N° Ordre......../FT/UMBB/2023

PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA

MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH

UNIVERSITY M'HAMED BOUGARA – BOUMERDES

Faculty of Technology

Doctorate Thesis

Presented by:

Boualem IKHLEF

To obtain a **Doctorate in Science**:

Field: Electrical Engineering

Specialty: Automatic and industrial informatics

Detection and identification of defects in gearbox systems using artificial intelligence based techniques

Before the jury composed of :

Mr ADJERID	Smail	Professor	UMBB	President
Mr RAHMOUNE	Chemseddine	M.C.A	UMBB	Thesis Director
Mr BENAZZOUZ	Djamel	Professor	UMBB	Thesis Co-Director
Mr MEDDOUR	Ikhlas	M.C.A	ENST Alger	Examiner
Mr BENFDILA	Arezki	Professor	UMMTO	Examiner
Mr KARA	Redouane	Professor	UMMTO	Examiner
Mr NAFA	Fares	M.C.A	UMBB	Invited

Academic years 2022/2023

Dedication

I dedicate this work:

- *To my lovely family, my wife and my Daughters (AYA, ILHAM, WIAM, NIHAL and SIDRA) who always encourage me with passion and endless support. I am so Lucky to have a woman who loves me so much and stands beside me.*
- *To my faithful father, brothers and sister, who have always helped me and believed that I could do it...*
- *In the memory of my mother who did everything for me and whom I will never forget*
- *To My Supervisor RAHMOUNE CHEMSEDDINE.*
  Who spare nothing towards my success ...

Acknowledgment

All praise and thanks giving to Allah the most powerful and most merciful who give me the ability and patience to accomplish this work.

This thesis has been carried out at the solid mechanics and systems laboratory of Faculty of Technology. I would like to express my sincere thanks (with an honorary bow) to my thesis director RAHMOUNE Chemseddine for his patient guidance and endless support throughout my Ph.D. study. His immense knowledge, professional suggestions and constructive comments have been essential for this dissertation.

Dr. Rahmoune chemssedine has been a great advisor who was always available to discuss and support all the technical problems and he provides a very pleasant research environment. I benefitted a lot from his guidance not only in the academic life but also in the common everyday life.

I also thank Professor BENAZZOUZ who agreed to be co-director of my thesis and accepted me into his laboratory as a PhD student.

I also thank my friend Dr AKROUM for advising me to pursue the doctorate and also for his beneficial directions

I would like to extend my appreciation to all the members of the solid mechanics and systems laboratory for their help.

Résumé :

Les boîtes de vitesses sont massivement utilisées dans les industries d'aujourd'hui en raison de leur énorme importance dans la transmission de puissance ; par conséquent, leurs défauts peuvent fortement affecter les performances des machines. Par conséquent, de nombreux chercheurs travaillent sur la détection et la classification des défauts des boîtes de vitesses. Cependant, la plupart des travaux sont effectués dans des conditions de vitesse constante, tandis que les engrenages fonctionnent généralement dans des conditions de vitesse et de couple variables, ce qui rend la tâche plus difficile. Dans ce travail, nous proposons une nouvelle méthode de surveillance de l'état des boîtes de vitesses qui est capable de révéler efficacement le défaut à partir des signatures vibratoires dans des conditions de fonctionnement variables. Tout d'abord, le signal de vibration est traité avec la transformation de paquets d'ondelettes discrètes à chevauchement maximal (MODWPT) pour extraire les modes. Ensuite, les caractéristiques du domaine temporel sont calculées à partir de chaque mode. Ensuite, l'ensemble de caractéristiques est réduit à l'aide de l'algorithme d'optimisation des colonies de fourmis (ACO) en supprimant les paramètres redondants et sans importance qui peuvent induire en erreur la classification. Enfin, un algorithme d'apprentissage d'ensemble Random Forest (RF) est utilisé pour former un modèle capable de classer le défaut en fonction des caractéristiques sélectionnées. L'aspect innovant de cette méthode est que, contrairement aux autres méthodes existantes, ACO est capable d'optimiser non seulement les caractéristiques mais aussi les paramètres du classificateur afin d'obtenir la plus grande précision de classification. La méthode proposée a été testée sur un ensemble de données réelles de conditions de fonctionnement variables composé de six boîtes de vitesses différentes. Dans le but de prouver les performances de notre méthode, celle-ci a été comparée à d'autres méthodes conventionnelles. Les résultats obtenus indiquent sa robustesse et sa stabilité de précision pour gérer le problème des conditions de fonctionnement variables dans la détection et la classification des défauts des boîtes de vitesses avec une grande efficacité.

Mots clés: Machines rotatives, boîtes de vitesses, détection de défauts, extraction de fonctionnalités, sélection de fonctionnalités, optimisation, classification

ملخص :

تُستخدم علب التروس على نطاق واسع في الصناعات الحالية نظرًا لأهميتها الهائلة في نقل الطاقة ؛ لذلك ، يمكن أن تؤثر عيوبها بشدة على أداء الآلات. لذلك يعمل العديد من الباحثين على كشف وتصنيف أعطال علبة التروس. ومع ذلك ، يتم تنفيذ معظم العمل في ظل ظروف سرعة ثابتة بينما تعمل التروس عادةً في ظل ظروف متغيرة السرعة وعزم الدوران مما يجعل المهمة أكثر صعوبة. في هذا البحث ، نقترح طريقة جديدة لمراقبة حالة علب التروس قادرة على الكشف بشكل فعال عن الخطأ من توقيعات الاهتزاز في ظل ظروف تشغيل مختلفة. أولاً ، تتم معالجة إشارة الاهتزاز باستخدام الحد الأقصى لتحويل الموجات المنفصلة المتراكبة (MODWPT) لاستخراج أوضاع (modes). ثم يتم حساب خصائص المجال الزمني من كل وضع. بعد ذلك ، يتم تقليل مجموعة الميزات (features) باستخدام خوارزمية مستعمرة النمل (Ant Colony Optimization) عن طريق إزالة المعلمات الزائدة وغير المهمة التي قد تضلل التصنيف. أخيرًا ، يتم استخدام خوارزمية تعلم مجموعة غابة عشوائية Random Forest لتدريب نموذج قادر على تصنيف الخلل بناءً على الميزات المحددة. يتمثل الجانب المبتكر في هذه الطريقة في أنه ، على عكس الطرق الأخرى الموجودة ، فإن ACO قادرة على تحسين ليس فقط الميزات ولكن أيضًا معلمات المصنف من أجل الحصول على أعلى دقة تصنيف. تم اختبار الطريقة المقترحة على مجموعة بيانات حقيقية لظروف تشغيل مختلفة تتكون من ستة علب تروس مختلفة. من أجل إثبات أداء طريقتنا ، تمت مقارنتها بالطرق التقليدية الأخرى. تشير النتائج المتحصل عليها إلى متانتها واستقرار دقتها للتعامل مع مشكلة ظروف التشغيل المتغيرة في الكشف عن أعطال علبة التروس وتصنيفها بكفاءة عالية.

الكلمات المفتاحية:

الآلات الدوارة، علب التروس، اكتشاف الأخطاء، استخلاص الميزات، اختيار الميزات، التحسين، التصنيف

Abstract:

Gearboxes are massively utilized in nowadays industries due to their huge importance in power transmission; hence, their defects can heavily affect the machines performance. Therefore, many researchers are working on gearboxes fault detection and classification. However, most of the works are carried out under constant speed conditions, while gears usually operate under varying speed and torque conditions, making the task more challenging. In this work, we propose a new method for gearboxes condition monitoring that is efficiently able to reveal the fault from the vibration signatures under varying operating condition. First, the vibration signal is processed with the Maximal Overlap Discrete Wavelet Packet Transform (MODWPT) to extract the modes. Next, time domain features are calculated from each mode. Then the features set are reduced using the Ant colony optimization algorithm (ACO) by removing the redundant and unimportant parameters that may mislead the classification. Finally, an ensemble learning algorithm Random Forest (RF) is used to train a model able to classify the fault based on the selected features. The innovative aspect about this method is that, unlike other existing methods, ACO is able to optimize not only the features but also the parameters of the classifier in order to obtain the highest classification accuracy. The proposed method was tested on varying operating condition real dataset consisting of six different gearboxes. In the aim to prove the performance of our method, it had been compared to other conventional methods. The obtained results indicate its robustness, and its accuracy stability to handle the varying operating condition issue in gearboxes fault detection and classification with high efficiency.

Keys words: Rotary machines, gearboxes, fault detection, feature extraction, features selection, optimization, classification

Table of contents

List of Figure

List of table

Acronyms

ACO	Ant Colony Optimization
CTL	Chipped Tooth in Length
CTW	Chipped Tooth in Width
CWT	Continuous Wavelet Transform
DBN	Deep Belief Network
DWPT	Discrete Wavelet Packet Transform
DWT	Discrete Wavelet Transform
EEMD	Ensemble Empirical Mode Decomposition
EMD	Empirical Mode Decomposition
EWT	Empirical Wavelet Transform
FCF	Fault Characteristic frequency
FLS	Fuzzy Logic System
FFT	Fast Fourier Transform
FM	Frequency Modulated
G	Good
GSW	General Surface Wear
HHT	Hilbert-Huang Transform
HLMD	Hilbert Local Mean Decomposition
HMODWPT	Hilbert Maximal Overlap Discrete Wavelet Packet Transform
HT	Hilbert Transform
Hz	Hertz
ICA	Independent Component Analysis
LMD	Local Mean Decomposition
MODWPT	Maximal Overlap Discrete Wavelet Packet Transform
MODWT	Maximal Overlap Discrete Wavelet Transform
MPE	Multi-scale Permutation Entropy
MT	Missing Tooth
RMS	Root Mean Square
STD	Standard Deviation
STFT	Short-Time Fourier Transform
TRC	Tooth Root Crack
VMD	Variational Mode Decomposition
WPEE	Wavelet Packet Energy Entropy
WPT	Wavelet Packet Transform
WT	Wavelet Transform

General Introduction

The efficiency of rotating machines is affected by faults and degradation occurring on their gearboxes, these defects can be generated by various factors such as poor lubrication, incorrect assembly, corrosion and overloading [1] and they can lead to substantial economic losses and serious risks and dangers on the working staff.

For this reason, gears fault diagnosis has always been a research subject where many techniques and approaches were proposed to address this issue in order to increase the efficiency and reliability of the addressed system.

For fault diagnosis, it has been noticed that the vibration signal has wide use and shows dominance over current, temperature, and acoustic emissions [2]. This is due to the simplicity of its acquisition and the importance of the information it provides concerning the source and the gravity of the fault [3].

Once the fault occurs, impulses periodically appear in the vibration signal whenever the damaged part of gear comes in contact, this repetition is known as Fault Characteristic Frequency (FCF).

In gear fault detection, FCF extraction is applied based on many signal processing techniques, for example in [4], Yang et al propose a precise gearboxes diagnosis method based on multi-feature and BP-AdaBoost,[5] used Radial Basis Particle Filter as an extracted signal denoising technique, to pretreat it for further diagnostic classification,[1] introduced an effective fault component separation method that integrates ensemble empirical mode decomposition (EEMD; an adaptive signal decomposition method in time-frequency domain) with independent component analysis (ICA; a blind source separation technique)

The efficiency of the previous techniques in fault detection is proved at constant speed and torque, but gears generally operate in severe environments under non-stationary working conditions, therefore the FCF changes, which makes the applicability of such methods impossible and non-accurate.

For real diagnosis and under varying operating conditions, several approaches have been developed to diagnose gear faults in recent years. Many research groups exploit a phase reference signal obtained from an encoder or a tachometer to remove speed variation effect [6]. Hun et al [7] introduced a deep belief network (DBN) algorithm for gear fault diagnosis based on wavelet packet energy entropy (WPEE) and multi-scale permutation entropy (MPE).

Zheng and Yang [8], worked on the extraction of gearbox fault feature of wind turbine under variable speed condition using improved adaptive variational mode decomposition (VMD). In the research work published by [9], the decomposition of the initial signal into different modes is done using EWT, then the most significant modes are selected to reconstruct a new signal relying on Kurtosis, in the next step, time domain features are extracted; and the Fuzzy Logic System (FLS) is utilized for the classification of bearings faults. This technique showed a high performance in fault classification. However, for EWT decomposition, an inappropriate selection of the modes number may lead to disagreeable decomposition results. Moreover, the linearity of the wavelet filtering bandwidth makes it non-adaptive for all the cases [10]. Furthermore, many condition monitoring and fault diagnosis works preferred DWT as a signal decomposition tool [12], where the analyzed data is decomposed with a band-pass filter in time and frequency domains into a collection of signals with a particular frequency band [13]. Unfortunately, the dyadic step in the down-sampling process seems to be the main limit of DWT, this issue is addressed using the Maximal overlap discrete wavelet transform (MODWT) as an optimized version of DWT to overcome the down-sampling process[14] [15], like DWT, this technique also has poor frequency resolution, to have a better resolution, we use the maximum overlapping discrete wavelet transformation (MODWPT), in this method, the complicated signal is broken down into several simple components and uniform frequency bandwidths are provided, this allows the reconstruction of the original signal and maintain the necessary information.

Nevertheless, for the purpose of reducing the number of classification process entries and in order not to lose important information by excluding parameter-based modes, we will use an optimization to select the modes to retain and the modes to exclude. In some studies, optimization is found to be used as a main step in condition monitoring considering the improvement it imports to the classification performance.

In this work, a new method is proposed to diagnose and monitor gears in various operating conditions. The innovative aspect about this method is that, unlike other existing methods, ACO is able to optimize not only the features but also the parameters of the classifier in order to obtain the highest classification accuracy.

The structure of this thesis is organized into five chapters as follows:

In Chapter 1, we present a general overview of maintenance and different failures of gear boxes.

In Chapter 2, we present the theoretical foundations of signals process and different statistical techniques.
In Chapter 3, we present overview about artificial intelligence techniques and its applications
In Chapter 4, we present the features selection and the optimization algorithm.
In the last Chapter, we present the different experiences about diagnostic and classification of the faults.
Finally, we will end this work with a conclusion

.

Chapter 1

Basic principle of maintenance

I.1. Introduction

Early detection of potential anomalies in the industrial system is critical to good performance in order to avoid minor deviations from nominal conditions and achieve desired functionality. Proper maintenance is an important factor for good performance avoiding small deviations from nominal conditions and corresponding to the expected energy yield, As well as avoiding the loss of equipment, and in some cases even endangering the safety of workers or users.

In this chapter, we will introduce the terminology of industrial maintenance and its different types and we will also present the different types of gearbox system failures.

I.2. Definition

I.2.1. Fault

A fault is a running anomaly in the machine it is any deviation between the theoretical value and the measured value, this deviation is preferably zero with the absence of fault; it can seem in sensors, actuators or inside the machine itself; however, a fault does not always result in a failure [18].

I.2.2. Degradation

Degradation represents a loss of performance of one or several of the functions performed by one or several equipment (if the performance is below the shutdown threshold defined in the functional specifications, there is no more degradation but failure). [19]

I.2.3. Failure

Said failure when a fault appears, the state of the system no longer corresponds to normal operation; which means there is a gap between the measured characteristic and the system's theoretical characteristic, that is, the system is no longer able to fully perform its functions [19].

I.2.4. Breakdown

A failure is the inability of an entity (component or system) to perform a required function, the difference between Breakdown and failure is that the failure corresponds to an event and the Breakdown to a state. In time, the failure is a date and the Breakdown is a time between the occurrence of the failure and the repair end date [19].

I.3. Maintenance

Maintenance is defined as a set of technical actions, administrative and management services making it possible to maintain or restore an asset to a specific state or able to provide a specific service [AFNOR X60-010 standard]. Maintenance therefore means carrying out operations that preserve the potential of the material to ensure production with efficiency and quality. In [19], the maintenance function is presented as a set of activities grouped into two subsets: predominantly technical activities and predominantly managerial activities (see Figure 1).

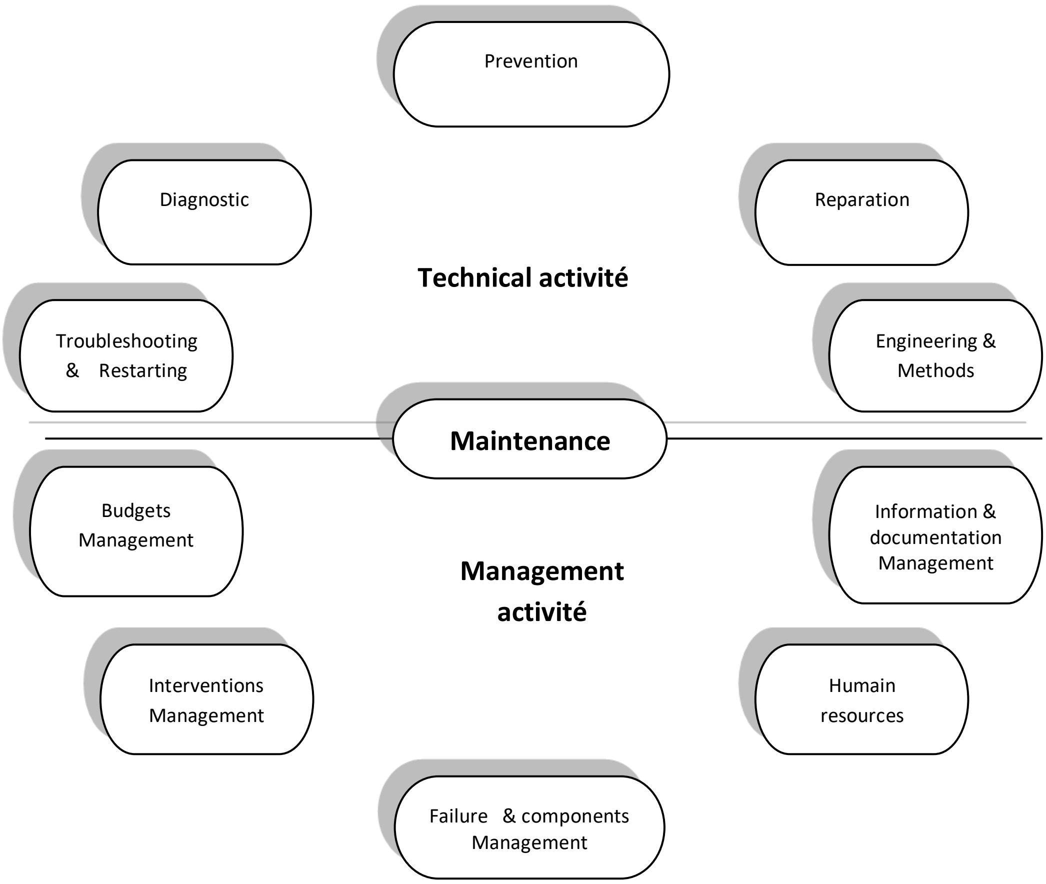


Figure 1: Maintenance function diagram

I.3.1. Maintenance types

The maintenance divided into two main categories [20]:

- The preventive maintenance.
- The corrective maintenance.

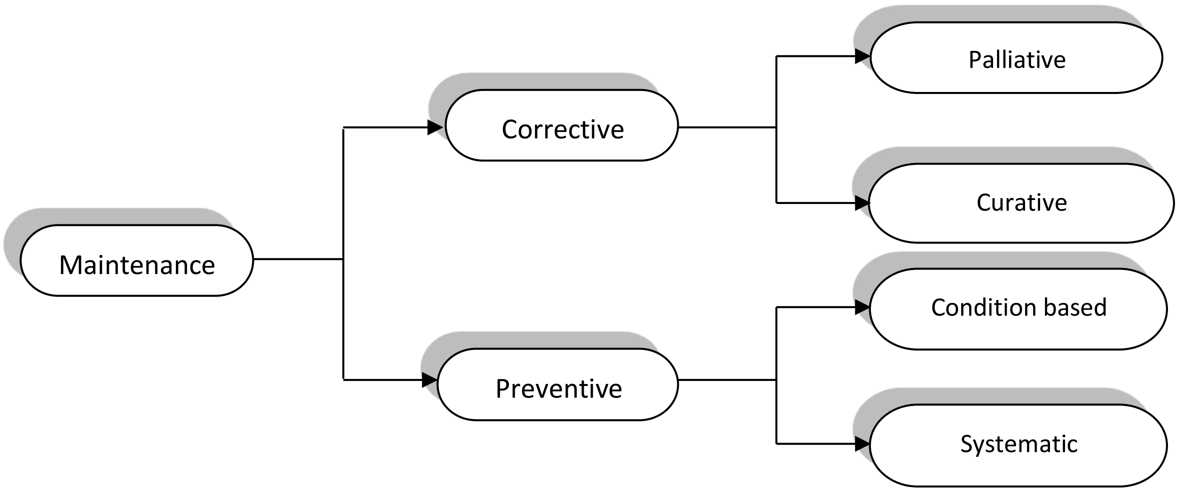


Figure 2: Maintenance types

I.3.1.1. Corrective maintenance

It is a maintenance that applied after the breakdown; a set of activities performed after a failure is detected; it's intended to return the machine or system to a working state. Thus, the application of corrective maintenance actions comes in direct response to the occurrence of a failure [20].

This corrective maintenance can be:

- Palliative.
- Curative.

I.3.1.1.1. Palliative maintenance

A set of corrective maintenance activities intended to allow an asset to perform temporary a function or part of a function, it is commonly referred to as troubleshooting [20].

I.3.1.1.2. Curative maintenance

A set of corrective maintenance activities designed to restore an asset to a specified state or to enable it to perform a required function, the results of the activities carried out must be permanent [20].

I.3.1.2. Preventive maintenance

Preventive maintenance carried out at predetermined intervals or according to prescribed criteria and intended to reduce the probability of failure or degradation of the system functioning, and carried out before the detection of any failure [20]. The service Continuity thus ensured. This maintenance type allows:

- ✓ The decrease in downtime due to breakdowns.
- ✓ Increasing the equipment lifetime.
- ✓ Reduction or even cancellation failures service.
- ✓ Elimination of accidents causes due to troubleshooting.
- ✓ The decision to revision operations at work stoppages time.

Preventive maintenance techniques can be:

- Systematic.
- Condition-based.

I.3.1.2.1. Systematic preventive maintenance

According to the standard (AFNOR X 60-10), Systematic maintenance is carried out at pre-established time intervals or according to a schedule fixed on the basis of the minimum lifetime of machine components [21]. It can also depend on the time elapsed or the number of units produced, but without prior control of the state of the equipment; in other words, the interventions are carried out at dates planned in advance; with a periodicity depending on different wear factors of the equipment (see Figure 3).

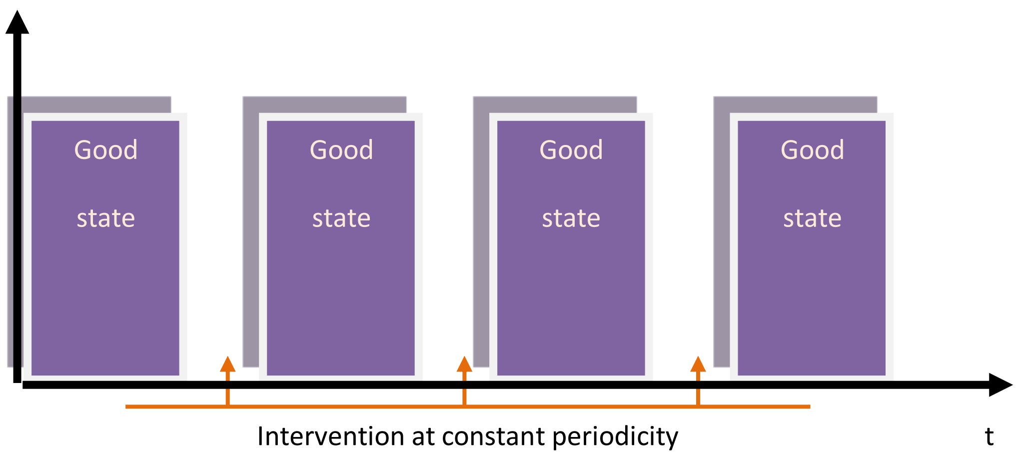


Figure 3: Systematic preventive maintenance.

I.3.1.2.2. Condition-based maintenance

Conditional maintenance is more concerned with the state of the system. Like the previous one, executed as a preventive measure. This type of maintenance relies on process monitoring.

That is, it’s subordinated to a predetermined type of event (Auto diagnostic, sensor information, wear measurement) revealing the degradation state of the system [22]. This definition is illustrated in Figure 4.

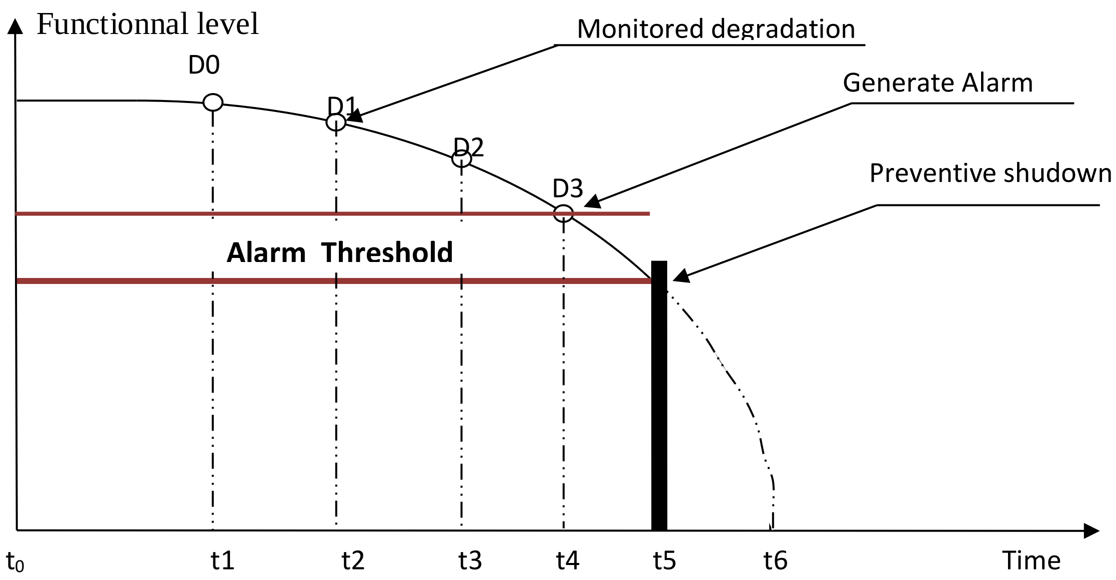


Figure 4: Condition based maintenance

The following diagram will better illustrate the difference between maintenance methods (corrective and preventive):

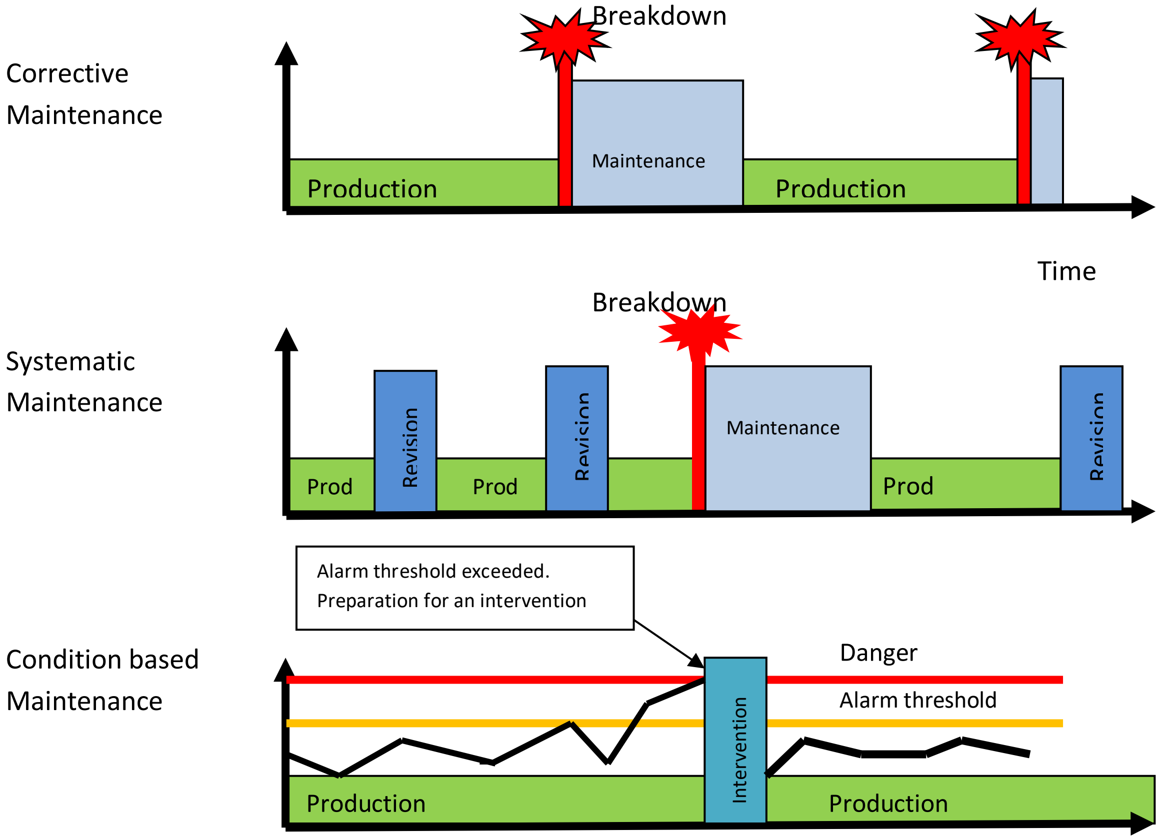


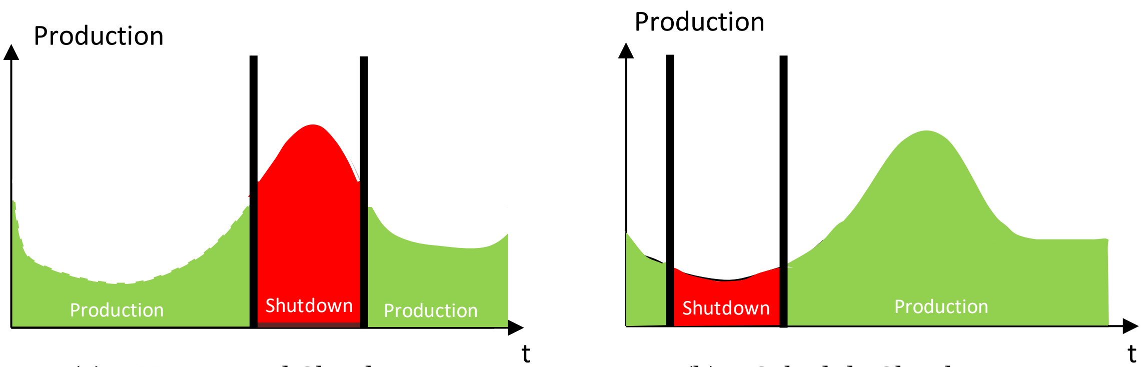
Figure 5: Comparison of different maintenance methods.

I.3.1.2.3. Main advantages of conditional preventive maintenance

When it comes to critical processes and/or complex systems, preventive maintenance is the optimal procedure and among advantages of conditional preventive maintenance [21-22] we find:

- ✓ Repairing machines only when they require it.
- ✓ Increase the availability equipment.
- ✓ Elimination of systematic stoppages.
- ✓ Limiting the repairs severity.
- ✓ More targeted intervention (prior fault location).
- ✓ Minimizing repair costs.
- ✓ Improved safety.
- ✓ Spare parts Supply according to actual needs.

The optimal choice of maintenance intervention time and its impact on production are illustrated in Figure 6.



(a) Unexpected Shutdown (b) Schedule Shutdown

Figure 6: Optimal choice of intervention maintenance time.

I.3.2. How to choose type of maintenance?

It is sometimes economically wise to wait for a breakdown to use curative maintenance with the prior mastery of the latter, while it can be more profitable in other cases to favor prediction and prevention.

For this, you should choose the right choice from the different types of maintenance according to multiple technical, economic, and other internal or external factors [21][23] like:

- The cyclical frequency or random failures equipment.
- Policies and work organization methods.
- The competitive position in the market.
- Products…

The choice of the maintenance implementation in a company is summarized in the diagram of Figure 7 [23][24].

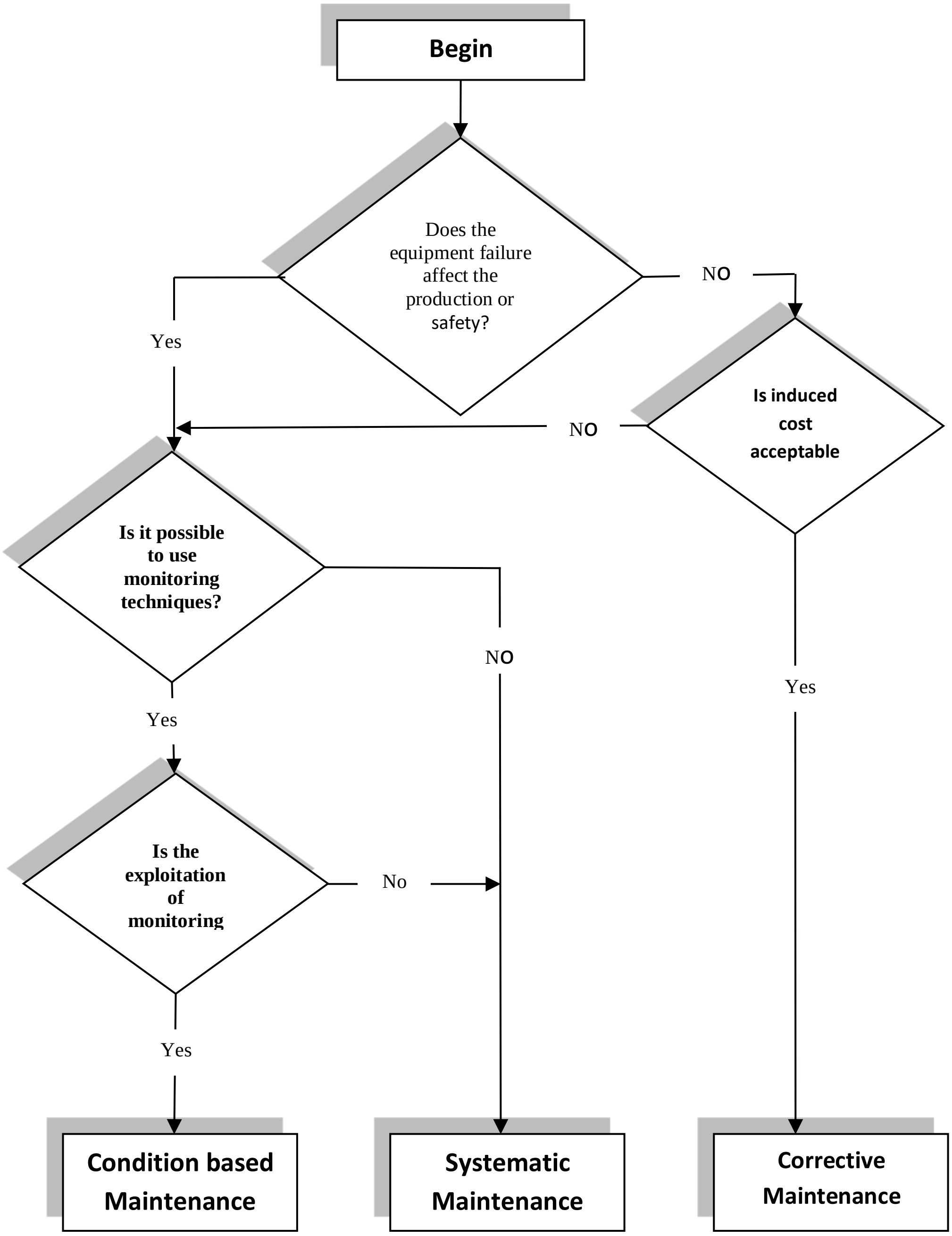


Figure 7: maintenance type choices.

I.4. Monitoring

Monitoring is a conditional preventive maintenance element, and the process that receives information which analyzes the state of the system and gives indications. It works for detecting and classifying failures by observing system evolution and conditions without shutting down the production line, then diagnose them by locating the faulty elements and identifying the root causes. Therefore, the objective at this level is to improve availability and security, as well as the minimization of loss in production [23-24].

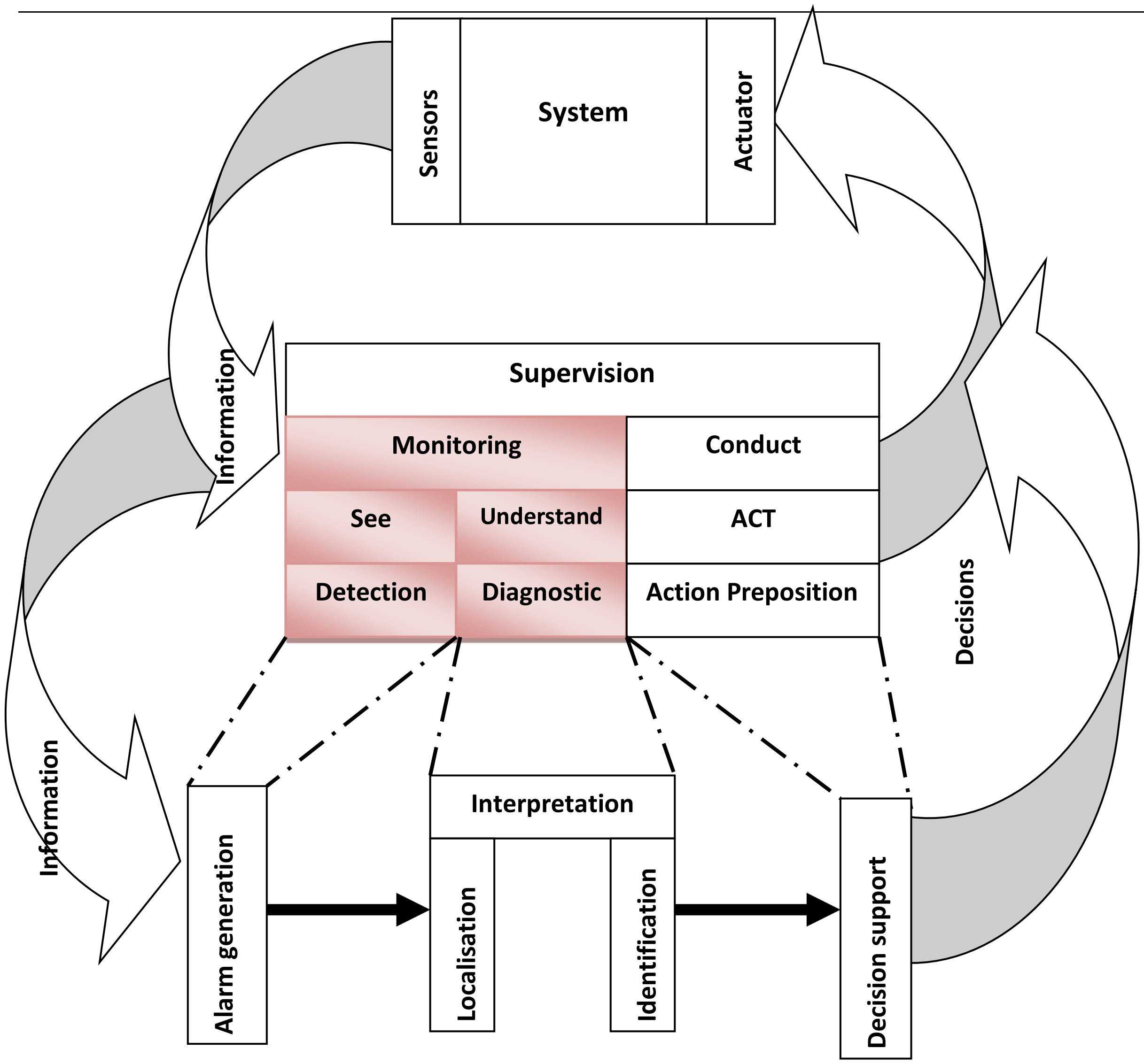


Figure 8: General process of system's supervision.

I.4.1. Monitoring process

The monitoring process consists of two main functions, which are:

I.4.1.1. Detection

The detection is considered as a separate element from the diagnosis and rather sees it is a monitoring entity [25]; it is the level that decides if the system contains a fault or not.

To detect the system's failure, it is not enough to test the no nullity of the residues to decide the appearance of faults, because the measured values are tainted with noise, and the system always insecure to disturbances. Therefore, it must be capable to classify observed situations as normal or abnormal.

I.4.1.2. Diagnosis

Diagnostic is identification of probable causes that led to fault, or its occurring in a system using proper logical reasoning, based on observations gathered on the same system by inspection, by control or by tests, [AFNOR].

This diagnostic consists of:

- **Localization**: The localization task allows determining the place of the detected fault. It represents the process of analyzing events in order to determine where the faulty components of the system are located by means of indications relating to the default element [26].
- **Identification**: The identification of fault is the last phase of the diagnostic procedure; it's the estimation of the fault's amplitude and temporal evolution in order to best explain the system's behavior. Which it allows determining exactly the cause of these symptoms by identifying the nature of the fault [27].

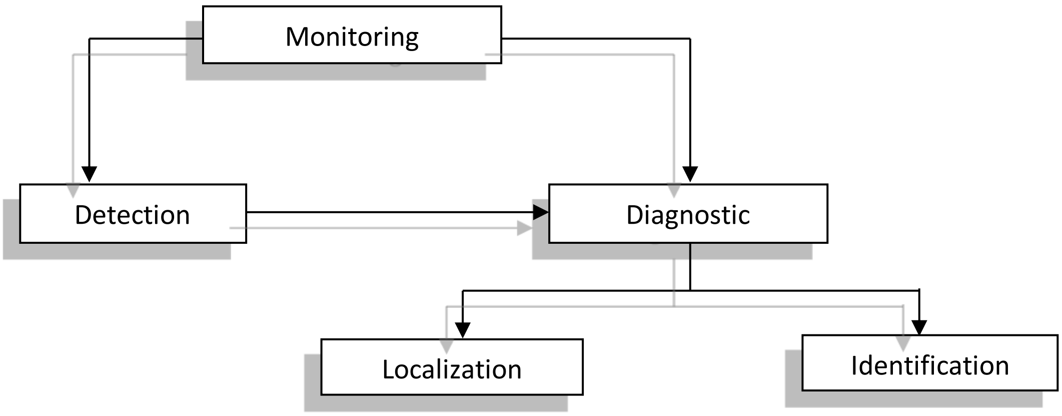


Figure 9: Monitoring process.

I.4.1.2.1. Diagnostic methods

Many methods have been developed to carry out a diagnostic approach, to a best prevent degradation of the monitoring system [28].

These methods are divided into two categories [29]:

✓ **Model-based methods:** which represent methods based on quantitative and/or qualitative models, Such as:

- State observation.
- State estimation.

- Analytical redundancy.

✓ **Model-free methods:** This type of methods is based either on knowledge or empirical and/or signal processing methods.

These methods are developed to be able to study effectively the dynamics of a system, for which the mathematical model is difficult to establish even if non-existent. As based on the analysis of measurements from the instrumentation chains installed on the system grouped as a whole of signatures, under the name 'data base'. which that used in diagnostic procedure by analyzing and classifying a number of characteristics extracted from these signatures, using signal processing and/or artificial intelligence to associating them with a mode of operation of the system [29].

I.4.2. Monitoring Types

We can distinguish two basic types of system monitoring, which requires an on-line monitoring system with diagnostic functions or an off-line monitoring system, such as simple machines.

I.4.2.1. On-line Monitoring or Continuous Monitoring

Online monitoring is used continuously in specific systems, which constantly monitor operating conditions through the continuous measurement of physical parameters, especially those that change rapidly such as temperature, traction, vibration etc. Therefore, immediately send an alarm signal when a sudden change in machine status to the control station [30], then the diagnostic process is carried out to find out the location and identify the cause of the fault.

I.4.2.2. Off-line Monitoring

In this type of monitoring, measurements, samples and checks are carried out at regular intervals. Like infrared thermography or oil analysis, we find this type of monitoring in the evolution of the physical parameter which is rather slow, in such a situation periodic monitoring is necessary. [30].

I.4.3. How to choose the monitoring type?

The choice of the monitoring type depends on the both of system and the failure type that we want to detect; the questions about the choice of the type of monitoring are summarized in Figure 10 [30]

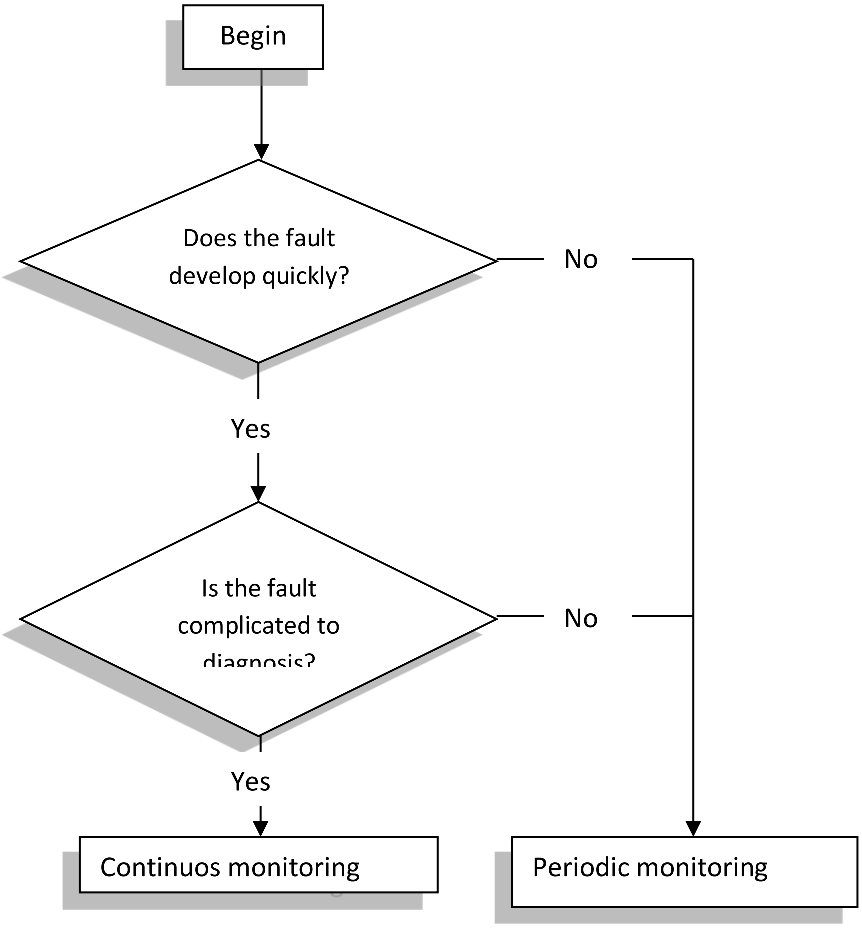


Figure 10: How to choose monitoring types.

I.5. Gears Overview Types and Function

In industry, gears are compact, positive engagement and power transmission elements that determine the speed, torque and direction of rotation of the driven machine elements. Gears are toothed mechanical parts that engage with a second toothed mechanism to change the speed or direction of the transmitted motion. In any gear pair, the smallest part is called the pinion and the largest part is called the immaterial gear. When the pinion is the driver, the output speed decreases and the torque increases, when the gear is the driver the output speed increases and the torque decreases. The speed reduction ratio of the whole unit is equal to the number of pinion teeth divided by the number of gear teeth.

Gears are classified by shape to spur gears, helical gears, bevel gears, worm gears, etc. and by shaft positions as parallel shaft gears, intersecting shaft gears, and non-parallel and non-intersecting shaft gears.

I .5.1. Spur gear

Spur gears have a parallel shaft with cylindrical pitch surfaces, they have straight cut teeth as shown in Figure 11; the tooth line is straight and parallel to the shaft axis, spur gears can achieve a high accuracy with relatively easy production processes, the limitation of spur gears is the occurrence of audible noise in high speed applications [31].

Figure 11: Spur gear

I.5.2. Helical gear

Similar to spur gears, helical gears are cylindrical gears with winding tooth lines used for parallel shafts. Their teeth are cut at an angle to the gear front face (Figure 12). Helical gears offer a lower noise level, less oscillation and more strength than spur gears. They also have better teeth meshing than spur gears because the contact starts at one end of the tooth and gradually extends as the gear rotates until both teeth are fully engaged [31]. They are quieter and can transmit higher loads, making them suitable for high speed applications. Helical gears are designed with right and left twisting which requires an opposite hand gears for a meshing pair.

Figure 12: Helical gear

I.5.3. Bevel gear

Bevel gears have a conical shape to transmit power between two perpendicular shafts as shown in Figure 13. Bevel gears are classified into two types according to the shape of their teeth, which are either straight or spiral. Spiral bevel gears have curved and oblique teeth that allow high performance and high-speed applications, while straight bevel gears have straight and tapered teeth. This straight design is used for lower speed applications. However, bevel gears are not suitable for parallel shafts and can be noisy at higher speeds [31].

(a) Spiral bevel gears (b) Straight bevel gears

Figure 13: Bevel gear

I.5.4. Worm gear

A worm gear as shown in Figure 14, is specific gear composition in which a screw (gear in the form of a screw) meshes with a worm gear (similar to a spur gear). The set-up drive is usually used when a high speed reduction is needed between the driving and driven shafts. Worm gears are also used to transmit the power between two non-parallel, non-intersecting shafts. In addition to that, worm gears can be used for high loads and to transfer a rotational motion to a translational motion [31].

Figure 14: Worm gear

I.5.5. Internal gear

Internal gears are spur gears that are turned "inside out”. This means that they have teeth cut inside the cylinders while the outside is retained smooth (Figure 15). This concept is widely used for planetary gear drives to achieve a higher speed reduction ratio. The rotation directions of the internal and the external mesh gears are the same, whereas they are opposite when two external gears are meshing [31].

Figure 15: Internal gear

I.6. Gears Failure

Gear failure may be caused by maintenance errors (e. g. lack of lubrication, incorrect reassembly after maintenance), manufacturing errors (e. g. insufficient strength of materials used in gear construction) or extremely demanding conditions [31], the two most common types of gear malfunctions are distributed and localized, a distributed defect could be corrosion that affects the gear's teeth, conversely, local flaws are more dangerous because although they begin small, they increase rapidly and have a negative and significant impact on the transmission process [32].

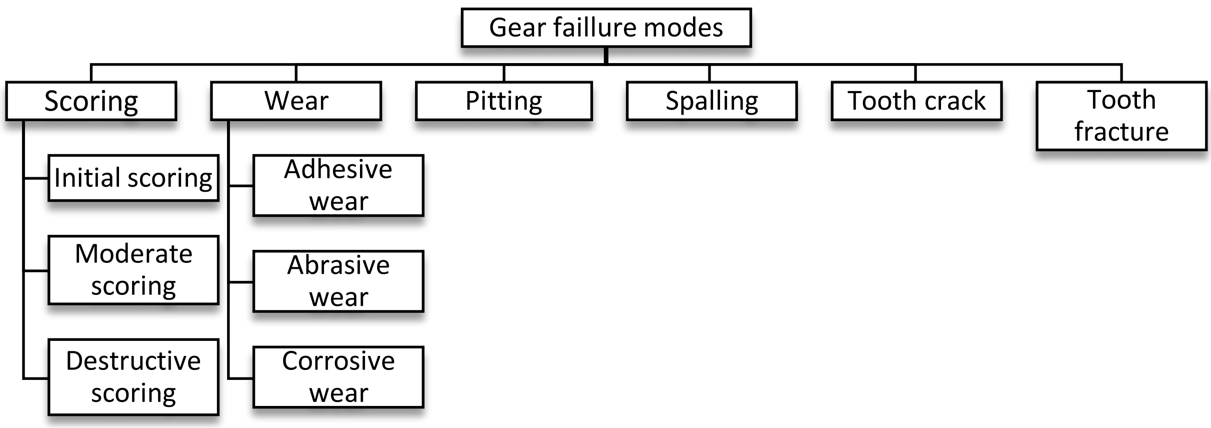


Figure 16: Gear defects classification

I.6.1. Scoring

Scoring occurs for several distinct reasons: inadequate lubrication, misalignment or overload, these factors are the main reasons for the direct metal-to-metal contact; this defect can lead to the welding of the meshing surfaces in addition to the increase in temperature [33]. As scoring progresses, the gear teeth may be distorted or even close to fracture, which reduces the system performance and decreases the gear's service life [33].

This defect can be classified as initial, moderate and destructive.

I.6.1.1. Initial scoring

Initial scoring is the scratches which appear at the teeth highest spots due to the poor lubrication (Figure 17), once these high spots are removed, stress decreases as the load is distributed over a larger area, initial scoring will then stop if the load and speed in addition to the oil temperature remain unchanged or reduced.

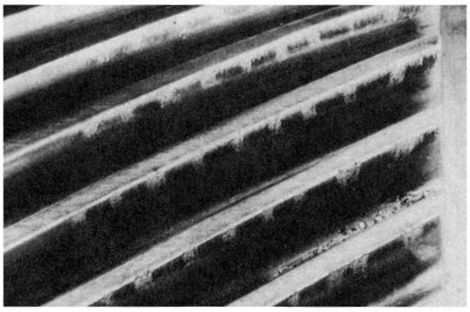


Figure 17: Initial scoring

I.6.1.2. Moderate scoring

If the load, speed or oil temperature increases, the scoring will extend to a larger area as shown in Figure 18; this scoring progress is called moderate scoring.

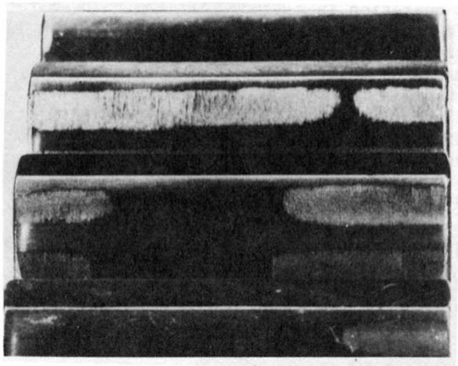


Figure 18: Moderate scoring

I.6.1.3. Destructive scoring

In the case when the load, speed or oil temperature appreciably increases, the torn parts quickly extend, as shown in Figure 19, to become a destructive scoring.

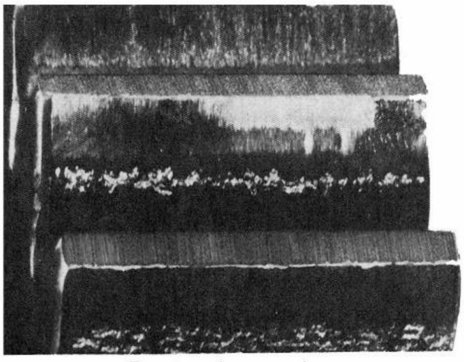


Figure 19: Destructive scoring

I.6.2. Wear

Gear wear could be defined as the progressive removal of metal from the tooth surface [48]; as a result, the tooth becomes thinner and weaker and even close to fracture. The absence of oil film, the ingress of abrasive particles into the oil and the oil composition itself are the three most frequent causes of the wear defect.

Wear defects can be classified as adhesive, abrasive and corrosive wear.

I.6.2.1. Adhesive wear

Adhesive wear can take millions of cycles to be noticeable making it very difficult to detect, adhesive wear is the material transfer from the surface of one tooth to another by tearing and welding (Figure 20) [34].

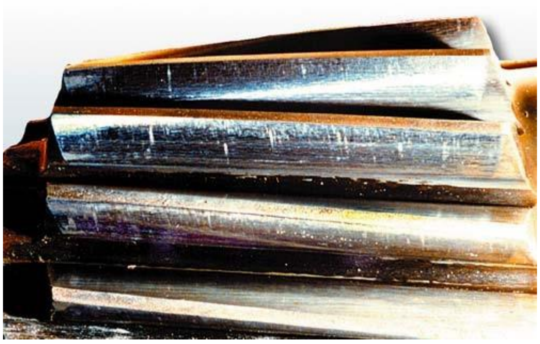


Figure 20: Gear adhesive wear

I.6.2.2. Abrasive wear

Abrasion wear is a consequence of the ingress of abrasive particles into the meshing area, these particles can be cut metal pieces or dust that enters through worn joints [33], the abrasive particles will form a series of grooves on the gear tooth surface in the sliding direction

(Figure 21); in this case, a high wear rate will quickly reduce the tooth thickness, which will certainly lead to a total failure.

Figure 21: Gear abrasive wear

I.6.2.3. Corrosive wear

Corrosive wear occurs when gears operate in an unfavorable chemical environment, such as a contaminated lubricant. Rust is a familiar type of chemical corrosion when water reaches the gear housing. Rust will immediately devour the tooth surface (Figure 22). Corrosive wear is usually associated with an increase of the vibration ratio [32].

Figure 22: Gear corrosive wear

I.6.3. Tooth pitting

Pitting is the formation of craters on the gear tooth surface; it happens when the contact stress exceeds the fatigue tolerance limit under a variable charge; small metal parts are removed in the fatigue area and a cavity is formed on the surface as shown in Figure 23. This cavity will cause a stress concentration on the mesh zone and soon the pitting will extend to the adjacent area until the entire surface is covered [33].

Figure 23: Tooth pitting

I.6.4. Spalling (chipping)

In general, there are two types of surface contact fatigue, pitting and spalling (chipping). Pitting appears as shallow craters on contact surfaces with a maximum cavity depth of ≈10µm, while chipping leaves deeper cavities on the contact surfaces with a depth of 20 µm to 100 µm as shown in the figure 24 [35].

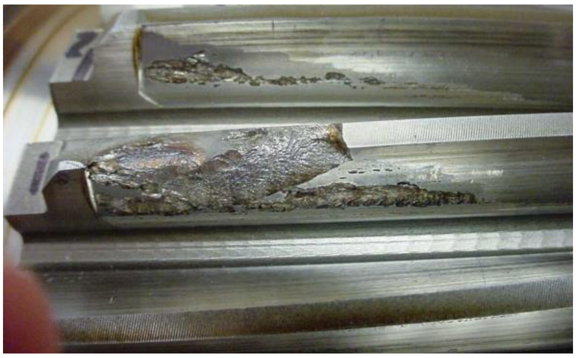


Figure 24: Spalling defect

I.6.5. Tooth crack

There are two categories of cracks, one occurs in the production process and the other from the usage. Microscopic cracks appear in the tooth when plastic deformation occurs in stress areas (notches and inclusions). The microscopic defect grows horizontally to the maximum tensile stress until it causes a sudden tooth rupture.

Figure 25: Tooth crack

I.6.6. Tooth fracture

Tooth fracture is the most dangerous gear failure because, a broken tooth can create extra serious problems with other gearbox parts like bearings, shafts, etc. which will certainly affect the entire transmission system; tooth rupture usually begins with a small crack that extends over the entire tooth [33].

Figure 26: Tooth fracture

I.7. Conclusion

In this chapter, we have presented an overview of industrial maintenance; as well as the monitoring concept with both of process levels and types, and highlighted the concept of diagnosis and its often-used methods, we also introduce different faults which can touch the pinion of the gearbox system.

Vibration diagnostic techniques rely on the vibration signal to detect, locate and diagnose defects. However, signals often contain interference from various components and background noise, emphasizing the critical importance of choosing the right signal decomposition technique.

Chapter 2 Maximum Overlap Discrete Wavelet Package Transform

II.1. Introduction

All the most important industrial systems are equipped with gearboxes, including cars, helicopters and wind turbines …etc. In a power transmission system, a gearbox must operate under different loads and speeds, if the working conditions are improper, the probability of gear failure will increase significantly, which will ultimately have a negative effect on the entire mechanism.

In 1999, statistics were carried out on the causes of accidents in gearbox systems and the result showed that the accidents were directly caused by mechanical failures, the most frequent cause being a defective gear [34], so, it's imperative to detect a gear defect early on so as to avoid any malfunction of the mechanism.

Vibration analysis is still commonly used as a predictive maintenance strategy; vibration signals contain a significant amount of valuable information that can provide insight into the status of an operating machine, when the shaft is connected to the gearbox, frictional and rotational forces are generated, the vibrations caused by these forces are transferred to the gearbox housing via the bearings.

Under typical operating conditions, each individual machine component generates its own relatively stable vibration patterns that are referred to as signatures of vibration [35], these signatures are taken as a benchmark to detect any increase in vibration which is most likely the result of gear failure.

To detect, locate and diagnose the affected component, vibration monitoring techniques use the altered vibration signature of the affected component, however, machine signals are often corrupted by contributions from several different parts in addition to background noise, and as a result, the selection of an appropriate technique is very important.

In this chapter we will present the MODWPT technique which we used to decompose the captured data.

II.2. Vibration Signal Data Processing and Feature Extraction

Many techniques are applied to record and process vibration signals for fault detection and diagnosis of rotating machine, the most popular include frequency domain and time domain in addition to time-frequency domain.

II.2.1. Time domain analysis

Time domain methods using vibration signals are the most conventional approaches in the rotating machine condition monitoring, vibration signal amplitude and energy distribution generally change in the event of a gear or bearing failure, hence, time domain techniques are applied directly to time data to analyse those impulsive patterns caused by the defective part. Some of the most commonly used statistical indicators in the time domain are as follows:

II.2.1.1. Mean (Average)

The mean μ is a statistician's jargon for the average value of a data signal; it is the sum of all data samples divided by the number of entries.

Mathematically the mean is defined as:

$$\mu = \frac{1}{N}\sum_{i=0}^{N-1} x_i \tag{1}$$

Where $x_i (i = 0, .., N - 1)$ is $\boldsymbol{i^{th}}$ sampling point of the signal $\boldsymbol{x}$ and $\boldsymbol{N}$ is the signal length.

II.2.1.2. Standard deviation (STD)

The STD is a measure used to quantify the variation in a time data set, the standard deviation is calculated by averaging the squares of the differences for each sample with the mean, then, the square root is taken to compensate for the initial squaring.

The formula for the standard deviation is defined as:

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=0}^{N-1}(x_i - \mu)^2} \tag{2}$$

II.2.1.3. Root mean square (RMS)

The root mean square (RMS) is described as the square root of the sum average of the squared signal samples.

Mathematically RMS is given by Eq. (3).

$$RMS = \sqrt{\frac{1}{N}\sum_{i=0}^{N-1}(x_i)^2} \tag{3}$$

RMS was designed to reflect the signal energy, only a periodic series of high energy events will increase the overall level of vibration, thus increasing the RMS value.

II.2.1.4. Shannon entropy

Entropy is originally implemented by Shannon to provide a useful criterion for analysing the complexity and uncertainty of a probability distribution [34][35]; it is also used to assess the irregularity and self-similarity of time series [34][35].

For a given time series$x_i\{x_1, x_2, \ldots, x_N\}$, the Shannon's entropy ***H(x)*** can be defined as:

$$H(x) = -\sum_{i=1}^{N} \rho\,(x_i)\, log_2(\,\rho(x_i)) \tag{4}$$

Where ρrepresents the probability density of $\{x_i\}$, and $log_2(\,\rho(x_i))$ represents the binary coding length.

Mathematically, Shannon's entropy is actually the expectation of the shortest average coding length according to the probability distribution of its state. It is the expectation of the information quantity and this can be considered as an indicator to measure the quality of this information.

II.2.1.5. Kurtosis

Kurtosis is widely used in the condition monitoring of mechanical systems such as gears; a defective gear tooth produces a series of peaks during meshing with another tooth; Kurtosis is used as an indicator to measure these impulsive peaks which are generated by the failure of the gear.

Mathematically, Kurtosis is defined as the fourth moment of a given signal x in which the fourth moment is normalized by the square of the variance, this is expressed by Eq. (5).

$$Kurtosis = \frac{\sum_{i=1}^{N}(x(t_i)-\mu)^4}{\left[\sum_{i=1}^{N}(x(t_i)-\mu)^2\right]^2} \tag{5}$$

II.2.1.6. Crest factor

Crest factor is the ratio between the instantaneous amplitude of the waveform and its RMS value; in other words, the crest factor indicates how extreme the peaks are in the waveform; this feature is used to detect changes in the signal pattern due to impulsive vibration sources caused by a defective gear; the Crest factor expression is given by Eq. (6).

$$CF = \frac{max|x_i|}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i)^2}} \tag{6}$$

II.2.2. Frequency (Spectral) domain analysis

One of the most used tools to detect or identify a fault in the gear is the frequency or spectral analysis; the components of a working machine emit vibration signals composed of frequencies which do not change even if their level changes.

The defective tooth generates a periodic signal with a unique characteristic frequency [36]; and by following the characteristic frequency of the fault, frequency analysis, unlike time analysis, can locate the position of the fault.

Fast Fourier Transform (FFT) is one of the most widely used signal processing methods; FFT transforms the discrete samples of a continuous time series signal into a frequency domain representation; it generates a description of the energy distribution within the signal as a function of frequency; frequency domain analysis makes it possible to isolate a specific frequency that is related to machine components or defects.

To obtain the signal spectrum ***X(f),*** the Fourier transform (FT) is applied:

$$X(f) = \int_{-\infty}^{+\infty} x(t)e^{j2\pi ft}dt \tag{7}$$

The gear fault signature (e. g. cracking, chipping and breaking) is generally manifested by periodic transient pulses when the gearbox shafts rotate, however, FFT cannot handle the non-stationary signals, which are often very frequent in the event of machine part failure.

II.2.3. Time-Frequency domain analysis

Time-frequency domain methods allow a simultaneous vibration signals analysis in both time and frequency domains, vibration signals consist of three important elements: "a sinusoidal component due to time varying loading, a broad-band impulsive component due to impact, and random noise [37]. Time-frequency analysis has been developed for these non-stationary signals, Non-stationary signals are better described by a time-frequency distribution to show the signal's energy distribution over a two-dimensional space (time and frequency).

Several time-frequency analysis methods have been developed to monitor the state of health of rotating machine.

Short-time Fourier transform (STFT), Wigner Ville distribution (WVD), wavelet transform (WT) and Hilbert– Huang transform (HHT) are the most commonly used time-frequency techniques.

II.2.3.1. Short-Time Fourier Transform

Unlike FFT, STFT provides a time and frequency resolution by moving a short time window along the signal to apply the Fourier spectrum [38], the STFT method makes it possible to monitor the frequency content evolution as a function of time, this is a useful technique to detect gear defects by investigating the energy distribution of the signal over a time-frequency domain [38], STFT is defined by Eq. (8).

$$S(f,\tau) = \int_{-\infty}^{+\infty} x(t)\omega(t-\tau)e^{(-j2\pi f t)}dt \tag{8}$$

In which, $\boldsymbol{\omega(t)}$ is the window that moves along the signal and $\boldsymbol{\tau}$ is the time location of this window.

However, STFT provides a constant resolution due to the fixed window size used to analyze the data signal; the window size determines the accuracy of the time, for example, a large window offers a great frequency resolution but a lower time resolution and vice versa.

II.2.3.2. Wavelet Transform (WT)

Wavelet Transform (WT) analysis uses a family of “wavelets” with fixed shape, but they can be shifted and dilated independently over time, which is particularly suitable for non-stationary signals [43][44], WT provides a time-frequency representation through the time window and scale functions, this technique has two benefits: the first is the time and frequency resolutions are independent, making WT useful for describing the local signal behavior, the second is the signal frequency can be analyzed without losing any useful information in the time domain, continuous wavelet transform (CWT) is the most used time-frequency technique to deal with non-stationary vibration signals.

CWT is a classic time-frequency analysis approach which has been developed to overcome the STFT limitation, similar to STFT, the signal is multiplied by a window function and the transformation is calculated separately for different signal elements in the time domain, the STFT window width duration is constant while WT uses a self-adaptive window given by a wavelet function whose duration varies within the frequency level; this is probably the most important advantage in the wavelet transform.

The signal is decomposed into a family of "wavelets" with a fixed shape, but they can be translated and dilated by time, the expression of the CWT is given by Eq. (9).

$$CWT(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(t)\psi^*(\frac{t-b}{a})dt \quad (9)$$

Where, ψ is the mother wavelet, translated by b and dilated by factor a.

II.2.4. Signal Decomposition methods and techniques

Among all the signal decomposition techniques that exist, we can cite:

- Hilbert-Huang Transform.
- Empirical Mode Decomposition (EMD)
- Local Mean Decomposition (LMD)
- Wavelet techniques

In this work we have used the one of the wavelet techniques for the decomposition of the signal.

II.2.4.1 Discrete wavelet transform (DWT)

The Discrete Wavelet Transform (DWT) is an integration of the wavelet transform using the discrete set of wavelet scales and translations, DWT decomposes the signal into a set of mutually orthogonal wavelets, this is the main difference with the CWT, DWT is a very effective tool because it considerably reduces the calculation time, the wide quantity of computation wavelet coefficients is the main limit of CWT.

For a continuous time data **x**, $X = [X_0, X_1, \dots, X_{N-1}]^T$ is the column vector of sampled sequences, and N is a power of **2**, DWT is an orthogonal transformation that operates through recursive filters using the pyramid algorithm given in Figure 27 [43], to perform DWT of the sampled vector, the low-pass filter $\{g_l : l = 0, \dots, L-1\}$ and the high-pass filter $\{h_l : l = 0, \dots, L-1\}$ are used, respectively, as scale and wavelet filters, These even-length filters satisfy:

$$\sum_{l=0}^{L-1} g_l^2 = 1, \sum_{l=0}^{L-1} g_l g_{l+2n} = \sum_{l=-\infty}^{\infty} g_l g_{l+2n} = 0 \quad (10)$$

With ***L*** is the filters width, At the first level (***j=1***), the wavelet coefficients $w_{1,t}$ and the scale coefficients $v_{1,t}$ are determined as follows:

$$w_{1,t} = \sum_{l=0}^{L-1} g_l \, x_{2t+1-l \, mod \, N-1} (t = 0, \ldots, N-1) \tag{11}$$

$$v_{1,t} = \sum_{l=0}^{L-1} h_l \, x_{2t+1-l \, mod \, N-1} (t = 0, \ldots, N-1) \tag{12}$$

The wavelet coefficients $w_{2,t}$ and the scale coefficients $v_{2,t}$ at level ***2*** are calculated from the scale coefficients $v_{1,t}$ at level **1** as follows:

$$w_{2,t} = \sum_{l=0}^{L-1} g_l \, v_{1,2t+1-l \, mod \, N-1} (t = 0, \ldots, N-1) \tag{13}$$

$$v_{2,t} = \sum_{l=0}^{L-1} h_l \, v_{1,2t+1-l \, mod \, N-1} (t = 0, \ldots, N-1) \tag{14}$$

For the level j_{th} the wavelet and scaling coefficients are given by:

$$w_{j,t} = \sum_{l=0}^{L-1} g_l \, v_{j-1,2t+1-l \, mod \, N-1} (t = 0, \ldots, N-1) \tag{15}$$

$$v_{j,t} = \sum_{l=0}^{L-1} h_l \, v_{j-1,2t+1-l \, mod \, N-1} (t = 0, \ldots, N-1) \tag{16}$$

where mod means the modulus after division.

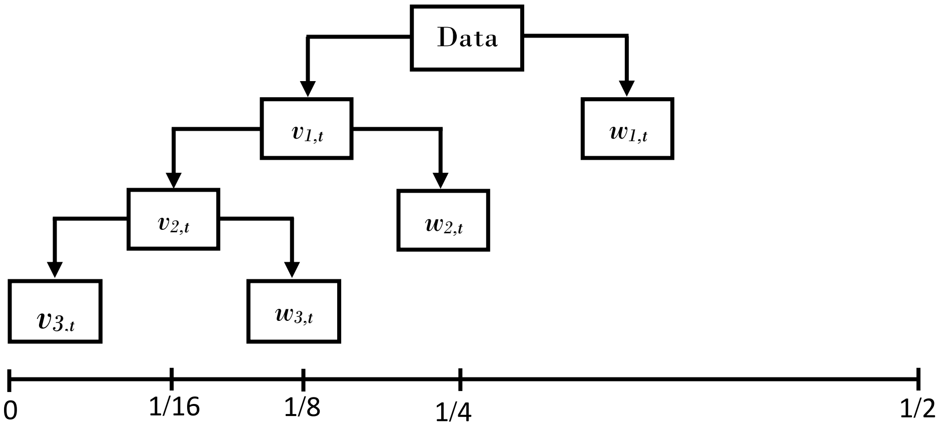


Figure 27: Decomposition process using DWT

II.2.4.2. Maximal overlap discrete wavelet transform (MODWT)

MODWT has several advantages over DWT, for example, MODWT can be properly defined for any sample sizes, while DWT is limited to a signal length with an integer multiple of a power of two (dyadic step). Unlike the DWT, MODWT does not decimate the coefficients and thus the number of scale and wavelet coefficients at each level is equal to the number of sample. MODWT is also called non-decimated DWT.

To avoid the down-sampling problem, MODWT uses new filters by insuring $2^{j-1}-1$ zeros between the elements of $\tilde{h}_l$ and $\tilde{g}_l$, the pyramid algorithm of MODWT produces the scaling coefficients$\{v_{j,t}^{M}\}$and the MODWT wavelet coefficients $\{w_{j,t}^{M}\}$as it is defined in:

$$w_{1,t} = \sum_{l=0}^{l-1} \tilde{g}_l \, x_{t-l \, mod \, N-1} \tag{17}$$

$$v_{1,t} = \sum_{l=0}^{l-1} \tilde{h}_l \, x_{t-l \, mod \, N-1} \tag{18}$$

$$w_{j,t} = \sum_{l=0}^{L-1} \tilde{g}_l \, v_{j-1,t-l \, mod \, N-1} \tag{19}$$

$$v_{j,t} = \sum_{l=0}^{L-1} \tilde{h}_l \, v_{j-1,t-l \, mod \, N-1} \tag{20}$$

The wavelet and scaling filters ($\tilde{g}_l$ and $\tilde{h}_l$) are rescaled as:

$$\tilde{g}_l = \frac{g_l}{\sqrt{2}} \tag{21}$$

$$\tilde{h}_l = \frac{h_l}{\sqrt{2}} \tag{22}$$

II.2.4.3. Maximal overlap discrete wavelet packet transform (MODWPT)

MODWT is an optimized version of the DWT to overcome the down-sampling process, but it still suffers from poor a frequency resolution like the DWT. The maximal overlap discrete wavelet packet transform (MODWPT) replaces MODWT and DWT for a better resolution. MODWPT provides a uniform frequency bandwidth and overcomes the time-varying transformation and also allows the reconstruction of the original signal without losing any information. MODWPT is non-orthogonal without the down sampling step and supports any sample size. In the MODWPT, both the scale and wavelet coefficients are subjected to high-pass and low-pass filtration when calculating the scale and wavelet coefficients of the next level. In standard transformations, the scale coefficients describe the frequency band ***[0, 1/2^{j+1}]***, while the wavelet coefficients describe the frequency band ***[1/2^{j+1}, 1/2^{j}]*** at level ***j***. On the other hand, the MODWPT partition the entire frequency band into frequency bands with the same length. For example, at a given level j, we have 2^jseparated portion with an equal length [43]. For perfect resolution at high frequencies, MODWPT is introduced and $v_{j,n}= \{v_{j,n,t}, t= 0,\ldots,N-1\}$ is the sequence of MODWPT coefficients at level ***j*** and frequency-index **n**. To produce $\{v_{j,n,t}\}$ we use:

$$v_{j,n,t} = \sum_{l=0}^{l-1} \tilde{f}_{n,l}\, v_{j-1,[n/2],(t-2^{j-1}l)\, mod\, N} \qquad (23)$$

With $\tilde{f}_{n,l} = \tilde{g}_l$ when nmod4=0 or 3, while $\tilde{f}_{n,l} = \tilde{h}_l$ when nmod4=1 or 2.

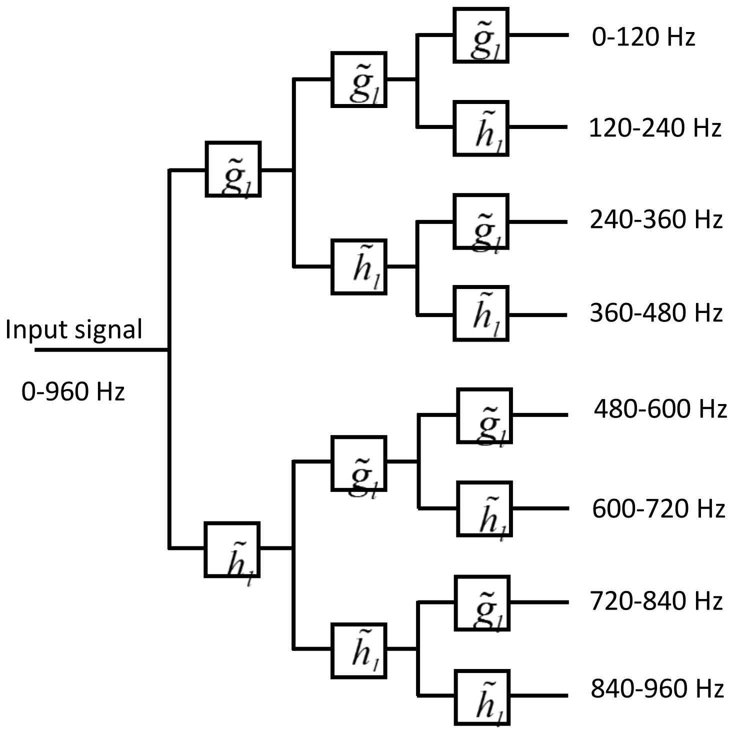


Figure 28: Three levels of the MODWPT decomposition

II.3. Conclusion

In this chapter, the theoretical foundation for traditional gear defect diagnosis methods is expounded upon. Initially, a succinct overview of conventional techniques for processing time, frequency, and time-frequency data was provided, along with a discussion of the advantages and disadvantages of each approach. Additionally, several frequently employed decomposition methods, as well as dimensional reduction and artificial intelligence techniques, were presented. Finally, a state-of-the-art analysis was included, which showcased various signal processing algorithms to be utilized in subsequent chapters.

Chapter 3 **Artificial intelligence and Machine Learning**

III.1. Introduction

The world witnessed a considerable development in the field of technology and computer science in the late twentieth century, where it actually started to rely on a new approach of intelligent algorithms based on artificial intelligence, which made possible the creation of intelligent machines based on natural intelligence simulation, which allows them to learn, make decisions, and predict. Artificial intelligence was and still used in many domains such as monitoring and diagnosing industrial (machines), that is what contributed to the development of the new industrial revolution called industrial revolution 0.4.

In this chapter, we will define artificial intelligence and its applications, and highlight one of its branches called machine learning, which we rely in our work.

III.2. Artificial intelligence

The AI is the study of computations, which make a machine capable of, perceive reason and act through the simulation of human intelligence in machines that are programmed to think like humans and mimic their actions, the term may also be applied to any machine that exhibits traits associated with a human mind such as learning and problem-solving [46].

III.3. Applications of Artificial Intelligence

The applications for artificial intelligence are endless, the technology can be applied to many different fields, including the following:

- **Games**

Artificial intelligence includes computers that play chess, this machine must weigh the consequences of any action it takes, as each action will impact the end result, in chess, the end result is winning the game.

- **Self-driving cars**

Self-driving cars are vehicles capable of driving without a driver on an open road, the computer system must account for all external data and compute it to act in a way that prevents a collision, it based on computer algorithms capable of learning by itself and thus performing tasks so far impossible to accomplish, these devices gradually appear in real traffic.

- **Search engines**

A search engine is an application allowing a user to carry out a local or online search, to be able to work, it's needed artificial intelligence, it is deployed on the web in the form of robots which analyze the different sites for Index, classify them and determine the order of presentation according to the terms of research, for example: Google search engine.

- **Medical**

In medicine, it concerns clinical and fundamental research, hospital practices, medical examinations and hospital care, the AI used in the healthcare industry for dosing drugs and different treatment in patients, and for surgical procedures in the operating room, in addition, help to identify and classify medical images.

- **Industry**

The AI has given a great impulse to industry 4.0, by the intelligence algorithms, the data generated by a factory can be used to increase the efficiency of industrial processes, in conditional based maintenance, AI is used to monitor the production process, and reduce failures, as well as energy consumption and help automate quality control to ensure high production and better planning.

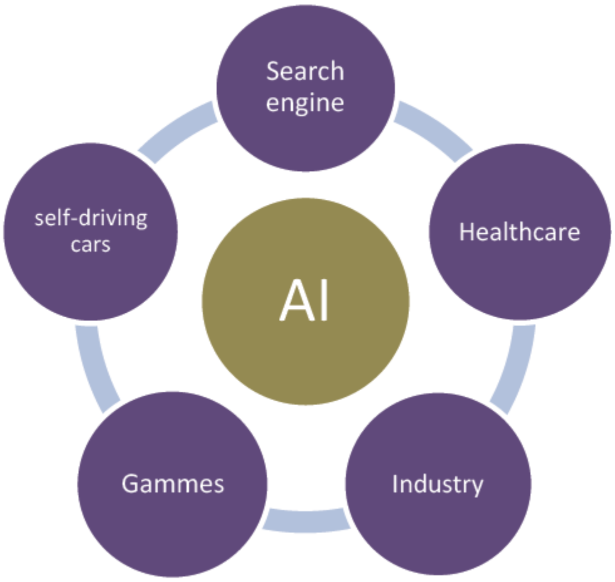


Figure 29: Artificial intelligence fields.

III.4. Machine Learning

Machine Learning (ML) is one of artificial intelligence fields, it's a set of techniques that gives the computer the ability to learn and make decisions for improving their performance tasks and to solve problems without being programmed explicitly [45].

Other definition: According to Arthur Samuel, 1959. « Give machines the ability to learn without explicitly programming them » [46].

III.4.1. Machine Learning types

Machine learning types can be divided into four categories according to their objective, the main categories are as follows [47]:

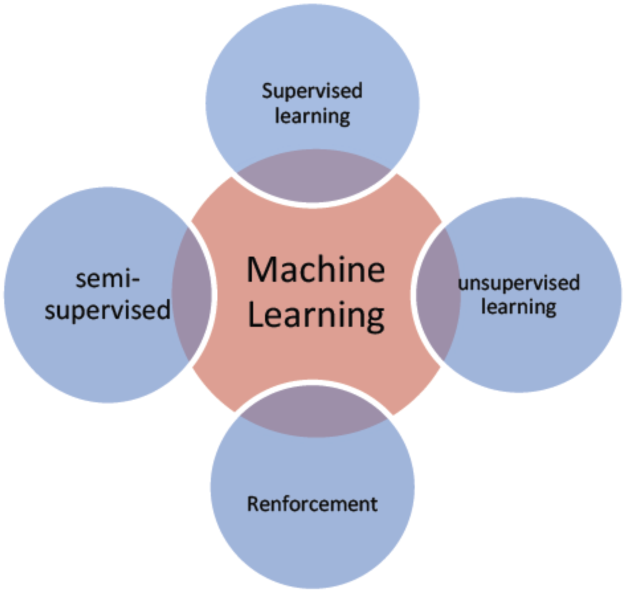


Figure 30: Machine-Learning types.

II.4.1.1. Supervised Learning

Supervised learning is creating a model capable of predicting response values for a new data set, the input data must be categorized into (features/labels) before start training, then algorithms use it to predict results in order to classify a new data later when it's no longer categorized.

A testing data set is often used to validate the model [48], with the Supervised Learning, we can solve two types of problems:

III.4.1.1.1. Classification problem

Classification is used for discrete response values, where data can be divided into specific classes, its objective is to classify the data, which can be identified and categorized according to their categorical features [49].

III.4.1.1.2. Regression problem

Regression is used for continuous response values (continuous variable), from the supervised learning of a variable data model, the machine is supposed to predict the different variations and fluctuations of real data [50].

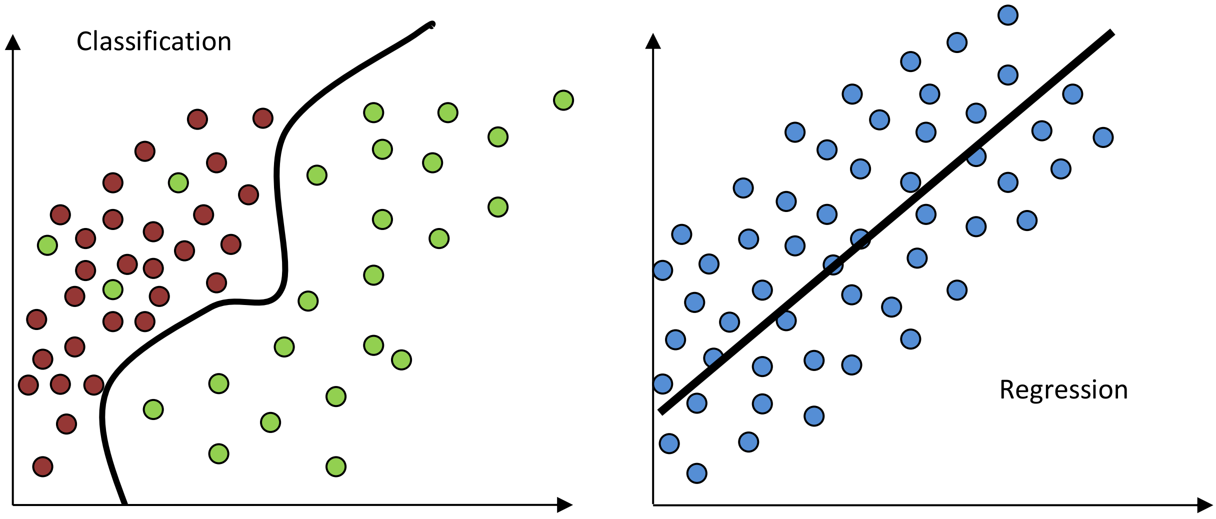


Figure 31: Regression and classification.

III.4.1.2. Unsupervised Learning

Unsupervised Learning involves providing a database to an algorithm that will divide a heterogeneous data group and group them into subgroups so that the most similar data between them is associated within the same homogeneous group. The most common unsupervised learning method is "clustering", which is used to carry out an exploratory analysis of the data in order to find groups in the data, the clusters are created by the means of similarity measure defined by metrics such as the Euclidean distance or the probabilistic distance [52].

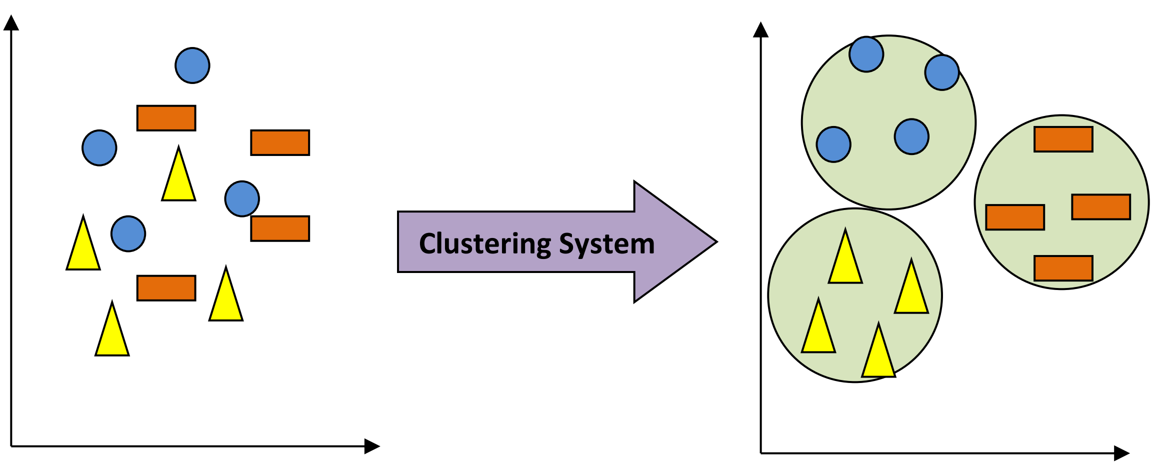


Figure 32: Clustering.

III.4.1.3. Semi-Supervised Learning

The combination of supervised and unsupervised learning produces semi-supervised machine learning, it works with a large amount of unlabeled data and a small amount of labeled data, this gives the advantages of unsupervised and at the same time supervised learning but using a small amount of labeled data, that means you can train a model to label data without having to use as much labeled training data [52].

III.4.1.4. Reinforcement Learning

Reinforcement learning (RL) is part of the science of decision making consists of learning an optimal behavior by interaction and observation with an environment to have the best response, an example of this is a child when he explores his environment and learns the best behavior to achieve his goal.

In this type of learning, artificial intelligence is faced with a situation of uncertainty, the program uses trial and error to find a solution to the problem, to get the machine to do what the programmer wants, the artificial intelligence gets either rewards or penalties for the actions it performs, its goal is to maximize the total reward, this type of learning is often used for the perception and movement of the robot, in game theory and in autonomous vehicles [52].

III.5. Classification Algorithm types

III.5.1. K-Nearest Neighbors

KNN Is a supervised classification algorithm and one of the simplest of non-parametric algorithms, which does not require model construction, and based on entire data [51], each observation of training dataset is represented in n-dimensional space, where ***n*** is the number of predictive variables.

For a new observation that we want to predict the class that belongs to, the algorithm will find the closest k instances of the new observation by calculating the Euclidean distance with the training data, then, it chooses the majority class among its closest k neighbors.

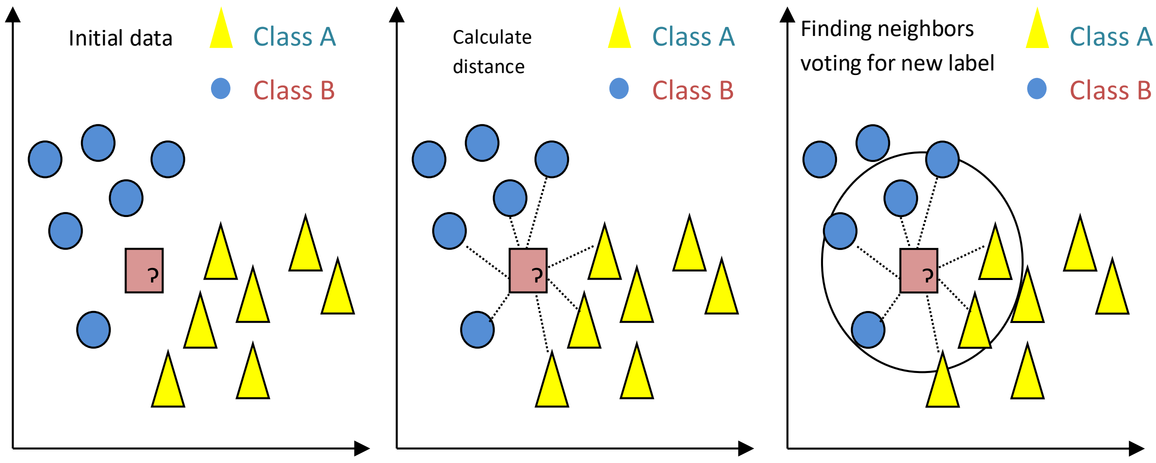


Figure 33: KNN classification algorithm.

III.5.2. Decision tree

Decision tree algorithm is based on a recursive partitioning method, it can be used for both classification and regression, it consisted of a tree structure representation [48], where each node represents a test of individual characteristics and each level represents a class. Decisions are based on conditions on any of the features, the internal nodes represent the conditions and the leaf nodes represent the decision based on the conditions [48].

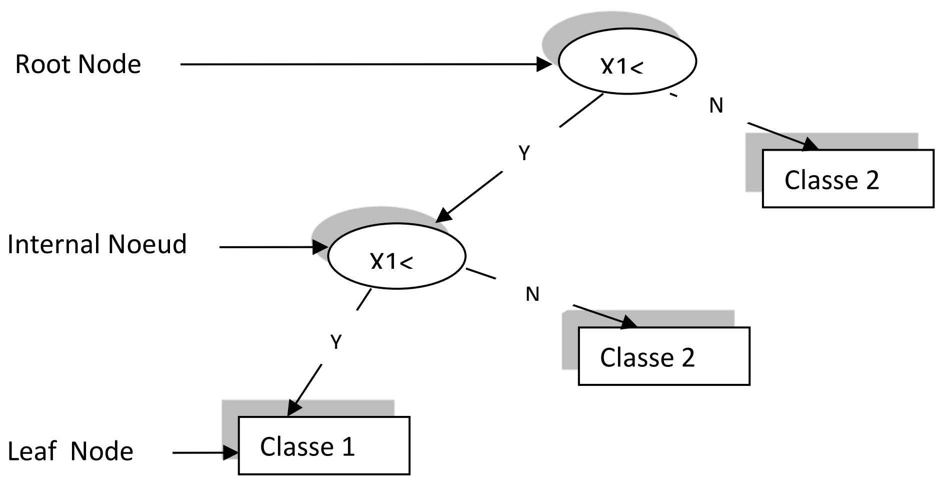


Figure 34: Decision Tree algorithm.

III.5.3. Random Forest

Leo Breiman (2001) proposed random forest algorithm, through the idea of adding a random component, in order to make the trees of aggregation more independent by adding chance in the choice of variables that intervene in models [48].

The random forest consists of several decision trees that work as a whole, each individual tree in the random forest gives a class prediction and the prediction model is given by a vote of all the results obtained by each tree.

(see figure below).

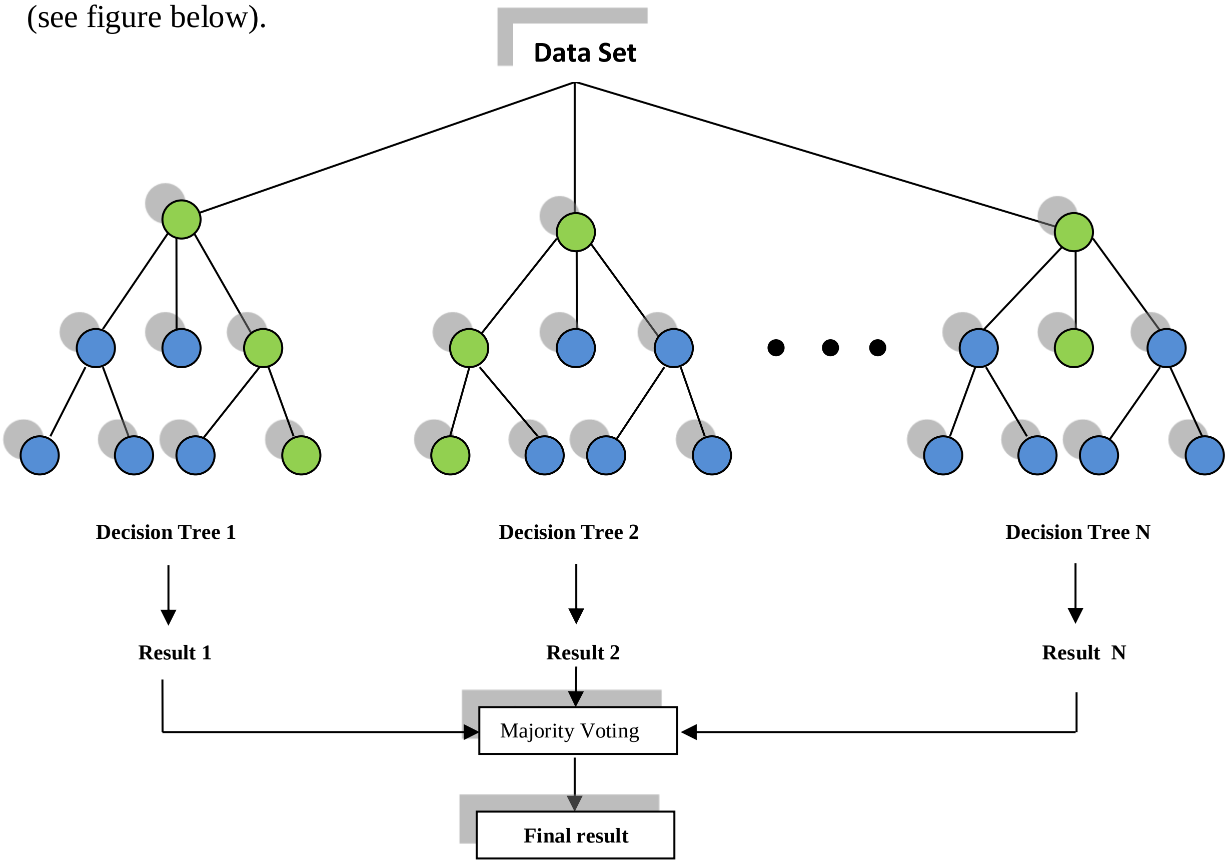


Figure 35: Random Forest algorithm.

III.5.5. Naive Bayes

Naive Bayes is a supervised probabilistic algorithm based on Bayes Theorem, it allows to calculate conditional probabilities providing solutions to different kinds of problems, it assumes that the existence of a characteristic of a class is independent of the existence of other characteristics, this algorithm can predicts membership probabilities for each class such as the probability that given record or data point belongs to a particular class, the class with the highest probability is considered as the most likely class [47].

III.6. Performance evaluation of classification models

There are many performance evaluation tools when it comes to finding a classification model, these tools will be used to the classification in the most appropriate way to choose the perfect classification model.

III.6.1. Confusion matrix

The confusion matrix is a size ***n x n*** table to view the results of predictive models; it is a tool for measuring the performance of classification models to **2** or more classes.

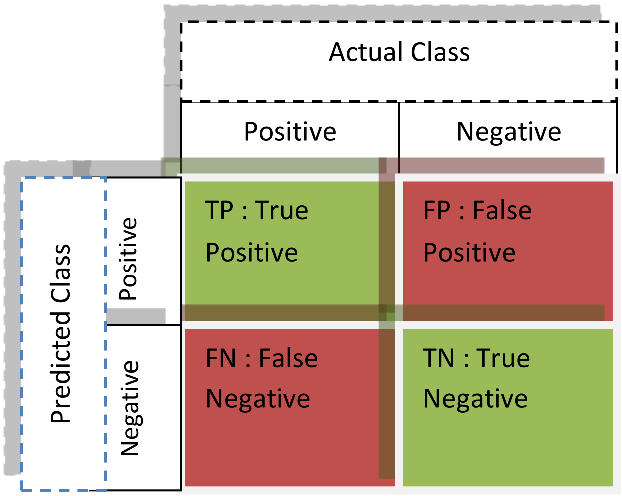


Figure 36: Confusion Matrix.

In confusion matrix of binary classification, we get four possibilities:

- True Positive (TP): is a result when the model most correctly predicts the positive class.
- True Negative (TN): is a result when the model most correctly predicts the *negative* class.
- False positive (FP): is a result when the model most incorrectly predicts the *positive* class.
- False negative (FN): is a result when the model most incorrectly predicts the negative class.

To calculate some measures and evaluate the classification model obtained, we use these four results as follows:

- Accuracy:

This term tells us about the number of good classifications that have been made on all the classification. It can be written as:

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FN+FP}} * 100 \tag{24}$$

- Error

$$\text{Error} = \frac{FP+FN}{FN+FP+TP+TN} * 100 = 1 - \text{Accuracy} \tag{25}$$

- Precision: Out of all that were marked as positive, how many are actually truly positive.

$$\text{Precision} = \frac{TP}{TP+FP} \tag{26}$$

- Recall sensitivity: Out of all the actual real positive cases, how many were identified as positive.

$$\text{Sensitivity} = \frac{TP}{TP+FN} \tag{27}$$

- Specificity: Out of all the real negative cases, how many were identified as negative.

$$\text{Specificity} = \frac{TN}{TN+FP} \tag{28}$$

III.7. Conclusion

In this chapter, we gave an overview about artificial intelligence and we highlighted the field of machine learning with the main of its techniques especially for classification task. In addition, we have seen different evaluation tools or measures with which one can find the best classification model. The notions of optimization and selection of features will be discussed in the next chapter.

Chapter 4

Optimization and features selection

IV.1. Introduction

Machine learning requires large amounts of data for training, which increases the time to calculate during the process, and can make it difficult to achieve optimal performance.
To avoid this problem, it is always preferable to make a selection of discriminant features to choose the best subset of them that lead to obtain the best performance in classification.

In this chapter, we will present the concept of optimization and how to see the problem of selection as an optimization problem, and we will show the general process of features selection based on artificial Ant colony algorithm.

IV.2. Optimization problem

An optimization problem is a formal mathematical model that is used to solve a real problem by seeking to maximize or minimize its objective function under constraints to find the best solution for that given problem. [49].
Optimization problems are generally presented in the following form:

$$\begin{cases} \min f(x) \\ g_i(x) \leq 0 \quad i \in I = \{1,2,..,m\} \\ h_i(x) = 0 \quad i \in I = \{1,2,..,m\} \\ x \in R^n \end{cases} \tag{29}$$

Where ***f*** represents the objective function, ***x*** represents the set of decision variables, the ***g*** and ***h*** functions represent the constraints of inequality and equality related to the problem. There are several types of optimization problems, many classifications exist, including classifications by type of variables (continuous, discreet, and mixed) or by type of problems [49].

IV.3. Optimization applications

Optimization used in many areas:

• **Physics:** The goal of physical optimization is often the minimization of the consumed energy.

• **Finance:** In this field, the major problem is that we want to optimize the costs under certain constraints.

• **Industry :**

- ✓ Optimizing the size and location of electronic components on a circuit currently occupies the attention of researchers.
- ✓ Minimize the falls when cutting elements into a piece of fabric.
- ✓ Manage stocks optimally.
- ✓ Data analysis: The main use of optimization is to pass a curve with certain properties (for example with a positive slope) as close as possible to certain points.

IV.4. Feature selection

Features selection is a very active research topic in different fields, where it can include many applications that requires processing data described by a very large number of features such as machine learning [51].

In this work, we only deal with the selection of attributes made for the supervised classification.

IV.4.1. Feature selection problem

In the presence of hundreds or thousands of features, there is a great chance that the features are correlated and express similar information, that's what make them redundant, on the other hand, the features that provide the most information for classification will be said to be relevant [51].

Feature selection consists of choosing a subset of relevant features from a set of large features by eliminating redundant and irrelevant or noisy features that have little or no influence on the information that we wish to predict.

Generally, the feature selection problems can be defined by:

Suppose ***f = {F1, F2, F3,…. ,N}*** a set of size ***N*** features, where ***N*** represents the total number of features studied, either **EV** a function that makes it possible to assess a subset of features, we assume that the maximum value of EV is obtained for the best subset of features.

$$\mathrm{Ev}(\mathrm{F}') = \max(\,\mathrm{Ev}(\mathrm{Z})) \text{ , } \mathrm{Z} \subset \mathrm{F} \tag{30}$$

Where ***|Z| = N′*** and **N′** is either a predefined number by the user or is on trolled by one of the methods of generation of subsets that we will see them later.

IV.4.2. Feature selection objective

In many areas, the problem-solving system is based on a large set of variables, which can create difficulties at several levels such as, complexity, computation time as well as the deterioration of the solving system in the presence of noisy data.

The goal of selection is to minimize the large database to an optimal subset of variables that consists of relevant variables and it should seek to avoid redundant variables, in addition, this set makes it possible to increase the precision of classification and the calculating time else even the comprehensibility of the proposed classifier [51].

IV.4.3. How can see Selection features as an Optimization Problem

We are going to focus on a phenomenon that is observed when the dimension of the variable space grows so fast that the data it includes become sparse and remote.

The selection of a subset of features can be seen as a search in a space of hypotheses (called a set of possible solutions) [51].

To find an optimal subset of features in an initial set “***X***” of “***n***” features requires to examine $\boldsymbol{2^n-1}$ possible subsets.

The selection process determines a subset of features from the initial set of features that it considers the most relevant, then, this subset is subjected to an evaluation procedure, which makes it possible to evaluate the performance and relevance of the subset, after that, depending on the evaluation procedure result, a criterion for stopping the process that will be noted “***J***” determines whether the subset of variables can be subject to the classification phase, if so, the selection process stops, otherwise another subset of variables is generated.

IV.5. Feature Selection Procedure

The various methods proposed in the literature for attribute selection can be described by a figure below, in which the following key elements can be found:

- A procedure for generating candidate subsets that determines by exploring the search space.
- An evaluation function giving the quality of the candidate subsets.
- Stop criterion.
- A validation process to check if the desired objective is achieved.

We detail below the important steps of this diagram:

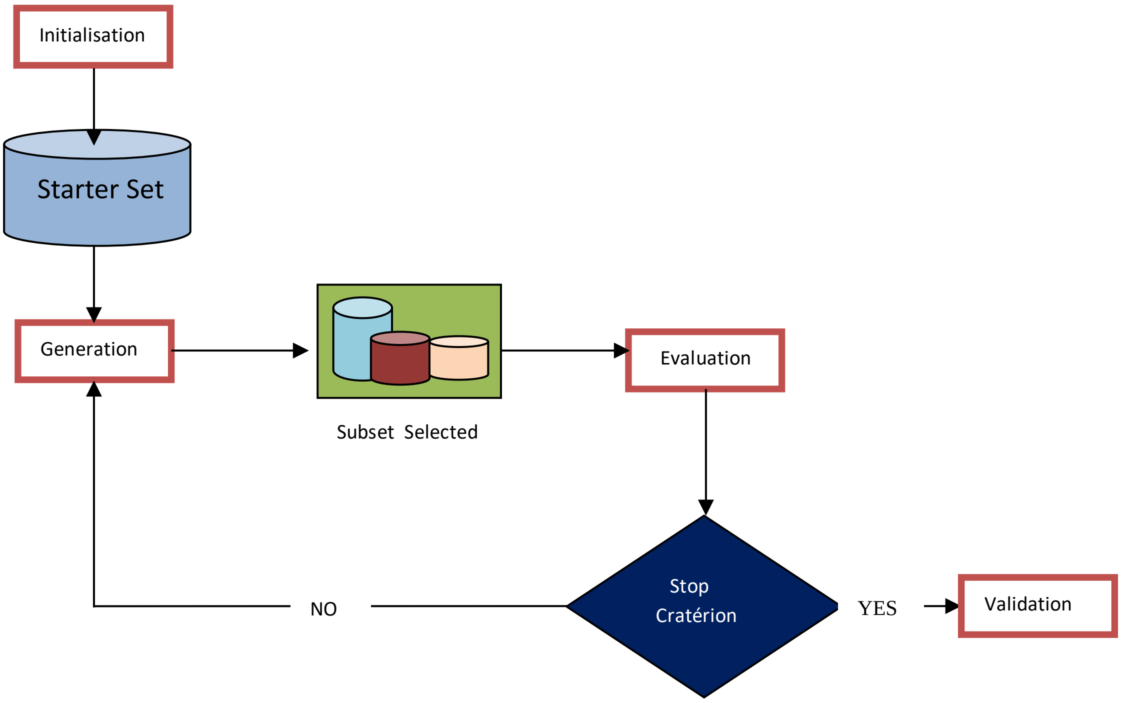


Figure 37: Features Selection procedure.

Before starting to apply the search procedure, it is necessary to define a starting point (or research direction). We can do by three ways:

- We can start with an initialization point in the form of an empty set of features and continue by the successive add to each iteration of one or more features.
- We can start with the whole of all the features and continue with the sequential deletion, with each iteration of the least relevant features.

- Another way is to start research with any subset of features, once the starting point is well chosen; a search procedure must use to generate subsets of features.

Generally, research strategies can be classified into three complete, heuristic and random categories, the figure below gives a general overview of the selection methods of variables based on the search strategy.

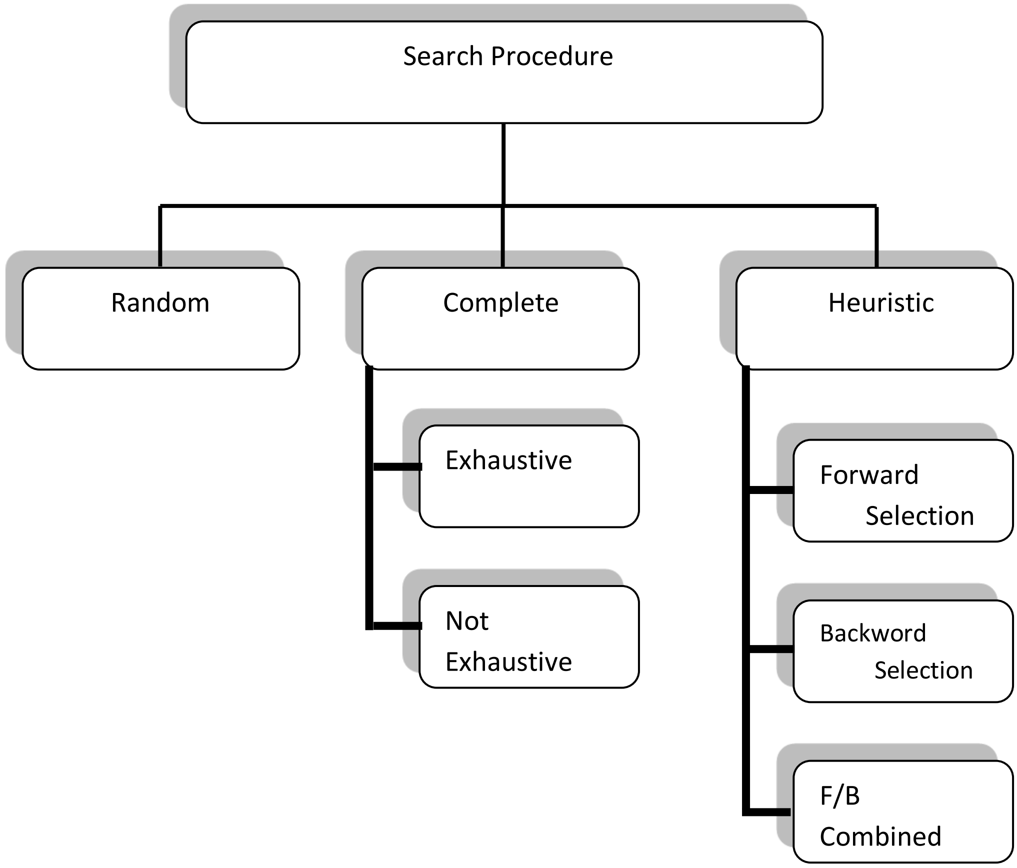


Figure 38: Search procedures.

IV.6. Feature selection methods

The Methods used to evaluate a subset of features in selection algorithms can be classified into three main categories: filter, wrapper, and embedded.

IV.6.1. Filter method

Feature selection is independent of any machine learning algorithm; instead, features are selected based on various statistical tests; the filter method is considered as a preprocessing step (filtering) before the learning phase; it evaluates the relevance of a feature according to measures that are based on the properties of the training data.

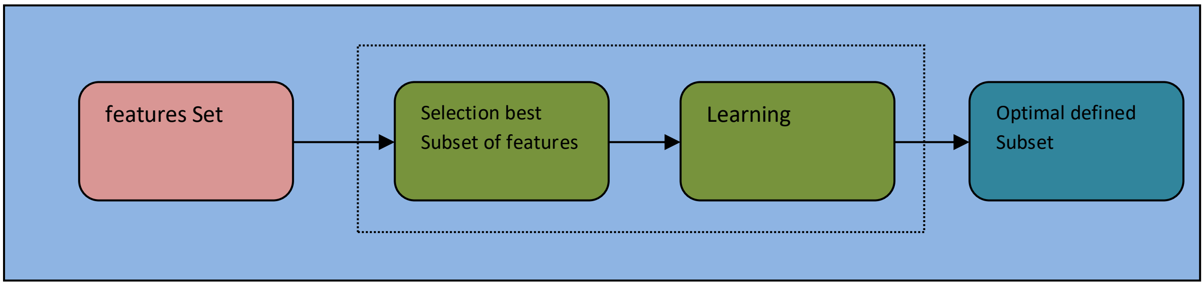


Figure 39: Filter method.

IV.6.2. Wrapper method

Wrapper methods perform a feature subset search, guided by the model result, so that they integrate the machine learning algorithm into the feature selection procedure, use a certain error as an evaluation criterion. These methods provide good performance; they often have better results than other methods [58]. The procedure of the "wrapper" method is illustrated in Figure.

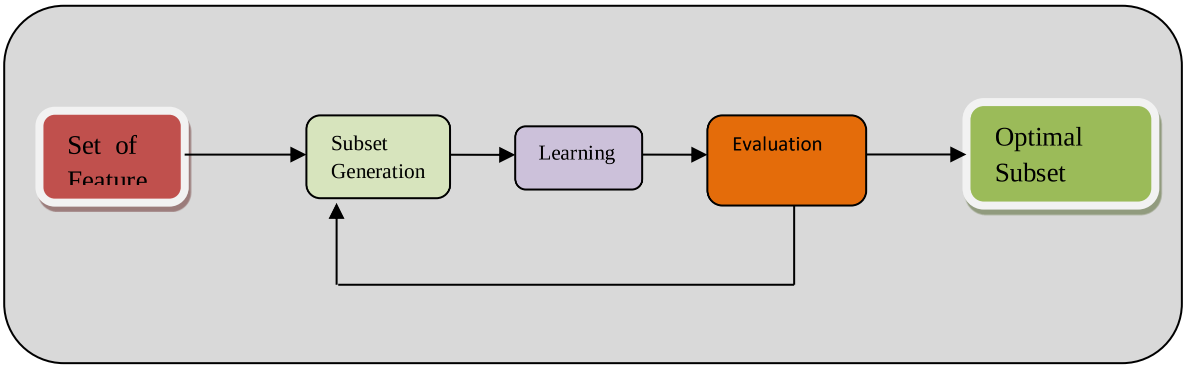


Figure 40: Wrapper method.

IV.6.3. Embedded method

Embedded method combines the two previous types of methods; the main advantage of this technique is that it is faster than the ''Wrapper'' method. The procedure of the ''Embedded'' method is illustrated in the Figure below.

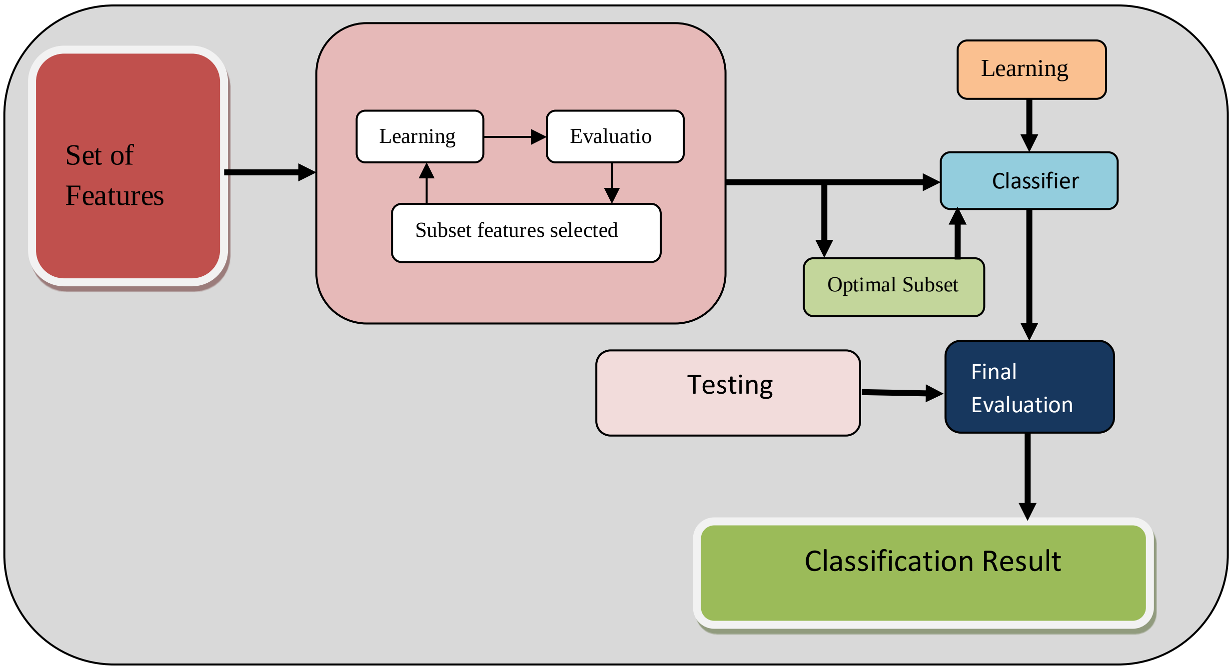


Figure 41: Embedded method [59].

IV.7. Swarm Intelligence

IV.7.1. Main Idea

Suppose you are part of a group that is looking for treasure. You know the approximate location of the treasure, but you don't know exactly where it is.

Each is equipped with a metal detector and can transmit the signal and the current position to the nearest neighbors. Everyone is therefore aware if one of his neighbors is close to the treasure than him. The person who found the treasure receives a higher reward than all the others, and the others are rewarded based on their distance from the treasure at the time the first finds the treasure. By exploiting the information, you will receive from your neighboring friends and acting on it, you increase your chances of finding the treasure. [53].

This is one of the examples of the benefits of cooperation in situations where you do not have global knowledge of an environment. The interact of the individuals with the group can be helpful to solve the global objective by exchanging locally available information.

Figure 42: Ant colonies.

Figure 43: Bird flocks.

Take for example, Insects that live in colonies, ants, bees, wasps, and termites, each insect in a colony has its own agenda, however, insect colonies seem so organized even seamless integrations of all individual activities don't seem require supervisor [52].

We are going to talk about Ant colony and the behavior of the ants when they search the food source, as well as the optimization based on artificial ant colony algorithm.

IV.7.2. Ant Colony Algorithm (ACO)

Among all the algorithms that imitate the behavior of living creatures, the ant colony algorithm is considered an optimization algorithm. The "swarm" is a collection of individuals who indirectly communicate with each other and change the local environment to solve problems through this coordination. The word "intelligence" in this case means that an element which behaves in an intelligent way with respect to the rest of the group due to the coordination of the whole group. The so-called “ants” is considered the basic unit of the swarm and hence algorithm [53].

The communication between the ants is done thanks to a substance which is called pheromone, to find their food ants mark their way by depositing pheromones on the ground. The basic principle is that the other ants go through the path that has more pheromone left by the previous ants, so the pheromone of this path will be reinforced more by their own pheromone. The fact that ants leave their pheromone on this path, then the probability that it will be followed increases progressively, the process can be expressed as a positive feedback loop based on the preceding steps [55]. when the ants arrive at the food source, the ants return on the same path to their nest, once the ants have completed their work and abandon this path, the pheromone dissipate, this is the phenomenon of evaporation.

The illustration of the based positive feedback loop identification of the shortest path is shown in Figure 44 The stronger pheromone trail left by the preceding ants will be chosen as the shortest path compared to the other one ad more ants will reinforce it as explained in Fig44-b.

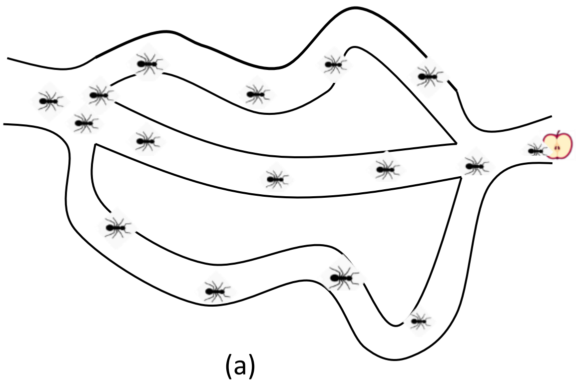


Figure 44: Basic explanation of the ant colony algorithm.

IV.7.2.1. Mathematical Description of ACO

ACO was firstly introduced by M. Dorigo and his colleagues in the early 1990s [55].It was considered a new nature-inspired metaheuristic method intended to solve TSP problems. The TSP problem can be presented as the distance optimization problem when a traveler visits multiple cities and each city is visited only once. The use of ACO in the TSP solution will be described in what follows.

We must consider the update of the pheromone at each round: this includes the amount of pheromone evaporated per unit length.

The vaporization mechanism will allow the ant to abandon the previously selected wrong path. Thus, early local optimization will be eliminated.

ρ is the evaporation coefficient. When the evaporation rate is equal to 1, there is no pheromone evaporation, therefore it is difficult to obtain convergence. But setting ρ too much lower than 1, is likely to get a better local response.

The intensity of pheromone on path-*ij* at time *t+1* is given by Eq. 31 [54]:

$$\tau_{ij}(t+1) = (1-\rho).\tau(t) + \sum_k \Delta\tau_{ij}(t,t+1) \tag{31}$$

Where $\Delta\tau_{ij}^{k}$ is the quantity per unit of length of pheromone laid on edge *(i,j)* by the *k-th* ant between time *t* and *t+1*.

ANT-quantity of deposited pheromone:

$$\Delta\tau_{ij}^{k}(t,t+1) = \begin{cases} \frac{Q}{L^k} \\ 0 \end{cases} \tag{32}$$

Where Q is a constant; d_{ij} is the Euclidean L^k is the distance between i and j; of the k^{th} ant.

The transition probability for the k-th ant from town i to town j as Eq. 33 [55]:

$$\rho_{ji}^{k} = \frac{[\tau]^{\alpha}[n]^{\beta}}{\sum[\tau]^{\alpha}[n]^{\beta}} \tag{33}$$

IV.8. Conclusion

In this chapter, we presented the principles of optimization, and feature selection with these concepts and selection methods. Then, we explained how the problem of selection can be seen as an optimization problem. We also mentioned a brief description on the behavior of ants in the search for food sources, As well as the optimization by the algorithms of ants colony.

In the next chapter, we will present the gearbox test bench on which we will carry out our experiment, and we will discuss the results of our method and compare it with other techniques.

Chapter 5 # Presentation of Results and Discussion

V.1. Introduction

To show the efficiency of the proposed method, we applied it on the gearbox system which is presented in the next section, our technique consists at the first to proceed at the extraction of features from the data acquisition with the MODWPT method, then the features obtained are optimized with the ACO algorithm whose the fitness function is the accuracy of Random Forest classifier, at the end the optimized features are classified with the Random Forest technique, the following Flowchart resume the method.

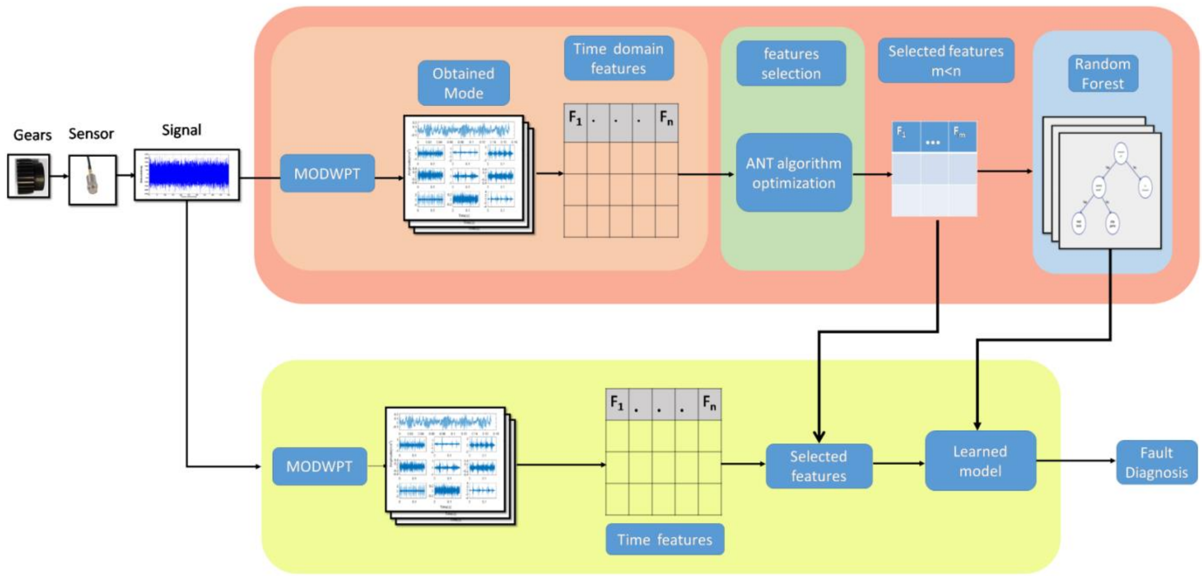


Figure 45: Flowchart of proposed method

V.2. Experimental Description

In this section, a speed reducer with a gear ratio of 25/56 is considered as a test bench (Figure.45). A nominal speed of 3,600 r/min electric dc motor is considered as a source of motion between the two shafts and different resistive torques are generated by a magnetic power brake that is coupled to the output shaft [56] [57].

The efficiency of the suggested method is tested using six pinions with different health states. The first one is a faultless pinion, and it is referred as good (G), while the rest have various types of defects, such as a tooth root crack (TRC), a chipped tooth in length (CTL), a chipped tooth in width (CTW), a missing tooth (MT), and general surface wear (GSW) (Figure.48).[56] [57].

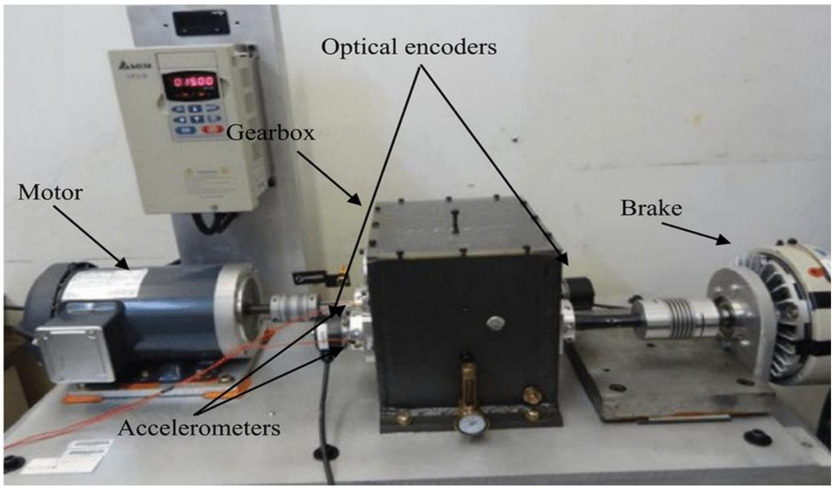


Figure 46: The gearbox test bench.

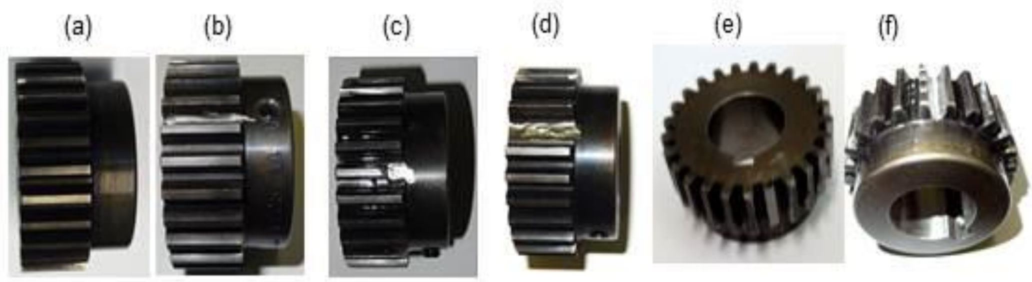


Figure 47: Six pinion states: G (a); TRC (b), CTL (c), CTW (d), MT (e) and GSW (f).

Table 1: Geometrical parameters of the gear system

Parameter	Pinion	Gear
Number of teeth	25	56
Pressure angle,°	14.5	
Face width, mm	19.05	
Diametric pitch	12	
Material	1140 carbon steel	
Outside diameter, mm	57.2	122.7582
Pitch diameter, mm	52.9082	101.6

Three pinions are installed simultaneously on the input shaft of the gearbox; with a simple axial movement of the wheel of the output shaft, the engagement of each of them is achieved (Figure 49). Two accelerometers (sensitivity: 100 mV/g) are radially installed to record vibration signals, in horizontal and vertical positions on the bearing case of the output shaft. The accelerometer channels time sampling frequency is equal to 125 kHz, the sampling frequency of the anti-aliasing filter is 27 kHz, and the acquisition duration is equal to 30 s.

The accelerometer signals have been collected for several operating conditions under different loads and different rotation speeds (see Table 2). Figure.49 shows the acceleration vibration signals recorded from pinions with different gear state for an operating speed equal to 900 r/min with 11-Nm load. From Figure.49, it can be clearly seen that the five different defects do not show a significant signature in the vibration signal. No significant increase in the energy of the temporal signal is noticed.

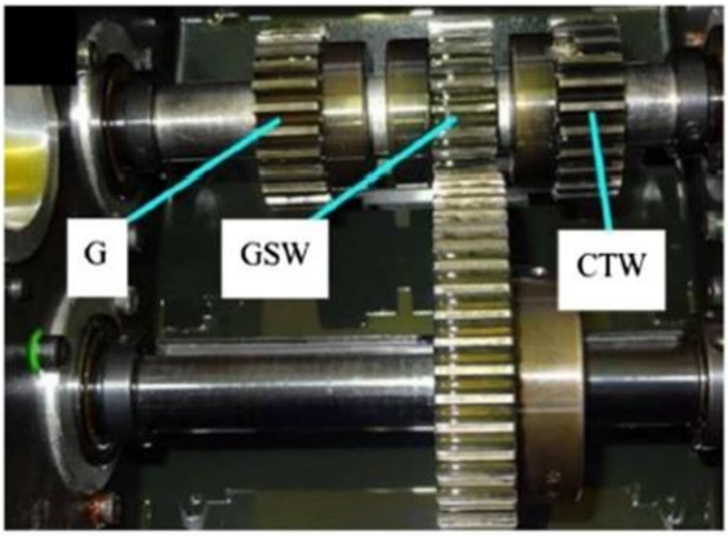


Figure 48: Single stage gear pinions location in the gearbox.

Table 2: Operating conditions

Gear State	Speed, r/min	Load, Nm
G	900, 1200, 1500, 1800, 2400	0, 8, 11
MT	900, 1200, 1500, 1800, 2400	0, 8, 11
TRC	900, 1200, 1500, 1800, 2400	0, 8, 11
GSW	900, 1200, 1500, 1800, 2400	0, 8, 11
CTW	900, 1200, 1500, 1800, 2400	0, 8, 11
CTL	900, 1200, 1500, 1800, 2400	0, 8, 11

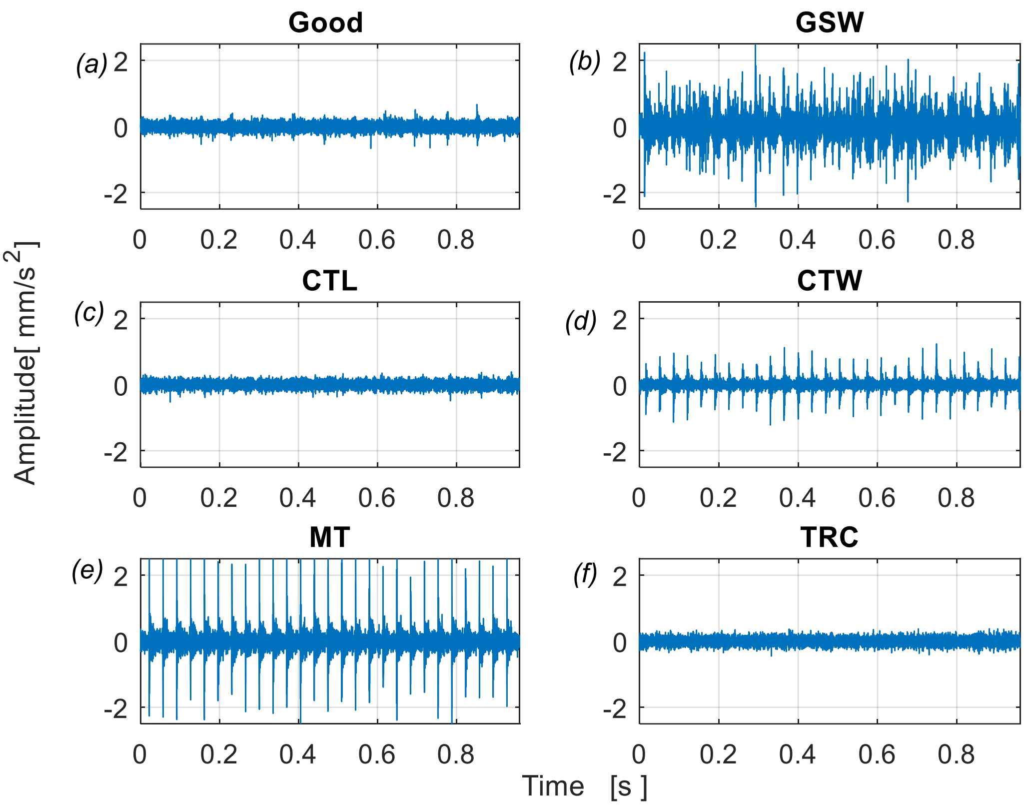


Figure 49: Pinion vibration signals of gearbox with different faults (Speed: 900 r/min, Load: 11 Nm).

The database used in this work is obtained from experiences carried out at the laboratories LaM-CoS (INSA-Lyon, France).

V.3. Features extraction:

In order to do a supervised classification, we must have the entries and the labels of each class of the faults, the entries in our case are the features of faults and to obtain them, we execute the following steps:

- We take the vibration signal which is coming from data acquisition, we decompose it by MODWT method in sixteen modes.
- We apply the statistical functions showed below on the resulting signals to extract the features:

Table 3: Table of statistical tools used to extract features.

Name	formula
Root mean square	$\sqrt{\frac{1}{n}\sum_{i=1}^{n}\lvert x_i\rvert^2}$
Crest factor	$C = \frac{\lVert x\rVert_\infty}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\lvert x_i\rvert^2}}$
Peak to peak	$Max(x) - Min(x)$
Skewness	$E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right]$
Kurtosis	$E\frac{(x-\mu)^4}{\sigma^4}$
Entropy	$-\sum_i p_i\, log_2(p_i)$
Mean	$\frac{1}{N}\sum_{i=1}^{N} A_i$
Std	$\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\lvert A_i-\mu\rvert^2}$
Var	$\frac{1}{N}\sum_{i=1}^{N}(x_i-\tilde{x})^3$
Root sum square	$X_{rss} = \sqrt{\sum_{n=1}^{N}\lvert X_n\rvert^2}$
Max	$Max\lvert x_i\rvert$
Min	$Min\lvert x_i\rvert$

we disposed the results as shown below:

$$\mathrm{feature}_{f1} = [[\mathrm{RMS}(m_1)\ \dots \mathrm{RMS}(m_{16})]\ [\mathrm{CF}(m_1)\ \dots \mathrm{CF}(m_{16})] \dots\dots [\mathrm{Min}(m_1)\ \dots \mathrm{Min}(m_{16})]]$$

where:

$m_1 \ldots m_{16}$: are the modes

we did the same thing for all experiences with six classes of faults at four different speed, at the end we got a 4800x192 matrix of features and each feature is composed of 200 samples and each class of faults is composed of 800 samples as shown below:

$$\text{Features} = \begin{array}{c} \begin{bmatrix} x_{0001,1} \cdots \cdots x_{0001,192} \\ x_{0800,1} \cdots \cdots x_{0800,192} \\ \vdots \\ x_{4001,1} \cdots \cdots x_{4001,192} \\ x_{4800,1} \cdots \cdots x_{4800,192} \end{bmatrix} \end{array} \begin{array}{c} \text{class 1} \\ \vdots \\ \text{class 6} \end{array}$$

where:

class 1: is the healthy state

class 2-6: is the faulty classes

V.4. Optimization and selection:

Removing irrelevant, redundant or noisy features and selecting those that contain the maximum amount of useful information will ensure satisfactory accuracy for the prediction or classification of gear faults. Additionally, this leads to improved learning accuracy and fault classification process.

In this work, the proposed method is based on the use of the ACO whose objective function depends on the RF classification algorithm.

several simulations are carried out to show the efficient of the proposed method, at the first we proceed with the Random Forest classifier and to show its performance compared to other classifiers, we have carried out others simulations with three other classifiers which are: K-Nearest Neighbor number, Naive Bayes and Decision Tree.

V.5. Application of proposed method

V.5.1. Number features equal to 50

Our method was firstly tested with fifty features; the obtained results are represented in Figure 51, It can be clearly noticed in the Confusion matrix that the classification accuracy is 99.375%.

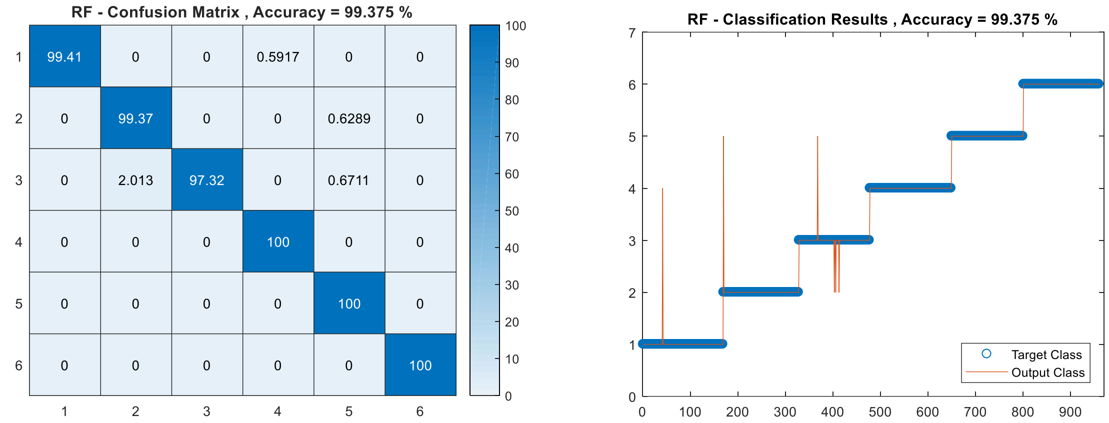


Figure 50: RF Confusion matrix with 50 features.

V.5.2. Best number of features

For this first simulation we randomly chose the number of features at 50, but it is not necessary the good choice, however, in order to choose the best number of features, we have carried out several tests with different numbers of features.

The classification accuracy was saved in each test aiming to determine the optimal number of the selected features and the result is shown on figure. 51.

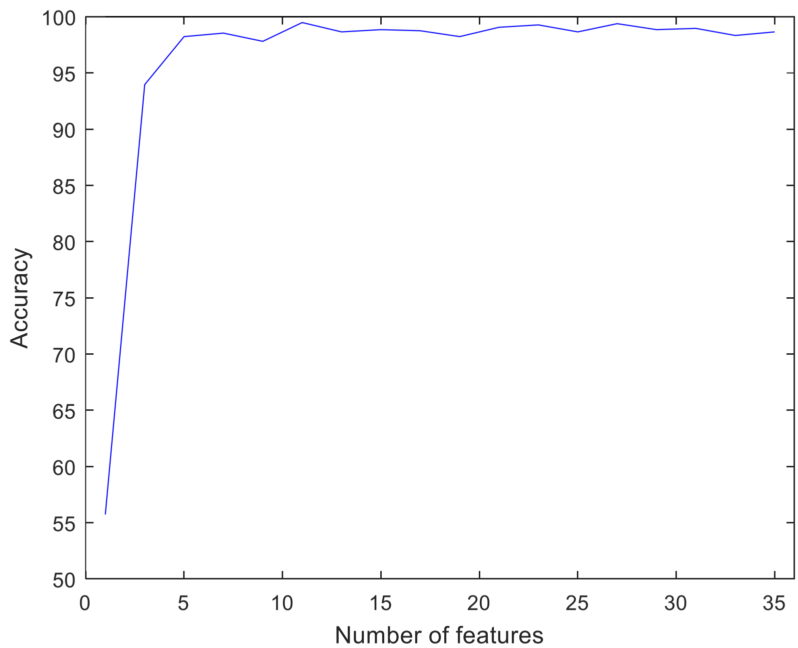


Figure 51: Accuracy classifications (RF) versus number of selected features.

Figure 52 represents the classification accuracy variation according to the number of features. It can be concluded that the best classification accuracy equals to 99.48 and it is obtained when the features number is 11, we also noticed that from this value, the accuracy became stable, so we repeated the simulation with this new number of features and the results are shown below:

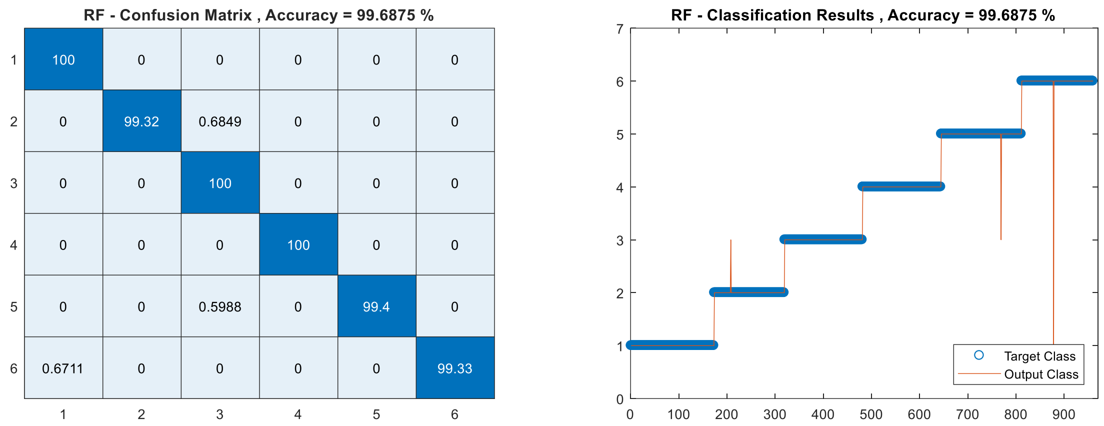


Figure 52: Confusion matrix with 11 features with Random Forest technique

From the figures, we noticed that the accuracy is 99.6875 which means that the proposed method is very good, and also from the figure we noticed that there are only three misclassified observations, the one is predicted in class three, whereas it belongs to the two class, the second observation is predicted in the third class whereas it also belongs to the fifth class, the last one is predicted in the first class whereas it belongs to the sixth.

V.6. Comparison with others techniques

In order to prove the efficiency of the proposed method, it was put in comparison with other techniques which are KNN, DT and CNB. The number of features is fixed at 11 for all methods, the result is shown below.

At the first we compared it to the K-nearest neighbor; we fix the nearest neighbor number of KNN at 5, the results are shown below:

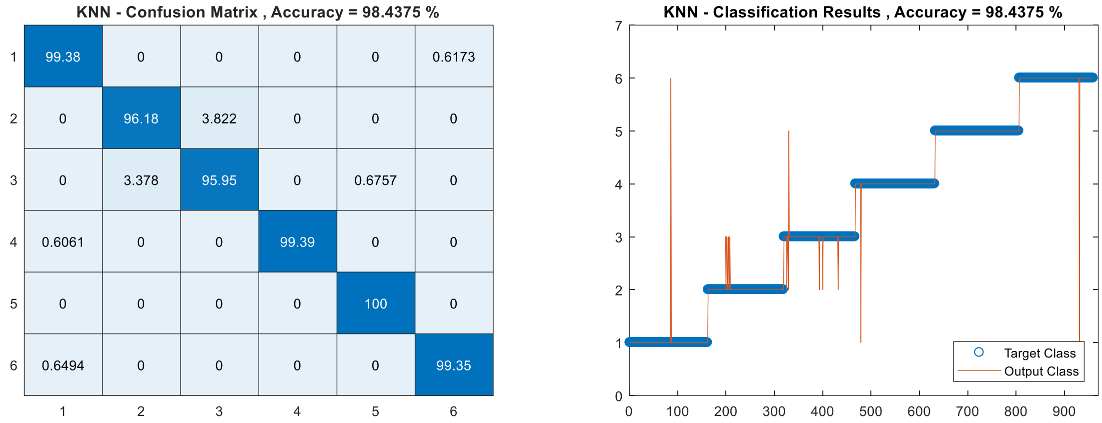


Figure 53: Classification result with 11 features with KNN

We noticed on the figures that the accuracy is 98.4375 which is smaller than the accuracy of Random Forest, there are twelve misclassification observations, one is predicted in the sixth class whereas is belongs to the first class, three observations are predicted in the third class whereas it belongs to the second class, six others observations are predicted in the second class whereas it belongs to the third class, one observation is predicted in the first class whereas it belongs to the fourth class, the last one is predicted in the first class whereas it belongs to the sixth class.

At the second, we compared it to the naive bayes technique, we keep the same number of features, the results are shown below:

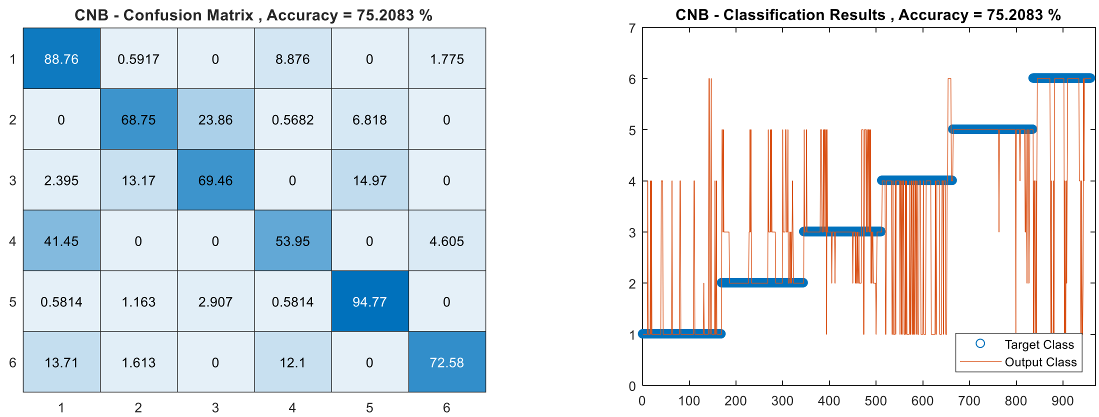


Figure 54: Classification result with 11 features with CNB

The figures show that there are many misclassified observations, and the accuracy is only 78.2083%, so this technique is not good for this type of classification.

At the third we compare it to the Decision Tree technique, the minimum number of father nodes of DT is 5, the results are shown below:

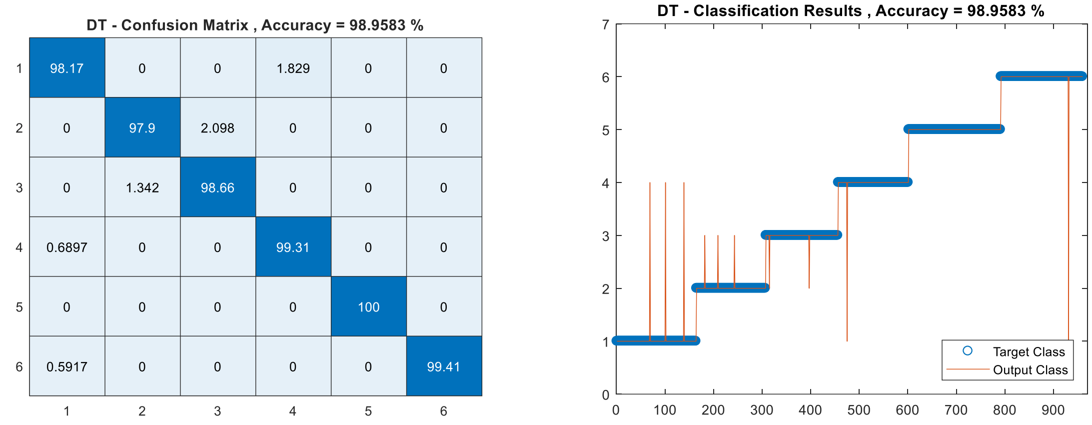


Figure 55: Classification result with 11 features with DT

The figures of this last technique show that there are 10 misclassified observations, there are three predicted in class four whereas they belong to the first class, there are three predicted in the third class whereas they belong to the second one, there is two predicted in the second class whereas it belongs to the third class, there is also one predicted in the first class whereas it belong to the fourth class, the last one is predicted in the fist class whereas it belongs to the sixth class.

V.7. Stability of the techniques:

To study the stability of the proposed method compared to the others, we repeated 15 tests and at each time we calculate and record the mean, the standard deviation, the maximum and the minimum of the classification accuracy, the results are shown on table 4.

It worth mentioning that experiment in this work is conducted under the condition of randomly select samples, where the average value of 15 experiments is taken, Furthermore, the standard deviation of classification accuracy is considered to analyze the stability of the classification method.

The classification results are shown in Table 4 and in Figure 56.

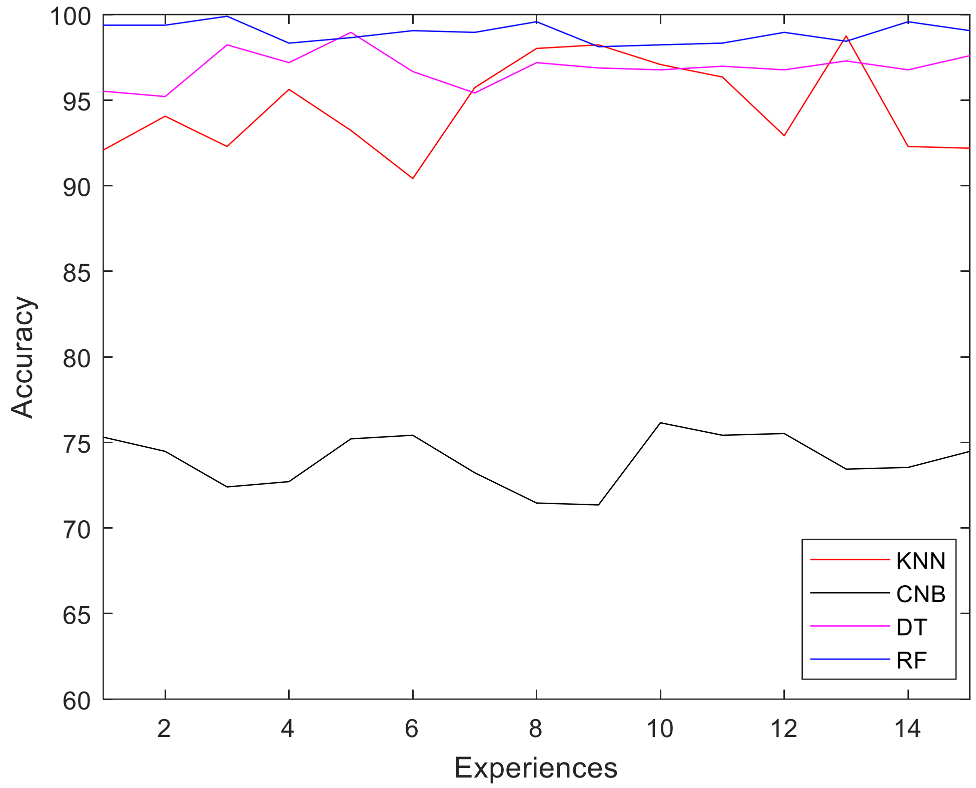


Figure 56: Accuracy variation according to the experiences

This figure 56 shows that the Random Forest is the best classification technique for this type of faults, it has the higher accuracy at each experience.

Table 4: Classification results of the methods.

	RF	KNN	DT	CNB
Max	99.90	98.75	98.96	76.15
Min	98.12	90.42	95.21	71.35
Mean	98.93	94.62	96.90	74.01
STD	0.56	2.66	1.00	1.55

The result in the Table 4 shows that the random forest is very stable than the other techniques, such as the standard deviation of the random Forest is smaller than the standard deviation of the others technique which means that the accuracy is always near to the mean.

To show that the proposed method is better than the others even if we vary the number of features, we have performed several simulations by varying the number of features, the result is shown in the following figure:

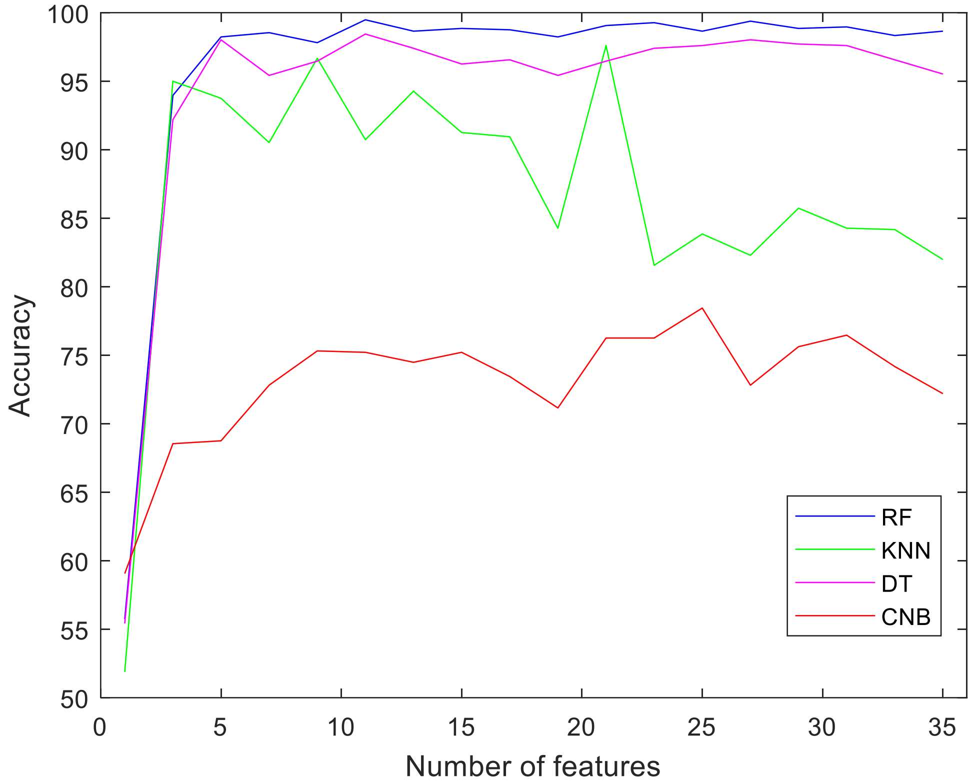


Figure 57: Classification accuracy variation according to the selected number of features.

The Figure 57 represents the variation of the accuracy according to the number of features for different classifier that is KNN, DT, CNB and RF. By comparing RF with other classifiers, we can see that RF always gives the best accuracy regardless of the number of features, even we see that at some points the KNN is better but it is very instable, so we can't take it.

V.8. Conclusion

From the results, we see that the CNB has the lowest classification accuracy rate namely only 76.15%, which means that the CNB method is not good for this type of fault classification; it is followed by KNN technique with 98.75% classification accuracy rate and DT with 98.96%. The best accuracy is obtained with the RF technique with 99.90% accuracy rate.

The KNN and DT have good accuracy but not better than RF and also those methods are not stables as we have seen in figure 56 and 57, however, RF is very stable which means that FR is the best classification technique for this type of fault.

General Conclusion

Rotating machines fault diagnosis is becoming increasingly important through time because of the fast development of industries. This is to help operators in troubleshooting by identifying and locating underlying problems, especially since any delays or misdiagnoses can endanger human safety, the condition of the machine and the environment, in spite of the great interest this field has gained, only a little focus was devoted for machines diagnosing under variable operating condition which represents the practical case in most industries. In this work, a new gearbox fault diagnosis method based on MODWPT-ANT and RF classifier is proposed to diagnose various faults of gearbox under variable operating condition. The effectiveness of the proposed method is validated by recognizing six fault types of gearbox. Compared with other existing classification methods, the obtained experimental results using RF classifier indicate that the proposed method provides an alternative way for gearbox health monitoring.

Under the premise of the same input, the RF classifier is always higher than that of DT, and the classification effect is better. By comparing with CNB and KNN, the proposed method has higher classification accuracy and can be better used for gearbox fault diagnosis, and the classification accuracy reaches 99.89%.

The results obtained with our proposed method show that the number of false alarms is almost zero, which is a significant improvement and a plus for the decision to intervene or not on the equipment.

With zero false alarm, we consider this proposed method is very efficiency and will provide a great improvement to decision-making in the industrial field that uses gearboxes systems.

We conclude that we achieved our fixed objective which was to develop a new method that from a vibration signal it can detect and classify six types of the industrial defects under operation varying condition.

This new technique is based on wavelet transform for the feature extraction and on the Ant Colony Optimization algorithm combined with Random Forest classifier for the feature selection which is the best combination for this application.

Perspectives:

As perspectives of this work, we can use neural network instead of Random Forest to classify the features, we can also use the bee algorithm instead the Ant Colony optimization and to apply this method on another system.

REFERENCES

[1] Wang, Gao and Yan, "***Integration of EEMD and ICA for wind turbine gearbox diagnosis*** », WIND ENERGY 2014; **17**:757–773, https://doi.org/10.1002/we.1653

[2] Hamadache and Lee, "***Rotor Speed-Based Bearing Fault Diagnosis (RSB-BFD) Under Variable Speed and Constant Load***", IEEE Transactions on Industrial Electronics 2015, https://doi.org/10.1109/TIE.2015.2416673

[3] Goyal and Pabla, "***The vibration monitoring methods and signal processing techniques for structural health monitoring: a review***". Archives of Computational Methods in Engineering, 23(4):585–594, 2016. https://doi.org/10.1007/s11831-015-9145-0

[4] Yang, Liu, Wu and Deng, "***Gear Backlash Detection and Evaluation based on Current Characteristic Extraction and Selection***", IEEE Access (Volume: 8) 2020, https://doi.org/10.1109/ACCESS.2020.2999478

[5] Yang and Chen, "***Fault diagnosis of gearbox based on RBF-PF and particle swarm optimization wavelet neural network***", Neural Computing and Applications 2018, https://doi.org/10.1007/s00521-018-3525-y

[6] Lu Wang, Jiawei Xiang, and Yi Liu. "***A time–frequency-based maximum correlated kurtosis deconvolution approach for detecting bearing faults under variable speed conditions***". Measurement Science and Technology, 30(12):125005, 2019b. https://doi.org/10.1088/1361-6501/ab3678

[7] D. Han, X. Guo and P. Shi, "***An Intelligent Fault Diagnosis Method of Variable Condition Gearbox Based on Improved DBN Combined With WPEE and MPE***" in *IEEE Access*, vol. 8, pp. 131299-131309, 2020, https://doi.org/10.1109/ACCESS.2020.3008208.

[8] Zheng X, Wang S, Qian Y. "***Fault feature extraction of wind turbine gearbox under variable speed based on improved adaptive variational mode decomposition***". *Proceedings of the Institution of Mechanical Engineers, Part A: Journal of Power and Energy*.2020;234(6):848-861. https://doi.org/10.1177/0957650919885720

[9] Gougam et all, "***Bearing fault diagnosis based on feature extraction of empirical wavelet transform (EWT) and fuzzy logic system (FLS) under variable operating conditions"*** . Journal of Vibro engineering, Vol. 21, Issue 6, 2019, https://doi.org/10.21595/jve.2019.20092

[10] P. Shi, W. Yang, M. Sheng, and M. Wang, "***An Enhanced Empirical Wavelet Transform for Features Extraction from Wind Turbine Condition Monitoring Signals***" Energies 10, no. 7 (July 2017): 972–985. https://doi.org/10.3390/en10070972

[11] Z. Li, Z. Ma, Y. Liu, W. Teng, and R. Jiang, "***Crack Fault Detection for a Gearbox Using Discrete Wavelet Transform and an Adaptive Resonance Theory Neural Network***," Journal of Mechanical Engineering 61, no. 1 (January 2015): 63–73 https://doi.org/10.5545/sv-jme.2014.1769.

[12] J. Sanz, R. Perera, and C. Huerta, "***Gear Dynamics Monitoring Using Discrete Wavelet Transformation and Multi-layer Perceptron Neural Networks***" Applied Soft Computing 12, no. 9 (September 2012): 2867–2878. https://doi.org/10.1016/j.asoc.2012.04.003

[13] Chen, Rhee and Liu, "***Empirical mode decomposition based on Fourier transform andband-pass filter***", International Journal of Naval Architecture and Ocean Engineering, Volume 11, Issue 2, July 2019, Pages 939-951. https://doi.org/10.1016/j.ijnaoe.2019.04.004

[14] Y. Yang, Y. He, J. Cheng, and D. Yu, "***A Gear Fault Diagnosis Using Hilbert Spectrum Based on MODWPT and a Comparison with EMD Approach***," Measurement 42, no. 4 (May 2009): 542–551. https://doi.org/10.1016/j.measurement.2008.09.011

[15] P.-W. Shan and M. Li, "***Nonlinear Time-Varying Spectral Analysis: HHT and MODWPT***" Mathematical Problems in Engineering 2010 (March 2010): 618231. https://doi.org/10.1155/2010/618231

[16] Huang and Baddour, "***Bearing fault diagnosis under unknown time-varying rotational speed conditions via multiple time-frequency curve extraction***", Journal of Sound and Vibration Volume 414, 3 February 2018, Pages 43-60 https://doi.org/10.1016/j.jsv.2017.11.005

[17] S.TOUAF, «***Diagnostic logique des systèmes complexes dynamiques dans un contexte multiagent*** », Thèse Université Joseph Fourier, Grenoble 1, 2005.

[18] H. Bendjama, «***Contribution to faults detection and diagnosis: Application to dynamic systems and materials*** ». PHD Thesis, National Polytechnic School.

[19] BDIRINA.KH, «***Diagnosis of gear faults by vibration analysis***» Thesis of Magisterium in Electronic Engineering University of M'sila (2006).

[20] Auberville J-Marie « ***Maintenance industrielle, De l'entretien de base à l'optimisation de la sûreté*** », Collection Technosup, 285 pages, 27/07/2004

[21] O. Djebili, "***Contribution to predictive maintenance by vibrational an alysis of rotating mechanical components. Application to ball stops subject to treading tiredness***" University doctoral thesis of Reims Champagne Ardenne, France, 2013.

[22] A. Kabouche, "***Contribution to optimizing maintenance in a distributed context***" PhD thesis from the University of Grenoble, France, 2011.

[23] B. Liang, S.D. Iwnicki, Y. Zhao, "***Application of power spectrum, cestrum, higher order spectrum and neural network analyses for induction motor fault diagnosis***" 2013 Mechanical Systems and Signal Processing 39(1-2):342-360 https://doi.org/10.1016/j.ymssp.2013.02.016

[24] J. Slavic, A. Brkovic, and M.R. Bolteza, "***Typical Bearing-Fault Rating Using Force Measurements-Application to Real Data***," Journal of Vibration and Control, 2164-2174, 2012. https://doi.org/10.1177/1077546311399949

[25] B.DARKHOVSKI, M.STAROSWIECKI, «**A game-Theoretic Approach to Decision in FDI**», IEEE Transactions on automatic control, Vol. 48, N°5, 2003. https://doi.org/10.1109/TAC.2003.811267

[26] P. M.FRANK, «***Fault Diagnosis in Dynamic Systems Using Analytical and Knowledge based Redundancy-A Survey and Some New Results***», Automatica, Vol. 26, N° 3, pp. 459- 474, 1990. https://doi.org/10.1016/0005-1098(90)90018-D

[27] R.J. Patton and J. Chen. "***An review of parity space approaches to fault diagnosis".*** In 1e IFAC Symposium on Fault Detection Supervision and Safety for Technical Processes SAFE PROCESS, Baden Baden, Germany, vol. 1, 239-255, 1991. https://doi.org/10.1016/S1474-6670(17)51124-6

[28] Rabah FELLOUAH, «***Producer Research Consortium, Safe Production Systems***», University doctoral thesis of the University of Toulouse, issued by INSA of Toulouse. December 2007.

[29] G. ZWINGELSTEIN, «***Diagnostique des défaillances***», diagnostic & maintenance Broché, Hermes Science Publications (21 juin 1995).

[30] M. Zemouri, " ***contribution to the monitoring of production systems using dynamic neural networks: application to e-maintenance*** ". These of the University of Franche-Comte, France, 2003.

[31] Dudley, Darle W. "***Dudley's gear handbook***." Published by Tata McGraw-Hill Education, 1991. ISBN-13 : 978-0070179035

[32] Elbarghathi, Fathalla. "***Condition monitoring of helical gearboxes based on the advanced analysis of vibration signals***."Diss. University of Huddersfield, 2016. https://eprints.hud.ac.uk/id/eprint/30647/

[33] Wang, W. J., and P. D. McFadden. "***Application of orthogonal wavelets to early gear damage detection.***" Mechanical Systems and Signal Processing 9.5 (1995): 497-507. https://doi.org/10.1006/mssp.1995.0038

[34] Ding, Yan, and Neville F. Rieger. "***Spalling formation mechanism for gears***." Wear 254.12 (2003): 1307-1317. https://doi.org/10.1016/S0043-1648(03)00126-1

[35] Randall, Robert Bond. "***Vibration-based condition monitoring: industrial, aerospace and automotive applications***". John Wiley & Sons, 2011. ISBN: 978-1-119-47755-6

[36] Tandon, Naresh, and Achintya Choudhury. "***A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings***." Tribology international 32.8 (1999): 469-480. https://doi.org/10.1016/S0301-679X(99)00077-8

[37] Samuel, Paul D., and Darryll J. Pines. "***A review of vibration-based techniques for helicopter transmission diagnostics***." Journal of sound and vibration 282.1-2 (2005): 475-508. https://doi.org/10.1016/j.jsv.2004.02.058

[38] Mustapha, Merzoug, Khalid Ait-Sghir, Abdelhamid Miloudi, Jean Paul Dron, and Fabrice Bolaers. "***Early detection of gear failure by vibration analysis***." Conference on Multiphysics Modelling and Simulation for Systems Design. Springer, Cham, 2014. https://doi.org/10.1007/978-3-319-14532-7_8

[39] Cohen, Leon. "***Time-frequency distributions-a review***." Proceedings of the IEEE 77.7 (1989): 941-981. https://doi.org/10.1109/5.30749

[40] Tang, Baoping, Wenyi Liu, and Tao Song. "***Wind turbine fault diagnosis based on Morlet wavelet transformation and Wigner-Ville distribution***", Renewable Energy 35.12 (2010): 2862-2866. https://doi.org/10.1016/j.renene.2010.05.012

[41] Yifei Yu, Linqing Luo, Bo Li, Kenichi Soga, and Jize Yan. "***Quadratic time-frequency transforms-based Brillouin optical time-domain reflectometry***." IEEE Sensors Journal 17.20 (2017): 6622-6626. https://doi.org/10.1109/JSEN.2017.2736606

[42] Liu, Yu, Junhong Zhang, and Liang Ma. "***A fault diagnosis approach for diesel engines based on self-adaptive WVD, improved FCBF and PECOC-RVM***." Neurocomputing 177 (2016): 600-611. https://doi.org/10.1016/j.neucom.2015.11.074

[43] Peng, Z. K., and F. L. Chu. "***Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography***." Mechanical systems and signal processing 18.2 (2004): 199-221. https://doi.org/10.1016/S0888-3270(03)00075-X

[44] Kwon, D.W., Ko K., Vannucci M., Reddy A.L.N., and Kim S."***Wavelet methods for the detection of anomalies and their application to network traffic analysis***." Quality and Reliability Engineering International 22.8 (2006): 953-969. https://doi.org/10.1002/qre.781

[45] Norden, E. Huang, Zheng Shen, Steven R. Long, Manli C. Wu, Hsing H. Shih, Quanan Zheng, Nai-Chyuan Yen, Chi Chao Tung, and Henry H. Liu Huang. "***The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis***." Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences Vol. 454, No. 1971 (Mar. 8, 1998), pp. 903-995 (93 pages)

[46] J. R Quinlan, "***C4.5: Programs for machine learning***", Morgan Kaufmann Publishers, San Mateo, 1993. https://doi.org/10.1007/BF00993309

[47] I. Witten, E. Frank, "***Data Mining: Practical Machine Learning Tools and Techniques with Java implementations***", Morgan Kaufmann Press, 2nd Edn., 2005. ISBN: 978-0-12-374856-0, https://doi.org/10.1016/C2009-0-19715-5

[48] L. Breiman "***Random Forests***", Machine Learning **volume 45**, pages 5–32 (2001) https://doi.org/10.1023/A:1010933404324

[49] MENASSEL Rafik, "***Optimization metaheuristic design for image compression***" PHD Thesis 2018, university BADJI MOKHTAR-ANNABA.

[50] Alan Chauvin. "***Contribution to the overall optimization for the dimensioning and energy management of hybrid electric vehicles based on a combinatorial approach***". PHD Thesis, 19 Sep 2016. https://tel.archives-ouvertes.fr/tel-01368509.

[51] A. Jain and d. Zongker. ***"Feature selection: evaluation, application, and small sample performance".*** Ieee trans pattern anal. Mach. Intell, 1997 https://doi.org/10.1109/34.574797

[52] A.j. Wijs, M. Torabi Dashi**, "*Extended beam search for non-exhaustive state space analysis*"** , The Journal of Logic and Algebraic Programming, 2012 https://doi.org/10.1016/j.jlap.2011.06.002

[53] E. Bonabeau, M. Dorigo, and G. Theraulaz. "***Swarm intelligence: From natural to artificial systems***". published by Oxford University Press, 1999. https://doi.org/10.1093/oso/9780195131581.001.0001

[54] Liu Yang SONG, "***Application of Ant Colony Algorithm in Fault Diagnosis of Roller Bearing***", *.Advanced Materials Research Vols. 291-294 (2011) pp 1957-1960*
https://doi.org/10.4028/www.scientific.net/AMR.291-294.1957

[55] M. Dorigo, C. Blum, "***Ant colony optimization theory: A survey"***, Theoretical Computer Science 344 (2005) 243 – 278, https://doi.org/10.1016/j.tcs.2005.05.020

[56]Fedala. S, Rémond. D, Zegadi. R, & Felkaoui, A « ***Contribution of angular measurements to intelligent gear faults diagnosis***". Journal of Intelligent Manufacturing, 29(5), 1115–1131. https://doi.org/10.1007/s10845-015-1162-1

[57] C. Rahmoune, B. Merainani, D, Benazzouz and S, Fedala "***Gear fault feature extraction and classification of singular value decomposition based on Hilbert empirical wavelet transform***". Journal of Vibro engineering, Volume 20 1603-1618, 2018;
https://doi.org/10.21595/jve.2017.18917

[58] Ghosh, M., Guha, R., Sarkar, R. *et al.* « ***A wrapper-filter feature selection technique based on ant colony optimization".*** *Neural Comput & Applic* **32**, 7839–7857 (2020).
https://doi.org/10.1007/s00521-019-04171-3

[59] Liping Xie, Zilong, Yihan Zhou, Li Yiliu He, Jiaxin Zhu « ***Computational Diagnostic Techniques for Electrocardiogram Signal Analysis*** » November 2020, Sensors

[60] Dash, M. and Liu, H. "***Feature Selection for Classification***". Intelligent Data Analysis: An International Journal 1997, 1, 131-156.
http://dx.doi.org/10.1016/s1088-467x(97)00008-5