure sound in one axis direction, and omnidirectional microphone, capable of recording all type of vibrations generated over its membrane.
- *Shaft voltage;* the shaft to ground voltage is measured.
- *Hoeben HE244 Analog Hall sensor*: place in the pole head, capable of measuring up to five Tesla. Three sensors were symmetrically installed to measure the magnetic flux crossing the rotor pole.

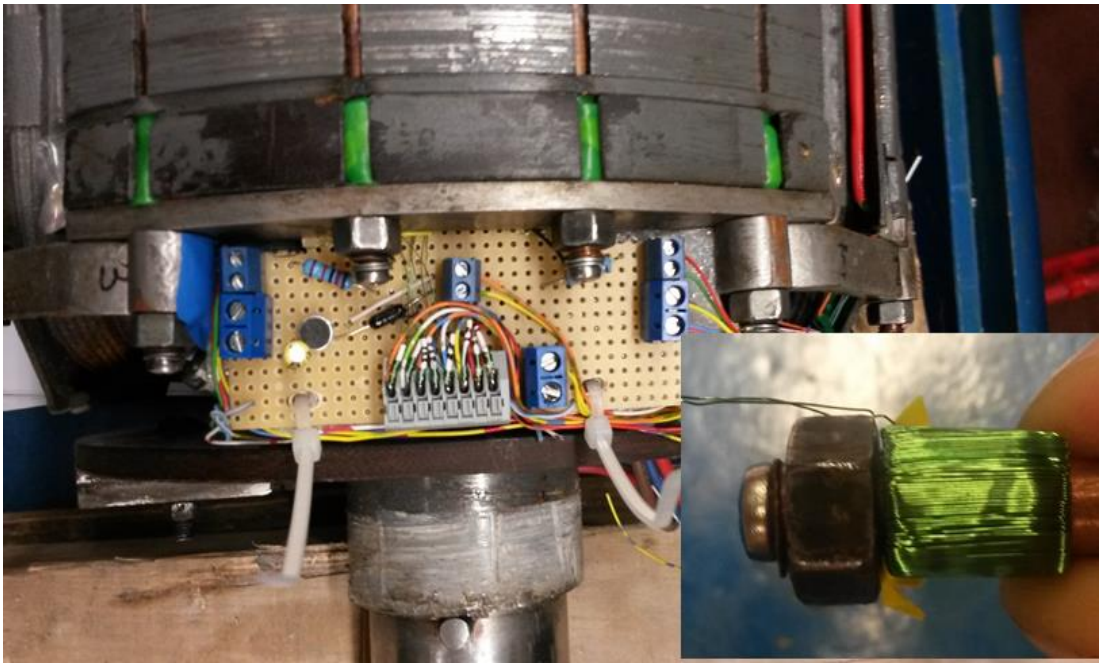


Figure 2-11 Rogowski coils assembly

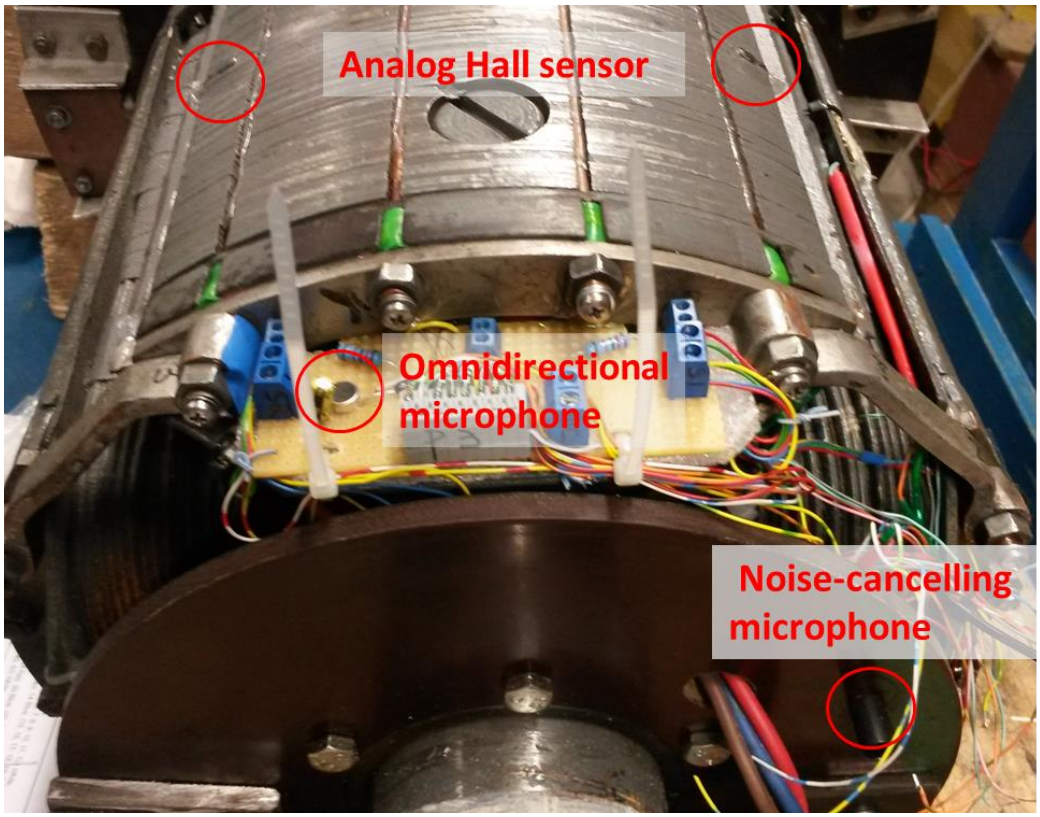


Figure 2-12. Rotor-mounted sensors

2.5 REGULATED LOAD OF THE SYNCHRONOUS MOTOR.

Connected through specially selected and configured Jaw couplings a DC machine running as a generator is used as a regulated load by controlling the armature current using a PWM controller.

During the starting process, the DC machine is used as a motor to speed-up the SM up to synchronous speed. When the generator has equal line voltage, frequency, and phase sequence as the power grid, a synchronizing unit, LUMEL KS3, closes the circuit allowing the SM run freely as a motor. At this moment, the power is removed from the DC machine and switched to generator mode.

Table 2-2 DC machine main parameter

Power [kW]	10
Voltage [V]	220
Phase inductance [H]	23
Current [A]	53
Rotational speed [rpm]	1500

2.5.1 DC MACHINE, ARMATURE CURRENT REGULATION

To precisely regulate different load levels, it was designed and built a PWM controller for the DC-machine field current, while running as a generator. The generated dc voltage is supplied to a fixed group of resistors, so it is possible to simulate different load levels up to 110% of the SM nominal power.

The pulses were created using an STM32 F4 Discovery evaluation board and connected to the computer using a UART port allowing on-line control. The design of this PWM controller was implemented in a PCB as seen in Figure 2-13. This PWM DC motor controller was regulated using Matlab SIMULINK processor-in-the-loop connectivity.

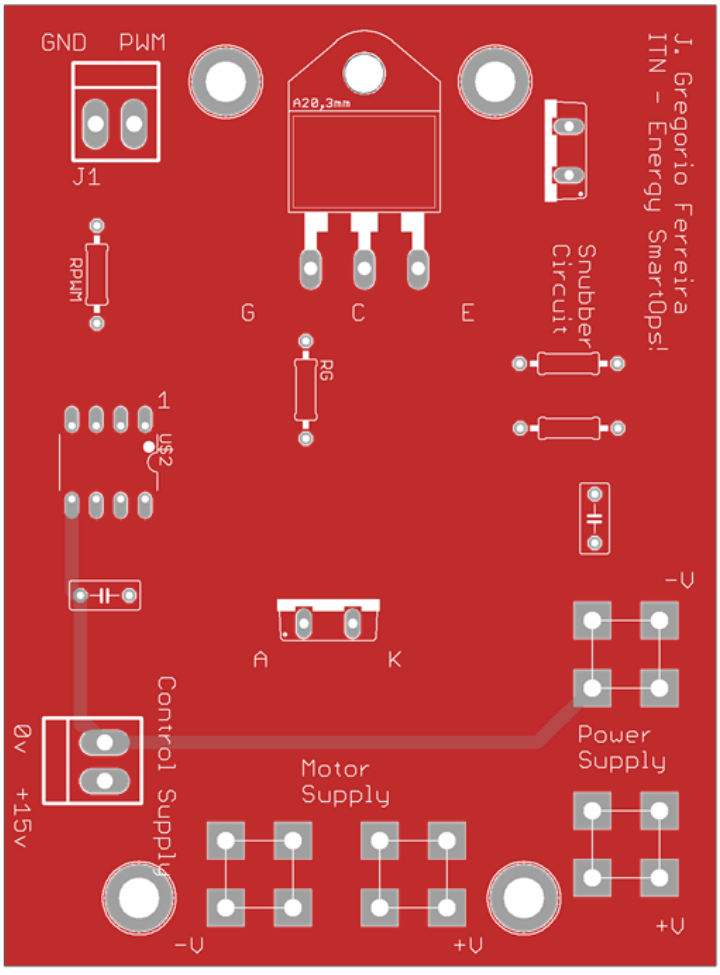


Figure 2-13. PWM circuit board

The value for the required duty cycle is on-line transmitted and written into the microprocessor. At the same time, the temperature is continuously monitored through the analog inputs available on the Discovery board. Main controller components can be seen in the Figure 2-14.

The final configuration was proven to be effective to handle high repetitive reverse voltages as a consequence of fast switching the field current flowing through the DC machine.

- Ultrafast IGBT 1,2 kV. TO247-Infineon
- STM32F407VGT6 microcontroller featuring 32-bit ARM Cortex-M4F core, 1 MB Flash, 192 KB RAM in an LQFP100 package
- HCPL-3120 gate optocouplers drive
- USB to UART, FTDI Dev Board FT231XQ
- Fast protection DIODE, snubber circuit: IXYS Semi 1.2kV.

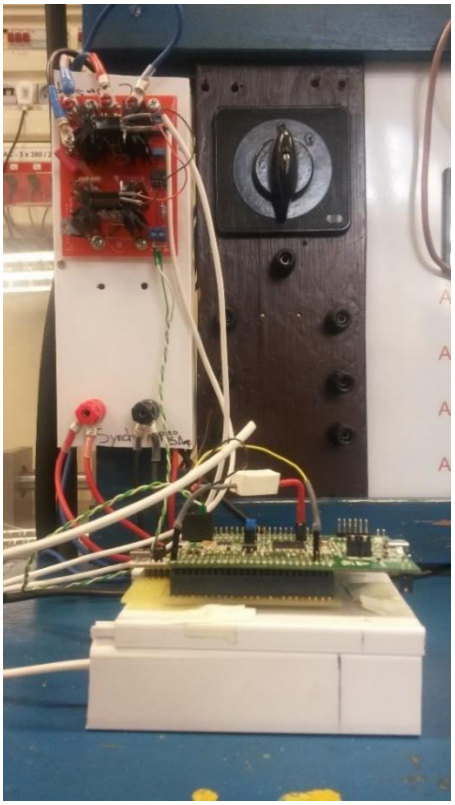


Figure 2-14. PWM DC machine controller

3 PRELIMINARY CALCULATIONS BASED ON MATHEMATICAL MODELS.

Accurate mathematical models require the resolution of differential equations defining electro-mechanical energy conversion phenomena. Which in the case of complex geometries, such as salient pole configuration, might require accepting simplifications or define assumptions that not necessarily represent the study of practical cases. Hence, the motivations for this research, multicriteria analysis, are justified as the methodology capable of identifying the condition of synchronous machines over the time

3.1 DESCRIPTION OF SM STATOR WINDING

The winding function theory and the FEM model are explained in following sections. It is important to establish and explain the physical interactions between the current flowing through the windings and the magnetic flux crossing the air gap.

The harmonics components of the stator voltages and currents are undesirable, for what techniques have been developed to suppress the unwanted harmonic components from the voltages and currents of a machine. One important method to suppress the harmonics is the use of fractional-pitch windings.

Space and time harmonics are reduced by proper distribution of winding coils in the slots. The stator winding of the SM under study was identified as fractional-pitch winding with parallel branches.

Table 3-1 SM winding parameters

Stator Double-layer fractional slot winding	
Number of phases	$m_s = 3$
Number of pair of poles	$p = 2$
Number of slots	$N_s = 42$
Pole pitch	$\tau = \frac{42}{2 * p} = 10.5$
Number of slots per pole and phase	$q_s = \frac{42}{2 * p * m}$
Number of parallel branches	$a = 2$
Number of coils per phase and parallel branch	$N_C = q_s * a$
Number of turns in each coil = Number of turns in each layer of slots	$N_t = 12$
Number of turns in one phase and in one parallel branch	$W_s = N_c * N_t$
Coil pitch	$y_s = 9$
Skewing factor	$k_{sf} = 1$

The winding layout is represented in Figure 3-1

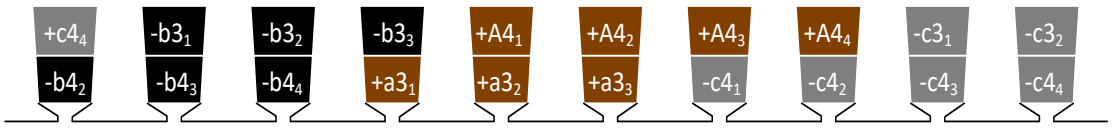
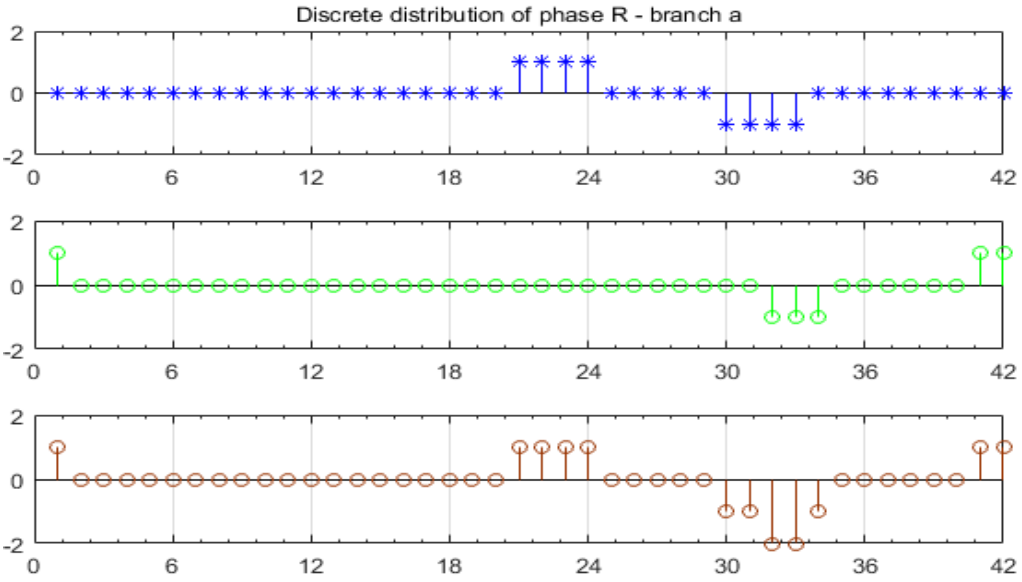


Figure 3-1. Winding layout

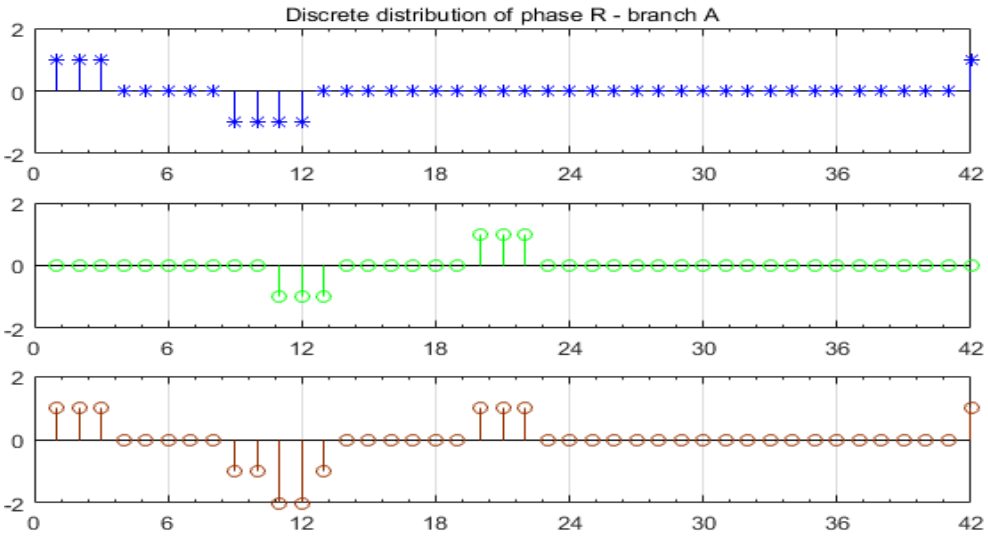
The type of the winding belongs to the classification of **double-layer fractional lapped winding, shorted and coil side shift;** having one slot overlapped with the previous zone and two with the following. This type of winding creates MMF nearly sinusoidally distributed along the air gap of the machine.

3.2 CALCULATION OF MMF SPATIAL HARMONICS SPECTRA

The calculation of the MMF spatial harmonic spectrum uses an equivalent winding located into the machine's air gap. From the numbers of turns, the winding layout over the stator circumference and stator core geometry, it is possible to draw the winding function. The winding function $wn(x)$ expresses the distribution of turns of individual winding along circumference of the air-gap.



(a)



(b)

Figure 3-2 a,b. Winding function for one turn coils. The width of the impulse function is omitted. The winding function of parallel branches "A" and "a" are the sum of them.

The way of introduction of the winding function is documented in **[38]** and in **[64]**. The function has non-zero values in slot opening area only and is equal to the

number of the turns in the slot, wounded coils, divided by the width of the area. This function can also be used to determine how much flux links a winding.

The MMF distribution is calculated according to the formula 3-1.

$$\boldsymbol{MMF(x,t) = i_n \int w_n(x)\, dx} \tag{3-1}$$

After zero-centering the turns function the winding function of stator equivalent phase winding is shown in Figure 3-2 a,b. For simplicity, the number of turns in each layer of the slots was introduced equal to one, and the width of the impulse function was omitted.

The total distribution of phase R winding is specified as the sum of the parallel branch from by A_4 and A_3, and the parallel branch from by a_4 and a_3, Figure 3-3.

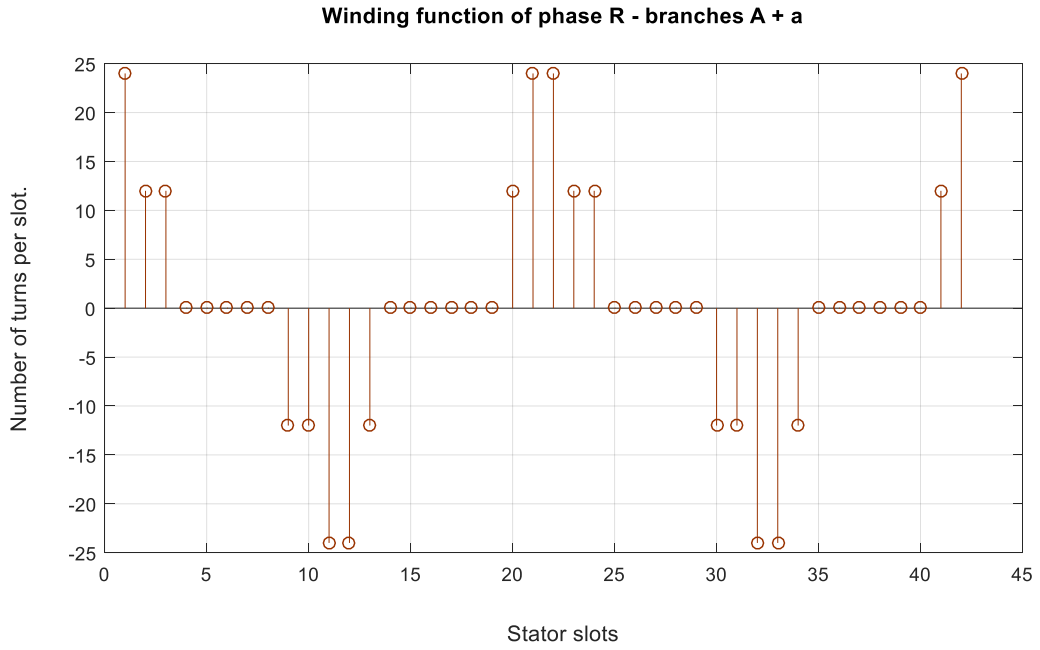


Figure 3-3 Total winding function of phase R including the numbers of turns in each slot. The width of the impulse function is omitted.

It can be seen in Figure 3-4, the maximum and minimum values of MMF are different, but the constant component cannot appear because windings are always built in a manner ensuring the condition:

$$\int_{x}^{x+2\pi} \boldsymbol{MMF(x,t)\, dx = 0} \tag{3-2}$$

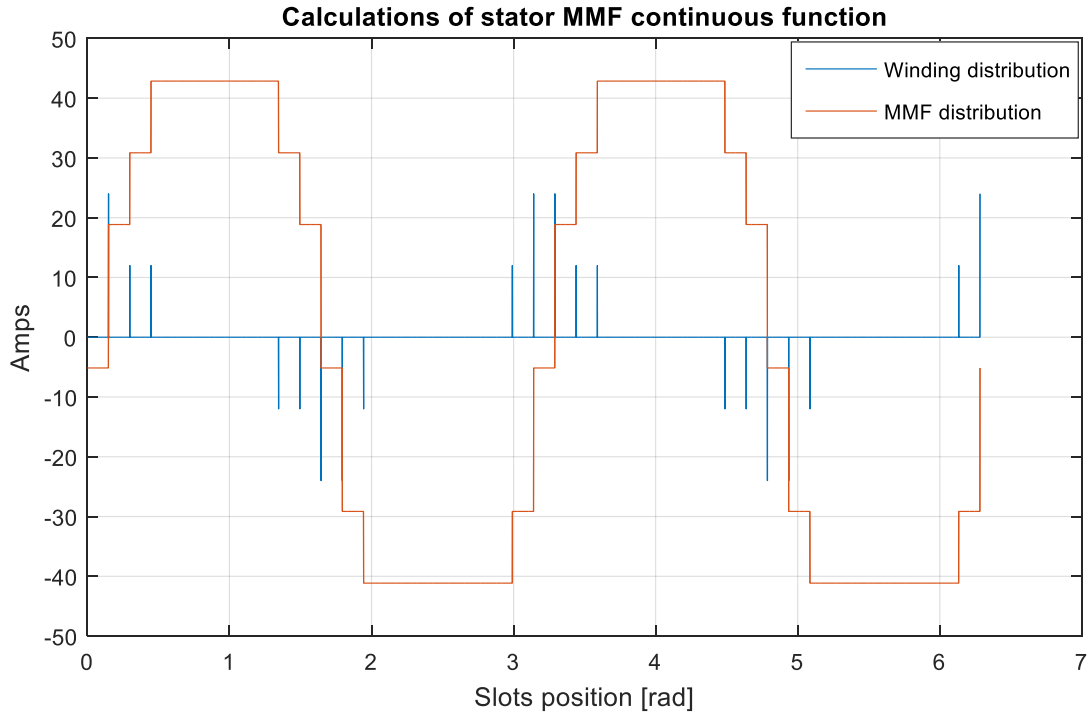


Figure 3-4. Winding and MMF distribution of phase R at current equals one Amp.

The Fourier spectra of MMF calculated using FFT procedure was obtained as can be seen in Table 3-2 and Figure 3-5.

Table 3-2 FFT for one phase MMF

Spatial harmonic order p=2	Spatial harmonic amplitude of one phase MMF in Amps at current equal to 1 Amp.
0	0.00
1p	49,83
2p	0.83
3p	8,95
4p	0.78
5p	0.91
6p	1.38
7p	0,00
8p	0.56
9p	0.59
10p	0.41

In the MMF spectra for one phase, the higher odd and higher even harmonics versus fundamental p-harmonic occur as the result of the fractional pitch of winding. It can be proven that dominated higher harmonics of order 3p, 6p and also 9p

disappear in total MMF generated by three-phase sinusoidal currents of the stator. For more precise calculations a pulse turns function should be taken into account.

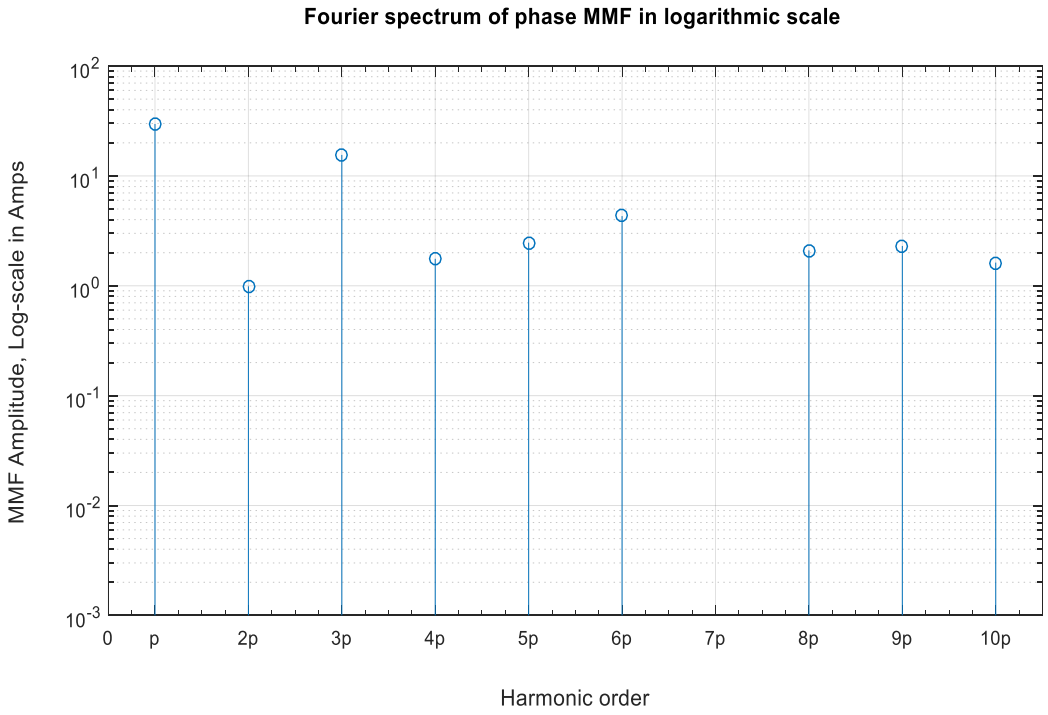


Figure 3-5. Fourier spectrum of phase MMF in logarithmic scale

The rotating magnetic field in the air-gap of the machine is produced by the total stator and rotor MMF.

The distribution of magnetic field density along the air gap circumference is also formed by the shape of stator and rotor surface. It is expressed by the formula:

$$\boldsymbol{B(x,\varphi,t) = \lambda(x,\varphi)\cdot MMF_t(x,\varphi,t)} \tag{3-3}$$

When the condition (3-3) is fulfilled:

$$\int_{x}^{x+2\pi} \boldsymbol{\lambda(x,\varphi)\cdot MMF_t(x,\varphi,t)dx = 0} \tag{3-4}$$

$\lambda(x,\varphi)$ means so called function air gap permeability including magnetic permeability $\boldsymbol{\mu_0}$.

It is essential that the voltage of stator winding is induced only by the space harmonics distribution of magnetic field density which number equals to the number of MMF harmonic produced by the winding currents.

3.3 DESCRIPTION OF SM FIELD-CIRCUIT MODEL

For this research, it is available the software MagNet 2D from Infolytica. It is intended to build a functional model to be used as a development environment to test machine configurations and to prototype the experiments, **[10], [11], [32], [41]**.

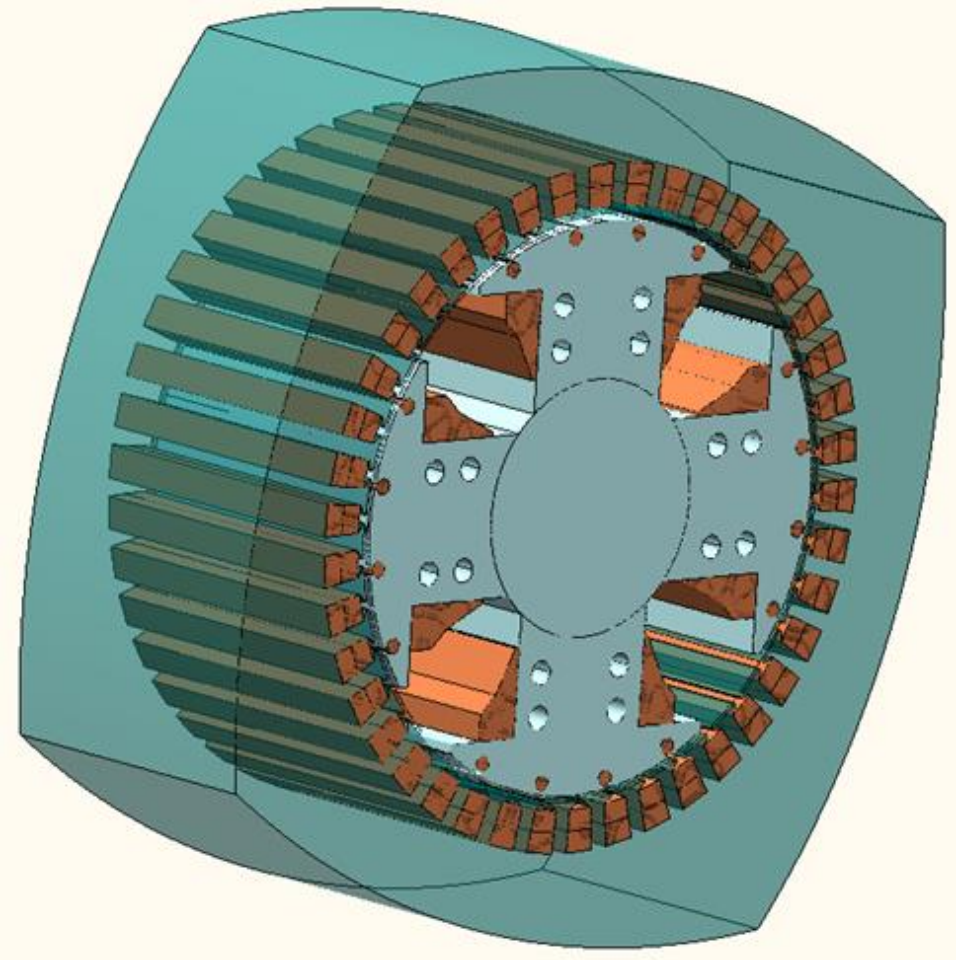


Figure 3-6. 3D visualization of the SM magnetic core without end regions.

The stator and field coils are defined as stranded wire and the rotor cage as solid bars.

A 2D finite element formulation is believed to be sufficient and utilized in this work. The contribution of the end region effects, over the electromagnetic-electrodynamic steady-state performance of such SM, is subtle; the resistances and inductances of end regions of the stator winding, field windings and front and end ring embracing the damper bars connections are neglected.

The 2D model is used to perform transient with motion analysis. The magnetic core is considered to be linear, and the circuit elements are connected to the equivalent schema represented in Figure 3-7.

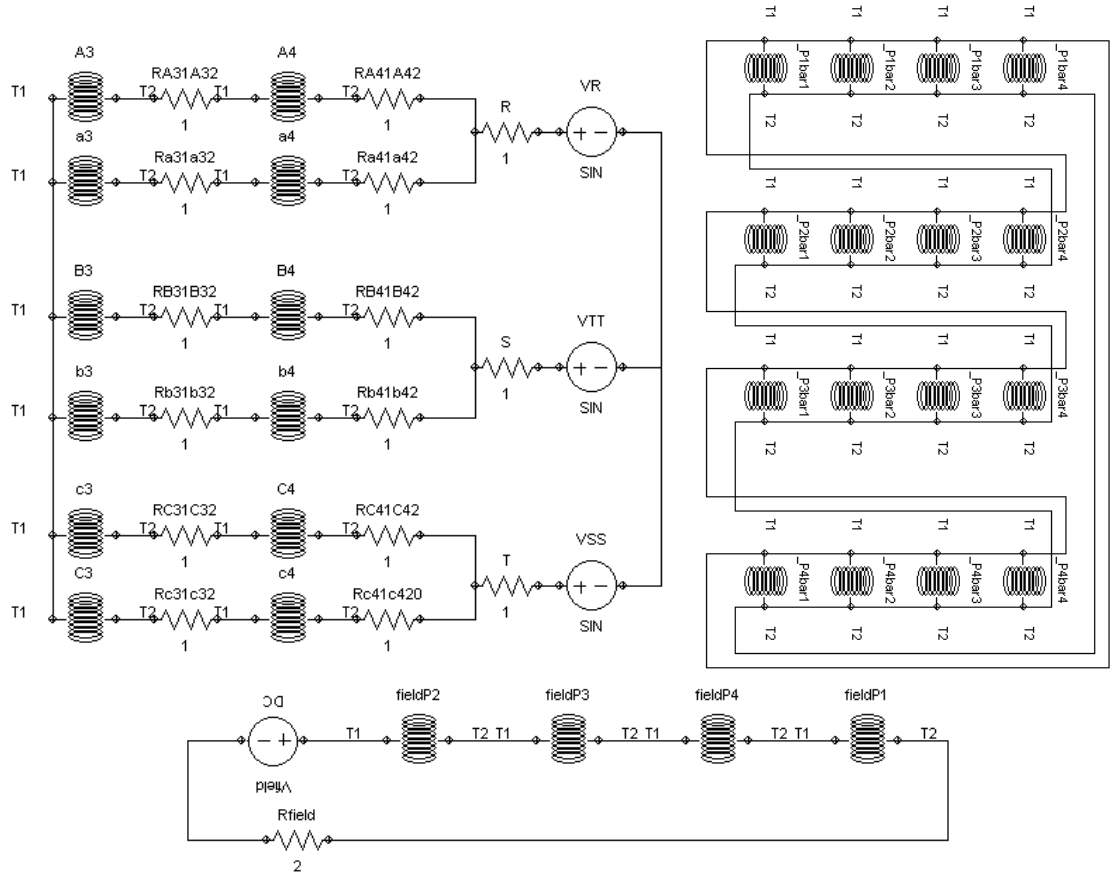


Figure 3-7. Schema of the stator and rotor coil connections presented by MagNet software

3.3.1 EQUATIONS OF FIELD-CIRCUIT MODEL

The magnetic field in this kind of electromagnetic object is usually described in classic A-V method. In this method, a general field equation for the transient states with motion has the following form:

$$\boldsymbol{curl}(\boldsymbol{\mu^{-1}} \cdot \boldsymbol{curl\, A}) + \boldsymbol{\gamma} \cdot \frac{\boldsymbol{\partial A}}{\boldsymbol{\partial t}} - \boldsymbol{\gamma} \cdot \boldsymbol{v} \cdot (\boldsymbol{curl\, A}) = -\boldsymbol{\gamma} \cdot \boldsymbol{grad\, V} \qquad 3\text{-}5$$

where A – magnetic vector potential, $curl\, \boldsymbol{A} = \boldsymbol{B}$ - field density vector, $\boldsymbol{\mu}$ - magnetic permeability, $\boldsymbol{\gamma}$ - electric conductivity, $\boldsymbol{V}$ - electric static potential, $\boldsymbol{v}$ - linear speed vector

For 2D models, the general equations get simplified to the scalar form. The component, modeling the influence of rotor movement with the speed $\boldsymbol{v}$, disappears when the rotor mesh is rotating versus stator one.

$$\frac{\boldsymbol{\partial^2 A_z}}{\boldsymbol{\partial x^2}} + \frac{\boldsymbol{\partial^2 A_z}}{\boldsymbol{\partial y^2}} + \boldsymbol{\mu\gamma} \frac{\boldsymbol{\partial^2 A_z}}{\boldsymbol{\partial t}} = -\boldsymbol{\mu\gamma} \frac{\boldsymbol{U}}{\boldsymbol{l}} \qquad 3\text{-}6$$

The equation for conductors, considered as thick conductors (massive), is as follows:

$$\boldsymbol{u_t = Ri + R \iint_S \gamma \frac{d}{dt} A \, dS} \qquad 3\text{-}7$$

where: $\boldsymbol{l, S}$ **-** indicate the length and cross-sectional area of the conductor**,** $\boldsymbol{U}$ **-** is the voltage between the end of conductors**,** $\boldsymbol{R}$ - is the dc resistance **[75].** This equation expresses that tension over a thick conductor is related to a sum of the voltage drop across the dc resistance ($\boldsymbol{Ri}$) and the voltage drops due to eddy currents ($\mathrm{R} \iint_{\mathrm{S}} \boldsymbol{\gamma} \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{t}} \mathrm{A}\, \boldsymbol{dS}$)

The general mechanical equation describing the motion of the rotor has a form:

$$\boldsymbol{J_m \frac{d\omega_m}{dt} - D_m \cdot \omega_m = T_e - T_L} \qquad 3\text{-}8$$

where $\boldsymbol{J_m}$**-** is the rotor inertia moment, $\boldsymbol{D_m}$ **–** is the friction damping coeficient, T_e – is the electromechanical torque, T_L- is the load torque acting on the machine axis and $\boldsymbol{\omega_m}$ **-** is the rotor speed. The electromechanical torque, $\boldsymbol{T_e}$**,** is usually calculated by virtual work method or by Maxwell stress tensor method. The MagNet program formulated, discretized and solved the set of field, circuits and mechanical equations.

3.4 FEM MODELLING

Now that the set-up has been explained, in this section, the results of the simulation are described and displayed. The CAD software used is designed to do calculations of transient states including movement of the rotor.

The type of simulation is transient 2D with motion. The solver was set to Newton-Raphson method of polynomial order two. This method is an extension of the Newton method, adapted to matrix systems of equations.

The 2D mesh is automatically generated. The stator has a static mesh; invariant from one-time step to the next. The mesh for moving components created shaft and rotor poles rotates with the rotor.

The sliding cylindrical surface between stator and rotor mesh, necessary for modeling the movement of the rotor, was introduced in the center of the air gap. The

rotor's rotating mesh slides with respect to the invariant stator mesh on the surface. The air gap mesh is re-calculated at every iteration to guarantee the continuity of the total mesh during dynamic simulation.

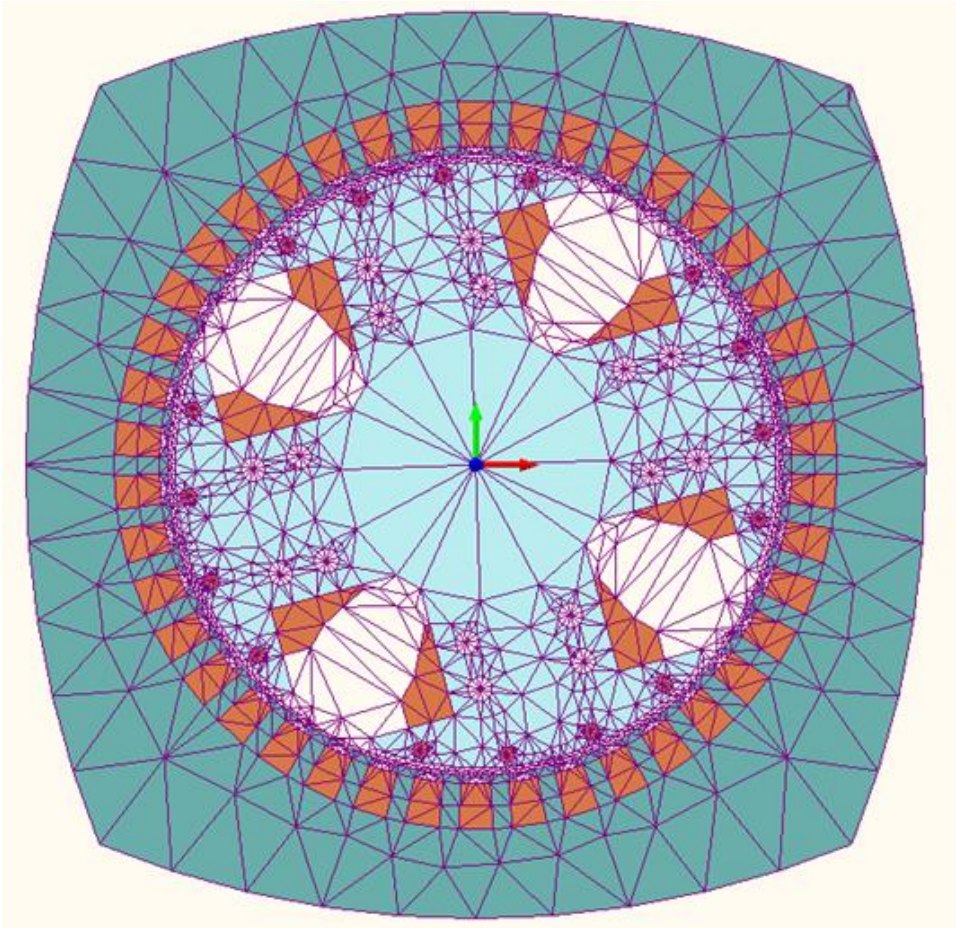


Figure 3-8. Initial, 2D automatically generated mesh.

Additional model parameters:

- The rotor moment of inertia set up to 0.3 $Kg * m^2$

The external mechanical and electrical forces in motor regimen were introduced as follow:

- The stator winding was supplied by balanced three-phase voltage system with nominal value.
- The field winding was supplied using an ideal DC voltage source.
- Rotor motion type: rotary load driven.
- Three different set-points were recorded. The first set point is under no load, 1 Nm simulating mechanical frictions. The next two set-points are configured to represent 50% and 100% of the nominal load 7,5kW.

The initial conditions for the simulation are as follow:

- Initial stator currents values $Ia = 10\ Amps$ and $Ib = Ic = -5Amps$.
- The speed at start-up: 9000 degrees per second. (Synchronous speed of SM).
- The rotor position at startup is 15 degrees.

Each set point was recorded for at least 0,8 seconds. To guarantee the steady-state operation, only the last 0,5 seconds for each set point are used for the calculation of the spectrum.

3.4.1 STATOR CURRENTS, CALCULATIONS OF TIME HARMONICS

All the data were exported as flat files, then loaded in MATLAB and each set-point was identified and extracted in different files. After that, all the processing was incorporated into the main processing algorithm, and the frequency domain features (harmonic content) were calculated. The three-phase currents and the spectrum for the current phase R are represented in Figure 3-9.

It can be observed the current higher time harmonics of order $(1 + 2n) = 1, 5, 7, \ldots$ already indicated in **[12]**. Generally according to study done in**[64]**, the frequencies of current harmonics at sinusoidal supply are $(\Omega_0 + 2np\Omega_m)/2\pi$. $(\Omega_0)/2\pi$ the frequency of supply voltage, Ω_m - the mechanical speed of the rotor.

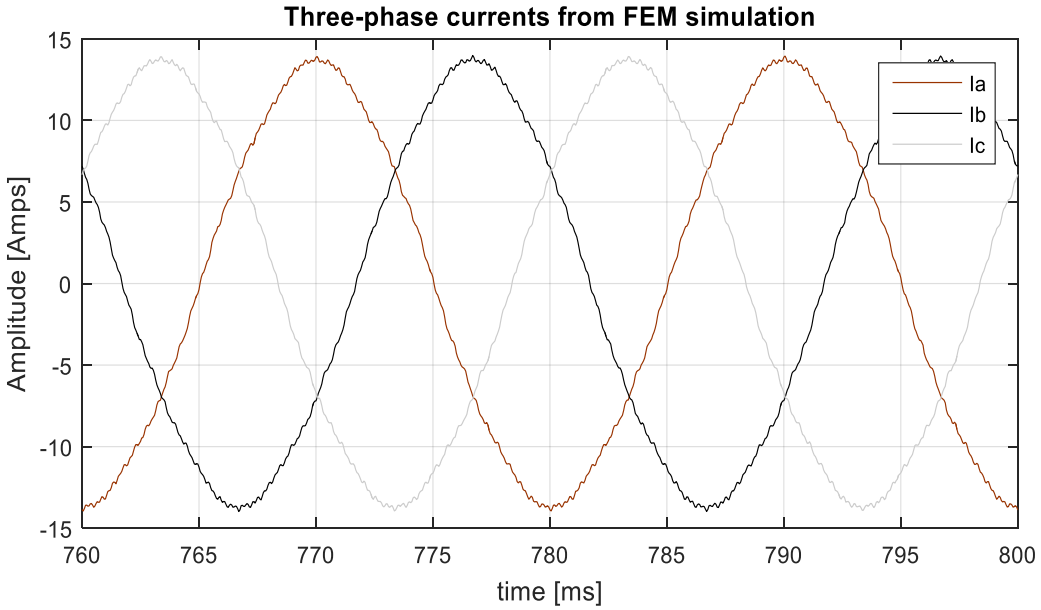


Figure 3-9 FEM simulation three-phase current in steady state 50% of nominal, field current $I_f = 10.8\ Amps$

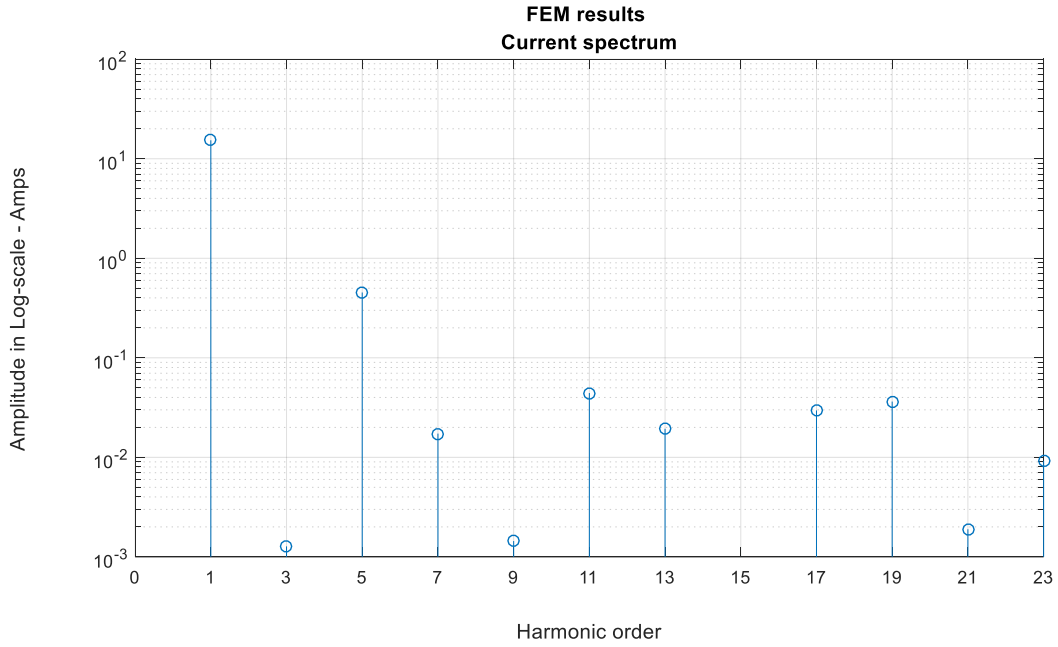


Figure 3-10. FEM simulation results, current spectrum phase R.

The fifth harmonic stator rotating field produced by sinusoidal currents rotates with speed five times slower than fundamental harmonic, but in the opposite direction (reverse) of the fundamental. Then, in the field current, a sixth time-harmonic is induced. This produces the oscillating field static related to the rotor.

The oscillating field is equivalent to a pair of positive and negative pair of fields rotating in the opposite direction versus the rotor with speed six times greater than synchronous speed.

The positive component induces in the stator winding a seventh time-harmonic, and the negative component induces a fifth time-harmonic. The same explanation applies to the eleventh and thirteenth harmonics.

These effects require a deeper understanding of the interactions between space and time harmonics into the machine. Among other factors, the presence of the rotor-cage in the studied SM attenuates these effects.

3.5 COMPARISON OF FEM CALCULATION RESULTS AND MEASUREMENT RESULTS.

As explained in subsection 2.4, different variables were measured. In order to compare the results obtained from the FEM model, with the real signals (see Figure 3-11) from the same modeled machine, the real spectrum for the phase currents was calculated, Figure 3-12

This comparison is necessary to identify observed inherent asymmetries;

- the different time-harmonics that might arise from the input supply
- the intrinsic space-harmonics generated by the machine winding distribution
- other geometrical/manufacturing asymmetries.

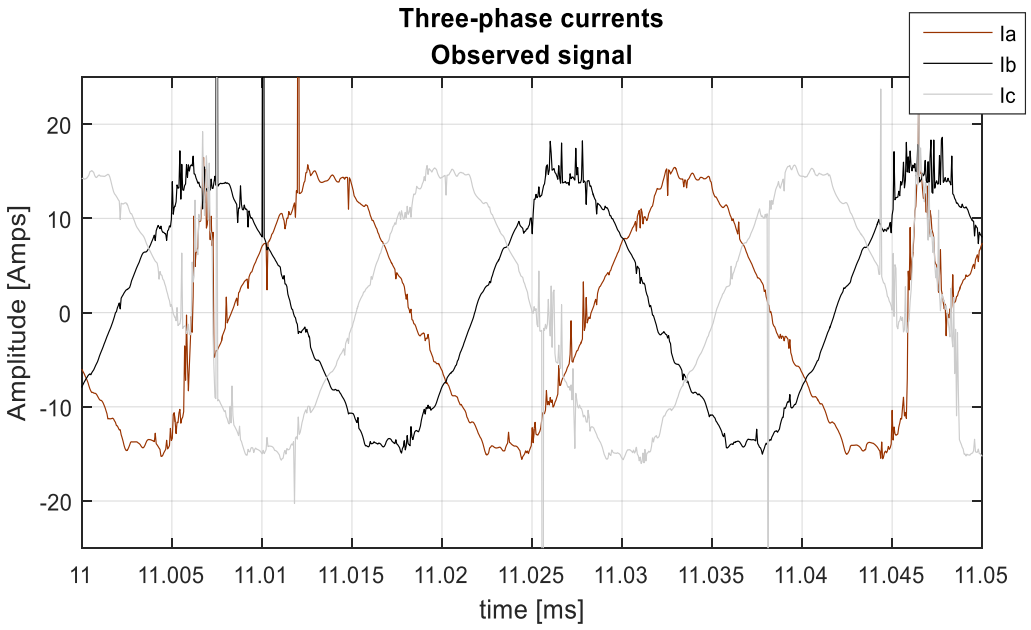


Figure 3-11 Three-phase SM observed currents in steady state 50% of nominal, field current $I_f = 10.7\ Amps$

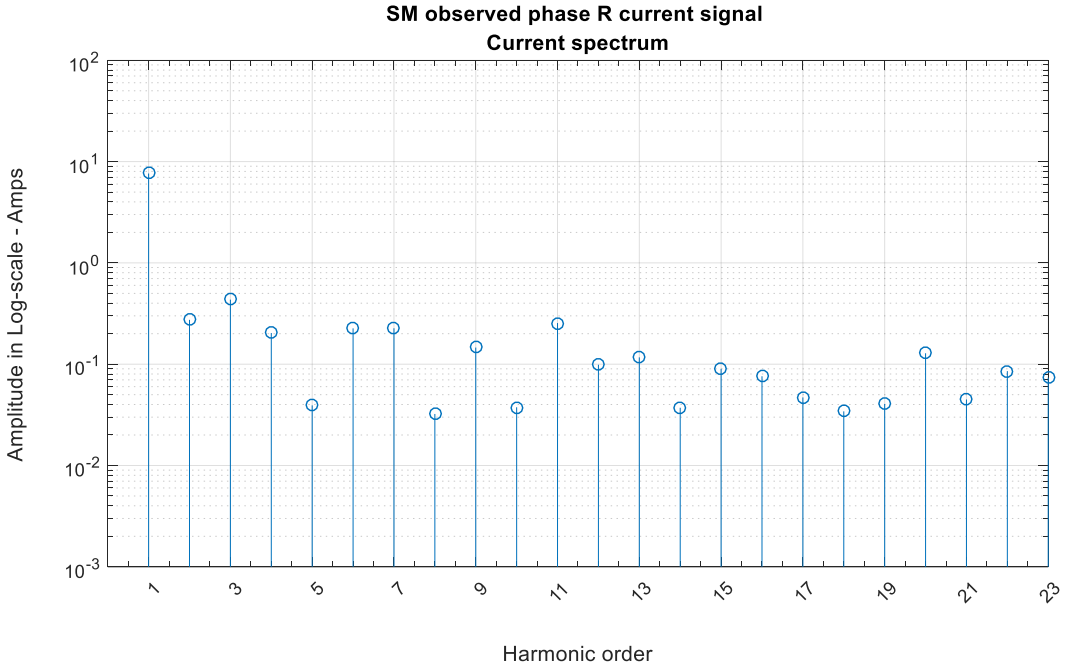


Figure 3-12 Observed data, current spectrum phase R.

As shown in Figure 3-13 the additional resistances in series with the coils can be used to create asymmetries in the phase current of the model. Similar asymmetries were identified directly from the machine by measuring the impedance and resistance of each of the parallel branches.

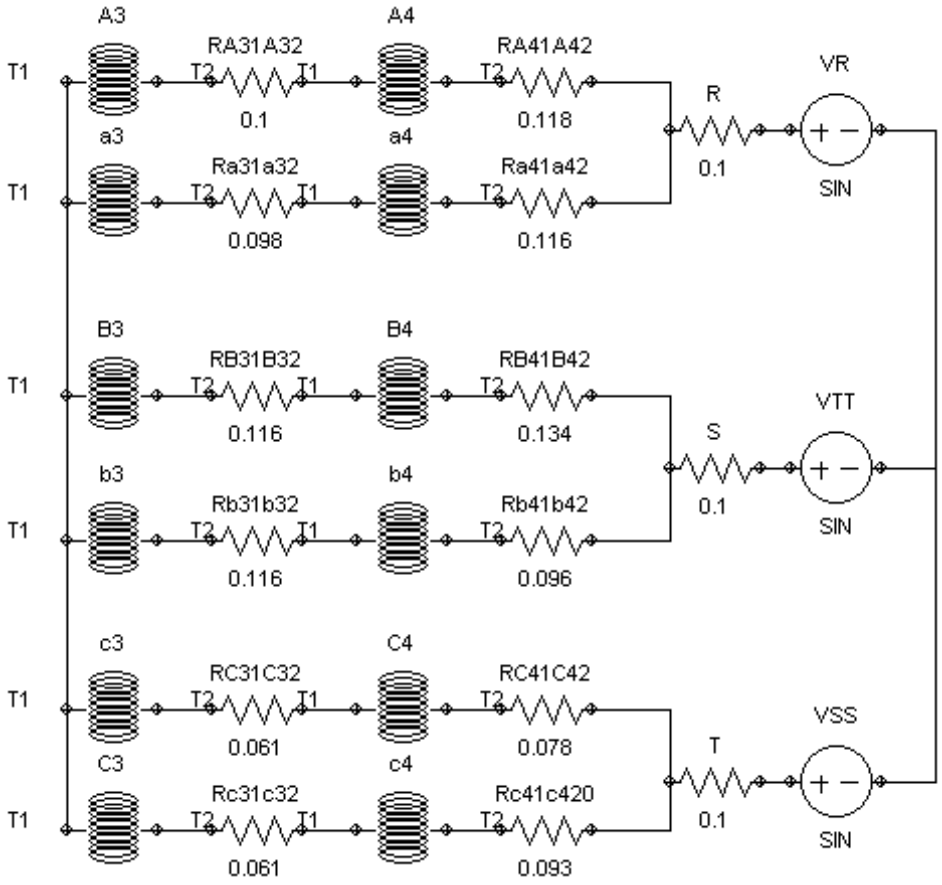


Figure 3-13 FEM model coils corrections

A correction of the FEM model coils resistances was made after identifying the differences in the spectrum content. Considering that the FEM model was supplied with 3-phases symmetrical source, it can be considered that the harmonic content is due to space harmonics as a consequence of the type of stator winding.

Since a direct comparison is not possible, the harmonic content obtained from the SM data is included as a reference. Moreover, the presence of noise or imbalances generates even harmonics in the SM that are not present in the analytical models.

In this practical implementation, the first non-ideality is the presence of harmonics in the supply. These time harmonics appear in the stator winding currents modifying the harmonics induced in the rotating magnetic field.

Included for completeness, Table 3-3 contains the current harmonic order referenced to the first harmonic, for:

- FEM model of symmetrical machine
- FEM model of asymmetrical machine (constructional asymmetries)
- Research machine from measurements

Table 3-3 Current spectrum comparison referred to the first harmonic order

Harmonic order	FEM model	FEM constructional Asymmetries	SM data
1	100.00%	100.00%	100.00%
2	0.00%	0.35%	3.59%
3	0.01%	0.31%	5.68%
4	0.00%	0.13%	2.66%
5	2.96%	3.00%	0.51%
6	0.00%	0.07%	2.96%
7	0.11%	0.11%	2.97%
8	0.00%	0.05%	0.42%
9	0.01%	0.04%	1.91%
10	0.00%	0.04%	0.49%
11	0.29%	0.28%	3.34%
12	0.00%	0.04%	1.29%
13	0.13%	0.12%	1.53%
14	0.00%	0.03%	0.49%
15	0.00%	0.01%	1.18%
16	0.00%	0.03%	1.00%
17	0.19%	0.19%	0.62%
18	0.00%	0.02%	0.45%
19	0.24%	0.23%	0.53%
20	0.00%	0.02%	1.68%

Comparing theoretical versus analytical models either based on related knowledge such as winding function or numerical simulation requires a deep understanding of the object under consideration. In most of the cases, the complexity of such approaches makes the implementation unpractical.

The methodology implemented here using FEM models seems to be not accurate enough. The results are qualitatively correct but for the healthy or evidently

damaged machine. The detailed explanation of the generated effects needs also established knowledge concerning circuital multi-harmonic models of synchronous machine. This analytical approach using spectra is recognized and was presented in [13][24][65][84]

Calculations presented in this chapter were used to acknowledge the problems that appear when the diagnostic of the machine is based on the analysis of stator and rotor currents spectra.

The synchronous machine available in the laboratory makes possible the acquisitions of many different signals. For this reason, in the next chapters, a multicriteria diagnostic methodology is defined and explained

4 MACHINE LEARNING FOR SIGNAL PROCESSING: SM FAULT CLASSIFICATION TASK

In the previous Chapters, the motivations driving this research and all the information necessary to guarantee the repeatability of the experiments were specified. In this Chapter, it is explained and described all the followed steps to faults detection using machine learning for signal processing techniques. To take advantage of machine learning algorithms, a significant amount of data is needed.

Once the experimental platform is already described, it is possible to concentrate all the efforts developing reliable datasets and extract meaningful information to be used as fault indicators. The number of recorded signals is 38, more than 250 GB of raw data; using different algorithms and strategies, the relevance of the various types of calculated features is quantified and explained.

In this Chapter 4, a methodology based on basic electrical machine knowledge and data-driven modeling is introduced and explained. As stated in Chapter 1, Table 1.1, the number of devoted publications on Synchronous machines fault diagnosis, using computational efficient data-driven models, is scarce or nonexistent. However, some of the publications used as inspiration are specified below.

In [4] the authors have proven that Linear Discriminant analyses methods give greater accuracy in pattern recognition for turbo-generators than k-Nearest Neighbors and Support Vector Machines. A method based on vibration signals to generate original features used for rotating machine fault diagnosis was introduced in [12]. The fault diagnosis of direct current motors was studied in [23], in which acoustic signals are analyzed using reflection coefficient and K-Nearest Neighbor classifier.

Focused on induction motors bearing fault detections [94] explains how Artifical Neural Networks are used to identify patterns in time domain features. Time and frequency dependent parameters are used in [6] to build up a pattern vector. Then, using k-Nearest Neighbors rules or linear discriminant functions, broken bars and stators unbalance in induction motors are classified.

The most modern language of machine learning and the fundamental concepts employed in Data Science to assess and describe the condition of the machine is used.

Signal processing enables the extraction of specific signatures existing in the signals and allows taking action mostly by subject-matter expert prior knowledge/experience but not by the learning.

In statistical signal processing is intended to identify those patterns that represent known boundaries, symbols sets or else known kinds of signal domains. On the other hand, machine learning is directed towards detecting or learning to detect previously unknown patterns.

So the idea is to get an adequate pattern representation of real-world signals using signal processing, then feed these patterns to machine learning algorithms to learn the scenario.

4.1 QUANTITATIVE OBSERVATIONS, QUALITATIVE RESPONSES

Digital signal processing is characterized by the unique type of data it uses: signals. The data here is treated as sampling points, but it is widely accepted in electrical engineering that when talking about signals, is related to time-varying waveforms.

In this research, signal processing refers to traditional techniques that require subject-matter experts and often a set of assumptions are established to define the applicability of the analysis. In other words, data signal processing refers to the ability to create a tidy dataset from the raw sensor data.

Each signal is kept as a variable, in numerical form, stored all together as columns in table form. Each row in the table is a sample taken for a given machine condition. Then, four categorical variables were assigned describing the machine condition, set-point, fault type and severity. The variable *Class* codifies the other four categories, as all the machine conditions combinations, shown in Table 4-1.

Table 4-1. Initial training class definition, 20 different machine conditions.

Categorical variable	Description	Values
Condition	This category register if the state of the machine is healthy or faulty.	0 – Healthy 1 – Faulty
Set-point	Refers to the level of load at which the machine is working.	1 – Sp1 (50%) 2 – Sp2 (100%)
Fault type	The simulated stator winding short-circuit, by deviating current from the winding through an external regulable resistor.	0 – Healthy 1 – A31A32 2 – A31A42 3 – A41A42
Severity	The amount of current deviated in each fault type.	0 – Healthy 1 – 500 mA 2- 1000 mA 3 – 2000 mA
Class	Aggregation of the other categorical variables to be used during supervised training.	The combination of all others to define one class.

The Table 4-1 contains all the categorical variables and the numerical coding used to define the qualitative responses. The class used to train the model uses the following format:

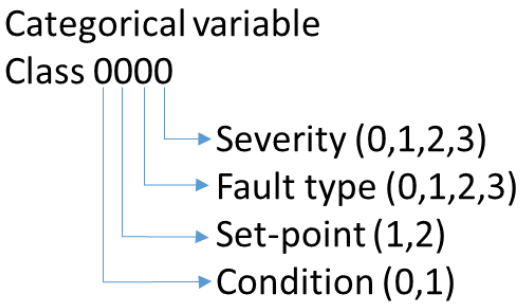


Figure 4-1 Formatting used for the categorical variable Class.

Finally, these datasets are ready for pre-processing in which the information is extracted into features, as a result of considering shorter time intervals or observations, each, describing the condition of the machine winding.

4.2 Data collection and Tidy data

The target is to transform the sensor data and turn it into a tidy dataset **[87]**, which then can be used to do the downstream analysis. It is necessary to manage each column in the tidy dataset to correspond to one feature.

The signals from the sensors are collected as a time series with sampling ratio 19 kSamples/seconds. To have workable file's size, each experiment's dataset was collected in data frames of a timespan of 2 minutes, six data sets for each case under consideration. So enough information is gathered to train, test and validate different machine learning models. Each file is comprised of thirty-eight variables (thirty-seven sensors and the timestamp), 2.280.000 data points per each signal ($19000\frac{Samples}{Second} * 120\ seconds = 2.280.000$)

The variables are:

- Current of phase R - I_R, I_A, I_a and its two parallel branches, the same applied for phase S - I_S, I_B, I_b and T -I_T, I_C, I_c;
- Field current - I_f;
- The phase voltages - V_{AN}, V_{BN}, V_{CN};
- Voltage between machine neutral and ground - V_{n_pe};

- Voltage between machine shaft and ground point - V_{shaft};
- Mechanical torque - $Torq$;
- Rotational speed - ω_1;
- A Rogowski coil (**RG**) per bar (**b**) and pole (**P**), only two consecutive poles, identified as P2 and P3:
    - $P2RG_{b1}, P2RG_{b2}, P2RG_{b3}$ and $P3RG_{b1}, P3RG_{b2}, P3RG_{b3}, P3RG_{b4}$;
- Voltage between the ends of each bar per pole, only two consecutive poles - $P2_{b1}, P2_{b2}, P2_{b3}, P2_{b4}$ and $P3_{b1}, P3_{b2}, P3_{b3}, P3_{b4}$;
- Noise-canceling membrane microphone - Mic_{NC};
- Omnidirectional membrane microphone - Mic_{OM};
- Analog hall sensors (**H**) in pole 3 – $P3H_1, P3H_2$ and pole 2 – $P2H_2$.

Refer to section 2.4 “Data acquisition system” for details of the installation of these sensors.

4.3 KNOWLEDGE EXTRACTION

Learning a rule from data, such as the thresholds of a fault indicator, also allows knowledge extraction. Knowledge extraction will be achieved by enabling machine learning over the dataset under consideration.

Concretely, consider the task of assessing the condition of a synchronous machine. A set of examples is given to a learning algorithm, where each case contains a condition as specified by a set of *features* (e.g. electrical power, spectrum, and other latent variables) along with the *class* (a qualitative response) for the intended machine’s condition.

The fundamental assumptions of learning require that the distribution of training examples to be identical to the distribution of test examples (including future unseen examples). In practice, this assumption is often violated to a certain degree.

For an algorithm to learn from data, the dataset must be formatted. Pre-processing and cleaning data are necessary steps that typically must be conducted before a dataset can be used effectively for machine learning. Raw data is often noisy and unreliable. In this sense, feature engineering is required, and this is where the

implementation of DSP techniques is needed consistently and programmatically to prepare/format the data to be fed into machine learning algorithms.

However, with the recent availability of Big Data platforms and access to more data processing power, deep learning algorithms are capable of modeling high-level abstractions in data, using multiple layers artificial neural networks. The features are automatically extracted instead of manual extraction as in feature engineering.

The implementation of deep learning requires large volumes of data, representing every possible condition of the machine, and a compromise between processing capacity and processing time.

Artificial neural networks, using deep learning architecture will become more affordable, but for now, a more traditional methodology based on DSP as a pre-processing step is under consideration to classify different fault conditions used together with supervised machine learning algorithms.

4.4 FILTERING

The first step is to perform a visual inspection of the data. Figure 4-2 displays six different signals at different time intervals; unwanted transients, or spikes are observed.

Taking into account that all the signals were collected under steady state condition, no local maxima or peak values are expected.

The filtering is applied in two steps. Firstly, the spikes are identifying as those data-points with an absolute value greater than 1,5 standard deviations. This approach is based on the so-called three-sigma rule of thumb, but considering a more restrictive interval of 86.64% of the available data points.

Then, the removed points are replaced by Piecewise Cubic Hermite Interpolating Polynomial (**PCHIP**). Finally, a median filter, considering only three adjacent points, is applied to smooth the signals, represented in Figure 4-3.

A justification for this type of filtering is mainly related to the nature of the signals under consideration. In an electrical machine, directly supplied from the grid, short duration events such as spiking or transients and electronic noise are most probably caused by external factors.

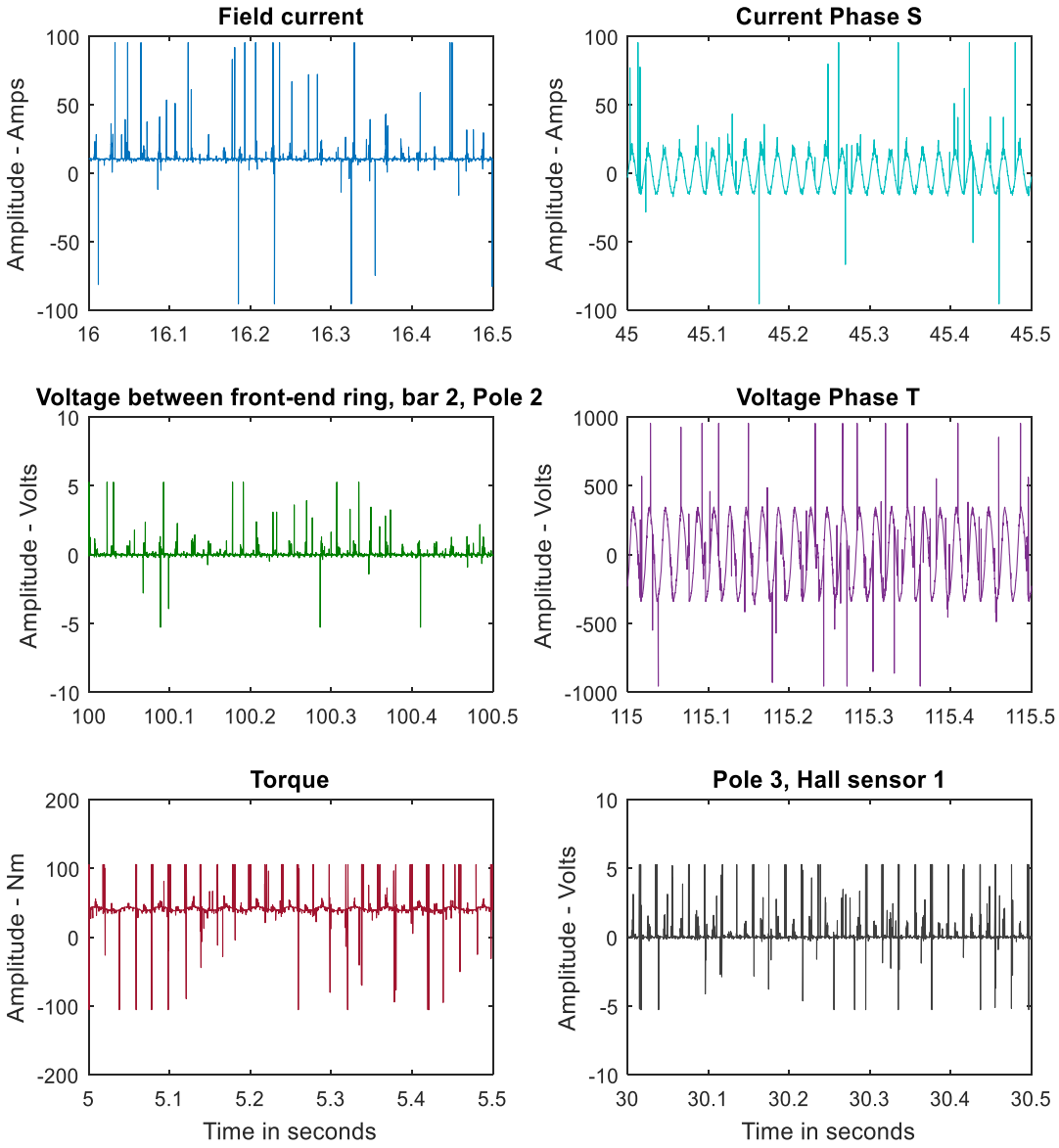


Figure 4-2 Raw signals, before filtering.

The observed attenuation of the signals is negligible. Even when the theoretical limit according to the Nyquist-Shannon sampling theorem, for the chosen sampling frequency, allows the consideration of up to 10kHZ, the frequencies that will be evaluated while training the machine learning models are the frequencies between the fundamental (50Hz) and the slot harmonic (2,1kHz). The selected frequencies range is below the filtering cutting frequencies which are significant for frequencies higher than 3kHz.

These frequencies are considered to contain the highest amount of information inherent in the operation of the synchronous machine under symmetrical condition and stator or rotor winding fault.

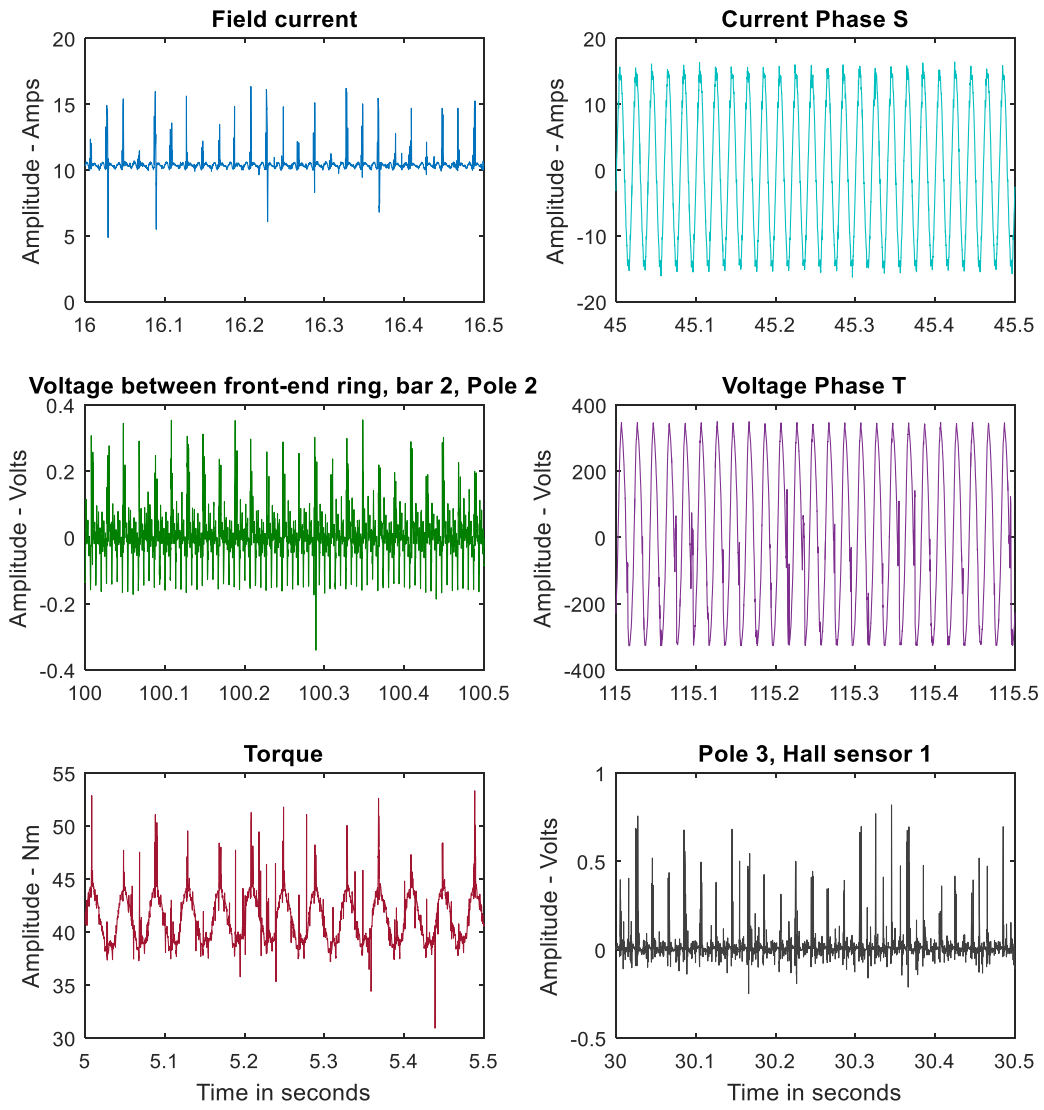


Figure 4-3 Signal after PCHIP interpolation of the 1,5-σ rule for outliers, and median filter taking into account 3 neighboring data points.

The filtering pseudocode is as follow:

Algorithm 1: **Filters**

1.1 *Load* dataset for a given machine condition.
1.2 **n** columns.
1.3 **nstd** standard deviations, range.
1.4 *for* Each **n** column *ii* = 1…**n** *do*
1.5 Calculate 50% trimmed mean
1.6 Calculate the thresholds.
1.7 Identify outliers; data points outside the thresholds.
1.8 Substitute the outliers by its PCHIP interpolated value.
1.9 Apply median filter considering 3 adjacent points.
1.10 End

4.5 FEATURES CALCULATION, MODELING OF THE SIGNALS.

Each collected dataset contains 120 seconds for a given machine condition, sixty datasets in total are available; 6 per each machine condition, equivalent to 12 minutes of data collected on different days, intended to capture any impact of different external disturbances.

To select the number of samples per observation, the main criteria applied was to get enough resolution and low processing cost using a second-order Goertzel algorithm to compute the multiple harmonics from $50\,Hz$ up to $2{,}5\,kHz$. The Goertzel algorithm performs a more efficient Fast Fourier Transform in computing than N-point DFT.

The frequencies of the harmonics in the DFT depend on the length of the transform, *N*, and they are integer multiples of the fundamental frequency. Therefore, the chosen length is $N = 7600$, which gives a minimum FFT resolution of 2,5Hz and comprises 20 cycles of the fundamental harmonic.

Inspired by P. Welch's methodology, time averaging estimation of the spectrum [86] and contrasted with different windows functions as in [26]. The selected window of 7600 samples, is overlapped by a factor of 10% (2 cycles) and multiplied by a Hamming window, so the effect of considering incomplete cycles is minimized.

The validity of this method has been published as part of the initial research for this dissertation in [18] and re-evaluated in Chapter 5 using machine learning.

Algorithm 2: **Features calculation**

1.1 *Load* filtered dataset for a given machine condition.

1.2 Calculate *Tt* iteration sequence from the *t* timestamp.

1.3 $Tt = from\ \min[(t) + \frac{22\ \text{cycles}}{50Hz}]\ to\ [\max(t)],\ step = \frac{22\ \text{cycles}}{50Hz}$

1.4 *for* Each *jj* along *Tt do*

1.5 Subset data having t >= $Tt(jj) - \frac{22\ \text{cycles}}{50Hz}$ *AND* t <= $Tt(jj)$

1.6 Normalize data sub-set.

1.7 Calculate statistical time domain features.

1.8 Calculate frequency domain features.

1.9 Calculate electric power features.

1.10 Calculate Clarke transformation features.

1.11 Calculate Park transformation features.

1.12 End

In summary, each observation is comprised of a subset of 8360 samples, to which different features, grouped in five different domains are calculated. After these transformations, five datasets containing 2990 observations each are generated.

Each calculated dataset corresponds to a specific group of features, calculated from different related functions to five different domains or transformations. In the following tables the number of resultant features after applying the specified function is expressed inside "( )"; in most of the cases, the function is directly applied over the data frame containing all the signals.

4.5.1 FREQUENCY-DOMAIN FEATURES:

A periodic function can be represented using Fourier series, so the first 50 harmonics, from 50 Hz to 2500 Hz are calculated. A feature is created containing the magnitude and phase information from each harmonic, referred to the fundamental. The extracted features are.

Table 4-2 List of frequency-domain features:

Amplitude at 50 *Hz* (37)
Amplitude referred to the fundamental[1] (1813)
Phase of fundamental (37)
Phase referred to the fundamental[2] (1813)

[1] The amplitude of each harmonic is divided by the amplitude of the fundamental

[2] The phase of each harmonic minus the phase of the fundamental.

The amplitude and phase angle are calculated for each variable; the output *freq-dataset* contains 2990 observations and 3700 features as a result of applying the Goertzel algorithm for the first 50 harmonics of each variable.

4.5.1 STATISTICAL TIME-DOMAIN FEATURES:

These are the most common statistical functions that are considered to have a meaningful outcome for this type of data, the relevance of each, is evaluated when used to train different classification algorithms.

Table 4-3 List of statistical features:

Mean, μ (37)	Standard deviation, σ (37)
Harmonic mean (37)	Range (37)
Interquartile range(37)	Mean absolute deviation (37)
Median absolute deviation (37)	$\hat{\sigma}$ from the mean[1] (37)
$\hat{\sigma}$ from the median[1] (37)	Skewness (37)
Kurtosis (37)	ℓ^2-norm (37)
Band power (37)	Spurious-free dynamic range (37)
Signal to noise and distortion ratio (37)	Signal to noise ratio, 2[2] (37)
Signal to noise ratio, 5[2] (37)	Signal to noise ratio, 10[3] (37)
Tot. harmonic distortion,5[3] (37)	Tot. harmonic distortion,8[3] (37)
Tot. harmonic distortion,12[3] (37)	

[1] The mean and median are used as a consistent estimator for the estimation of the standard deviation. The difference between the estimated and calculated σ is used as a feature.

[2] The signal to noise ratio (**SNR**) is calculated at three different noise levels, excluding the power contained in the n lowest harmonics; the SNR for harmonics order greater than 2, 5 and 10 are calculated.

[3] Total harmonic distortion (**THD**) in dBc of the real-valued sinusoidal signal with n harmonics included in the calculation.

There is no limitation or association to the type of variable to which the selected statistical features can be applied. The *stats-dataset* contains 2990 observations and 777 features.

4.6 CLARKE TRANSFORMATION FEATURES:

Clarke transformation is often used to simplify the analysis of three-phase circuits from a stationary *a-b-c* reference system to a system with coordinates $\alpha, \beta, 0$ **[50]** and no so often used to detect faults in electric motors **[76][74]**. Where

coordinates α, β are orthogonal to each other, and coordinate 0 corresponds to the zero-sequence component.

This is an algebraic transformation, which also produces a stationary reference system. The electric power components are then calculated from voltages and currents in the $\alpha, \beta, 0$ coordinates. Here used as features, so the variability of these will be meassured and quantify as a fault indicator.

Table 4-4 List of Clarke transformation features:

FFT, Amplitude at 50Hz (12)	Phase of fundamental (12)
FFT, Amplitude referred to the fundamental (588)	Phase referred to the fundamental (588)
Power from dot product, sequence 0. (6)	Reactive power from dot (1)
Power from dot product (6)	RMS [1](12)
$\frac{P_{Clarke}}{P_{phase\_dot}}$ (1)	$\frac{P_{Clarke}}{P_{phase\_rms}}$ (1)
$\frac{Q_{Clarke}}{Q_{phase\_rms}}$ (1)	

[1] Applied to $V_0, V_\alpha, V_\beta, I_0, I_\alpha, I_\beta$

The *Clarke* dataset contains 1224 features. 1187 of these, are features in the frequency domain, similar to the features calculated in the *freq-dataset*, but the variables used are obtained from the Clarke transformation. The other features are similar to the ones computed for the *VI_power dataset*.

4.6.1 PARK TRANSFORMATION FEATURES:

Park transformation has the unique property of eliminating all the time-varying inductances from the circuital equations of the monoharmonic model of the balanced three-phase AC machines.

Under a fault condition, the inductances of the machine might vary, these variations are expected to be kept during the transformation. The Park_dataset contains 817 features. 800 of these features are in the frequency domain.

Similar to the features extracted for the *Clark-dataset*, the Park transformation features are:

Table 4-5 List of Park transformation features

FFT, Amplitude at $50Hz$ (8)	Phase of fundamental (8)
FFT, Amplitude referred to the fundamental (392)	Phase referred to the fundamental (392)
Power from dot product [2] (3)	Reactive power from dot[2] (3)
RMS [1] (8)	$\frac{P_{Park}}{P_{phase\_dot}}$ (1)
$\frac{P_{Park}}{P_{phase\_rms}}$ (1)	$\frac{Q_{Park}}{Q_{phase\_rms}}$ (1)

[1] Applied to the Park magnitudes $V_{dqs}, I_{dqs}, I1_{dqs}, I2_{dqs}$

[2] Calculated with the main and the parallel currents.

4.6.2 ELECTRIC POWER FEATURES, VIPOWER-DATASET:

Using calculated vectorial magnitudes or the harmonic content extracted after FFT transformation, the relationship between currents and voltages is evaluated.

Table 4-6 List of electric power features, current and voltages magnitude, and phase:

Voltage amplitude ratio[1]	$\boldsymbol{V_{i,j} = \frac{\max(fft(V_i))}{\max(fft(V_j))}}$
Lag from FFT[1]	$lag\_V_{i,j} = \frac{phase(V_i)}{phase(V_j)}$
Lag from dot product[1]	$dotlag\_V_{i,j} = \frac{\cos^{-1}(V_i \cdot V_j)}{norm(V_i) * norm(V_j)}$
Lag difference[1]	$lagDiff\_V_{i,j} = abs\left(lag_{V_{i,j}} - dotlag_{V_{i,j}}\right)$
Parallel branch current lag[2]	$lag\_I_{phase,branch} = phase(I_{ph,branch}) - phase(I_{ph})$

[1] Calculated for the main currents and for each parallel branch.

[2] Only the lag between each parallel branch for a given phase and the main phase current is calculated.

Table 4-7 List of electric power features; active, reactive power, and efficiency. Only fundamental harmonic is being considered.

Mechanical power	$\boldsymbol{P_{out} = Torque * \omega}$
Phase power dot product 1	$P_{phase\_dot} = \frac{sum(V_a \cdot I_{a1})}{N_{samples}}$
Phase power RMS cos φ 1	$S_{rms} = RMS(V_{phase})\, RMS(I_{phase\ or\ branch})$ $\varphi = cos\left(phase(V_{phase}) - phase(I_{phase})\right)$ $P_{phase\_rms} = S_{rms}\ cos(\varphi)$
Reactive power RMS cos φ 1	$Q_{phase\_rms} = S_{rms}\ sin(\varphi)$
η_{phase} dot product	$\eta_{phase} = \frac{P_{out}}{3 * P_{phase\_dot}}$
η_{phase} from RMS power	$\eta_{phase_{rms}} = \frac{P_{out}}{3 * P_{phase_{rms}}}$
Diff. main-parallel, dot	$P_{sum,dot} = P_{phase\_dot} - (P_{branch1,dot} + P_{branch2,dot})$
Diff main-parallel, rms	$P_{sum,rms} = P_{phase\_rms} - (P_{branch1,rms} + P_{branch2,rms})$
Power phase ratio, dot	$K_{i,dot} = \frac{P_{i,dot}}{P_{a,dot} + P_{b,dot} + P_{c,dot}}$
Power phase ratio, rms	$K_{i,rms} = \frac{P_{i,rms}}{P_{a,rms} + P_{b,rms} + P_{c,rms}}$
η_{Total} from the dot product	$\eta_{Total,dot} = \frac{P_{out}}{sum(P_{phase\_dot})}$
η_{Total} from rms	$\eta_{Total,dot} = \frac{P_{out}}{sum(P_{phase\_rms})}$
Averaged $Q_{instantaneous}$	$Q_{instant} = sum\left(\frac{(V_a - V_b)I_c}{\sqrt{3}} + \frac{(V_c - V_a)I_b}{\sqrt{3}} + \frac{(V_b - V_c)I_a}{\sqrt{3}}\right)\frac{1}{N_{samples}}$
$Q_{phase,ratio}$	$Q_{phase,ratio} = \frac{Q_{phase,rms}}{Q_{instant}}$

[1] All these functions applied to all phases and neutral combinations such as $\boldsymbol{V_{ba}}, \boldsymbol{V_{ca}}, \boldsymbol{V_{cb}}, \boldsymbol{V_{na}}, \boldsymbol{V_{nb}}, \boldsymbol{V_{nc}}$ for voltages and main and parallel currents I_{ab}, $I_{ac}, I_{cb}, I_{a1,a}, I_{a2,a}, I_{b1b}, I_{b2b}, I_{c1c}, I_{c2c}$. Fundamental harmonic only.

In this dataset only the supply voltage, the main and parallel branches currents and the neutral voltage are considered, no rotor mounted sensor is considered.

The features calculated here intend to exploit the asymmetries between phases that might arise during abnormal behavior as a consequence of winding faults.

The validity of these as fault indicator is under consideration. The *VI_power* dataset contains 114 features in total. These features are specific to the domain of electrical machines, and the same magnitudes can be calculated using different mathematical approaches.

As seen in Table 4-7, different mathematical approaches (dot product or root mean squared) can be applied to calculate the variability of the electric power, as a function of the voltage and current in each phase. It is believed that any winding shortening or short circuit might affect these quantities.

4.7 MACHINE LEARNING USING *R*:

So far all the development of this research has been done using MATLAB, which is widely used for signal processing. However, even when Mathworks has recently introduced a new MATLAB toolbox to train models and classify data using supervised machine learning, called Classification learner, is still on earlier stage and only allows the implementation of ready, built-in algorithms and with limited options for tuning the model. For this reason, the chosen tool to train and build a classification model to assess the condition of electric motors is *R*.

R is a free, high-level programming language that provides flexible objects and functions for statistical analysis, data mining, machine learning, data visualization among other data analysis capabilities **[72]**. *R* has a large, active and growing community of users, so many documentation, and new algorithm implementations, distributed as packages are freely available. *R* is not just a statistics package it is a language designed to operate in the way that problems are thought about. *R* is both flexible and powerful.

5 SUPERVISED LEARNING: FAULT CLASSIFICATION TASK

Regardless of whether the learner is a computer or a human, the core learning process is similar. It can be divided into the following three components;

- **Data input**: It utilizes observation, memory storage, and recall to provide a factual basis for further reasoning.
- **Abstraction**: represent the data into broader representations.
- **Generalization**: based on abstracted data, to form a basis for action.

Supervised learning is the most commonly studied learning task. It relies on a set of input features, target features, and a set of training examples where the mapping between the target features and the input features is given, so the task is to identify the new input features.

When the target features are continuous, this is called regression, and classification for when the target variables are discrete or categorical classes. Initially, there are twenty different categories; ten per each load level, within these ten, three distinct winding faults evaluated at three different severities and the healthy; Table 4-1. Initial training class definition, 20 different machine conditions.

5.1 PROBABLY APPROXIMATELY CORRECTLY LEARNING - PAC

The alternative to complex models is a data-driven model. However, for an algorithm to build a model, it is necessary to provide the conditions from which the task or learner can be thought.

As stated in the theory of the learnable **[77]**, instances are generated at random from X according to some target distribution D; generally D not known to the learner, D is stationary, D may be any distribution.

Let X be a set called the **instance space**. In our case, X is the dataset containing all the observation and extracted features. A c**oncept** over X is just a subset $c \subseteq X$ of the instance space. In the assessment of the machine, the concepts are the four categorical variables: the machine condition, set-point, fault type and severity encoded all in one *class* variable **[33]**.

A concept can thus be thought of as the set of all instances that positively exemplify some simple or interesting rule. We can equivalently define a concept to be a boolean mapping c: $X \rightarrow \{0, 1\}$, with $c(x) = 1$ indicating that x is a positive example of c and $c(x) = 0$ pointing out that x is a negative example.

For this reason, X it is also sometimes called the **input space**. A **concept class** C over X is a collection of concepts over X. Ideally, we are interested in concept classes that are sufficiently expressive for reasonably general knowledge representation.

In the SM fault detection model, a learning algorithm will have access to not only positive and negative examples (boolean case), but up to twenty classes, of an unknown **target concept** c, chosen from a known concept class C. The learning algorithm will be judged by its ability to identify a **hypothesis concept** that can accurately classify instances as examples of c.

LEARNER PERFORMANCE

Having D as any fixed probability distribution over the instance space X, if h is any concept over X, then the distribution D provides a subjective measure of error between h and the target concept c; namely, we define

$$\boldsymbol{error}(\boldsymbol{h}) = \boldsymbol{Pr}_{\boldsymbol{x}\in\boldsymbol{D}}[\boldsymbol{c}(\boldsymbol{x}) \neq \mathbf{h}(\boldsymbol{x})]$$

where the subscript $\boldsymbol{x} \in \boldsymbol{D}$ to $\mathbf{Pr}[\cdot]$ indicates that the probability is taken with respect to the random draw of x according to D.

This is a conceptualized definition of PAC taken from **[33]**. More learning frameworks and extended analysis and applications can be studied as part of the Computational learning theory, which is a sub-field of Computer Science.

5.2 SUPERVISED LEARNING

An abstract definition of supervised learning is as follows. Assume the following data is used to train a learner:

- a set of input features, X1, …, Xn
- a set of target features, Y1, …, Yk
- a set of training examples or training dataset, where the values of the input features and the target class are given for each observation/example.
- a set of test examples or testing data set, where only the values for the input features are given.

Table 5-1 The knowledge matrix from all the available dataset

Observations	Clark	Park	VI power	Freq	Stats	Class
obs-1	$()_{1x1224}$	$()_{1x817}$	$()_{1x114}$	$()_{1x3700}$	$()_{1x777}$	Machine condition
obs-2	$()_{2x1224}$	$()_{2x817}$	$()_{2x114}$	$()_{2x3700}$	$()_{2x777}$	Machine condition
⋮	⋮	⋮	⋮	⋮	⋮	⋮
obs-n	$()_{nx1224}$	$()_{nx817}$	$()_{nx114}$	$()_{nx3700}$	$()_{nx777}$	Machine condition

As seen in previous Chapters, the datasets created from the lab experiments, are all obtained in the same way. Each observation (for a given machine condition) represents the state of the machine for a timestamp of 0.44 seconds (22 AC cycles at 50Hz), and enough data was collected to balance the classes over all the considered machine conditions.

Hence, the supervised learning task is focused on multi-class problem used to detect different levels of fault conditions and severity levels, categorized in various target features as indicated in Table 4-1.

5.2.1 GENERATIVE AND DISCRIMINATIVE LEARNING

In the classification problem in which the goal is to assign *y* labels to a given *x* input, **Discriminant Learning** evaluates:

$$\boldsymbol{f(x) = \arg\max_{y} p(y|x)} \quad 5\text{-}1$$

which merely chooses what is the most likely class considering *x*; the decision boundary between the classes is modeled.

Generative Learning explicitly models the actual distribution of each class. Using Bayes' rules to calculate the posterior distribution of *y* given *x*

$$\boldsymbol{p(y|x) = \frac{p(x|y)p(y)}{p(x)}} \quad 5\text{-}2$$

and since the algorithm estimates the *arg max* to make predictions, it is not necessary to calculate the denominator **[47][39]**, therefore

$$\boldsymbol{f(x) = \arg\max_{y} p(x|y)\, p(y)} \quad 5\text{-}3$$

In summary, a generative algorithm models how the data was generated in order to categorize a signal. One of the advantages of generative algorithms is that it is possible to use *p(x,y)* to produce a new data similar to existing data. Discriminative algorithms do not distinguish how the data was generated, simply categorizes a given signal. Generally, provide better performance in classification tasks **[78]**.

5.2.2 PERFORMANCE METRICS FOR CLASSIFICATION MODELS

A common method for describing the performance of a classification model is the confusion matrix. In the table below, a simple version for a two-class model is shown. Where *TP* – true positives, *FP* – false positive, *FN* – false negative and *TN* – true negative.

Table 5-2 The confusion matrix for the two-class problem.

	Observed	
Predicted	Event	Non-event
Event	TP	FP
Nonevent	FN	TN

Given m classes (where $m \geq 2$), a confusion matrix is a table of at least size m by m. For a classifier to have good accuracy, most of the elements would be represented along the diagonal, with the rest of the entries being zero or close to zero.

That is, ideally, *FP* and *FN* are around zero. The details of its calculation, are provided in the following sections; in Table 5-3 is possible to see the results after training a Sparse Linear Discriminant Analysis used to extract most relevant features from the VIpower dataset.

The task in feature selection for fault classification and prediction is to maximize the classification accuracy; that is, to distinguish as unequivocally as possible a particular fault condition from others and the healthy state of the system at hand while providing maximum prediction accuracy and precision **[25]**.

Table 5-3 VIpower dataset sparseLDA confusion matrix, sp1.

CONFUSION MATRIX

	Reference									
Prediction	0100	1111	1112	1113	1121	1122	1123	1131	1132	1133
0100	70	0	2	0	0	1	0	0	0	0
1111	0	79	3	0	1	2	0	0	0	2
1112	0	2	55	0	0	3	0	2	0	2
1113	1	0	0	67	0	1	2	0	0	1
1121	0	0	0	0	77	0	0	0	0	0
1122	5	1	10	1	0	57	0	3	0	0
1123	0	0	0	2	0	0	71	0	0	0
1131	0	1	3	0	0	7	0	61	0	2
1132	0	0	0	1	0	0	1	0	71	0
1133	0	1	12	1	0	0	0	0	0	62

In other words, quantifying how strongly a feature is associated with a specific class, so that by knowing the value of a feature (or set of features) it is possible to predict or, in this case, identify the condition of the machine.

The confusion matrix for m classes consist of the intersection matrix between the predicted values with the reference values, so it is possible to estimate the accuracy of the model to identify each of the classes.

5.3 EVALUATION METRICS

Once the confusion matrix (**CM**) is calculated for the predicted values versus the testing dataset (reference values) and considering that the number of cases per set

point (machine conditions) under evaluation is ten, we get a matrix of 10x10 from which the accuracy, sensitivity, precision and $F1_{score}$ are calculated.

In addition to the evaluation metrics, classifiers can also be compared with respect the level of understanding and the information given by the classifier or predictor, known as **interpretability**, then the computational costs associated with the training and validation of a given classifier, **speed**.

Table 5-4. Variables to calculate Evaluation metrics

Number of observations	$n.obs = sum(CM)$
Number of classes	$n.class = nrow(CM)$
Correctly classified observations.	$diag = diag(CM)$
Observations per class	$obs.class = apply(CM, 1, sum)$
Predictions per class	$pred.class = apply(CM, 2, sum)$
Distribution of observation over the reference classes	$p = \frac{obs.class}{n}$
Distribution of observation over predicted classes	$q = \frac{pred.class}{n}$

It is possible now, to define some basic variables that will be needed to compute the evaluation metrics. These are part of the *R* code created to calculate the accuracy of each model, so it is possible to select the model with higher performance.

Accuracy: number of correctly classified observations, out of all the observations, expressed as percentage

$$\boldsymbol{Accuracy, ACC = \frac{sum(diag)}{n}} \qquad 5\text{-}4$$

Per class metrics: to assess the performance with respect to every class in the dataset, these are the common per-class metrics:

- ***Sensitivity*** or recall is the true positive rate as the fraction of observations of a class that was correctly predicted

$$\boldsymbol{Sensitivity = \frac{diag}{obs.class}} \qquad 5\text{-}5$$

- ***Precision*** is the fraction of correct predictions for a certain class.

$$\boldsymbol{Precision} = \frac{\boldsymbol{diag}}{\boldsymbol{pred.class}} \quad 5\text{-}6$$

- The ***F1_score*** is defined as the harmonic mean of precision and sensitivity.

$$\boldsymbol{F1_{score}} = \boldsymbol{2} * \boldsymbol{Precision} * \frac{\boldsymbol{Sensitivity}}{\boldsymbol{Precision} + \boldsymbol{Sensitivity}} \quad 5\text{-}7$$

5.4 LARGE P SMALL N DATA

In high-dimensional data, such as the knowledge matrix built in this research, with a relative to each class large number of "p" predictors and small "n" number of samples ($p \gg n$), selecting features raises the statistical limitation of identifying fault indicators from a multivariate dataset; estimators of the sample mean and covariance matrix are usually unstable **[81]**.

In order to obtain reliable classification models, the learner needs to distinguish between 10 different classes of each predictor, in other words, to produce a substantially different answer between 10^p different configurations of the n-dimentional vector. Therefore, it is unlikely that we could gather and process the required number of samples.

No rigid rule defining the number of necessary features needed for each model type is established. Classifiers that generalize easily, e.g. linear classifiers, Naïve Bayesian, allow a greater number of features since the classifier itself is less expensive **[27][34]**. Classifiers that tend to model non-linear decisions boundaries very accurately do not generalize well, so algorithms such as neural networks, decision trees, k-Nearest Neighbors classifiers, etc., are prone to overfitting the data.

It is important to highlight that overfitting occurs both when estimating several parameters in a lower dimensional space, and when in a higher dimensional space, relatively few parameters are estimated. Based on the above-explained, it has been decided to process each dataset independently. In the following sub-section, the extraction of the most important features per each set-point and dataset domain is explained.

5.4.1 FEATURES CORRELATION

Selecting the right features from the data is a necessary task. The performance of the algorithm will vary depending on the total number of features considered, in addition to the overall processing time. In datasets containing highly correlated features, it is desirable to remove them. Many classifiers perform better if highly correlated features are removed; an exception of this is partial least squares.

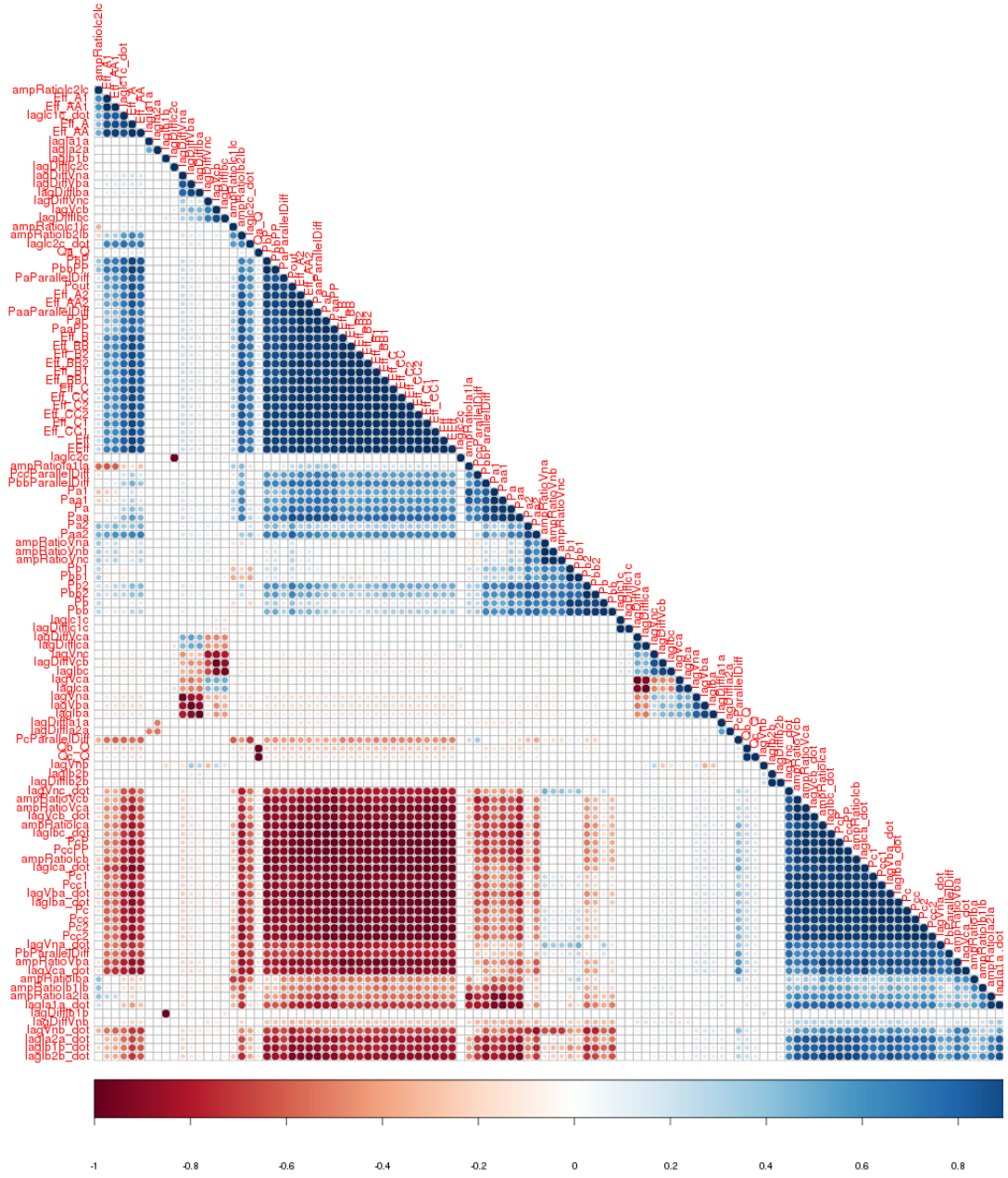


Figure 5-1 VI-power dataset, a matrix of correlation – ***generated in R, using corrplot package.***

In Figure 5-1, it is possible to see the matrix of correlation **[83]** for all the VI-power dataset features. The features are clustered based on how strong is the correlation between them. A simple option could be selecting one of the features

from each group and created a model from it, but this process does not guarantee that we keep all the information to assess the machine.

Due to the number of features in each features-dataset, using this paper format, it is only possible to represent the correlation matrix for the VIpower-dataset; only 114 features. This type of plot makes feasible to identify the features with correlation higher than a given threshold, then the mean absolute correlation of each variable is calculated, and the one with the largest value removed **[48]**.

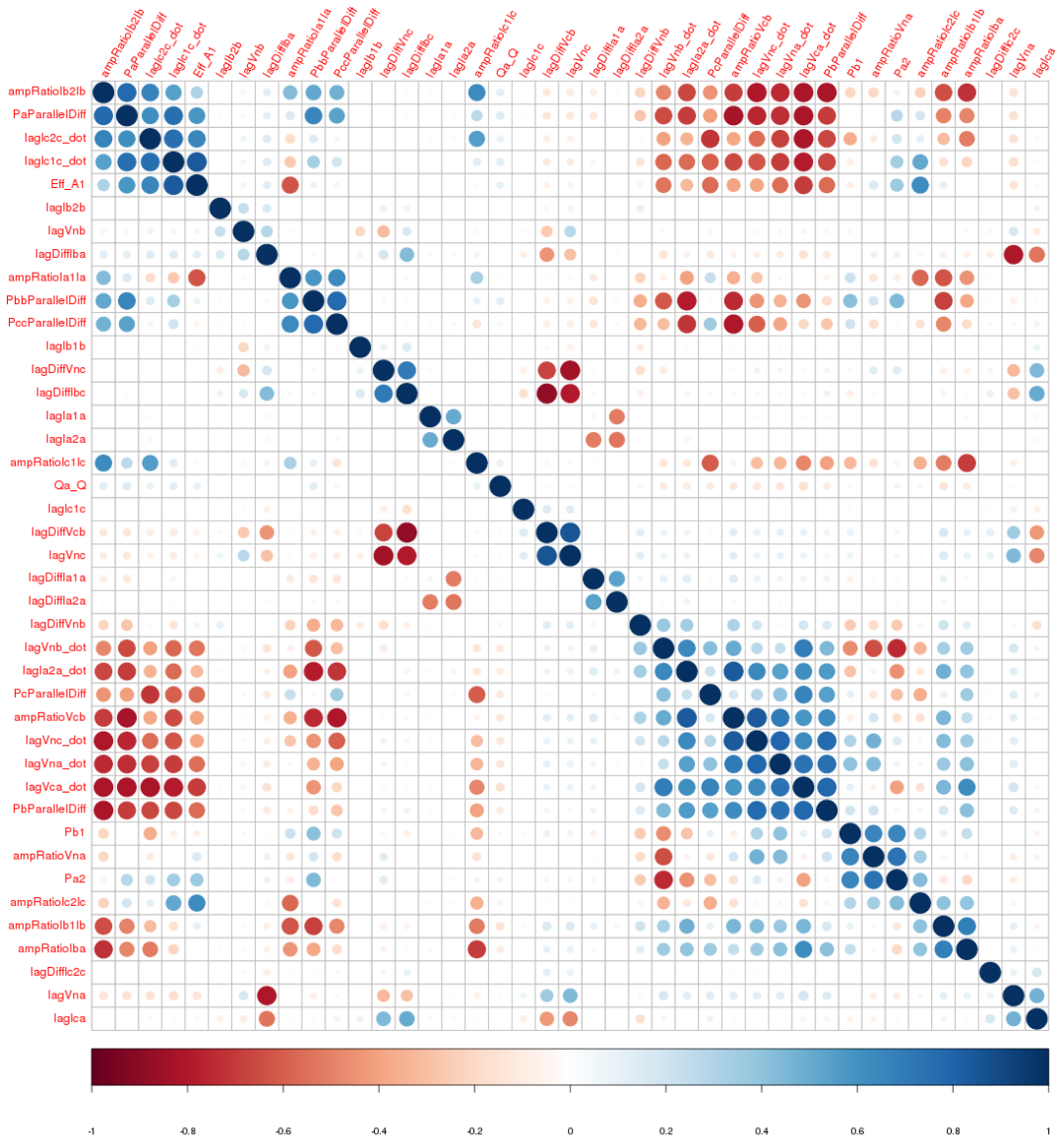


Figure 5-2 correlation matrix after removing features with correlation higher than 0,9 from the original dataset – *generated in R, using corrplot package.*

A fast and robust implementation of this is done in the R-package "caret," *findCorrelation()* function **[88]**. Choosing the threshold to identify highly correlated features it might be a complicated task that requires several iterations to build a model with high accuracy and no overfitting, in Figure 5-2 a threshold of 0.9 was

applied. This step can be avoided applying ℓ_2 penalty within model regularization, this penalty helps to mitigate the effect of correlated predictors.

It is also desired to implement models that are easy to interpret and the computation time, at least after the learning process, is fast enough to detect fault before it is catastrophic, without compromising the accuracy to assess the machine. Therefore, it is of interest, finding a classification algorithm that will let us identified, per class, the most important features from each dataset, so we can use it as feature extraction method and build a sufficient matrix of knowledge that will allow a reliable, fast and robust fault classification model.

5.5 FEATURES EXTRACTION, MULTICLASS PROBLEM.

Based on all the exposed it seems logical that the best classifier for Synchronous machine fault diagnosis learning task is Linear Discriminant Analysis (**LDA**). However, LDA cannot be directly applied because it is known to fail in cases where $p \gg n$. Moreover, when looking for options to simplify the laboratory set-up and make it more compatible to real industrial environments, it is desired to find a classifier that peforms feature selection as well. Hence, it will be possible to reduce the number of installed sensors and features extracted, without compromising the accuracy of the machine condition classification.

The available literature devoted to high-dimensional multiclass classification problem is concentrated on medical and chemometrics applications. Additionally to this, in recent years many sparse classifier methodologies were published. The most promising for the study of fault classification problem is Sparse Discriminant Analysis.

The authors in **[9]** have developed and implemented an R package for the sparse version of LDA. This model is a regularized version of LDA with two tuning parameters: Least Absolute Shrinkage and Selection Operator (**LASSO**) using the ℓ_1 penalty that eliminates unimportant predictors and hence provides feature selection, and Elastic Net using the ℓ_2 penalty that shrinks the discriminant coefficients towards zero. In **[89]** the theoretical rules of feature selection in LDA are studied, and a new method is proposed for Sparse Linear Discriminant Analysis (sLDA). **[20]** it's a complete review of Sparse Methods in regression and

classification with application to Chemometrics, and the multiclass problem for high-dimensional data using sLDA is presented in **[42]**.

5.5.1 SPARSE DISCRIMINANT ANALYSIS

The methodology used to extract the most relevant features per each dataset is based on the scientific development done in **[9]**. The code implemented in *R* is in Figure 5-3, this code is parallelized and the calculation executed across 32 cores. Then, taking advantage of the machine learning functions in the *caret* package, the sLDA method is evaluated for each dataset.

For a given n x p data matrix $\boldsymbol{X}$ and a vector of length n of the outcome y, then the LASSO solves the problem:

$$\boldsymbol{minimize_{\beta}}\{\|\boldsymbol{y}-\boldsymbol{X\beta}^{\boldsymbol{2}}\|+\ \boldsymbol{\lambda}\|\boldsymbol{\beta}\|_{\boldsymbol{1}}\} \quad 5\text{-}8$$

and the Elastic Net solves the problem

$$\boldsymbol{minimize_{\beta}}\{\|\boldsymbol{y}-\boldsymbol{X\beta}^{\boldsymbol{2}}\|+\ \boldsymbol{\lambda}\|\boldsymbol{\beta}\|_{\boldsymbol{1}}+\ \boldsymbol{\gamma}\|\boldsymbol{\beta}\|^{\boldsymbol{2}}\} \quad 5\text{-}9$$

The pseudo-code is below:

Algorithm 3: **Features extraction**

1.1	*For* Each features dataset
1.2	Create data partition, training 75% / testing 25%
1.3	Define control training as repeated 3 times, 5-fold Cross-Validation
1.4	Define number of selected variables and tune grid for lambda values
1.5	Declare parallel backend
1.6	Pre-processing training dataset: center and scale.
1.7	Fit sparseLDA model, metric Accuracy
1.8	Extract relevant predictors and importance score, append to previous iterations
1.9	Plot fitting results
1.10	*end*

```r
# My Feature extraction function

## Necessary Packages
require(foreach); require(doMC);
require(plyr); require(dplyr); require(caret);
require(corrplot);
## function
myVarImp <- function(dataset)
{
setwd("~/Yo/FeatiresExtract")

file <- paste("~/data/", dataset, ".csv", sep = "")
## Load dataset
df <- read.csv(file,
       stringsAsFactors=FALSE,
       row.names = NULL,
       header = TRUE,
       na.strings = c(NA, "", " "))
## Modify the classes to start with a character
   df$Class <- paste("C", df$Class, sep = "")
## Removing the classess
temp <- df %>%
    select(-Condition, -sp, -Fault, -Severity, -Class) %>%
    na.omit()
## Calculate matrix of correlation
cor.temp <- cor(temp)
## Remove high correlated variables
temp <- temp[, -findCorrelation(cor.temp, .9)]
## Seed to guarantee reproducible results
set.seed(0402)

# Preparin the data
## Creating data partition
inTrain <- createDataPartition(df$Class,
                               p = .75, list = FALSE)[, 1]

## Training(75%)
training <- temp[ inTrain, ]
  row.names(training) <- NULL
  trainingClass <- factor(df$Class[ inTrain ])
## External testing dataset(25%)
testing  <- temp[-inTrain, ]
  row.names(testing) <- NULL
  testingClass <- factor(df$Class[ -inTrain ])
```

```r
# Define and train the model
## Defining training control and grid
ctrl <- trainControl(method = "repeatedcv",
                     repeats = 3, number = 5,
                     verbose = TRUE,
                     classProbs = TRUE)
## Define sparse grid to tune the training algorithm
sparseLDAGridx <-  expand.grid(.NumVars = c(1:15),
                               .lambda = c(0, 0.01, .1, 1, 10, 100))

# Declare parallel backend
registerDoMC(cores=32)

time.1 <- Sys.time()

# Training the model
fit <- train(trainingClass ~ .,
             data = cbind(training, trainingClass),
             method = "sparseLDA",
             tuneGrid = sparseLDAGridx,
             trControl = ctrl,
             metric = "Accuracy", # not needed it is so by default
             importance=TRUE,
             preProc = c("center", "scale"))

## fitting model plot: Accuracy vs. Number of features
plot(fit)
## Extract the names of variables with higher accuracy per class
myPredictors0 <- data.frame(dataset = dataset,
                            SDA_Predictors = predictors(fit))

time.2 <- Sys.time()
print(round(time.2 - time.1, 2))

return(myPredictors0)
}
```

Figure 5-3 Features extraction R code.

By taking advantage of the model parameterization, it is possible to specify a vector or search grid establishing the possible combinations to try, To decrease the processing time and the number of iterations, the range of features to be evaluated per class is fixed to maximum fifteen, for each iteration for a given range between $(0\ to\ 100)$ of empirical values of lambda.

To avoid overfitting and validate the fitted model, the strategy chosen was to split the data into training, with 75% of all the data-points balanced for every class, and 25% to evaluate the performance of the model. Then, the sLDA training is validated using repeated k-fold Cross-Validation. Finally, before fitting the model, the data is standardized.

In a further step, and in order to confirm the results, the models will be re-evaluated considering observations from datasets taken on different days.

The information to evaluate the importance of all the features is taken from each model simulation, so the importance is tight to the model performance. The ability of each feature to correctly classify the condition of the machine is estimated for each class. The selected features can be repeated or considered in more than one class.

To extract the features and evaluate the performance of the model, two separate set points at 50% and 100% of nominal load are considered. Initially, each set-point is considered independently of the other; this gives us 20 classes in total. Later on, this is re-evaluated, and the ability of the model to correctly classify the status of the SM for different set-points simultaneously will be considered.

The following sub-sections contain the results from fitting the sLDA model to each features-dataset. The ability to classify healthy over faulty condition, and to classify up to ten machine condition is represented in a 3D scatter plot based on the top three more important features extracted from the model.

5.5.2 CLARKE-DATASET FEATURES EXTRACTION

The Clarke-dataset consist of 1224 features. After fitting the data to a sLDA model, we can observe that the optimal model is achieved with $\lambda = 1$ and 7 features per class for machine set to 50% of nominal load (sp1). A feature could be repeated in more than one class

The figure below contains the repeated (3 times) cross-validation (5 folds) for different combinations of accuracy and number of variables for a given lambda value. The processing time using 32 cores, at 2.2 GHz processor base frequency, was 9.89 hours.

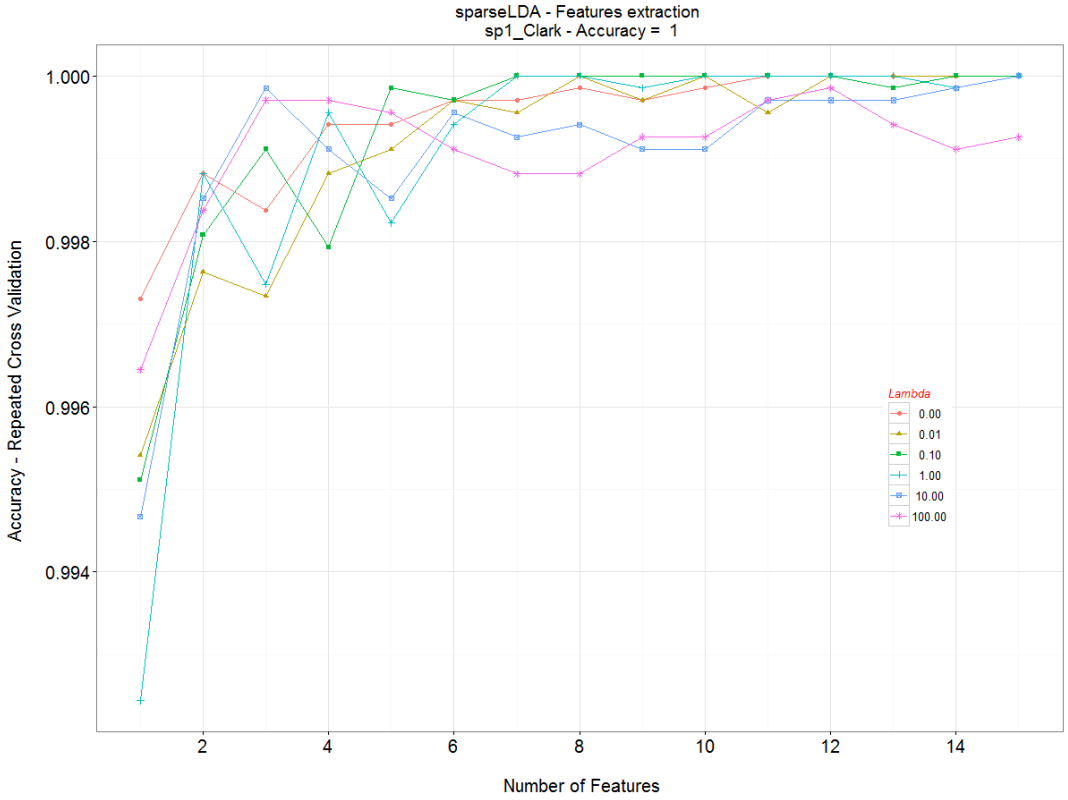


Figure 5-4 sp1 Clarke-dataset, sLDA features extraction as a function of lambda.

Considering the first three most significant variables Clark_Vab0_α $= \sqrt{\frac{2}{3}}\left(V_a - \frac{1}{2}V_b - \frac{1}{2}V_c\right)$, Clark_Iab0_1_α $= \sqrt{\frac{2}{3}}\left(I_{aA} - \frac{1}{2}I_{bB} - \frac{1}{2}I_{cC}\right)$, parallel branch 1 and Clark_Iab0_2_α $= \sqrt{\frac{2}{3}}\left(I_{aa} - \frac{1}{2}I_{bb} - \frac{1}{2}I_{cc}\right)$, parallel branch 2, for a balanced three-phases system, it is possible to verify that the α component is in phase with the *phase A*, and since the fault was seeded in this phase, seems reasonable that the variables that better identify the fault are directly related to the *phase A*.

The three most important features within the Clark-dataset are represented in the following 3D scatter plot (see Figure 5-5). Each observation is represented by a marker whose position depends on the relationship between the three features.

It is also possible, to identify how each of the conditions forms a cluster.

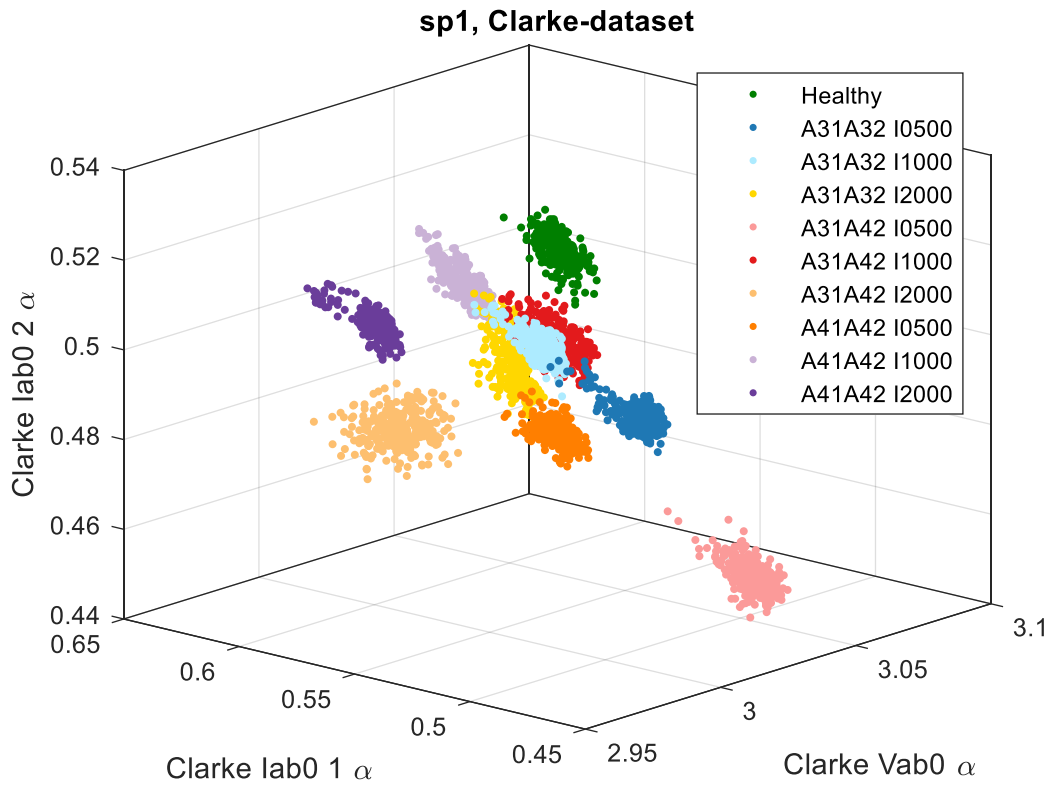


Figure 5-5 3D Scatter plot with the three most important variables, sp1 Clarke

Considering the machine set point at 100% of Nominal power (sp2), the maximum accuracy obtained is 95,14% for the optimal model where the number of features per class is 15 with $\lambda = 0$.

The processing time was 7.84 hours, and the total number of selected features is 117. Comparing the results from the sp1 and sp2 the performance for the sp1 modeling is superior, under the same generic model tuning.

One important factor to consider is that the accuracy of the model for Clarke sp1 dataset is higher than for the sp2 and only half of the number of selected features.

As an overall conclusion for the Clark-dataset, only considering the main three-phase current and voltages together with the current in the parallel branches, it is possible to unequivocally distinguish the healthy condition from all the other classes defining nine different fault conditions.

The various severities of *phase A* winding shortening are classified with a minimum accuracy of 95%. The Figure 5-6 displays the accuracy obtained for each iteration of lambda and number of features used to classify the various machine conditions.

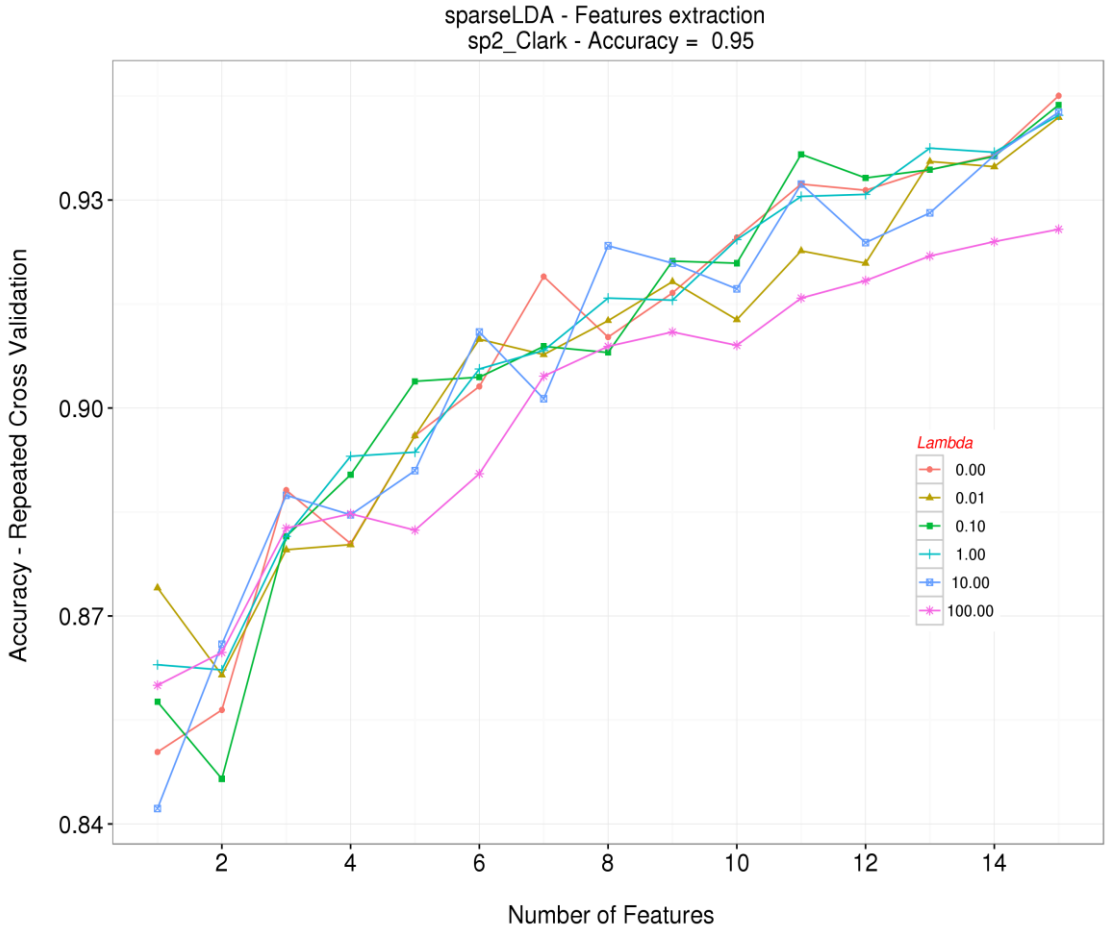


Figure 5-6 sp2 Clarke-dataset, sLDA features extraction as a function of lambda.

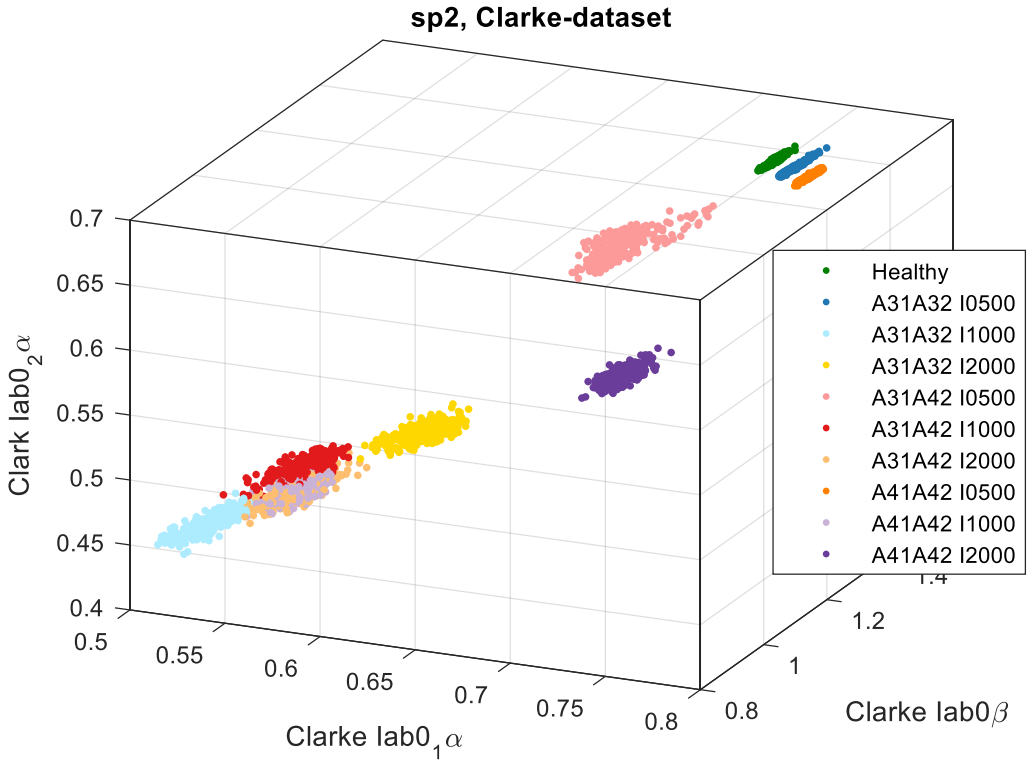


Figure 5-7 3D Scatter plot with the three most important variables, sp2 Clarke

Contrary to the initial intuition, severities of 0.5 Amps of deviated current (*I0500*) are easily identified for sp2 and for sp1 considering only the three top features for each set-point.

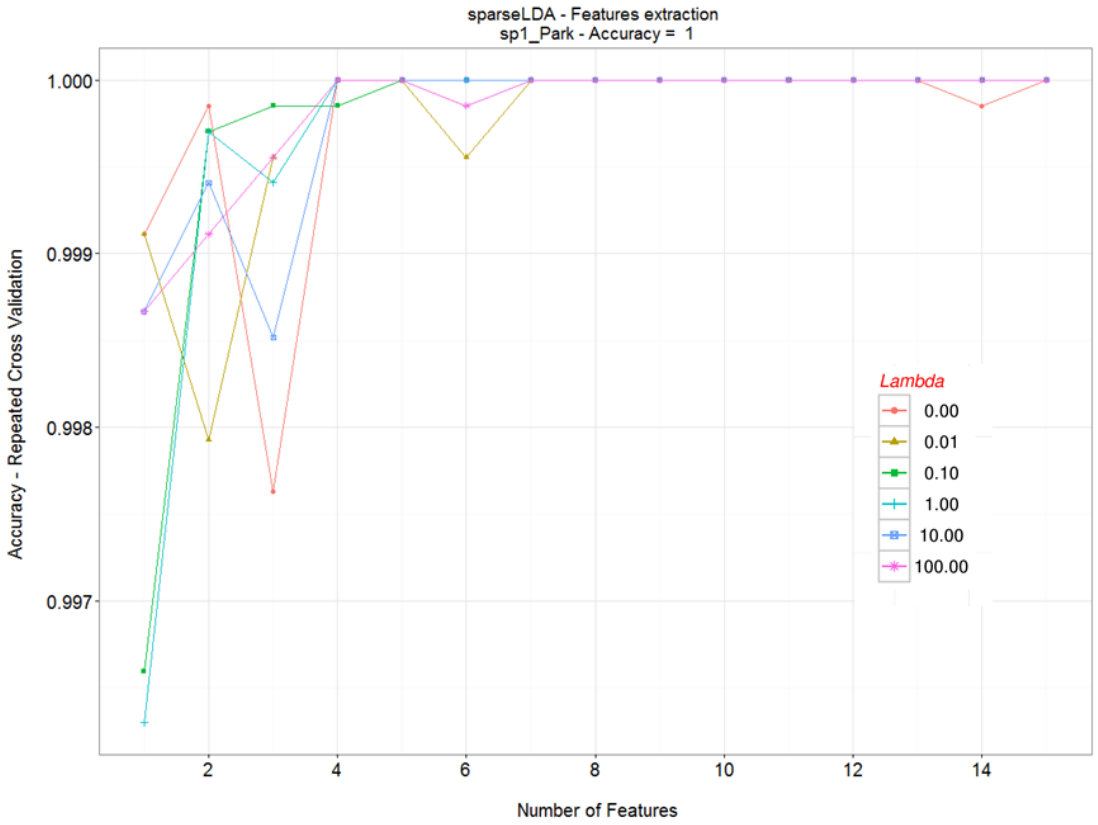


Figure 5-8 sp1 Park-dataset, sLDA features extraction as a function of lambda.

5.5.3 PARK-DATASETS SLDA FEATURE EXTRACTION

The Park-dataset consist of 817 features. After fitting the sLDA model to the sp1 Park was possible to select the optimal model with $\lambda = 100$ and 4 features identified as the most optimal for each class. The processing time was 5.55 hours in total.

The most important features selected from the Park dataset at sp1 are represented in Figure 5-9.

As in the Clark dataset for sp1, we have phase voltage-related feature in one of the dimensions of the 3D scatter plot, in this case, the d-axis Voltage, and in the other two dimensions, two currents; in this case the Park current in the direct axis for the parallel branch 1 ($Park\ I_d, branch\ 1$) and the for the parallel branch 2, ($Park\ I_d, branch\ 2$).

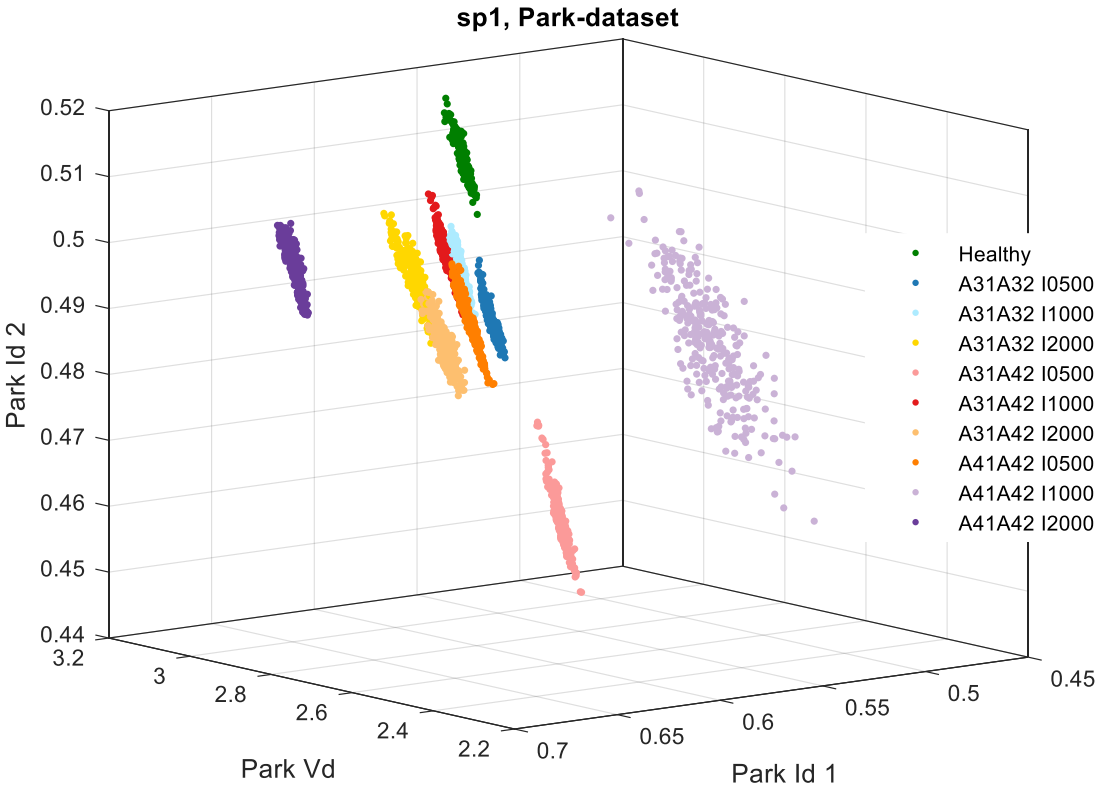


Figure 5-9 3D Scatter plot with the three most important variables, sp1 Park

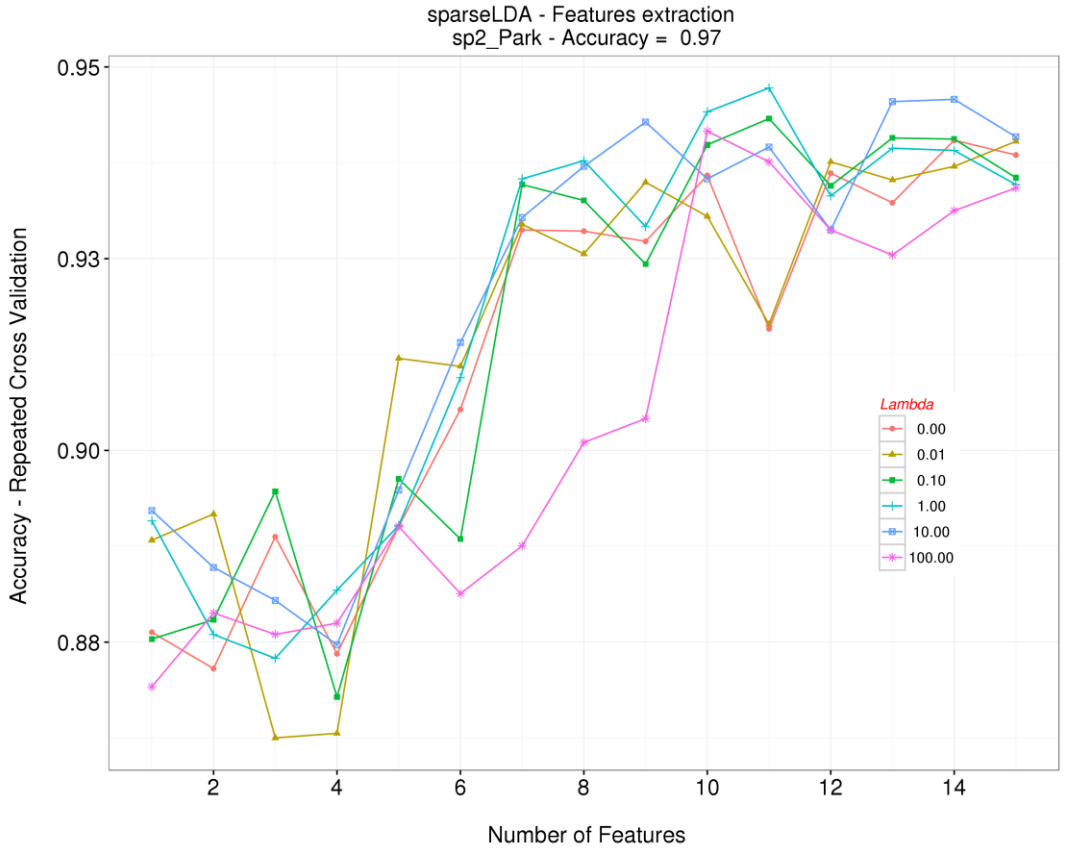


Figure 5-10 sp2 Park-dataset, sLDA features extraction as a function of lambda.

The sp2 Park dataset required less time, 3.75 hours, to find the optimal model having $\lambda = 1$ and 11 features. Even when the number of selected features is higher, the accuracy is also lower than for sp1.

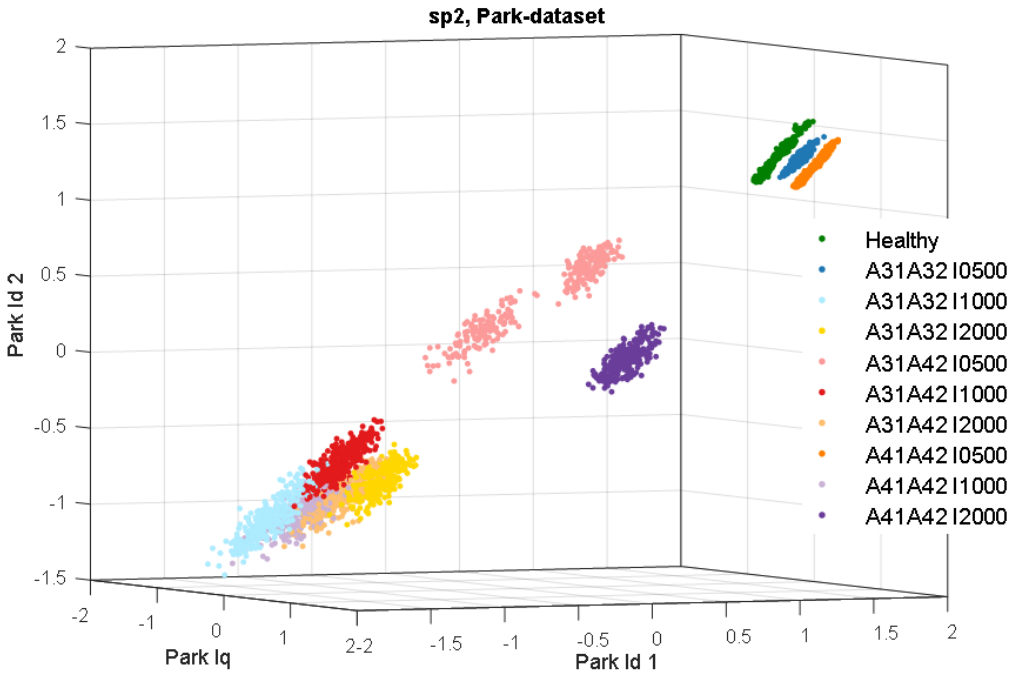


Figure 5-11 3D Scatter plot with the three most important variables, sp2 Park

From the 3D scatter plots, it is possible to see that both models are capable to positively differentiate the *Healthy* from all the conditions of the machine using the d-axis current and voltage vectors.

5.5.4 VIPOWER-DATASET, SLDA FEATURE EXTRACTION

VIpower-dataset contains different magnitudes related to the main harmonic power of electric machines. It is also the smallest dataset with 114 features. The sp1 was processed in 45.3 mins with $\lambda = 0.1$ and 9 features extracted for each machine condition.

The 3D scatter plot in Figure 5-13 contains the most significant features. The *ampRatio Vba* as the amplitud of the fundamental harmonic of the *phase A voltage* over the amplitud of the fundamental harmonic of the *phase B voltage*.

The other two features are the phase power calculated using the dot product, for each parallel branches:

$$P_{a1} = \frac{\sum(V_a \cdot I_{a1})}{N_{samples}} \text{ and } P_{a2} = \frac{\sum(V_a \cdot I_{a2})}{N_{samples}}$$

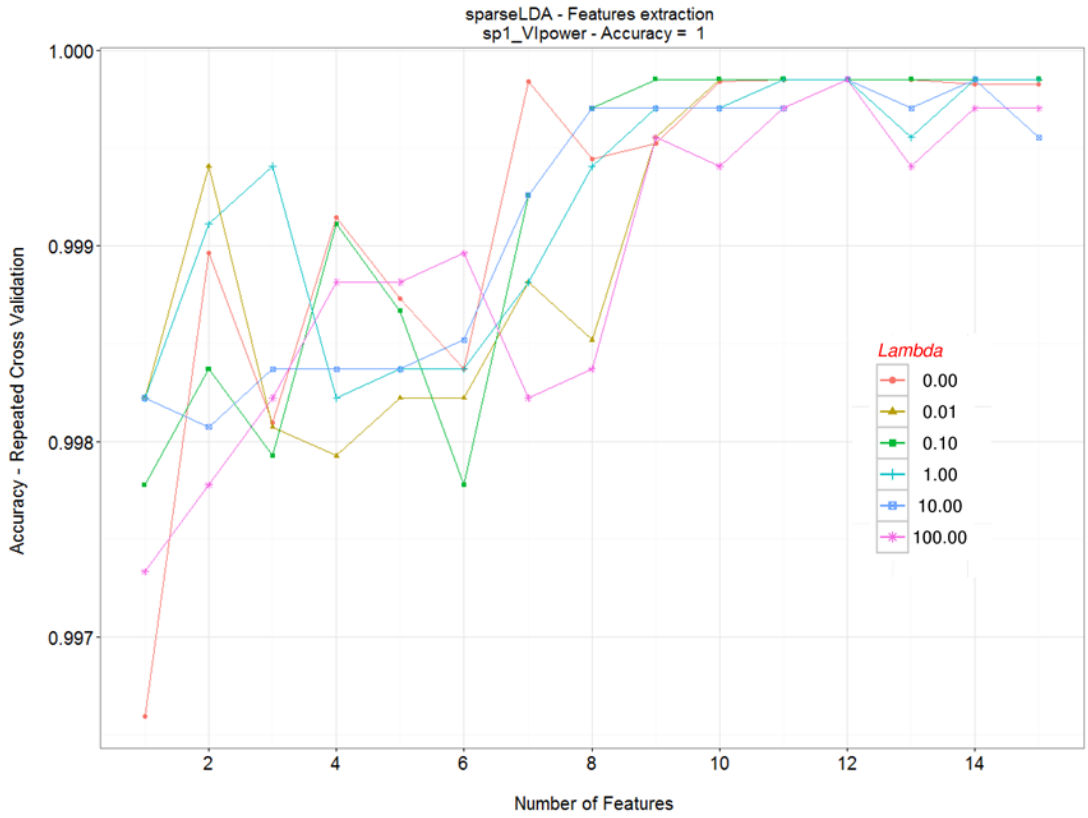


Figure 5-12 sp1 VIpower-dataset, sLDA features extraction as a function of lambda.

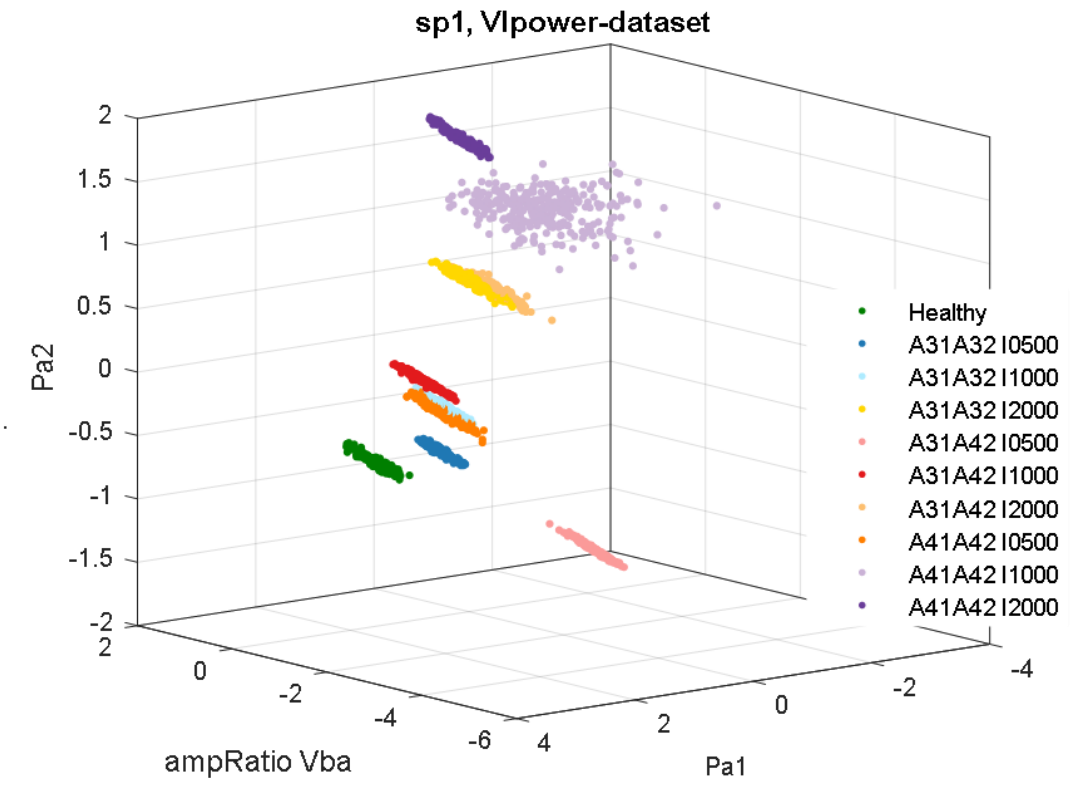


Figure 5-13 3D Scatter plot with the three most important variables, sp1 VIpower

All the three features selected by the sLDA algorithm are related to *phase A*, in which the fault was seeded.

A similar situation is for sp2, in which the most optimal model was obtained with $\lambda = 0$ and 15 selected features. To select the most important features the processing time was 40.52 minutes.

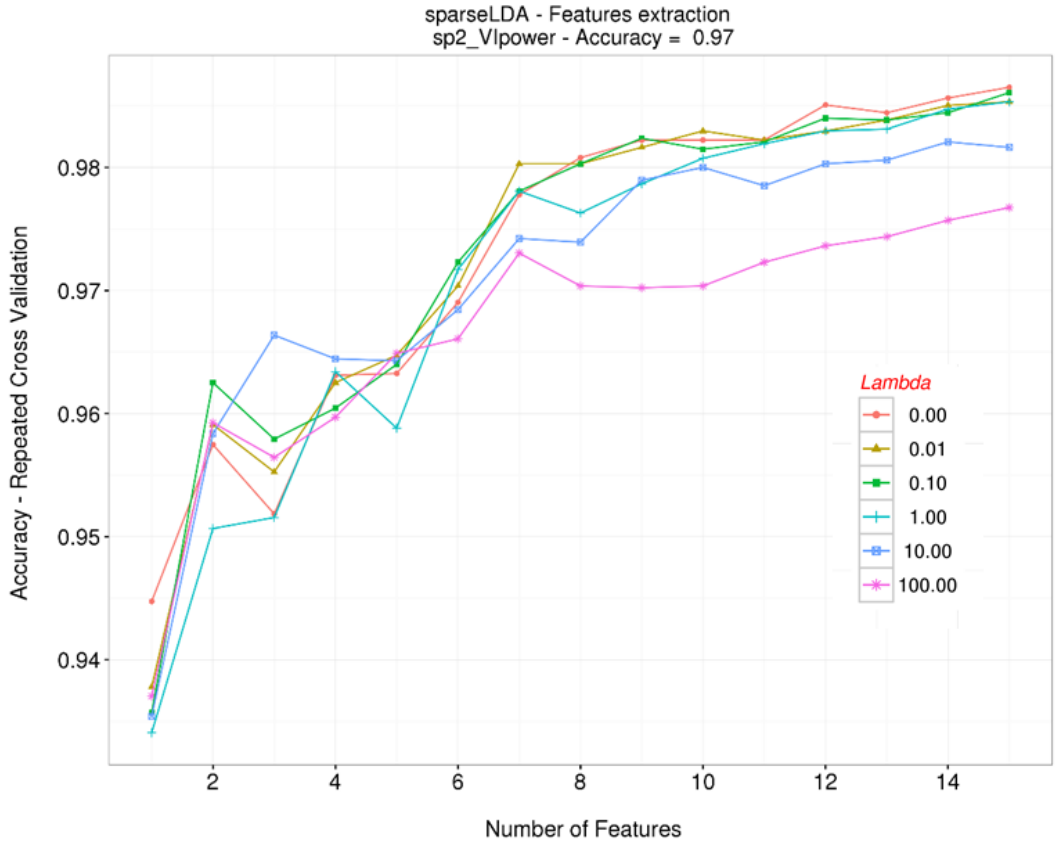


Figure 5-14 sp2 VIpower-dataset, sLDA features extraction as a function of lambda.

Comparing both scatter plots, it is possible to recognize that the models for sp1 and sp2 unequivocally distinguish between the healthy condition (in green) and the other nine conditions.

As in the previous datasets, using the three most important features extracted using sLDA it is also possible to distinguish between the three severities seeded between the three different coils combinations.

To differentiate between all the other six machine conditions, it is necessary to consider more features than only the three most important, and the visual representation of more than three dimensions is not possible, so it is necessary to build a methodology that recognizes the condition of the machine from all the source of information available.

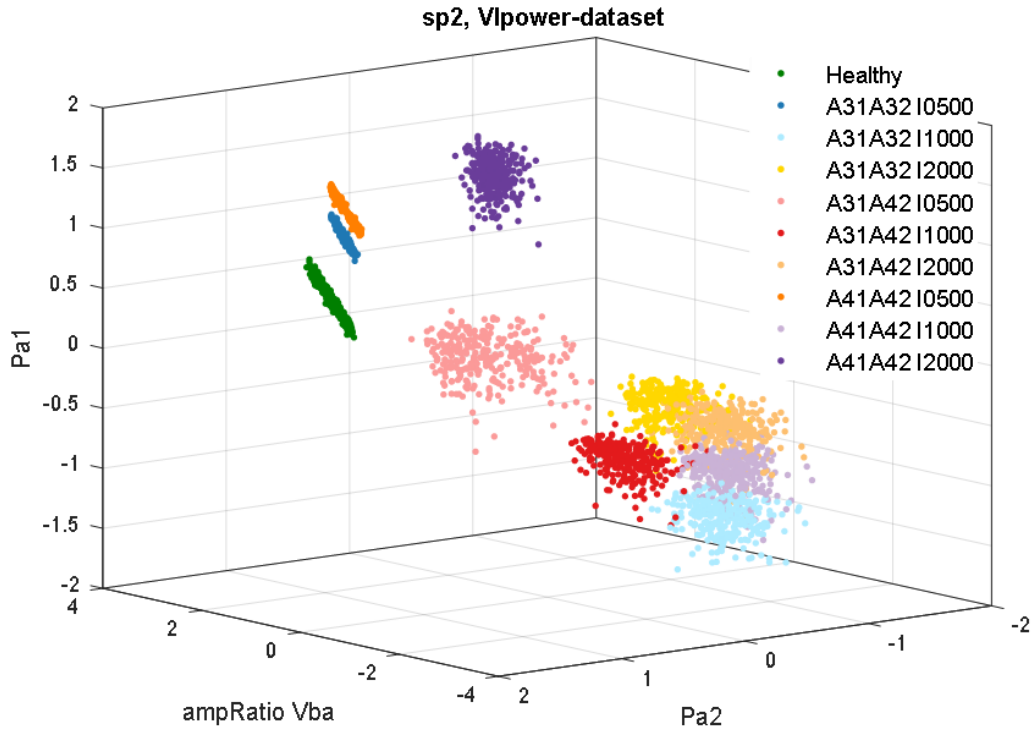


Figure 5-15 3D Scatter plot with the three most important variables, sp2 VIpower

5.5.5 FREQ-DATASET SLDA FEATURE EXTRACTION

Based on the harmonic content, from all the harmonics with base 50Hz, up to 2500 Hz for all the signals, 3700 features were extracted.

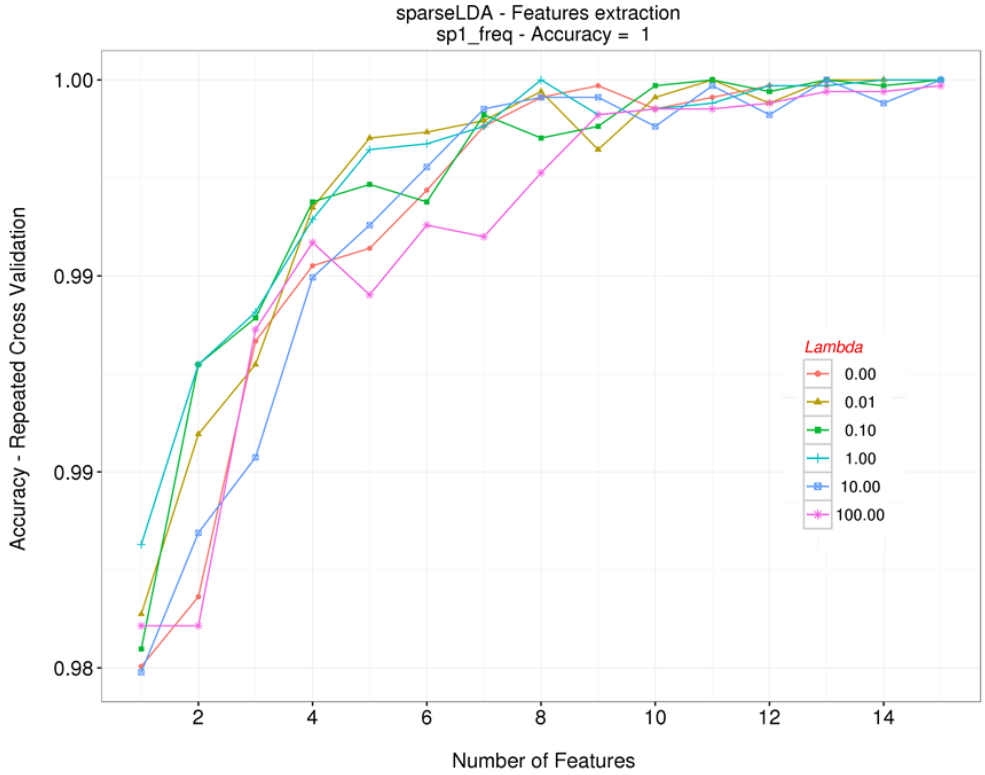


Figure 5-16 sp1 freq-dataset, sLDA features extraction as a function of lambda.

The time to process the sp1 frequency dataset was 1.4 days, with $\lambda = 1$ and 8 features selected per class, as shown in Figure 5-16.

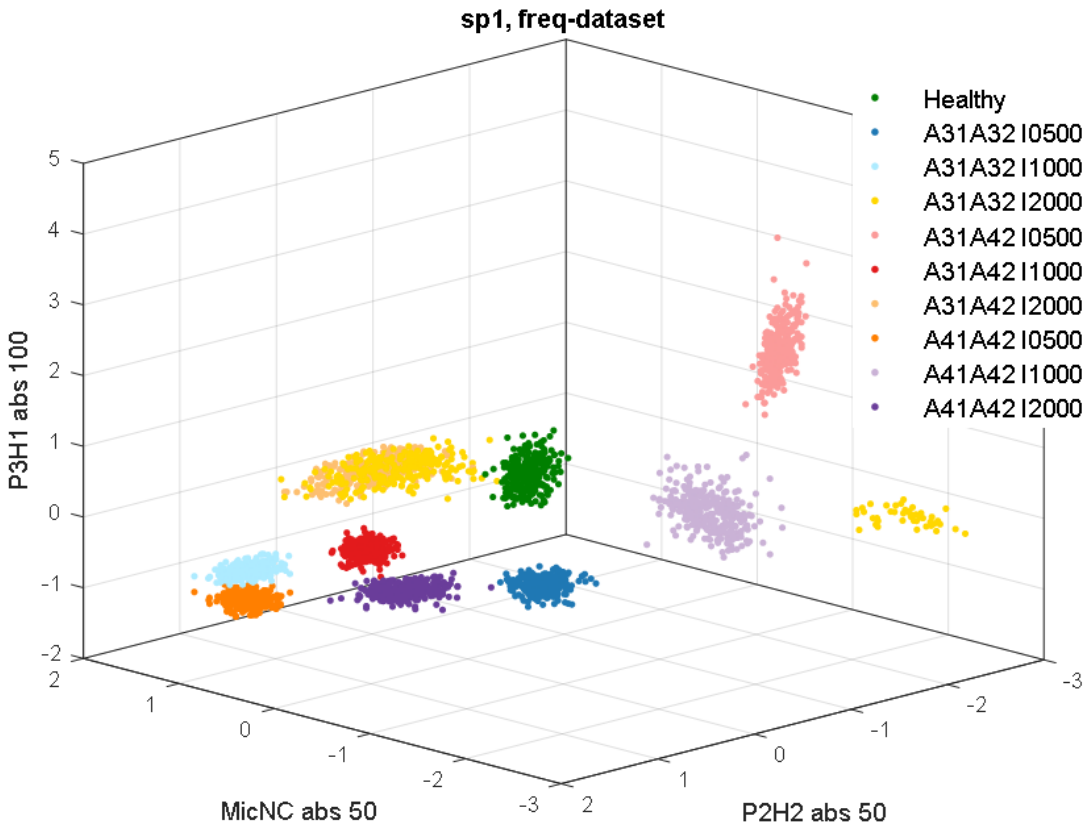


Figure 5-17 3D Scatter plot with the three most important variables, sp1 freq

So far, all the extracted most important features, are directly or indirectly related to the *phase A* in which the seeded fault was applied for different severities. This shows the ability of the model of recognizing the fault by learning from the data, but still, it is required to measure all the phases.

In contrast, using the amplitude at 50Hz for the noise canceling microphone *(MicNC abs 50)*, the amplitude at 50Hz of the second pole Hall sensor *(P2H2 abs 50)* and the amplitude at 100Hz for the Hall sensor installed in the adjacent pole *(P3H1 abs 100)*, in Figure 5-17, it is possible to appreciate that all the 10 conditions of the machine considered here, are unequivocally classified and clustered for the sp1.

For the sp2 freq-dataset, the optimal group of features was obtained for $\lambda = 1$ and up to 15 features extracted per class. The boundaries between classes are fuzzier, Figure 5-19. The time to process the initial 3700 features was 1.03 days, which is almost 9 hours lower than for sp1, but the accuracy obtained is 6% lower as well.

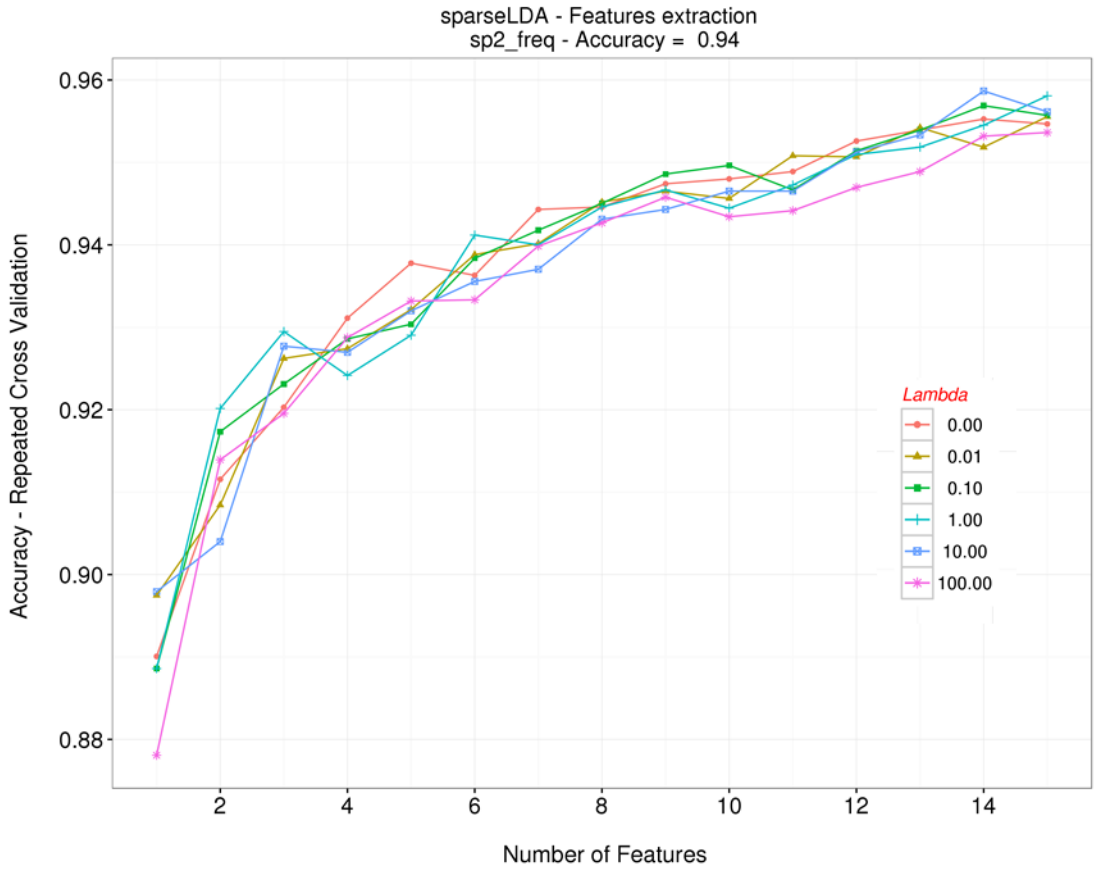


Figure 5-18 sp2 freq-dataset, sLDA features extraction as a function of lambda.

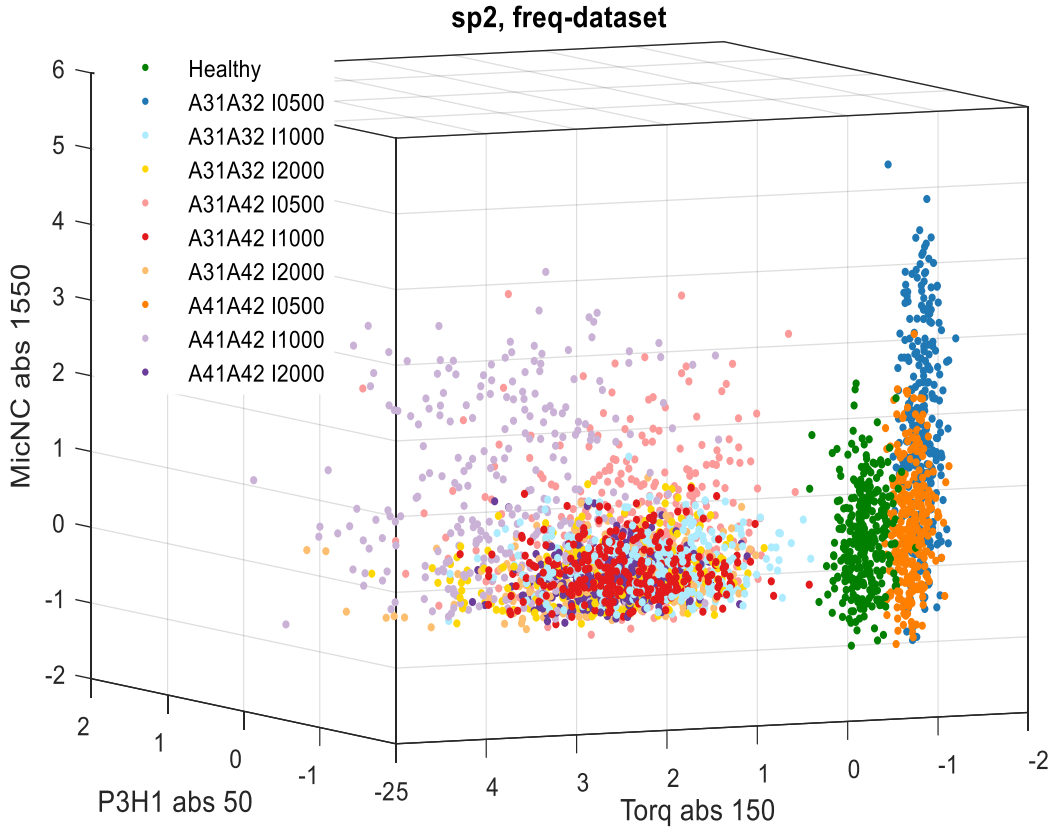


Figure 5-19 3D Scatter plot with the three most important variables, sp2 freq

The features represented in Figure 5-19 correspond to the amplitude at 1550 HZ of the noise canceling microphone *(MicNC abs 1550)*, the amplitude at 50Hz of the Hall sensor installed on the pole three *(P3H1 abs 50)* and the amplitude at 150Hz of the torque *(Torq abs 150)*.

5.5.6 STATS-DATASET SLDA FEATURE EXTRACTION

The number of optimal features extracted per class is 4, with $\lambda = 1$. It was necessary 6.97 hours to compute the sp1 dataset generated from the most common statistical applicable functions; 777 features in total. Using features that are not calculated directly from one of the voltages or phase currents, or as a combination of them was possible to choose three features that clearly differentiate all the classes as shown in Figure 5-21.

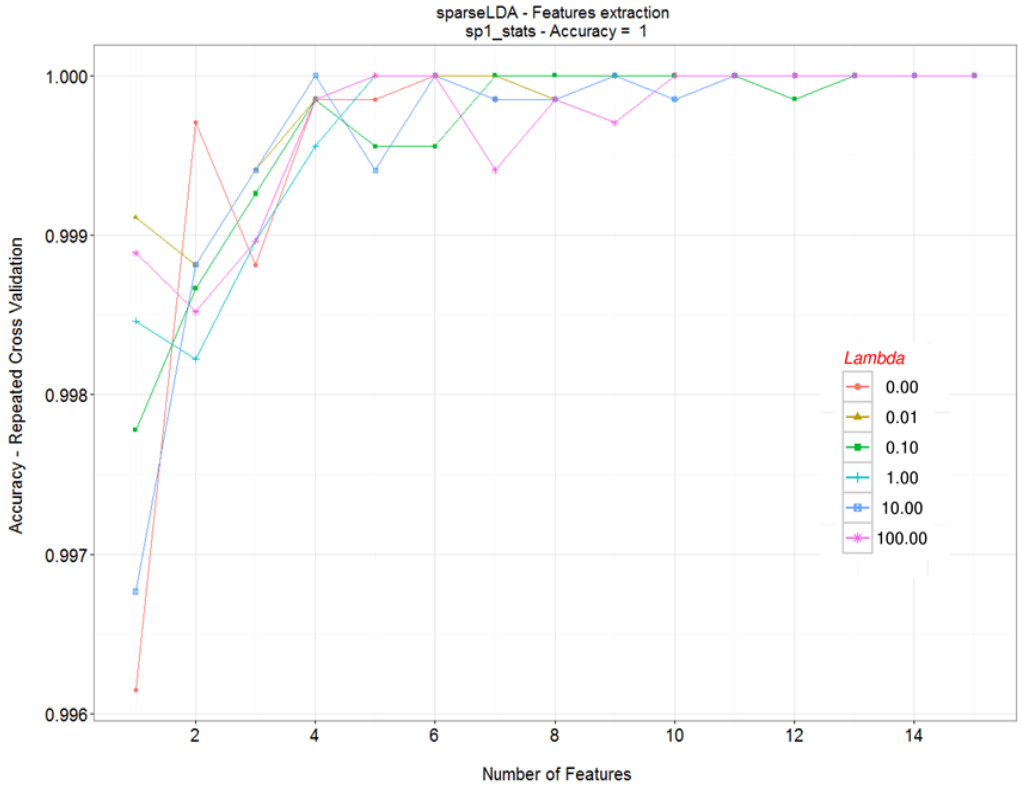


Figure 5-20 sp1 stats-dataset, sLDA features extraction as a function of lambda.

Considering the standard deviation from the rotational speed *(w1 sig)*, and the noise canceling microphone installed parallelly to the shaft axis *(MicNC sig)*, together with the spurious free dynamic range from the voltage between shaft and ground *(ShaftV sfdr)*, calculated as the difference in dB of the power between the fundamental frequency and the next largest frequency (spur), was possible to clearly assess the condition of the machine, under 50% of nominal load for all the classes.

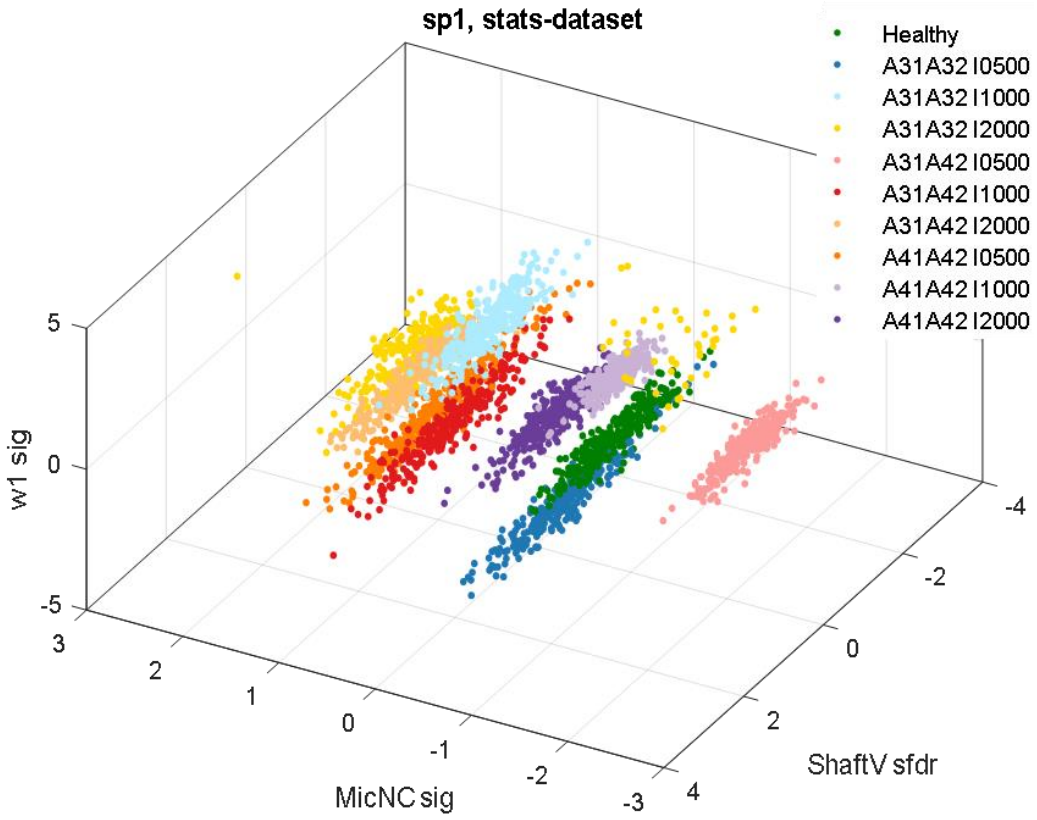


Figure 5-21 3D Scatter plot with the three most important variables, sp1 stats

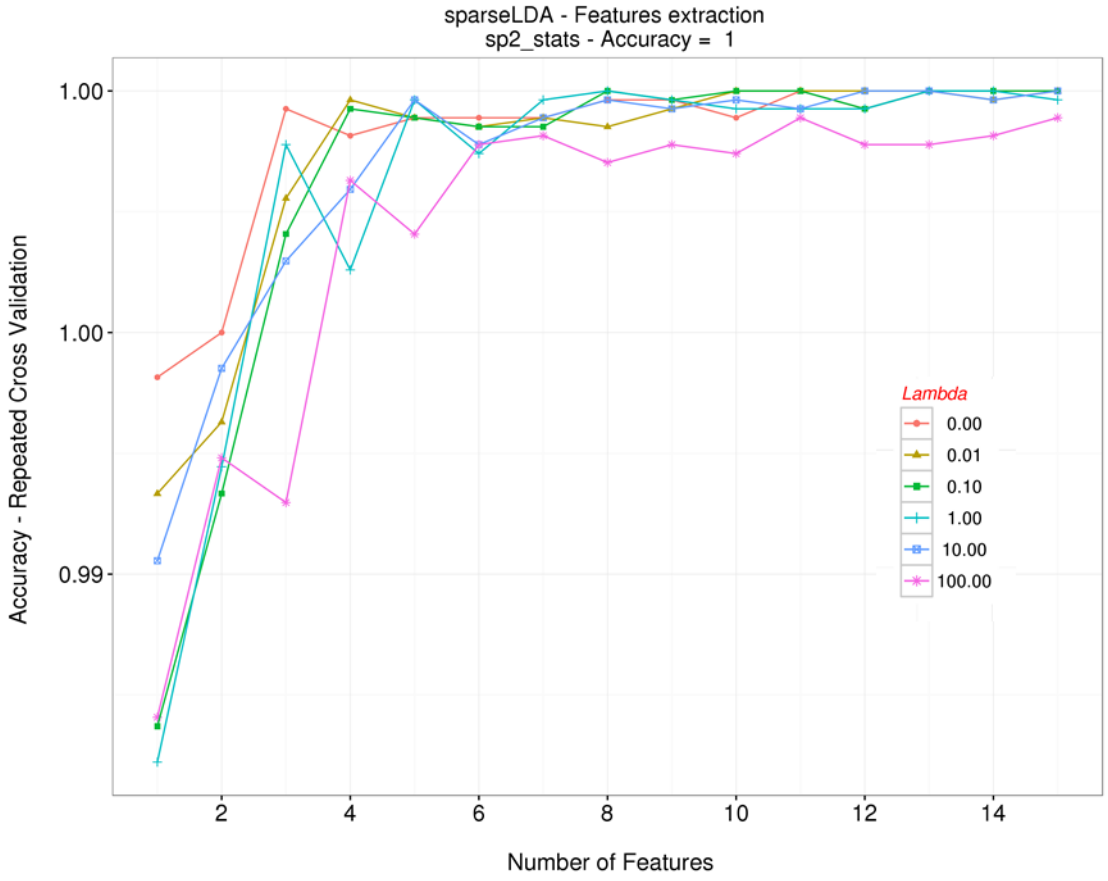


Figure 5-22 sp2 stats-dataset, sLDA features extraction as a function of lambda.

To process the sp2 stats-dataset, 5.42 hours were needed. The optimal group of features was obtained for $\lambda = 1$ and 8 features.

The best combination to create the 3D scatter plot, Figure 5-23, was using the standard deviation of the voltage drop between the front and end ring for the bar four in the pole three *(P3b4 sig)*, and the Rogowski coil installed on the pole two, bar one *(P2RGb1 sig)*. The third feature is the median absolute deviation of the rotational speed *(w1 absMedian)*.

However, finding a combination of three features to cluster all the classes in a three-dimensional space was not possible. Although it was possible to choose features that unequivocally differentiate between faulty and healthy condition, it will be necessary to consider more features to classify all the 10 classes correctly.

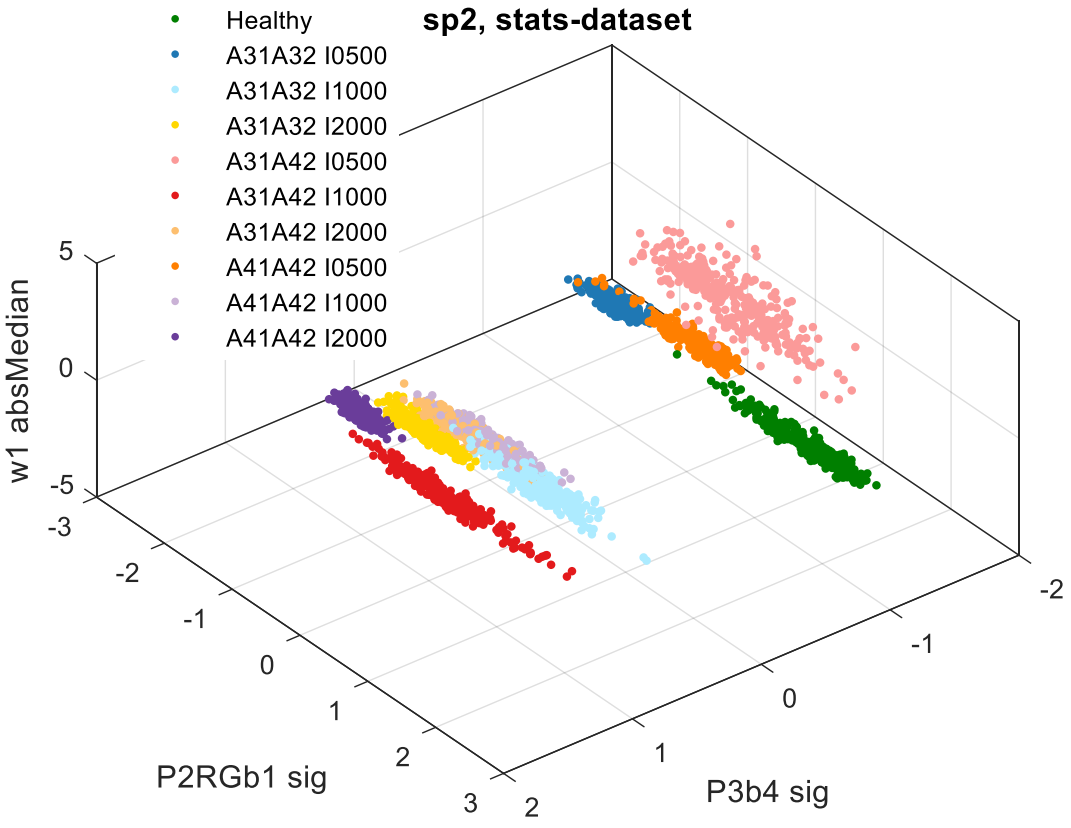


Figure 5-23 3D Scatter plot with the three most important variables, sp2 stats

So far it has only been considered each of the datasets separately, in the following subsection, the results of all of them will be compared and finally a new dataset, as the combination of all the essential features, will be used to train different classification algorithms.

5.6 COMBINED-DATASET sLDA FEATURE EXTRACTION

The total features considered in the above subsections was 6632 features. Using sLDA, was possible to narrow to the 688 most optimal features for the classification task of assessing the condition of the synchronous machine, for three type of faults at three different severities levels.

Comparing the differences between sp1 and sp2:

- The accuracy to unequivocally classify all the classes is greater in sp1.
- The time to process is 1.34 times higher for sp1 than sp2.
- λ is bigger for sp1 compared to sp2.
- The number of optimal features needed per class and the total of extracted features is twice bigger in sp2 than sp1.

The sLDA was set to select up to 15 features per class with regularization to allow the feature selection between variables that might be correlated. A summary table can be found below:

Table 5-5 Features extraction summary table

Total features	Dataset	Extracted features	Processing Time	λ	Features per class	sLDA Acc
1224	sp1 Clarke	59	9.89 h	1	7	1
	sp2 Clarke	117	7.84 h	0	15	0.92
817	sp1 Park	32	5.55 h	100	4	1
	sp2 Park	79	3.75 h	1	11	0.97
114	sp1 Vipower	47	45.3 mins	0.1	9	1
	sp2 Vipower	62	40.52 mins	0	15	0.97
3700	sp1 freq	71	1.4 days	1	8	1
	sp2 freq	121	1.03 days	1	14	0.94
777	sp1 stats	36	6.97 h	10	4	1
	sp2 stats	64	5.42 h	1	8	0.99
6632	**Totals**	**688**	**4.13 days**	**NA**	**95**	**9.79**

From the 100 features extracted (36 for sp1 and 64 for sp2) from the stats-dataset the frequency of occurrence is as follow:

Table 5-6 Stats-functions, ordered appearance list based on the occurrence.

Function	Occ.	Function	Occ.	Function	Occ.
range	14	sfdr	8	absSigMean	2
absMean	9	absSigMedian	7	mu	2
absMedian	9	bandpower	6	sinad	2
rssq	9	snr	6	thd	1
sig	9	kurtosis	4		
iqr	8	skewness	4		

Looking for the most relevant harmonics, we can see:

Table 5-7 Harmonics, ordered appearance list based on the occurrence.

Hz	Occ.	Hz	Occ.	Hz	Occ.	Hz	Occ.
100	27	450	12	1150	5	1650	2
50	26	600	12	1250	5	1900	2
150	24	650	11	1550	5	2300	2
500	23	2100	11	1950	5	2500	2
350	22	750	10	700	4	1400	1
250	20	1100	10	900	4	1500	1
2050	18	800	8	2200	4	1600	1
200	17	950	8	2350	4	1700	1
1000	17	1050	8	1200	3	1800	1
550	13	2150	8	1300	3	2400	1
300	12	1350	6	2250	3	2450	1
400	12	850	5	1450	2		

The most frequent variables

Table 5-8 Variables, ordered appearance list.

Var	Freq	Var	Freq
I	219	Mic	21
V	82	Torq	16
RG	47	w1	7
Hall	40	ShaftV	5

5.7 Data-Driven. Condition based maintenance

So far, all the followed steps explained in the previous Chapters were taken to define and set-up the data acquisition system, collect all possible datasets and prepare and define the minimum necessary features to train an algorithm capable of monitoring and identifying the condition of the machine.

Already introduced in Chapter 1, Figure 1-2 is replicated below to highlight the last step to implement a condition-based maintenance system capable of real-time diagnosis. The challenge now is to implement in one single function all the steps and with a processing time sufficient to identify the fault before it is catastrophic, or ideally as part of a real-time monitoring system.

Figure 5-24 Data-Driven, condition based maintenance.

Once the most important features are identified and the importance score extracted, we can train different classification algorithms, and based on various metrics, choose the most suitable and accurate. It is not the intention of this thesis to describe in detail each classification method.

The theory supporting the mathematical development and implementation is explained in courses of statistics and machine learning at the different level of engineering studies. Hence, it is only demonstrated the validity of such methods for the classification task of assessing synchronous machines, by means of various metrics and their capability of unequivocally differentiate between fault conditions or classes.

The combined dataset consists of features of different kinds; 688 in total. Trying to imitate real-world conditions, both set points are merged, so 20 classes in total are used to train and validate the models. As shown in Table 4-1. , other categorical variables within the classes are defined, these will be used to quantify the ability of the model to assess the different machine conditions as sub-groups of the *class* variable used during the training process.

5.8 CLASSIFICATION ALGORITHM NOTES

There are hundreds of machine learning algorithms available. One could try to explore all of them, and then publish the aggregated accuracy, but this approach is impractical under the defined conditions, thus the need for a basic pre-selection of possible candidates.

The ten selected models are explained. These classification algorithms have been chosen based on two criteria; first, identifying those algorithms that have been successfully used for fault diagnosis and the other criteria looking at strong candidates that best suit for the classification problem under consideration.

The first algorithm used is **Linear Discriminant Analysis**; it is a supervised linear transformation which ultimate goal is to retain the discriminant information of each class. LDA is a way to determine decision boundaries between classes directly. Therefore, LDA shows advantages for fault diagnosis over similar algorithms such as principal component analysis (PCA) **[6]**.

Different faults in turbogenerators were classified in **[4]** using LDA. The authors in **[28]** have developed a tool based on discriminative energy functions using Fisher discriminant analysis to identify discriminative frequency-domain regions. In **[82]** the limitations of LDA are explained, and a new variant Fisher discriminant analysis with orthogonal discriminant components is proposed. Another discriminant algorithm used is **Sparse Discriminant Analysis**, already introduced in subsection 5.5.1, and used as a feature extraction method.

Classification and regression trees (CART) is another algorithm under consideration. The algorithm used here is based on the sum of squared errors loss function. These type of algorithms split all the attributes on values that minimize the loss function. One advantage of CART, is that do not make any assumption on the classifier structure and spatial distribution.

No so many publications in the field of electrical machine fault diagnosis were found in which CART model was used to classify faults, in most of the cases, it was used to extract relevant features as in **[91]**. The authors in **[59]** have used a hybrid fuzzy min-max neural network with CART to perform online motor fault detection and diagnosis.

CART is just one type of decision trees. Other decision trees base model used for our purposes is **Random Forest**. In which many classifications trees are grown, so the model accuracy will be the aggregation of each tree classification during the training. The advantage of random forest is that does not overfit, and runs efficiently on large datasets with greater accuracy among current algorithms. Another decision tree with a greater efficiency is **C5.0** algorithm, in which the decision trees are created based on the combination that provides the maximum information gain. An application of this kind of algorithm compared with other classifiers based on decision trees has been published in **[52]**, in which the vibration of rotatory machinery is used to classify the status and detect faults.

As shown in Table 1-1, Support Vector Machine (**SVM**) is, after Neural Networks, the second most popular classification method used for fault diagnosis of electric machines. In **[7]** the authors have used a 2-dimensional multiclass SVM to predict the fault modes with an error rate of 1,48%.

In **[90]** a fault diagnosis approach using SVM is presented. This method has dealt with high dimensional data as well, and the solution given was through dimensionality reduction algorithms. **Support Vector Machine** using the **polynomial kernel** of d-dimension is selected since this type of kernel is most suitable for large datasets **[2]**.

The next model is **Naïve Bayes (NB)**. This classification model is one of the most efficient and effective inductive learning algorithms, and even when the modeling assumptions is that features are independent of each other, it has been proven that when some features seem correlated, NB classifier performs well **[95]**, and the processing time is relatively small compared with other classification models.

The final models considered are based on ensemble learning techniques. **Bagged CART** (**Treebag**) and **Bagged Flex discriminant analysis** (**bagFDA**), are methods that

use variations of samples to train base classifiers, in this case, CART and FDA, with the aim to decrease variance, not bias.

Stochastic Gradient Boosting (**gbm**) is another ensemble learning technique based on sequential ensemble, in this case, tree based. In each boosting iteration, every new subset contains the elements that were likely to be misclassified by previous trees.

In total, 10 models are being compared:

Table 5-9 Models' metrics for the categorical variable "*Class*"; 20 in total.

Classification Model	Avg Precision	Avg Recall	Avg F1	ACC	Training time	Predict time
LDA	17%	17%	17%	17%	40.5 secs	0.3 secs
CART	NA	50%	NA	50%	1.43 min	0.48 secs
treebag	12%	12%	12%	12%	6.86 min	0.95 secs
rf	12%	12%	12%	12%	27.28 min	0.44 secs
svmPoly	32%	33%	32%	33%	1.16 hours	22.11 secs
gbm	27%	27%	27%	27%	27.51 min	0.41 secs
bagFDA	47%	47%	45%	47%	1.03 hours	3.03 secs
nb	26%	28%	26%	28%	45.75 min	29.77 secs
sparseLDA	39%	39%	39%	39%	10.55 hours	0.31 secs
C5.0	NA	50%	NA	50%	26.74 min	3.9 secs

5.8.1 MODELS COMPARISON

As stated in section 5.3, precision, recall, and F1 are metrics calculated per class. For easiness of representation and comparison, these metrics are given per model and averaged per all categories.

The training dataset used is the combined dataset; 688 selected optimal features and 5980 observation. This gives 20 classes to be classified, the same ten machine conditions but for each set-point. As seen in

Table 5-9, the maximum accuracy is 50%. These values are considered to be low. Hence, it is necessary to review the causes.

To evaluate the ability for classifying and differentiating between the condition of the machine, healthy (2 categories, for sp1 and sp2) vs. faulty (18 different categories), it is possible to aggregate the original categorical variables into two categories defining the healthy or faulty condition.

The results are tabulated in Table 5-10, all models perform the same and are capable of differentiating between a healthy or faulty condition in 100% of all cases, for both set-points. So it is necessary to examine the other sub-classes and identify the source of the low accuracy for the combined categorical variable.

Table 5-10 Models' metrics for Categorical variable *Condition – Healthy vs. Faulty*.

Classification Model	Avg Precision	Avg Recall	Avg F1	ACC
LDA	1	1	1	1
CART	1	1	1	1
treebag	1	1	1	1
rf	1	1	1	1
svmPoly	1	1	1	1
gbm	1	1	1	1
bagFDA	1	1	1	1
nb	1	1	1	1
sparseLDA	1	1	1	1
C5.0	1	1	1	1

Considering the categorical variable *Severity*, there are four possible classes; healthy, and three different values of deviated current, 500 mAmps, 1 Amp, and 2 Amps. All the models equally perform with an accuracy of 100%. The same situation happens when considering the categorical variable *Fault type* ; healthy, and three different kinds of coils shortening.

Looking at the ability of the models to differentiate between set-points, the accuracy obtained in this case, explains the underperformance of the models when trying to classify the 20 classes in the categorical variable *Class*.

From the exposed above, it seems that the models are not capable of differentiating between different machine workloads. Therefore, in order to detect all the 20 different conditions, it is necessary to define a way to trigger the adequate model attending to the level of load.

Table 5-11 Models' metrics for the categorical variable Set-point (sp).

Classification Model	Avg Precision	Avg Recall	Avg F1	ACC
LDA	17%	17%	17%	17%
CART	NA	50%	NA	50%
treebag	12%	12%	12%	12%
rf	12%	12%	12%	12%
svmPoly	33%	33%	33%	33%
gbm	27%	27%	27%	27%
bagFDA	47%	47%	47%	47%
nb	28%	28%	28%	28%
sparseLDA	40%	40%	40%	40%
C5.0	NA	50%	NA	50%

5.9 FAULT CLASSIFICATION TASK EXAMPLES

Until now we have reviewed the sparseLDA model as feature extraction methodology, and the selected most significant features were used to compare and contrast the capability of different machine learning algorithms of classifying ten classes per machine set-point.

Before continuing to the Chapter 6, in this sub-section, used as an example, two different fault classification tasks are considered. The first step is to select the signals manually, extract all the features, and finally, train and validate the model.

5.9.1 ROTOR FAULT CLASSIFICATION TASK

As explained in subsection 2.2.2, with this laboratory set-up, it is possible to evaluate two types of rotors winding shortening. Already published in **[17], [19]**, using frequency analysis was feasible to identify the rotor fault considering the power spectral density (PSD) of one phase currents. A fault condition is detected when a peak in the frequency domain is detected outside the 95% confidence interval calculated based on P-Welch method, for the healthy condition.

Figure 5-25 contains the representation of PSD for a rotor and stator fault data sample. Here, the fault detection is done by comparing the current value with the expected (averaged) healthy machine's thresholds values. While this method is simple, it is especially sensitive to time harmonics causing false positive as part of online condition monitoring systems.

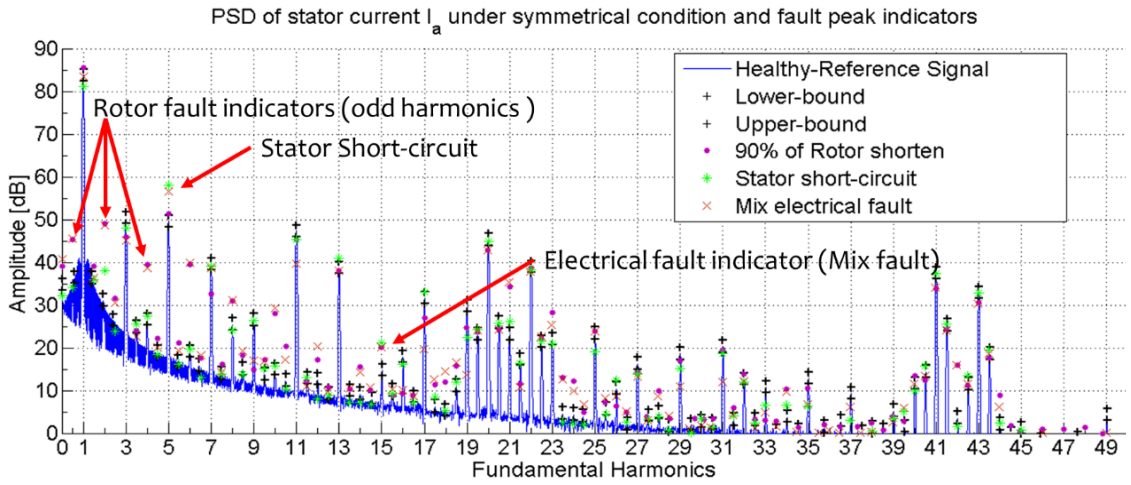


Figure 5-25 PSD of stator current I_a, displaying the peak values under rotor and stator fault condition. The 95% CI are represented as the boundaries for healthy machine's condition.

The problem under consideration is a two-class classification task. The accuracy obtained is 100% for most of the models. As an example, the confusion matrix for the sparseLDA model is:

Table 5-12 The confusion matrix for sparseLDA model.

	Observed	
Predicted	ok	fault
ok	6	0
fault	0	20

Considering the current of the phase R (*Ia*), and calculating the frequency features it is possible to analyze the performance of all the classification algorithms to detect faulty over healthy machine condition represented in Table 5-13. Identifying 90% of rotor winding shortening fault is possible by only looking at the frequency content of one of the phase current. In most of the algorithms, the new data points are classified with an accuracy of 100%.

Table 5-13 Rotor fault, classification algorithms metrics.

Classification Model	Avg Precision	Avg Recall	Avg F1	ACC
LDA	93%	75%	79%	88%
CART	100%	100%	100%	100%
treebag	100%	100%	100%	100%
rf	100%	100%	100%	100%
svmPoly	100%	100%	100%	100%
gbm	100%	100%	100%	100%
bagFDA	100%	100%	100%	100%
nb	NaN	50%	NaN	75%
sparseLDA	100%	100%	100%	100%
C5.0	100%	100%	100%	100%

The two classes are “ok” for machine condition healthy and “fault” for machine condition faulty. The number of data point used for training were only 26. This contrasts significantly with the number of observations needed to classify the stator fault.

By looking at the features selected after training the model, we can observe that the slot harmonic’s side-lobes and the 9^{th} and 11^{th} harmonics are clear fault indicators, in the presence of rotor shorten fault.

5.10 STATOR FAULTY/HEALTHY CLASSIFICATION TASK, FREQ DATASET

Using similar process as before it is also possible to analyze the performance of the algorithms under stator winding shortening fault. One important factor to highlight is that to detect the fault in the stator on every phase winding, it is necessary to consider the phase current for all the three phases. Since the three faults and severities were seeded in the same phase, only the frequency features of the current phase *Ia* are considered in this example.

This initial assumption introduces "prior knowledge" or constraints to the number of features under consideration. The variability of the data and the processing time is substantially reduced; a maximum accuracy of 92 % is obtained, Table 5-14 by random forest and support vector machine with polynomial kernel for the classification task of identifying faulty or healthy machine condition considering phase current *Ia* only.

Table 5-14 Stator Fault/Healthy condition using Ia only, classification algorithms metrics.

Classification Model	Avg Precision	Avg Recall	Avg F1	ACC
LDA	81%	78%	79%	88%
CART	68%	59%	61%	80%
treebag	88%	83%	85%	91%
rf	89%	85%	87%	92%
svmPoly	90%	83%	86%	92%
gbm	87%	79%	82%	90%
bagFDA	85%	85%	85%	90%
nb	75%	79%	77%	84%
sparseLDA	77%	73%	75%	85%
C5.0	88%	82%	84%	91%

When using the same assumption and considering ten different classes, the maximum accuracy is only 64% by bagged CART (treebag), Table 5-15. To differentiate between the location and severity of the fault, more information is needed, so the learning takes place, and higher accuracy can be achieved.

Table 5-15 Stator 10 classes fault detection using Ia only, classification algorithms metrics.

Classification Model	Avg Precision	Avg Recall	Avg F1	ACC
LDA	56%	54%	55%	54%
CART	NaN	40%	NaN	40%
treebag	66%	64%	65%	64%
rf	66%	62%	63%	62%
svmPoly	66%	62%	63%	62%
gbm	60%	57%	58%	57%
bagFDA	60%	60%	60%	60%
nb	56%	42%	42%	42%
sparseLDA	50%	49%	49%	49%
C5.0	62%	61%	61%	61%

In order to increase the accuracy and to avoid/decrease the need for prior knowledge or constraints, a multivariate methodology is proposed and described in Chapter 6.

6 PROPOSED METHODOLOGY AND FURTHER STEPS.

This thesis proposes a reference method for condition-based maintenance. Even when this methodology aims to be generic and not tied to a specific application, its validity is demonstrated by verifying the obtained accuracy while assessing a synchronous machine under different fault conditions. The most obvious use of this methodology is for manufacturers of electric motors with the capacity to perform similar tests. Classification algorithms can be trained to be used together with the motor in condition monitoring applications; perhaps as an extra associated product.

From the initial 38 recorded signals, 6632 features were calculated. Then, 245 optimal features were identified for the set-point 1 (50% of nominal load), and 443 optimal features for the set-point 2 (100% of nominal load). The initial results showed that the models underperform when considering the machine set-point as a categorical variable. In other words, a total of nine fault condition and the healthy condition were examined for two different set-points; 20 classes in total.

When evaluating all the machine conditions independently from the set-point, the accuracy of classifying the type and severity of the fault increases. This finding simplifies and reduces the time to train the models and makes the application of this methodology for online monitoring more robust for industrial environments.

Now, as a conclusion of this research, the final methodology is presented.

6.1 TRAINING DATA

In total, twelve minutes for each machine condition were recorded into six datasets of two minutes, collected on different days (intended to capture any impact of the various external disturbances). Each machine condition is defined by 27.36 millions of data points in its rawest form. After extracting all the features and defining the windows size of 8360 samples comprising one observation, there are 35880 observations available.

Each dataset was pre-processed and divide into 60% for training data (21600 observations), 20% for testing (7200 obs.) and the remaining 20% for validation (7080 obs.). The split of these subsets was performed maintaining all the classes balanced. Then, the training data for each of the five datasets is used as the input for the features extraction sLDA algorithm.

Table 6-1 Categorical variables, new categorization independent from the workload.

Categorical variable	Description	Values
Fault type	The simulated stator winding short-circuit, by deviating current from the winding through an external regulable resistor.	0 – Healthy 1 – A31A32 2 – A31A42 3 – A41A42
Severity	The amount of current deviated in each fault type.	0 – Healthy 1 – 500 mA 2- 1000 mA 3 – 2000 mA
Class	Aggregation of the other categorical variables to be used during supervised training.	The combination of the fault type and severity.

The maximum number of features to extract, per class, is set to 25 (features can be repeated between classes). As seen in Chapter 5, subsection 5.8.1, when trying to

classify each fault condition at two different load levels (20 different classes), the accuracy obtained was below the expected, despite the longer processing time.

After reviewing the performance of the models for each sub-class, it was possible to implement a major simplification by merging the categorical variable *Class* from 20 different categories into 10 categories independent from the working load, registered as sp1 or sp2, shown in Table 6-1.

Under this new categorization, the classification task is to assess the condition of the machine, independently of the workload and for different fault severities of shortening coils, in various winding-branches.

The ten classes and their meaning are:

Table 6-2 Machine condition codes used to identify each observation.

Code	Machine condition
C00	Healthy
C11	Fault between coil A31A32, severity of deviated current 500 mA
C12	Fault between coil A31A32, severity of deviated current 1000 mA
C13	Fault between coilA31A32, severity of deviated current 2000 mA
C21	Fault between coils A31A42, severity of deviated current 500 mA
C22	Fault between coils A31A42, severity of deviated current 1000 mA
C23	Fault between coils A31A42, severity of deviated current 2000 mA
C31	Fault between coil A41A42, severity of deviated current 500 mA
C32	Fault between coil A41A42, severity of deviated current 1000 mA
C33	Fault between A41A42, severity of deviated current 2000 mA

The identification of the coil in which the fault appears might be seen as redundant or unnecessary. In the presence of winding with shorted coil, not so many options can be considered to fix this fault and get back the motor working at 100% of its capabilities. After different types of troubleshooting, as covered by various standards such as NEMA MG 1-2011 (MG1), IEEE 1068-2009 (IEEE 1068),

ANSI/EASA AR100-2010 (AR100) and others, typically the solution is to rewind the entire stator.

In contrast, having identified the location and severity of the fault will provide sufficient information, and when possible, help to tackle the problem before is catastrophic or in the following scheduled maintenance. In the case of large motors, it might be feasible to repair the winding fault in an early stage.

6.2 SM FAULT DIAGNOSIS – METHODOLOGY

The methodology is proposed in three steps. **The first step** is to extract the most optimal group of features from each available dataset, in this case, the algorithm chosen is Sparse Discriminant Analysis. **The second** step is to aggregate all selected features into one dataset, based on the most important features. The **third** step is modeling and comparing the results from each model results.

The final methodology is highly dependent on the available processing resources, the environment of the system under consideration and the possibility of installing extra sensors.

In order to define a general methodology and to summarize the results, the criterion applied is to consider those features that are most important or with a higher score to get optimal models and with the higher accuracy possible.

In cases such as the fault detection studied in this research, the criterion to define the number of features might be influenced by the type of fault to detect and the availability of different signals.

6.2.1 FIRST STEP: FEATURES EXTRACTION

Using the function **varImp()** from the *caret package* in *R*, was possible to evaluate the performance between classes by ROC curve analysis on each predictor **[16][37]**. The analysis is carried out by decomposing the problem into pair-wise problems, and the area under the curve, AUC, is calculated by the trapezoidal rule for each class pair **[36]**.

The score obtained from running sLDA is scaled to maximum 100, so the comparison across datasets is possible. An example from the stats dataset can be seen in the table below.

Table 6-3 Example of features selection scores, stats dataset.

Feature	C00	C11	C12	C13	C21	C22	C23	C31	C32	C33
P3H1_absMean	50	90.8	100	50	100	0.1	54.9	75.9	90.3	90.8
P3H1_snr_10	100	100	100	99.9	100	99.2	96.1	99.9	98	100
P3RGb1_mu	99.4	100	100	92.8	100	46.4	70.5	100	100	100
P3H2_mu	63.6	90.8	100	64.1	100	2.7	84.4	75.9	87.7	90.8

The features listed in the table above are:

- **P3H1_absMean**: the mean absolute deviation of the Hall sensor 1, pole 3
- **P3H1_snr_10**: the signal to noise ratio excluding the power of the 10 lowest harmonics, of the Hall sensor 1, pole 3.
- **P3RGb1_mu**: mean of the voltage induced in the Rogowski coil installed in bar 1, pole.
- **P3H2_mu**: mean of the Hall sensor 2, pole 3.

6.2.2 SECOND STEP: DATA AGGREGATION

Once the feature importance is calculated for each dataset, it is necessary to group them into one single set of features. The proposed method is to multiply the score of the feature obtained when training the model by the accuracy of the sLDA model. Then, the top 25 features per class are considered. The final dataset contains 107 features. The total of features selected per dataset is represented in Table 6-4:

Table 6-4 Features selected per dataset

Dataset	count
stats	76
freq	22
VIpower	6
Clark	2
Park	1

The top 10 features are selected considering the total score across all classes. These features are chosen based on the amount of information contained to classify a new observation unequivocally.

However, other options to define the minimum number of necessary features to identify the different types of faults might be considering the type of sensors and the total accuracy or, selecting the group of features that best different between faulty or healthy condition only. Any option that minimizes the number of features and maximizes the classification task can be implemented as well.

Table 6-5 Top 10 extracted features.

Features	C00	C11	C12	C13	C21	C22	C23	C31	C32	C33	Total
ampRatioIa1Ia	93	93	93	93	93	93	93	93	93	93	927
P3RGb4_iqr	92	93	93	92	93	92	92	92	92	93	923
P3RGb4_absMedian	92	93	92	92	93	92	92	92	92	93	922
P3H1_snr_02	93	93	93	93	93	92	90	93	92	93	922
P3H1_sinad	93	93	93	93	93	92	90	93	92	93	922
P3H1_snr_05	93	93	93	93	93	92	89	93	91	93	921
P3H1_snr_10	93	93	93	93	93	92	89	93	91	93	920
Vn_pe_abs_300	91	92	93	91	93	91	91	92	91	92	917
MicNC_abs_150	92	92	93	92	92	92	92	92	92	84	915

From these results, it is possible to observe that the proportion of the absolute magnitude of the fundamental harmonic of the parallel branch Ia1 (the branch in which the fault was seeded) to the absolute magnitude of the fundamental harmonic of Ia is the feature with higher importance.

In fact, considering only this feature, it is possible to identify the condition of the machine with an accuracy of 96% using C5.0 algorithm; with a training time of 2.78 minutes and 0.16 seconds to evaluate a new data point.

The second most important variable is the voltage induced in the Rogowski coil installed in the bar 4, pole 3. The two extracted features from it are P3RGb4_iqr – interquartile range and P3RGb4_absMedian – mean absolute median.

The third most important variable is the Hall sensor 1, pole 3. The features extracted from it are the signal to noise ratio; P3H1_snr_02, P3H1_snr_05 and P3H1_snr_10, excluding the power of the 2, 5, 10 lowest harmonics respectively. Another feature extracted is the signal to noise and distortion ratio P3H1_sinad.

The other two features in the top 10 are the absolute magnitude of the motor's neutral Voltage to ground (Vn_pe) at 600 Hz; only possible in Y-winding connection. And the absolute magnitude of the harmonic at 150 Hz of the Noise-canceling Microphone, installed parallel to the shaft and rotating with it.

6.2.3 THIRD STEP: MODELLING

The total number of models considered in this research is 10, based on selecting different types of models that adapt better to large datasets or less affected by correlated features or just models that are faster; the details of each model can be found in section 5.8

Table 6-6 Results from the models considering a 10 classes classification task.

Classification Model	Avg Precision	Avg Recall	Avg F1	ACC	Training time	Predict time
LDA	96%	96%	96%	95.5%	13.88 secs	0.21 secs
CART	NA	40%	NA	39.1%	43.34 secs	0.25 secs
treebag	99%	99%	99%	98.6%	3.72 min	0.73 secs
rf	99%	99%	99%	99.2%	22.43 min	0.47 secs
svmPoly	99%	99%	99%	99.2%	1.21 hours	5.04 secs
gbm	100%	100%	100%	99.5%	18.87 min	0.57 secs
bagFDA	97%	97%	97%	97.2%	4.39 hours	5.79 secs
nb	81%	78%	77%	78.7%	20.49 min	7.15 secs
sparseLDA	91%	91%	91%	90.2%	16.66hours	0.21 secs
C5.0	100%	100%	100%	99.3%	11.06 min	3.44 secs

The final training dataset contains 21600 observations and 107 features. Models such as Stochastic Gradient Boosting or C5.0 gives close to 100% of accuracy, the faster model to predict a new observation is LDA with 95.5% of accuracy.

In general, the performance of the ten models to differentiate between the 10 classes is satisfactory. The ability to classify the condition of the machine, healthy vs. faulty, is achieved with an accuracy of 100%.

It is possible to deep-dive into the results from gbm, the parameter obtained from the optimal model were: number of trees = 150, interaction depth = 3, shrinkage = 0.1 and minimum number of observation in node = 10. From the confusion matrix, we can observe that gbm is capable of unequivocally identify the classes C00, C11, C21, and C31. In other words, it is possible to detect the winding in which the fault was seeded at 500 mAmps of deviated current. The model underperforms when differentiating the other two severities or deviated current.

Table 6-7 Confusion matrix, Stochastic Gradient Boosting.

	Reference									
Prediction	C00	C11	C12	C13	C21	C22	C23	C31	C32	C33
C00	708	0	0	0	0	0	0	0	0	0
C11	0	708	0	0	0	0	0	0	0	0
C12	0	0	707	0	0	0	0	0	1	0
C13	0	0	0	698	0	0	8	0	0	2
C21	0	0	0	0	708	0	0	0	0	0
C22	0	0	0	0	0	707	0	0	1	0
C23	0	0	0	9	0	0	697	0	2	0
C31	0	0	0	0	0	0	0	708	0	0
C32	0	0	2	0	0	0	2	0	704	0
C33	0	0	0	0	0	0	0	0	1	707

Evaluating the most important variables used in the optimal model after training the gbm, the results are shown in Table 6-8.

Table 6-8 gbm variable importance, optimal model

Feature	Importance
ampRatioIa1Ia	100
P3b3_range	33.24
lagIa1a	31.56
MicNC_abs_150	30.15
w1_bandpower	20.09

The score for the ampRatioIa1Ia triples the value of the next variable; P3b3_range calculated as the difference between the maximum and minimum of the current induced in the bar 3, pole 3.

The other features are lagIa1a as the difference between the lag of the parallel branch current Ia1 and the phase current Ia, MicNC_abs_150 as the power at 150 Hz of the Noise-cancelling microphone and the w1_bandpower as the average power of the rotational speed.

6.3 DISCUSSION AND FURTHER STEPS

This research embraced different fields of electrical engineering, power electronics, and computer science. The intention was to define a simple and flexible methodology, but robust, using practical implementation. Especial attention was given to providing all the necessary details to replicate all the followed steps so it can serve as a base for future research works.

After a great effort and time dedicated to building the rotor signal collector and extended literature review looking for previous attempts of using new types of signals applied to condition monitoring and looking for all the possible options and combination of signals using latest technological developments in sensors, was possible to test the validity of all the signals and quantify the importance or amount of information necessary to training different classification algorithms.

In Chapter 2 the configuration of the test bench was explained, and the signals were grouped into External sensors – traditional signals, section 2.4.1 and Rotor-mounted sensors, section 2.4.2. Using the same methodology considering only external sensor, it is possible to obtain 100% accuracy to identify the condition of the machine and a maximum of 96.3% of accuracy to identify the type of fault and severity. Both results are obtained using support vector machine with a training

time of 1.72 hours (21600 observations) and 11.1 seconds to evaluate 7080 observations.

Considering RSC signals only, it was possible to identify the condition of the machine with 100% of accuracy and up to 99.3% of accuracy to determine the type of fault using Stochastic Gradient Boosting with 18.5 mins of training time and 0.57 seconds for validation.

There are also many potential improvements; for example, consider different workloads and confirm that the model prediction does not change. Consider mechanical failures, so it is possible to train the algorithm to assess more machine conditions. Perhaps, the most important improvement is to generalize this methodology to more machines, synchronous and other types, so a cross-validated algorithm, over different motors, might be capable of extracting general features to all, and pursue a universal condition monitoring methodology. The implementation of such a kind of solution, in real industrial environments, will definitely be of value to increase the reliability of the installed based and the introduction of smarter equipment.

Some further steps should consider modifying the features selection algorithm to be capable of selecting the features per the type of variable or measured physical magnitude, so the process of defining the minimum number of signals to collect is automated and based on different criteria, for example, to detect a specific type of fault as quick as possible.

This kind of methodology can be used to create the so-called “Digital-twin” of an electric motor; data generated from a model together with sensor data, will allow the assessment of any system by Residual Evaluation of expected versus measured.

6.4 THE CHALLENGE OF CONDITION-BASED MAINTENANCE

Questions such as: how many different types of motors might be installed in the same factory? Will the motor manufacturer provide all the necessary details or will it be possible to purchase a condition monitoring system specially designed for each motor? Will this condition monitoring system be compatible with currently installed systems in a given industrial process.?

All these questions and others, make the implementation of condition-based maintenance strategies a challenge that is limiting the industry to evolve from corrective/reactive strategies to more robust and optimal fault detection systems, allowing proactive maintenance strategies and decreasing the downtime. So it is desirable the development and implementation of techniques compatible with different types of motors and configurations.

Author comments

As of December 2016, a similar methodology is under development applied to healthcare imaging products such as Magnetic resonance systems and Computerized tomography scans as part of my current role as Data Scientist in a global healthcare provider company. Based on the historical service data, together with the machine data, it is possible to define the knowledge matrix allowing the assessment of the systems contrasted with the population expected behavior.

7 REFERENCES

[1] ABDI-JALEBI, EHSAN ; MCMAHON, RICHARD: High-performance low-cost Rogowski transducers and accompanying circuitry. In: *Instrumentation and Measurement, IEEE Transactions on* vol. 56, IEEE (2007), Nr. 3, pp. 753–759

[2] AFIFI, ASHRAF: Improving the classification accuracy using support vector machines (SVMS) with new kernel. In: *Journal of global research in computer science* vol. 4 (2013), Nr. 2, pp. 1–7

[3] AKAR, M. ; TASKIN, S. ; SEKER, S. ; CANKAYA, I.: Detection of static eccentricity for permanent magnet synchronous motors using the coherence analysis. In: *Turk. J. Electr. Eng. Comput. Sci* vol. 18 (2010), pp. 963–974

[4] BACCHUS, ALEXANDRE ; BIET, MÉLISANDE ; MACAIRE, LUDOVIC ; LE MENACH, YVONNICK ; TOUNZI, ABDELMOUNAIM: Comparison of supervised classification algorithms combined with feature extraction and selection: Application to a turbo-generator rotor fault detection. In: *Diagnostics for Electric Machines, Power Electronics and Drives (SDEMPED), 2013 9th IEEE International Symposium on*, 2013, pp. 558–565

[5] BHATTACHARYA, ABHISEKH ; DAN, PRANAB K: Recent trend in condition monitoring for equipment fault diagnosis. In: *International Journal of System Assurance Engineering and Management*, Springer, pp. 1–15

[6] CASIMIR, R ; BOUTLEUX, E ; CLERC, G: Fault diagnosis in an induction motor by pattern recognition methods. In: *Diagnostics for Electric Machines, Power Electronics and Drives, 2003. SDEMPED 2003. 4th IEEE International Symposium on*, 2003, pp. 294–299

[7] CHEN, W-Y ; XU, J-X ; PANDA, SK: Application of artificial intelligence techniques to the study of machine signatures. In: *Electrical Machines (ICEM),*

2012 XXth International Conference on, 2012, pp. 2390–2396

[8] CHURCHILL, T. L.: *Development of a Rotor-Mounted Scanner for Hydrogenerators*, 1993

[9] CLEMMENSEN, LINE ; HASTIE, TREVOR ; WITTEN, DANIELA ; ERSBØLL, BJARNE: Sparse discriminant analysis. In: *Technometrics*, Taylor \& Francis (2012)

[10] DENG, F ; DEMERDASH, NAO: A coupled finite-element state-space approach for synchronous generators. I. model development. In: *Aerospace and Electronic Systems, IEEE Transactions on* vol. 32, IEEE (1996), Nr. 2, pp. 775–784

[11] DENG, F ; DEMERDASH, NAO ; VAIDYA, JG ; SHAH, MJ: A coupled finite-element state-space approach for synchronous generators. II. Applications. In: *Aerospace and Electronic Systems, IEEE Transactions on* vol. 32, IEEE (1996), Nr. 2, pp. 785–794

[12] DENG, LINFENG ; ZHAO, RONGZHEN: An intelligent pattern recognition method for machine fault diagnosis. In: *Ubiquitous Robots and Ambient Intelligence (URAI), 2013 10th International Conference on*, 2013, pp. 187–192

[13] DROZDOWSKI PIOTR, WEINREB KONRAD WEGIEL TOMASZ: Wykrywanie ekscentrycznosci wirnika w maszynach synchronicznych poprzez analize harmoniczna pradow. In: *SME* vol. 95 (1995)

[14] DUNN, SANDY: *Condition Monitoring in the 21st Century*. URL http://www.plant-maintenance.com/articles/ConMon21stCentury.shtml. - abgerufen am 2013-27–12

[15] EDMONDS, JS ; CHURCHILL, TL: Unique sensor applications for hydroelectric generator rotor-mounted sensor scanning technology. In: *Northcon/96*, 1996, pp. 73–77

[16] FAWCETT, TOM: An introduction to ROC analysis. In: *Pattern Recogn. Lett.* vol. 27. New York, NY, USA, Elsevier Science Inc. (2006), Nr. 8, pp. 861–874

[17] FERREIRA, JOSÉ GREGORIO ; SOBCZYK, TADEUSZ: Multicriteria diagnosis of synchronous machines—Rotor-mounted sensing system. Rotor signal collector construction. In: *Control (CONTROL), 2014 UKACC International Conference on*, 2014, pp. 450–455

[18] FERREIRA, JOSÉ GREGORIO ; SOBCZYK, TADEUSZ J ; WARZECHA, ADAM ; OTHERS: Multicriteria diagnosis of synchronous machine using the welch method. In: *Czasopismo Techniczne* vol. 2015, Portal Czasopism Naukowych Ejournals. eu (2015), Nr. Elektrotechnika Zeszyt 1-E (8) 2015, pp. 343–352

[19] FERREIRA, JOSE GREGORIO ; SOBCZYK, TADEUSZ ; WARZECHA, ADAM: Multicriteria diagnosis of synchronous machine using MCSA. In: *Internation*

Symposium on Electrical Machines SME (2014), Nr. 4/2014 (104), p. 313

[20] FILZMOSER, PETER ; GSCHWANDTNER, MORITZ ; TODOROV, VALENTIN: Review of sparse methods in regression and classification with application to chemometrics. In: *Journal of Chemometrics* vol. 26, Wiley Online Library (2012), Nr. 3-4, pp. 42–51

[21] FIORENZO FILIPPETTI, MEMBER IEEE ALBERTO BELLINI MEMBER IEEE G´ ERARD-ANDR´ E CAPOLINO FELLOW IEEE.: Condition Monitoring and Diagnosis of Rotor Faults in Induction Machines: State of Art and Future Perspectives. In: *Diagnostics for Electric Machines, Power Electronics \& Drives (SDEMPED), 2013 IEEE International Symposium on*, 2013

[22] GANCHEV, MARTIN ; KUBICEK, BERNHARD ; KAPPELER, HANSJOERG: Rotor temperature monitoring system. In: *Electrical Machines (ICEM), 2010 XIX International Conference on*, 2010, pp. 1–5

[23] GLOWACZ, ADAM ; GLOWACZ, WITOLD: Diagnostics of Direct Current motor with application of acoustic signals, reflection coefficients and K-Nearest Neighbor classifier. In: *Przeglkad Elektrotechniczny* vol. 88 (2012), Nr. 5a, pp. 231–233

[24] GLOWACZ, Z ; KOZIK, J: Detection of Synchronous Motor Inter-Turn Faults Based on Spectral Analysis of Park'S Vector/Detekcja Zwarc Zwojowych W Silniku Synchronicznym Bazujaca Na Analizie Spektralnej Wektora Przestrzennego Pradu Twornika. In: *Archives of Metallurgy and Materials* vol. 58 (2013), Nr. 1, pp. 19–23

[25] HAN, JIAWEI ; PEI, JIAN ; KAMBER, MICHELINE: *Data mining: concepts and techniques* : Elsevier, 2011

[26] HARRIS, FREDRIC J: On the use of windows for harmonic analysis with the discrete Fourier transform. In: *Proceedings of the IEEE* vol. 66, IEEE (1978), Nr. 1, pp. 51–83

[27] HASTIE, TREVOR ; TIBSHIRANI, ROBERT ; FRIEDMAN, JEROME ; FRANKLIN, JAMES: The elements of statistical learning: data mining, inference and prediction. In: *The Mathematical Intelligencer* vol. 27, Springer (2005), Nr. 2, pp. 83–85

[28] ILONEN, JARMO ; KAMARAINEN, J-K ; LINDH, TUOMO ; AHOLA, JERO ; KALVIAINEN, HEIKKI ; PARTANEN, JARMO: Diagnosis tool for motor condition monitoring. In: *IEEE Transactions on Industry Applications* vol. 41, IEEE (2005), Nr. 4, pp. 963–971

[29] ISERMANN, ROLF: Fault diagnosis of machines via parameter estimation and knowledge processing—tutorial paper. In: *Automatica* vol. 29, Elsevier (1993), Nr. 4, pp. 815–835

[30] ISERMANN, ROLF: *Fault-diagnosis applications: model-based condition monitoring: actuators, drives, machinery, plants, sensors, and fault-tolerant systems* : Springer, 2011

[31] JUÁREZ-CALTZONTZIN, LAURA L ; TRINIDAD-HERNÁNDEZ, GUSTAVO ; ASIA'\IN-OLIVARES, TOMÁS I ; RUIZ-VEGA, DANIEL: Theoretical and Experimental Analysis of the Short Circuit Current Components in Salient Pole Synchronous Generators. In: *Proceedings of the 11th Spanish Portuguese Conference on Electrical Engineering (11CHLIE)*. vol. 1, 2009

[32] KANERVA, SAMI ; SEMAN, SLAVOMIR ; ARKKIO, ANTERO: Inductance model for coupling finite element analysis with circuit simulation. In: *Magnetics, IEEE Transactions on* vol. 41, IEEE (2005), Nr. 5, pp. 1620–1623

[33] KEARNS, MICHAEL J ; VAZIRANI, UMESH VIRKUMAR: *An introduction to computational learning theory* : MIT press, 1994

[34] KORBICZ, JÓZEF: *Fault diagnosis: models, artificial intelligence, applications* : Springer, 2004

[35] KOWALSKI, CZESLAW T ; ORLOWSKA-KOWALSKA, TERESA: Neural networks application for induction motor faults diagnosis. In: *Mathematics and Computers in Simulation (MATCOM)* vol. 63 (2003), Nr. 3, pp. 435–448

[36] LACHICHE, NICOLAS ; FLACH, PETER: Improving accuracy and cost of two-class and multi-class probabilistic classifiers using ROC curves. In: *Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003)* (2003)

[37] LANDGREBE, THOMAS CW ; DUIN, ROBERT PW: Approximating the multiclass ROC by pairwise analysis. In: *Pattern recognition letters* vol. 28, Elsevier (2007), Nr. 13, pp. 1747–1758

[38] LIPO, THOMAS A: *Analysis of synchronous machines* : CRC Press, 2012

[39] LIU, BIN ; WEBB, GEOFFREY I.: Generative and Discriminative Learning. In: SAMMUT, CLAUDE AND WEBB, GEOFFREY I. (ed.): *Encyclopedia of Machine Learning*. Boston, MA : Springer US, 2010 — ISBN 978-0-387-30164-8, pp. 454–455

[40] LU, BIN ; LI, YAOYU ; WU, XIN ; YANG, ZHONGZHOU: A review of recent advances in wind turbine condition monitoring and fault diagnosis. In: *Power Electronics and Machines in Wind Applications, 2009. PEMWA 2009. IEEE*, 2009, pp. 1–7

[41] MA, XIAOHE ; SHEN, SONGHUA ; BI, LUYAN ; LI, JIANFENG: Finite Element Method modeling and analysis for aeronautic synchronous generator with damper windings on unloading and short-circuit conditions. In: *Industrial Engineering and Engineering Management, 2007 IEEE International*

Conference on, 2007, pp. 1848–1852

[42] MAI, QING ; YANG, YI ; ZOU, HUI: Multiclass Sparse Discriminant Analysis. In: *arXiv preprint arXiv:1504.05845* (2015)

[43] NANDI, S. ; TOLIYAT, H. A. ; LI, X.: Condition monitoring and fault diagnosis of electrical motors-a review. In: *Energy Conversion, IEEE Transactions on* vol. 20, IEEE (2005), Nr. 4, pp. 719–729

[44] NANDI, SUBHASIS ; TOLIYAT, HAMID A: Fault diagnosis of electrical machines-a review. In: *Electric Machines and Drives, 1999. International Conference IEMD'99*, 1999, pp. 219–221

[45] NETI, P. ; NANDI, S.: Stator inter-turn fault detection of synchronous machines using field current signature analysis. In: *Industry Applications Conference, 2006. 41st IAS Annual Meeting. Conference Record of the 2006 IEEE*. vol. 5, 2006, pp. 2360–2367

[46] NETI, PRABHAKAR ; NANDI, SUBHASIS: Stator interturn fault detection of synchronous machines using field current and rotor search-coil voltage signature analysis. In: *Industry Applications, IEEE Transactions on* vol. 45, IEEE (2009), Nr. 3, pp. 911–920

[47] NG, ANDREW Y ; JORDAN, MICHAEL I: On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes. In: *Advances in neural information processing systems* vol. 2, MIT; 1998 (2002), pp. 841–848

[48] NOES, T ; MEVIK, B-H: Understanding the collinearity problem in regression and discriminant analysis. In: *Journal of chemometrics* vol. 15, John Wiley & Sons (2001), Nr. 4, pp. 413–426

[49] OSCARSON, G. ; IMBERTSON, J. ; IMBERTSON, B ; MOLL, S.: *The ABC's of Synchronous Motors*, 2000

[50] PAAP, GERARDUS C: Symmetrical components in the time domain and their application to power network calculations. In: *Power Systems, IEEE Transactions on* vol. 15, IEEE (2000), Nr. 2, pp. 522–528

[51] PENMAN, J ; DEY, MN ; TAIT, AJ ; BRYAN, WE: Condition monitoring of electrical drives. In: *Electric Power Applications, IEE Proceedings B* vol. 133, IET (1986), Nr. 3, pp. 142–148

[52] POUREBRAHIMI, ALIREZA ; MOKHTAR, SAMIRA ; SAHAMI, SOHEILA ; MAHMOODI, MINA: Status detection and fault diagnosing of rotatory machinery by vibration analysis using data mining. In: *Computer Technology and Development (ICCTD), 2010 2nd International Conference on*, 2010, pp. 131–135

[53] PYRHONEN, J. ; JOKINEN, T. ; HRABOVCOVÁ, V.: *Design of rotating electrical machines* : Wiley, 2009

[54] RADZIK, MICHAL ; SOBCZYK, TADEUSZ J: Steady-state analysis of synchronous machines loaded by an angle depended torque. In: *Przeglkad Elektrotechniczny* vol. 89 (2013), Nr. 2b, pp. 158–161

[55] RAHIMIAN, MINA M ; BUTLER-PURRY, KAREN: Modeling of synchronous machines with damper windings for condition monitoring. In: *Electric Machines and Drives Conference, 2009. IEMDC'09. IEEE International*, 2009, pp. 577–584

[56] RASMUSSEN, JAMES R ; THOMPSON, DW: On-line generator condition assessment using rotor-mounted sensor technology. In: *Electric Machines and Drives, 1999. International Conference IEMD'99*, 1999, pp. 815–817

[57] RASMUSSEN, JAMES R ; TUTTLE JR, CB: Early detection of internal stator winding faults using rotor-mounted infrared sensors. In: *Electric Machines and Drives, 1999. International Conference IEMD'99*, 1999, pp. 800–802

[58] REINAUD, JULIA: CO2 allowance & electricity price interaction. In: *Impact on industry's electricity purchasing strategies in Europe. IEA information paper* (2007)

[59] SEERA, MANJEEVAN ; LIM, CHEE PENG: Online motor fault detection and diagnosis using a hybrid FMM-CART model. In: *IEEE transactions on neural networks and learning systems* vol. 25, IEEE (2014), Nr. 4, pp. 806–812

[60] SHUTING, WAN ; YONGGANG, LI ; HEMING, LI ; GUIJI, TANG: The analysis of generator excitation current harmonics on stator and rotor winding fault. In: *Industrial Electronics, 2006 IEEE International Symposium on*. vol. 3, 2006, pp. 2089–2093

[61] SMITH, DALE P.: *PdM centralization for significant energy savings*. URL www.reliableplant.com/Articles/Print/15492. - abgerufen am 2013-25–10

[62] SOBCZYK, TADEUSZ J ; WEGIEL, TOMASZ ; SULOWICZ, MACIEJ ; WARZECHA, ADAM ; WEINREB, KONRAD: A distributed system for diagnostics of induction motors. In: *Diagnostics for Electric Machines, Power Electronics and Drives, 2005. SDEMPED 2005. 5th IEEE International Symposium on*, 2005, pp. 1–5

[63] SOBCZYK, TADEUSZ J ; ZAJAC, MIECZYSLAW ; KICH, ZYGMUNT R ; SULOWICZ, MACIEJ: Signal processing for extraction of characteristic features of induction motor stator currents at rotor faults. In: *Diagnostics for Electric Machines, Power Electronics and Drives, 2005. SDEMPED 2005. 5th IEEE International Symposium on*, 2005, pp. 1–6

[64] SOBCZYK, TJ: Methodological aspects of mathematical modelling of induction machines. In: *WNT, Warsaw* (2004)

[65] SOBCZYK, TJ: Frequency analysis of faulty machines-possibilities and limitations. In: *Diagnostics for Electric Machines, Power Electronics and Drives, 2007. SDEMPED 2007. IEEE International Symposium on*, 2007, pp. 121–125

[66] SOBCZYK, TJ: Inductanceless model of salient-pole synchronous machines. In: *Power Electronics, Electrical Drives, Automation and Motion, 2008. SPEEDAM 2008. International Symposium on*, 2008, pp. 620–625

[67] SOBCZYK, TJ ; RADZIK, M: Direct determination of currents spectra of synchronous machines at load perturbations. In: *Diagnostics for Electric Machines, Power Electronics and Drives, 2009. SDEMPED 2009. IEEE International Symposium on*, pp. 1–6

[68] STONE, GC: Condition monitoring and diagnostics of motor and stator windings-A review. In: *Dielectrics and Electrical Insulation, IEEE Transactions on* vol. 20, IEEE (2013), Nr. 6, pp. 2073–2080

[69] TAVNER, PETER J ; GAYDON, BG ; WARD, DM: Monitoring generators and large motors. In: *Electric Power Applications, IEE Proceedings B* vol. 133, IET (1986), Nr. 3, pp. 169–180

[70] TAVNER, PJ: Condition monitoring-the way ahead for large electrical machines. In: *Electrical Machines and Drives, 1989. Fourth International Conference on (Conf. Publ. No.??)*, 1989, pp. 159–162

[71] TAVNER, PJ: Review of condition monitoring of rotating electrical machines. In: *Electric Power Applications, IET* vol. 2, IET (2008), Nr. 4, pp. 215–247

[72] TEAM, R CORE: *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2015

[73] THEODORE, WILDI ; OTHERS: *Electrical Machines, Drives And Power Systems, 6/E* : Pearson Education India, 2007

[74] TIPSUWANPORN, VITTAYA ; LEAWSOONG, MONGKOL ; NUMSOMRAN, ARJIN ; WONGRATANAPORNKUL, CHANWIT: Fault Detection In compressor Using FFT Algorithm. In: *Proceedings of the World Congress on Engineering and Computer Science*. vol. 1, 2013

[75] TOLIYAT, HAMID A ; NANDI, SUBHASIS ; CHOI, SEUNGDEOG ; MESHGIN-KELK, HOMAYOUN: *Electric Machines: Modeling, Condition Monitoring, and Fault Diagnosis* : CRC Press, 2012

[76] VAIMANN, T ; KALLASTE, A ; KILK, A: Using Clarke vector approach for stator current and voltage analysis on induction motors with broken rotor bars. In:

Elektronika ir Elektrotechnika vol. 123 (2012), Nr. 7, pp. 17–20

[77] VALIANT, LESLIE G: A theory of the learnable. In: *Communications of the ACM* vol. 27, ACM (1984), Nr. 11, pp. 1134–1142

[78] VAPNIK, VLADIMIR NAUMOVICH ; VAPNIK, VLAMIMIR: *Statistical learning theory*. vol. 1 : Wiley New York, 1998

[79] VOTZI, HELMUT L ; VOGELSBERGER, MARKUS ; ERTL, HANS: Low-Cost Current Sensor for Power Capacitors Based on a PCB Rogowski-Coil

[80] WAIDE, P. ; BRUNNER, C. U. ; OTHERS: *Energy-Efficiency Policy Opportunities for Electric Motor-Driven Systems*, 2011

[81] WANG, CHENG ; CAO, LONGBING ; MIAO, BAIQI: Optimal feature selection for sparse linear discriminant analysis and its applications in gene expression data. In: *Computational Statistics & Data Analysis* vol. 66 (2013), Nr. C, pp. 140–149

[82] WANG, WEI ; ZHAO, CHUNHUI ; SUN, YOUXIAN ; GAO, FURONG: A variant fisher discriminant analysis algorithm and its application to fault diagnosis. In: *Control Conference (ASCC), 2015 10th Asian*, 2015, pp. 1–6

[83] WEI, TAIYUN ; SIMKO, VILIAM: *corrplot: Visualization of a Correlation Matrix*, 2016. — R package version 0.77

[84] WEINREB, K. ; DROZDOWSKI, P.: Detection of Winding Faults of a Salient Pole Synchronous Machine by a Spectral Analysis of Currents. In: *Proc. ICEM*. vol. 94, pp. 5–8

[85] WEINREB, KONRAD ; SULOWICZ, MACIEJ: Noninvasive diagnostics of synchronous machine windings internal asymmetry. In: *Maszyny Elektryczne: zeszyty problemowe* (2007), Nr. 77, pp. 59–64

[86] WELCH, PETER D: The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. In: *IEEE Transactions on audio and electroacoustics* vol. 15 (1967), Nr. 2, pp. 70–73

[87] WICKHAM, HADLEY ; OTHERS: Tidy data. In: *Journal of Statistical Software* vol. 59, Foundation for Open Access Statistics (2014), Nr. 10, pp. 1–23

[88] WING, MAX KUHN. CONTRIBUTIONS FROM JED ; WESTON, STEVE ; WILLIAMS, ANDRE ; KEEFER, CHRIS ; ENGELHARDT, ALLAN ; COOPER, TONY ; MAYER, ZACHARY ; KENKEL, BRENTON ; TEAM, THE R CORE ; ET AL.: *caret: Classification and Regression Training*, 2015. — R package version 6.0-62

[89] WITTEN, DANIELA M ; TIBSHIRANI, ROBERT: Penalized classification using Fisher's linear discriminant. In: *Journal of the Royal Statistical Society: Series*

B (Statistical Methodology) vol. 73, Wiley Online Library (2011), Nr. 5, pp. 753–772

[90] XIE, YUAN ; ZHANG, TAO: A fault diagnosis approach using SVM with data dimension reduction by PCA and LDA method. In: *Chinese Automation Congress (CAC), 2015*, 2015, pp. 869–874

[91] YANG, BO-SUK ; OH, MYUNG-SUCK ; TAN, ANDY CHIT CHIOW ; OTHERS: Fault diagnosis of induction motor based on decision trees and adaptive neuro-fuzzy inference. In: *Expert Systems with Applications* vol. 36, Elsevier (2009), Nr. 2, pp. 1840–1849

[92] YANG, WENXIAN ; LITTLE, CHRISTIAN ; TAVNER, PETER J ; OTHERS: Data-driven technique for interpreting wind turbine condition monitoring signals. In: *IET Renewable Power Generation* vol. 8, IET (2013), Nr. 2, pp. 151–159

[93] YANG, WENXIAN ; TAVNER, PJ ; CRABTREE, CJ ; WILKINSON, MICHAEL: Research on a simple, cheap but globally effective condition monitoring technique for wind turbines. In: *Electrical Machines, 2008. ICEM 2008. 18th International Conference on*, 2008, pp. 1–5

[94] ZAREI, J.: Induction motors bearing fault detection using pattern recognition techniques. In: *Expert Systems with Applications* vol. 39, Elsevier (2012), Nr. 1, pp. 68–73

[95] ZHANG, H.: The optimality of naive Bayes. In: *A A* vol. 1 (2004), Nr. 2, p. 3

[96] ZHAOXIA, XIAO ; HONGWEI, FANG: Stator Winding Inter-Turn Short Circuit and Rotor Eccentricity Diagnosis of Permanent Magnet Synchronous Generator. In: *Control, Automation and Systems Engineering (CASE), 2011 International Conference on*, 2011, pp. 1–4