

# Application de l'Intelligence Artificielle et de l'Apprentissage Profond pour la Maintenance Prédictive des Equipements des Chantiers Pétroliers

# Résumé

Ce mémoire a exploré les applications des nouvelles techniques émergentes de l'intelligence artificielle et de l'apprentissage profond (réseaux de neurones en particulier) pour la maintenance prédictive, le diagnostic et le pronostic. De nombreuses architectures neurales telles que les réseaux de neurones entièrement connectés, convolutifs et récurrents ont été développées et testées sur des bases de données publiques telles que le C-MAPSS de la NASA, les bases de données de Case Western Reserve University et FEMTO afin de diagnostiquer l'état de santé des équipements et/ou de prédire la durée de vie utile restante (RUL) avant panne. De nombreuses procédures de traitement des données et d'extraction de caractéristiques ont été utilisées en combinaison avec des techniques d'apprentissage profond telles que la réduction de la dimensionnalité (analyse en composantes principales) et le traitement du signal (analyses de Fourier et d'ondelettes) afin de créer des caractéristiques plus significatives et plus robustes à utiliser en entrée des architectures de réseaux neuronaux. Ce mémoire a également exploré l'utilisation potentielle de ces techniques dans la maintenance prédictive au sein des plateformes pétrolières pour la surveillance des équipements critiques des champs pétrolifères afin de réduire les temps d'arrêt et les coûts de maintenance imprévus.

Mots clés: maintenance prédictive, pronostique, apprentissage profond, réseaux de neurones

# Abstract

This thesis explored applications of the new emerging techniques of artificial intelligence and deep learning (neural networks in particular) for predictive maintenance, diagnostics and prognostics. Many neural architectures such as fully-connected, convolutional and recurrent neural networks were developed and tested on public datasets such as NASA C-MAPSS, Case Western Reserve University Bearings and FEMTO Bearings datasets to diagnose equipment health state and/or predict the remaining useful life (RUL) before breakdown. Many data processing and feature extraction procedures were used in combination with deep learning techniques such as dimensionality reduction (Principal Component Analysis) and signal processing (Fourier and Wavelet analyses) in order to create more meaningful and robust features to use as an input for neural networks architectures. This thesis also explored the potential use of these techniques in predictive maintenance within oil rigs for monitoring oilfield critical equipment in order to reduce unpredicted downtime and maintenance costs.

Keywords: predictive maintenance, prognostics, deep learning, neural networks

# Contents

<b>Introduction</b>	<b>1</b>
<b>1 Vers une Approche de Pronostic Data-Driven</b>	<b>4</b>
1.1 Le Pronostic des équipements mécaniques . . . . .	4
1.2 Estimation de durée de vie utile restante . . . . .	5
1.3 Approches Physiques, Data-Driven et Hybrides . . . . .	5
1.3.1 Modèles Physiques . . . . .	5
1.3.2 Modèles Data-Driven . . . . .	6
1.3.3 Modèles Hybrides . . . . .	7
1.4 Pourquoi une approche Data-Driven? . . . . .	7
1.5 Conclusion . . . . .	8
<b>2 Les étapes d'une approche data-driven</b>	<b>9</b>
2.1 Aquisition de données . . . . .	9
2.2 Extraction de Caractéristiques . . . . .	10
2.2.1 Traitement du signal . . . . .	10
2.2.2 Analyse de Fourier . . . . .	11
2.2.3 Transformation en Ondelettes . . . . .	12
2.2.4 Reduction de la dimensionnalité . . . . .	15
2.3 Diagnostic . . . . .	16
2.4 Pronostic . . . . .	17
2.5 Décision de la maintenance . . . . .	17
2.6 Conclusion . . . . .	17
<b>3 Introduction au Réseaux de Neurones Artificiels</b>	<b>19</b>
3.1 La structure de réseaux de neurones artificiels . . . . .	19

3.2	Feedforward: de l'entrée vers la sortie . . . . .	20
3.3	Fonction d'activation . . . . .	21
3.4	Entraînement du réseau . . . . .	21
3.4.1	Fonction de coût . . . . .	22
3.4.2	Gradient Descent . . . . .	22
3.4.3	Convex and non-convex functions . . . . .	22
3.4.4	Les minima globaux et locaux . . . . .	24
3.4.5	Points selle . . . . .	24
3.4.6	Optimisation de réseaux de neurones . . . . .	24
3.4.7	Backpropagation . . . . .	25
3.5	Réseaux de neurones récurrents . . . . .	26
3.5.1	Long Short-Term Memory . . . . .	26
3.6	Convolutional neural networks . . . . .	27
3.7	Conclusion . . . . .	28
<b>4</b>	<b>Évaluation de l'État des Équipements par les Réseaux de Neurones</b>	<b>30</b>
4.1	Introduction à la base de données C-MAPSS . . . . .	30
4.2	Visualisation de la dégradation des turboréacteurs . . . . .	32
4.3	classification de l'état de santé des turboréacteurs . . . . .	34
4.4	Prédiction de RUL . . . . .	36
4.4.1	Modélisation de RUL . . . . .	36
4.4.2	Prédiction de RUL par un réseau de neurones . . . . .	37
4.4.3	Amélioration de la prédiction RUL en utilisant les réseaux LSTM	39
4.5	Application aux équipements des chantiers pétroliers . . . . .	41
4.6	Conclusion . . . . .	43
<b>5</b>	<b>Diagnostic et Pronostic des Roulements avec les Réseaux de Neurones</b>	<b>45</b>
5.1	Etude de cas: Diagnostique des roulements . . . . .	45
5.1.1	Génération de données à partir de vibration . . . . .	46
5.1.2	Diagnostic des défauts de roulements à l'aide de réseaux de neurones . . . . .	48
5.2	Etude de cas: Pronostique des roulements . . . . .	50
5.2.1	Introduction au base de données FEMTO . . . . .	50
5.2.2	Génération de données à partir des scaleogrammes . . . . .	52

5.2.3	Détection de la défaillance des roulements à l'aide de réseaux de neurones convolutifs . . . . .	53
5.2.4	Le besoin de nouvelles caractéristiques de pronostic . . . . .	56
5.2.5	Caractéristiques trigonométriques et descripteurs cumulatifs . . . . .	57
5.3	Application aux équipements des chantiers pétroliers . . . . .	60
5.3.1	Top Drive . . . . .	60
5.3.2	Les composants de Top Drive . . . . .	61
5.3.3	Proposed approach for Top Drive monitoring using neural networks	65
5.4	Conclusion . . . . .	66
	<b>Conclusion</b>	<b>67</b>

# List of Figures

1.1	Classification des approches de pronostic . . . . .	5
1.2	Taxonomie des approches de pronostic [1] . . . . .	6
1.3	Configuration hybride en série (Figure adaptée de la référence [2]) . . . . .	7
1.4	Configuration hybride parallèle (Figure adaptée de la référence [2]) . . . . .	7
2.1	Techniques de traitement du signal dans les différents domaines . . . . .	12
2.2	Signal dans le domaine temporel et sa transformation de Fourier rapide	13
2.3	Plan de résolution temps–fréquence . . . . .	13
2.4	Les différents types d’ondelettes . . . . .	14
2.5	Scaleogram d’un échantillon de vibrations . . . . .	14
2.6	Transformation en ondelettes discrètes (DWT) comme banque de filtres	15
2.7	Décomposition du signal de niveau 3 à l’aide de DWT . . . . .	16
3.1	Structure d’un réseau de neurones . . . . .	20
3.2	Différentes fonctions d’activation . . . . .	21
3.3	Condition de convexité . . . . .	23
3.4	Exemple de fonction convexe (gauche) et non convexe (droite) . . . . .	23
3.5	Fonction non convexe avec minima globaux et locaux . . . . .	24
3.6	Point selle . . . . .	25
3.7	Forward (gauche) et Backward (droite) passes . . . . .	26
3.8	Architecture de réseaux de neurones recurrents . . . . .	26
3.9	Cellule Long Short-Term Memory . . . . .	28
3.10	Appliquer des convolutions (en multipliant l’entrée par un noyau) de manière itérative à différentes régions (bleu foncé) de l’entrée 2D (carré bleu entier). Chaque itération donne une valeur numérique (vert foncé). Le résultat de toutes les itérations est la sortie 2D verte [3]. . . . .	28

4.1	Schéma simplifié d'un turboréacteur simulé dans C-MAPSS [4] . . . . .	30
4.2	Développement de 6 capteurs de sortie d'un des turboréacteurs (normalisé)	32
4.3	Dégradation de la santé de l'équipement (des couleurs plus claires indiquent l'avancement de la dégradation de la santé) de quatre turboréacteurs à partir de la base de données C-MAPSS . . . . .	33
4.4	Processus d'entraînement du classificateur de l'état des unités . . . . .	35
4.5	Courbe ROC du classificateur sur les données de test . . . . .	35
4.6	Différents modèles de RUL . . . . .	36
4.7	Processus d'entraînement du réseau utilisés pour prédire RUL . . . . .	38
4.8	Résultats de prédiction de RUL par un réseau de neurones . . . . .	38
4.9	Processus d'entraînement du réseau LSTM . . . . .	40
4.10	Résultats de prédiction du réseau LSTM . . . . .	41
4.11	Emplacement des capteurs dans un chantier pétrolier [5] . . . . .	42
5.1	Composants d'un roulement . . . . .	45
5.2	Génération de données par conversion du signal en images $64 \times 64$ . . . . .	46
5.3	Signaux convertis de différents types de défauts . . . . .	47
5.4	Entraînement de classificateur de défauts de roulements . . . . .	49
5.5	Matrice de confusion de la classification des défauts de roulements à l'aide de CNN . . . . .	50
5.6	Plateforme PRONOSTIA [6] . . . . .	51
5.7	Scalogrammes des différentes phases de la vie du roulement1_1 (vibrations horizontales) . . . . .	53
5.8	Scalogrammes des différentes phases de la vie du roulement1_1 (vibrations verticales) . . . . .	54
5.9	Entraînement du classificateur de l'état de santé des roulements . . . . .	54
5.10	Courbe ROC du classificateur de l'état de santé des roulements sur l'ensemble de test . . . . .	56
5.11	Caractéristiques trigonométriques du roulement1_1 . . . . .	58
5.12	Caractéristiques cumulatives classiques et trigonométriques . . . . .	58
5.13	Features fitness . . . . .	59
5.14	Top Drive Bentec 500-HT . . . . .	60
5.15	Sous-ensembles du Top Drive Bentec 500-HT . . . . .	62
5.16	Unité de forage du Top Drive Bentec 500-HT . . . . .	63
5.17	Les composants principaux de l'entraînement . . . . .	64
5.18	Palier de butée à rouleaux coniques SKF [7] . . . . .	64

5.19 Position des roulements dans le Top Drive [8] . . . . .	65
5.20 Approche proposée pour la surveillance du Top Drive à l'aide de réseaux de neurones . . . . .	66

# List of Tables

3.1	Définitions mathématiques de quelques fonctions d’activation . . . . .	21
4.1	Sortie de simulation C-MAPSS pour mesurer la réponse du système . .	31
4.2	Statistiques sur le nombre d’unités et la longueur des cycles dans la base de données C-MAPSS. . . . .	31
4.3	Architecture du classificateur de l’état des unités . . . . .	34
4.4	Métriques du classificateur sur les données de test . . . . .	36
4.5	Architecture d’un réseau de neurones pour la prédition de RUL . . . .	37
4.6	Architecture de réseau LSTM pour la prédition de RUL . . . . .	39
4.7	Résultats d’entraînement du réseauLSTM . . . . .	40
4.8	Liste des capteurs dans un chantier pétrolier [5] . . . . .	44
5.1	Dimensions des roulements CWRU et les fréquences de défauts . . . .	46
5.2	Nombre d’échantillons pour chaque classe de défauts . . . . .	47
5.3	Architecture CNN du classificateur de défauts de roulements . . . .	48
5.4	Résultats de l’entraînement du classificateur de défauts de roulements .	49
5.5	Differentes conditions de fonctionnement et roulements correspondants dans la base de données FEMTO . . . . .	52
5.6	Architecture du classificateur de l’état de santé de roulements . . . .	55
5.7	Résultats de l’entraînement du classificateur de l’état de santé de roulements . . . . .	55
5.8	Indicateurs supplémentaires pour la performance du réseau . . . . .	55
5.9	Caractéristiques trigonométriques et classiques de pronostic [9] . . . .	57
5.10	Différence de monotonicité entre les caractéristiques trigonométriques et classiques et leurs descripteurs cumulés . . . . .	59
5.11	Différence de trendabilité entre les caractéristiques trigonométriques et classiques et leurs descripteurs cumulés . . . . .	59

# Acronyms

**AUC** Area Under Curve.

**CBM** Condition-Based Maintenance.

**CM** Condition Monitoring.

**CNN** Convolutional Neural Networks.

**CWT** Continuous Wavelet Transform.

**DD** Data-Driven.

**DWT** Discrete Wavelet Transform.

**LCM** Life Cycle Management.

**LSTM** Long Short-Term Memory.

**PHM** Prognostics and Health Management.

**PM** Preventive Maintenance.

**RNN** Recurrent Neural Networks.

**ROC** Receiver Operating Characteristics.

**RUL** Remaining Useful Life.

“All models are wrong, but some  
are useful”.

---

George E. P. Box

# Introduction

Toutes les machines en général, et mécaniques en particulier, sont sujettes à la dégradation de leur état et à la détérioration de leurs performances avec le temps, ce qui entraîne éventuellement leur défaillance. Ces défaillances peuvent avoir des impacts négatifs sur les aspects économiques, humains, environnementaux, ... La défaillance d'une machine est une propriété intrinsèque de ces systèmes (en raison de leurs propriétés physiques inhérentes), elle peut être — afin d'éviter des résultats négatifs — partiellement prévenue, retardée et même prévue, mais elle ne peut jamais être totalement empêchée ou arrêtée, ceci est principalement réalisé par la maintenance.

La maintenance est définie comme un ensemble d'activités destinées à maintenir ou à restaurer une unité utilitaire dans un état dans lequel elle peut remplir une fonction requise [10].

La façon dont la maintenance industrielle est effectuée a évolué avec les progrès de la technologie depuis le début de la révolution industrielle jusqu'à aujourd'hui. Sa forme la plus élémentaire est l'entretien correctif non planifié qui est effectué après l'apparition d'une panne, afin de rétablir une unité fonctionnelle dans un état où elle peut remplir une fonction requise [10]. Avant la Seconde Guerre mondiale, la machinerie était relativement simple et la demande de production était modérée, de sorte qu'elle pouvait être maintenue après une panne. Après la guerre et avec la reconstruction de l'industrie, le marché est devenu plus compétitif et moins tolérant au temps d'arrêt, l'industrie s'est donc tournée vers la maintenance préventive (Preventive Maintenance: PM)<sup>1</sup> qui est effectuée à des intervalles prédéterminés ou selon des critères prescrits afin de réduire la probabilité de défaillance d'une unité fonctionnelle [10]. *La maintenance corrective et préventive sont toutes deux obsolètes.* C'était donc une question de temps avant que l'idée de la maintenance prédictive ou conditionnelle (Condition-Based Maintenance: CBM) émerge, ce qui diffère des autres formes en basant le besoin de l'intervention sur l'état réel de la machine plutôt que sur un calendrier préétabli [11]. Il y a un désaccord dans la littérature sur cette classification (de la CBM en tant que maintenance préventive ou une forme distincte) [12]. La différence de taxonomie n'est pas d'un intérêt réel pour la discussion actuelle.

Un nouveau domaine qui est apparu récemment est celui de pronostics et de gestion de la santé (Prognostics and Health Management : PHM). Le PHM s'est imposé comme une approche essentielle pour obtenir des avantages concurrentiels sur le marché mon-

---

<sup>1</sup>Les termes seront mentionnés en Français, traduits et abrégés (selon la littérature) en Anglais

dial en améliorant la fiabilité, la maintenabilité, la sécurité et l'abordabilité. Comme le CBM, le PHM est une discipline qui a émrgé dans l'industrie avec une présence modeste dans le domaine académique, en particulier dans les applications militaires [13] (e.g. le développement de l'avion de combat F-35 [14]). Les concepts et les composants du PHM ont été développés séparément dans de nombreux domaines tels que le génie mécanique, électrique et les sciences statistiques, sous des noms variés [15]. Alors que le CBM se concentre sur la surveillance du système, le PHM est une approche plus intégrée qui vise à fournir des lignes directrices pour la gestion de la santé du système. Il s'agit donc d'une philosophie de gestion du cycle de vie (Life Cycle Management: LCM) qui met l'accent sur la prévisibilité (c.-à-d. le pronostic) des défaillances et de la maintenance. Ceci est généralement réalisé par l'adoption d'une stratégie de surveillance, qui peut être une technique de CBM [13]. Le but principal du pronostic (aussi CBM) est l'estimation de la durée de vie utile restante (Remaining Useful Life: RUL) en étant donné l'état actuel de l'équipement et son profil opérationnel passé [16].

Le PHM agit à un niveau un peu plus élevé que le CBM, puisqu'il a l'ambition claire de permettre la gestion de la santé. Il s'agit d'une activité liée à LCM, ce qui signifie qu'une approche est suivie pour optimiser toutes les activités (de maintenance) pendant le cycle de vie complet du matériel. Cela comprend le choix d'une politique de maintenance appropriée, la définition de l'échéancier d'intervention et la décision sur le moment où un matériel doit être mis hors service. CBM n'offre pas ce soutien LCM étendu. Le domaine PHM ne prescrit ni un concept de maintenance spécifique ni une stratégie de surveillance. Cependant, dans les études typiques de PHM, des politiques de CBM ou d'autres politiques de maintenance sont adoptées et dans de nombreux cas, les techniques de surveillance d'état (Condition Monitoring: CM) sont appliquées [13].

Comme mentionné précédemment, tous ces efforts pour développer des stratégies de maintenance ont été principalement dus aux coûts élevés des pertes de production et de la maintenance réactive dans des industries à une forte intensité capitaliste comme les projets Onshore et Offshore dans l'industrie pétrolière et gazière sont principalement des investissements, qui peuvent avoir de graves conséquences financières et environnementales si une défaillance catastrophique se produit. Par conséquent, une approche efficace de la gestion de la maintenance est essentielle à la poursuite de la production de manière sûre et fiable [17]. Les organisations Offshore subissent en moyenne des impacts financiers de 38 millions de dollars par an en raison de temps d'arrêt imprévus (pour les organisations les moins performantes, 88 millions de dollars). Moins de 24 % des opérateurs décrivent leur approche de maintenance comme étant prédictive et basée sur des données et des analyses. Plus des trois quarts adoptent une approche réactive ou basée sur le temps. Les opérateurs utilisant une approche prédictive basée sur des données connaissent en moyenne 36 % de temps d'arrêt non planifié de moins que ceux qui utilisent une approche réactive. Cela peut se traduire, en moyenne, par une baisse de 34 millions de dollars par an du résultat net [18].

Dans ce mémoire, une approche de pronostic basée sur les données (Data-Driven) sera introduite, présentant les différentes étapes nécessaires depuis l'acquisition des données jusqu'à l'estimation de RUL. En raison de la sensibilité commerciale et technique des données dans le domaine pétrolier (et toutes les industries avec une grande valeur ajoutée) et donc de l'impossibilité de les obtenir, les bases de données publiques (NASA Ames Data Repository [19]) qui sont habituellement utilisées pour évaluer les

## *Introduction*

---

algorithmes de pronostic dans la littérature seront utilisées ici aussi. Une approche sera proposée pour transférer les connaissances obtenues à des applications concrètes dans l'industrie pétrolière.

Ce mémoire est divisé comme suit:

**Chapitre 01** Vers un Approche de Pronostic Data-Driven

**Chapitre 02** Les Etapes d'un Approche Data-Driven

**Chapitre 03** Introduction au Réseaux des Neurones

**Chapitre 04** Évaluation de l'État des Équipements par les Réseaux de Neurones

**Chapitre 05** Diagnostic et Pronostic des Roulements

**Conclusion**

# Chapter 1

## Vers une Approche de Pronostic Data-Driven

L'objectif de ce chapitre est de présenter les différentes approches pronostiques avec une taxonomie détaillée, les différentes étapes de toute approche pronostique seront décrites puis l'accent sera mis sur les méthodes fondées sur les données.

### 1.1 Le Pronostic des équipements mécaniques

Certains systèmes complexes surtout dans le domaine de pétrolière fonctionnent sous des conditions très sévères (offshore, désert... ) ce qui peut engendrer l'occurrence des défaillances et leur dégradation. Les pannes et les arrêts non planifiés causent automatiquement des pertes de production, ce qui peut avoir des conséquences économiques énormes. Avec ces contraintes économiques, il faut développer les programmes de maintenance pour minimiser la probabilité des défaillances et le coût. Comme discuté dans l'introduction de ce mémoire, ces programmes doivent être basés sur les principes Condition-Based Maintenance et Prognostics and Health Management.

Le pronostic et la gestion de la santé (Prognostics and Health Management) a deux aspects principaux[20]:

1. **Pronostic:** diagnostic prédictif, ce qui comprend la détermination de la durée de vie utile restante (durée de bon fonctionnement) d'une composante ou d'un bien.
2. **Gestion de la santé:** la capacité de prendre des décisions concernant les actions de maintenance en basant sur les informations du diagnostic/pronostic, les ressources disponibles et la demande opérationnelle.

## 1.2 Estimation de durée de vie utile restante

L'objectif principal du pronostic c'est l'estimation de la durée de vie utile restante (Remaining Useful Life) du système. RUL est défini selon l'équation 1.1:

$$RUL = t_f - t_c \quad (1.1)$$

Où  $t_f$  est le temps prédit pour l'occurrence de la défaillance et  $t_c$  est le temps actuel (le temps quand la prédiction est faite).

## 1.3 Approches Physiques, Data-Driven et Hybrides

Toute approche pronostique peut être basée sur des modèles physiques, modèles data-driven ou une combinaison hybride des deux (Figure 1.1).

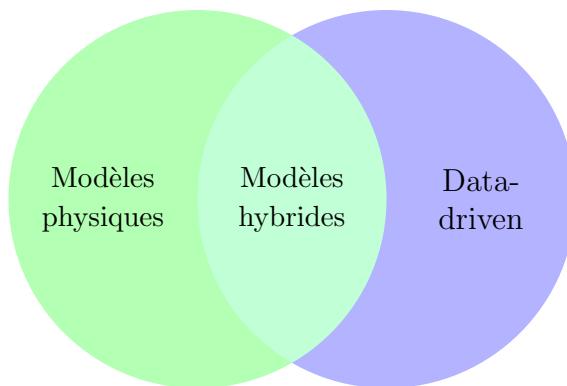


Figure 1.1: Classification des approches de pronostic

Ces trois catégories constituent une classification générale fondée sur l'approche suivie, chacune d'entre elles pouvant être subdivisée en sous-catégories. Une taxonomie détaillée est présentée à la figure 1.2.

### 1.3.1 Modèles Physiques

Les modèles physiques évaluent la santé du système en utilisant une formulation mathématique explicite (boîtes blanches) développée sur la base d'une compréhension scientifique et technique de son comportement. Cependant, le principal avantage de ces modèles physiques consiste à utiliser des modèles de dégradation pour prédire le comportement à long terme [21]. Les approches physiques sont capables de fournir une estimation précise de l'état de santé du systèmes si le modèle physique est développé avec une compréhension complète des mécanismes de défaillance et une estimation efficace des paramètres du modèle. Cependant, pour certains systèmes mécaniques complexes, il est difficile de comprendre la physique des dommages, ce qui limite l'application de ces approches [22].

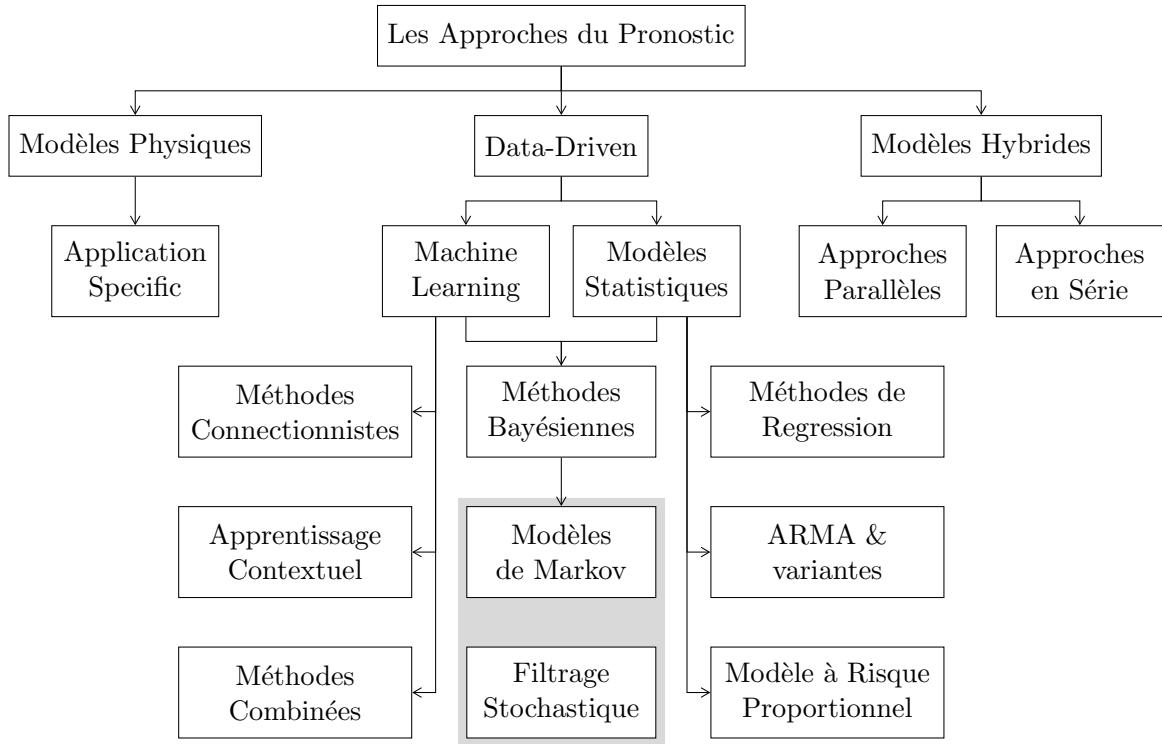


Figure 1.2: Taxonomie des approches de pronostic [1]

### 1.3.2 Modèles Data-Driven

Les modèles Data-Driven s'appuient sur des données collectées précédemment (données de surveillance, données sur les paramètres opérationnels, ...) pour établir un modèle capable d'évaluer la santé du système et de prévoir son comportement et sa dégradation. Contrairement aux modèles physiques, et comme leur nom l'indique, les modèles Data-Driven ne s'appuient pas sur les connaissances humaines mais principalement sur les données historiques collectées pour modéliser le processus de dégradation. Habituellement, ils sont considérés comme des boîtes noires.

#### Modèles Statistiques

L'approche statistique repose sur la construction et l'ajustement d'un modèle probabiliste en utilisant les données historiques sans dépendre d'aucun principe physique ou technique [23]. Si et al. [23] ont présenté une revue des approches statistiques. Selon cette revue, de nombreux modèles entrent dans cette catégorie tels que les modèles de régression (e.g. la régression linéaire), la moyenne mobile autorégressive et ses variantes, les techniques de filtrage stochastique (e.g. filtre de Kalman, filtre particulaire, ...).

#### Machine Learning

L'apprentissage machine (en anglais Machine Learning) est un domaine de l'intelligence artificielle qui a explosé ces dernières années et a fait des percées dans de nombreux domaines tels que Computer Vision et Natural Language Processing. Les modèles

d'apprentissage machine sont des modèles boîte noire qui permettent de découvrir des mappages même très complexes d'une entrée à une sortie. De nombreux types d'algorithmes entrent dans cette catégorie comme les méthodes connectionnistes (e.g. les réseaux de neurones artificiels), l'apprentissage contextuel (e.g. les machine à vecteurs de support). Différentes approches peuvent être combinées ensemble pour créer des modèles mixte qui peuvent être plus performants qu'un modèle unique.

### 1.3.3 Modèles Hybrides

Les modèles hybrides sont une combinaison d'un modèle physique et d'un modèle Data-Driven. Il existe deux types de modèles hybrides selon la façon dont les deux types de modèles sont combinés. Le modèle Data-Driven peut être intégré dans un modèle physique en configuration série (Figure 1.3) où il est utilisé pour ajuster les paramètres du modèle physique qui utilisé alors pour faire des prédictions.

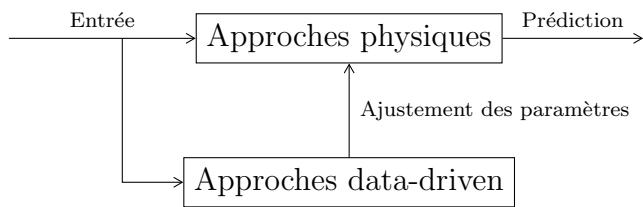


Figure 1.3: Configuration hybride en série (Figure adaptée de la référence [2])

Les deux types de modèles peuvent être combinés dans une configuration parallèle (Figure 1.4) où les deux modèles font des prédictions séparées qui peuvent être combinées pour obtenir l'estimation finale.

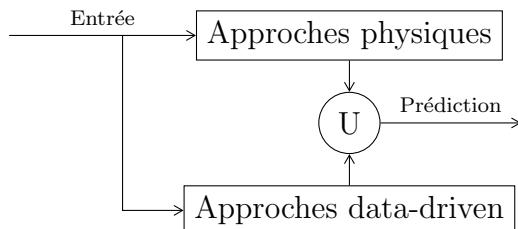


Figure 1.4: Configuration hybride parallèle (Figure adaptée de la référence [2])

## 1.4 Pourquoi une approche Data-Driven?

Comme mentionné précédemment, comprendre le processus de dégradation pour les systèmes très complexes est extrêmement difficile, c'est pourquoi le développement des modèles physiques est très problématique pour ces systèmes.

Les 20 dernières années ont vu de grands progrès dans le développement de nouvelles techniques de détection, de méthodes de pronostic/diagnostic et dans l'application des méthodes d'analyse informatisées.

Il est intéressant de noter que, lors de l'atelier de 2002 sur la maintenance conditionnelle organisé par Advanced Technology Program du National Institute of Standards and Technology (NIST) des États-Unis, les obstacles suivants à l'application généralisée des mesures de confiance ont été identifiés :

- L'impossibilité de prédire avec exactitude et fiabilité la durée de vie utile restante d'une machine.
- L'incapacité de surveiller continuellement une machine.
- L'incapacité des systèmes de maintenance à apprendre et à identifier les défaillances imminentes et à recommander les mesures à prendre.

Ces obstacles peuvent être redéfinis comme des déficiences au niveau des pronostics, de la détection et du raisonnement. Ces limitations et d'autres encore de la mise en œuvre actuelle des techniques de maintenance conditionnelle ont, bien entendu, été reconnues par d'autres et ont conduit à l'élaboration de programmes (e.g. dans le domaine militaire) visant à les surmonter [20].

Aujourd'hui, les capteurs dans l'industrie sont devenus peu coûteux et omniprésents, la puissance de calcul a augmenté exponentiellement ce qui a permis de développer des algorithmes et outils informatiques plus avancés : la révolution de l'intelligence artificielle et l'apprentissage automatique. L'industrie génère d'énormes quantités de données dans de nombreux domaines dont la majorité est inexploitée, l'exploitation de ces données en utilisant les dernières avancées technologiques peut augmenter les profits, réduire les coûts de manière drastique et s'avérer être un énorme avantage économique.

## 1.5 Conclusion

L'adoption de modèles de pronostic basés sur des données peut être utile, en particulier lorsque le comportement du processus de dégradation est ambigu et que l'élaboration de modèles basés sur la physique pour quantifier l'état de santé de systèmes complexes est compliquée et que leurs résultats ne sont pas fiables. Les récents développements des techniques de détection, l'augmentation de la puissance de calcul disponible (c'est-à-dire des unités de traitement plus rapides et moins coûteuses) et l'abondance de données de surveillance non exploitées ont fourni le cadre nécessaire à l'adoption de ces modèles basés sur les données, qui sont plus faciles à développer, à déployer et à automatiser que leurs homologues basés sur la physique.

# Chapter 2

## Les étapes d'une approche data-driven

Une approche de pronostic Data-Driven (une approche qui s'appuie sur des données opérationnelles historiques pour construire un modèle qui est utilisé pour prédire la durée de vie utile restante) doit passer par de multiples étapes—de l'acquisition de données jusqu'à l'estimation de la durée de vie utile restante. Dans ce chapitre, ces différentes étapes seront examinées en détail.

### 2.1 Aquisition de données

Un signal est une fonction qui transmet des informations sur le comportement d'un système ou les attributs d'un phénomène. les signaux se produisent naturellement et sont également synthétisés. Un signal n'est pas nécessairement une grandeur électrique. Cependant, pour effectuer des activités telles que la synthèse, le transport, l'enregistrement, l'analyse et la modification de signaux, il est souvent pratique d'utiliser un signal sous la forme d'une grandeur électrique [24].

L'acquisition de données est un processus de capture et de stockage de différents types de données de surveillance (signaux) provenant de divers capteurs installés sur l'équipement surveillé. C'est le premier processus de pronostic des machines, qui fournit des informations de base de surveillance d'état (Condition Monitoring) pour les processus suivants. Un système d'acquisition de données est composé de capteurs, de dispositifs de transmission de données et de dispositifs de stockage de données [22]. Les données de surveillance d'état sont très versatiles. Il peut s'agir de données sur les vibrations, de données acoustiques, de données d'analyse d'huile, de données sur la température, la pression, l'humidité, les conditions météorologiques ou l'environnement, etc. Les différents capteurs, tels que les micro-capteurs, les capteurs à ultrasons et les capteurs d'émission acoustique ont été conçus pour collecter différents types de données. Les technologies sans fil, telles que Bluetooth, ont fourni

une solution alternative à la communication de données à un prix avantageux [16].

Bien que la recherche sur des concepts avancés comme les réseaux de capteurs sans fil et la récolte d'énergie pour alimenter des capteurs autonomes soit en cours, l'acquisition de données (capteurs) et la manipulation sont aujourd'hui plutôt bien établies. Par conséquent, une grande partie de la recherche dans cette discipline se concentre sur l'analyse des données obtenues pour en extraire de l'information [13]. Parce que cette discipline est bien développée, beaucoup de nouvelles installations et techniques d'acquisition de données ont été conçues et appliquées dans les industries modernes. Ces installations puissantes et polyvalentes ont rendu l'acquisition de données pour la mise en œuvre du PHM plus pratique et plus faisable [25].

## 2.2 Extraction de Caractéristiques

L'approche de pronostic data-driven est principalement utilisée lorsqu'il est difficile de comprendre le comportement physique d'un système complexe. La compréhension du comportement et de l'interaction des différents éléments qui conduisent à la dégradation des machines est le point de départ pour développer un modèle physique pour les pronostics. D'autre part, cette approche utilise des données de surveillance des conditions pour modéliser implicitement son comportement. Les modèles utilisés dans l'approche data-driven utilisent des données de surveillance pour modéliser un comportement complexe et capturer des modèles complexes, mais ils sont considérés comme des boîtes noires : ils ne fournissent pas nécessairement un aperçu du processus. En général, la performance de ces modèles dépend de la qualité des données d'entrée (données). La partie humaine ne peut pas effectuer la tâche que le modèle effectue, mais le traitement des données d'entrée peut augmenter considérablement les résultats. Ce traitement est nécessaire parce que les données des capteurs sur lesquelles le modèle repose sont généralement redondantes, bruyantes et incomplètes, ces imperfections sont dues à de nombreuses raisons.

L'extraction de caractéristiques est une étape de prétraitement importante dans le processus de développement des modèles d'apprentissage machine et influence directement les performances du modèle. Par conséquent, cette étape doit être réalisée avec soin afin d'extraire des caractéristiques significatives des données brutes. Les données sur les vibrations contiennent des informations très utiles sur l'état du système, mais elles nécessitent un prétraitement important avant d'être utilisées comme données d'entrée pour un modèle spécifique. Ce chapitre décrit certaines des techniques de traitement du signal utilisées dans l'analyse vibratoire traditionnelle, mais dans ce contexte, elles seront utilisées comme extracteurs de caractéristiques pour une architecture de réseau de neurones.

### 2.2.1 Traitement du signal

Le traitement du signal est l'étude et l'analyse des signaux stockés afin de révéler leurs propriétés—qui peuvent ne pas être apparentes au premier abord, en utilisant un ensemble d'algorithmes et de techniques. Dans le contexte de la surveillance de l'état, ces propriétés révélées par le traitement du signal peuvent être révélatrices de santé

de la machine. Le traitement du signal est un sous-domaine bien établi et mature du génie électrique, avec de nombreuses techniques et algorithmes proposés dans la littérature.

Le traitement du signal peut être classé en trois catégories : **Analyse Temporelle**, **Analyse Fréquentielle** et **Analyse Temps–Fréquence**.

### Analyse Temporelle

Les mesures originales des signaux qui sont généralement échantillonnés de manière répétée entre des intervalles de temps prédéfinis sont sous forme de domaine temporel. Ainsi, l'analyse du domaine temporel est directement basée sur la mesure originale [25].

### Analyse Fréquentielle

L'analyse du domaine fréquentiel est basée sur les signaux transformés dans le domaine fréquentiel. L'avantage de l'analyse en domaine fréquentiel par rapport à l'analyse en domaine temporel est sa capacité à décomposer les signaux originaux en une série de composantes fréquentielles. L'analyse du domaine fréquentiel la plus utilisée est l'analyse du spectre au moyen de la transformée de Fourier rapide (FFT). L'idée principale de l'analyse spectrale est d'isoler et de localiser certaines composantes de fréquence d'intérêt relatives aux caractéristiques de défaut des machines [26].

### Analyse Temps–Fréquence

Le problème de l'analyse du domaine temporel et de l'analyse du domaine fréquentiel est que chacune d'elles ne dispose d'aucune information sur l'autre domaine (l'analyse du domaine temporel ne dispose d'aucune information sur le domaine fréquentiel, et l'analyse du domaine fréquentiel ne dispose d'aucune information sur la position dans le temps).

Ainsi, l'analyse du domaine temps-fréquence, qui étudie les signaux de mesure dans les domaines du temps et des fréquences, a été appliquée à l'analyse des signaux de mesure non stationnaires. L'analyse temps-fréquence décrit les caractéristiques des signaux de mesure dans les fonctions bidimensionnelles du temps et de la fréquence afin de mieux révéler les modes de défaillance des machines.

La figure 2.1 présente les différentes techniques utilisées pour chaque type d'analyse:

#### 2.2.2 Analyse de Fourier

L'analyse de Fourier, également appelée analyse harmonique, d'un signal périodique  $x(t)$  est la décomposition de la série en sommation de composantes sinusoïdales, où chaque sinusoïde a une amplitude et une phase spécifiques.

La transformation de Fourier (FT) d'un signal  $x(t)$  peut être mathématiquement donnée par l'équation 2.1 :

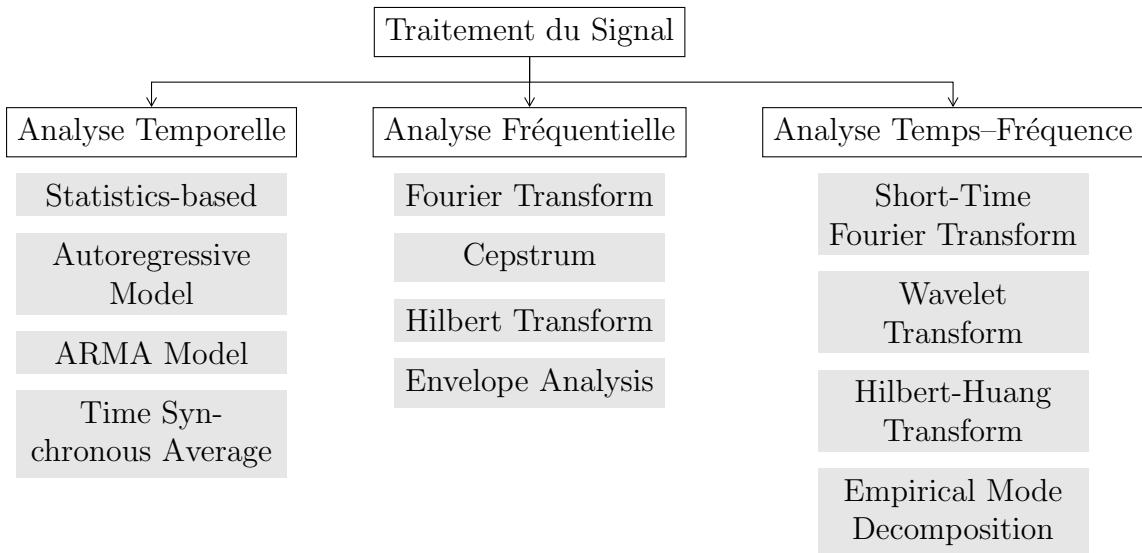


Figure 2.1: Techniques de traitement du signal dans les différents domaines

$$X(w) = \int_{-\infty}^{\infty} x(n)e^{-j\omega t} dt \quad (2.1)$$

Dans les applications pratiques du traitement numérique des signaux où les signaux sont discrets dans le temps plutôt que continus (par exemple l'analyse des vibrations), on utilise plutôt une version discrétisée appelée transformation de Fourier discrète (TFD), qui est exprimée mathématiquement par l'équation 2.2 :

$$X(w) = \sum_{-\infty}^{\infty} x(t)e^{-j\omega t} dt \quad (2.2)$$

La Fast Fourier Transform (FFT) est un algorithme efficace utilisé pour implémenter la TFD dans les ordinateurs. La figure 2.2 montre un signal dans sa forme d'onde (ou domaine temporel) et son spectre correspondant (domaine fréquentiel) obtenu à l'aide de l'algorithme FFT. Le spectre montre les composantes de fréquence présentes dans le signal:

### 2.2.3 Transformation en Ondelettes

La transformation en ondelettes est également un outil d'analyse spectrale, comme la transformation de Fourier. La principale différence est que la transformation de Fourier décompose le signal en composantes sinusoïdales, mais que la transformation en ondelettes le décompose en un ensemble de fonctions oscillatoires appelées **ondelettes**. Contrairement aux sinusoïdes, les ondelettes sont localisées dans le temps, ainsi la transformation en ondelettes ne fournit pas seulement des informations sur la fréquence présente dans un signal mais aussi sur le moment de leur apparition. La transformation en ondelettes est une bien meilleure solution que la transformation de Fourier lorsqu'on étudie des signaux non linéaires et non stationnaires (c'est-à-dire que ses composantes de fréquence varient avec le temps).

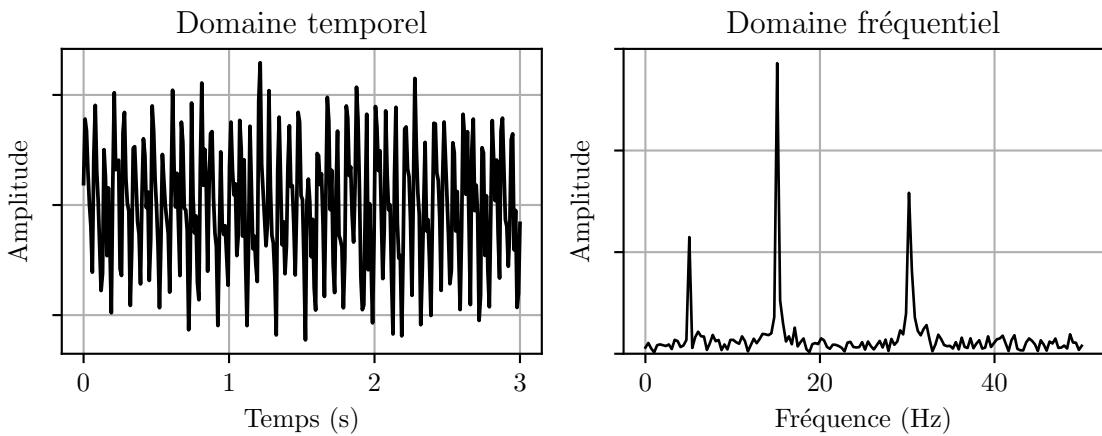


Figure 2.2: Signal dans le domaine temporel et sa transformation de Fourier rapide

La figure 2.3 montre la différence de résolution en temps et en fréquence entre les différentes méthodes. Dans la forme d'onde, le signal a une résolution absolue en temps et une résolution nulle en fréquence. La transformation de Fourier, au contraire, transforme totalement le signal dans le domaine fréquentiel, ce qui lui confère une résolution absolue en fréquence mais aucune résolution en temps. La transformation de Fourier à court terme est calculée à l'identique à la transformation de Fourier, mais elle est effectuée sur des segments séparés du signal original afin de préserver une certaine résolution dans le temps. La transformation en ondelettes présente une résolution temporelle élevée pour les hautes fréquences et une résolution haute fréquence pour les basses fréquences.

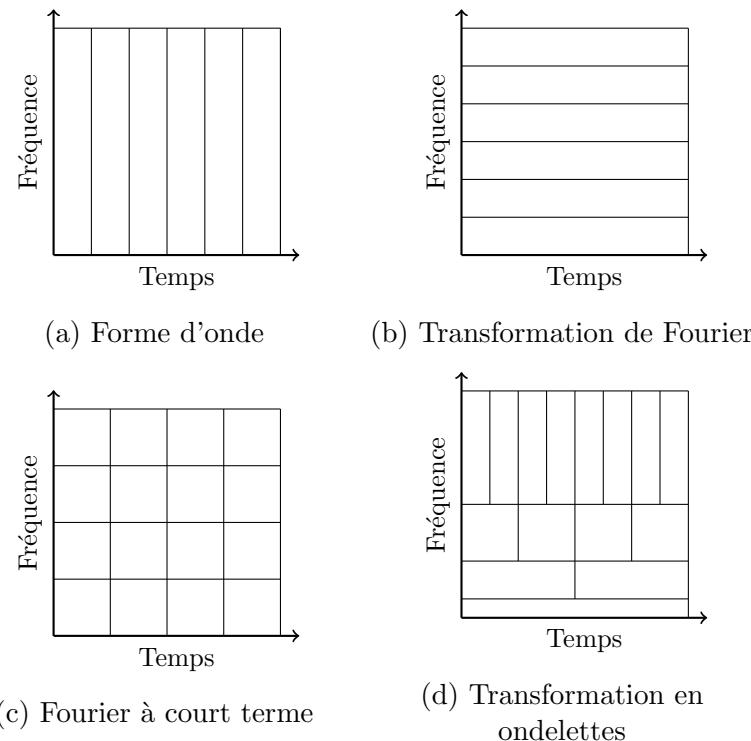


Figure 2.3: Plan de résolution temps–fréquence

Il existe une grande variété d'ondelettes qui servent à différentes fins comme l'ondelette de Morlet, l'ondelettes de Daubechies et bien d'autres.

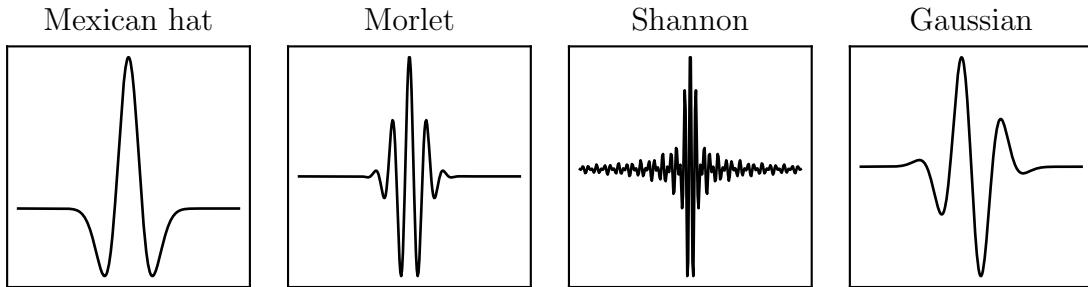


Figure 2.4: Les différents types d'ondelettes

### Transformée en ondelettes continue

Mathématiquement, la transformation en ondelettes continue est définie par l'équation 2.3:

$$CWT_x^\psi(\tau, s) = \frac{1}{\sqrt{|s|}} \int_{-\infty}^{\infty} x(t)\psi^* \left( \frac{t - \tau}{s} \right) dt \quad (2.3)$$

Où  $x(t)$  est le signal original,  $\psi^*$  est une fonction appelée **l'ondelette mère** ;  $s$  et  $\tau$  sont les facteurs de **dilatation** et **translation** respectivement. Le signal original est multiplié par l'ondelette mère qui est mise à l'échelle en utilisant différents facteurs de dilatation puis translatée sur le signal.

La sortie de CWT est un scaleogram comme celui de la figure 2.5 qui est un scaleogram (tracé de contour rempli) de données de vibrations sur une durée de 0.1s :

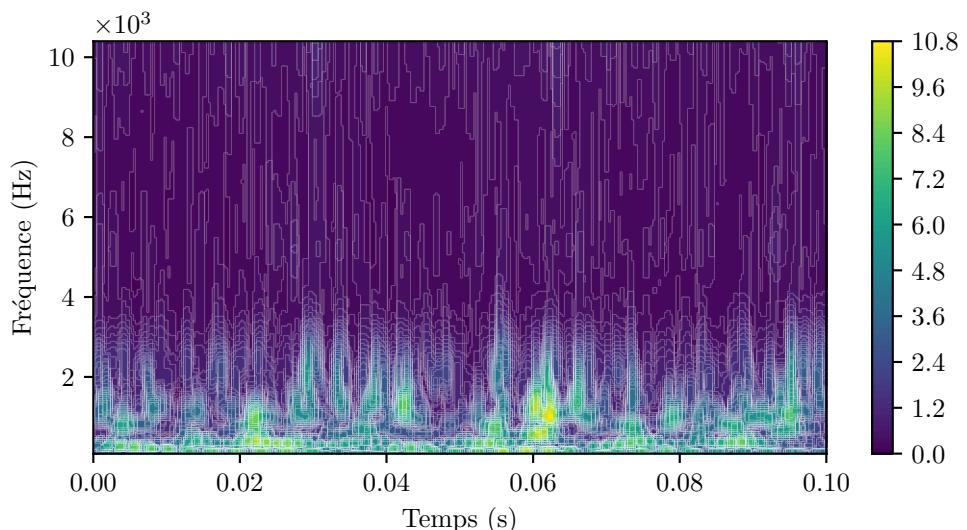


Figure 2.5: Scaleogram d'un échantillon de vibrations

Les axes x et y représentent respectivement le temps et la fréquence. Les différentes couleurs indiquent la hauteur (c'est-à-dire l'amplitude) de chaque fréquence (axe des y) à chaque instant-de-temps (axe des x), ce qui—différemment de la transformation de Fourier—fournit des informations sur les fréquences présentes dans le signal ainsi que sur les moments où ces fréquences sont présentes.

### Transformation en ondelettes discrète

Dans des applications pratiques, la transformation en ondelettes discrètes (Discrete Wavelet Transform: DWT) est implémentée comme une banque de filtres où le signal est passé à travers des filtres passe-bas et passe-haut pour obtenir des **approximation** et des **coefficients de décomposition**. La figure 2.6 montre un DWT avec 2 niveaux de décomposition qui donne des coefficients d'approximation et de décomposition du 2ème ordre :

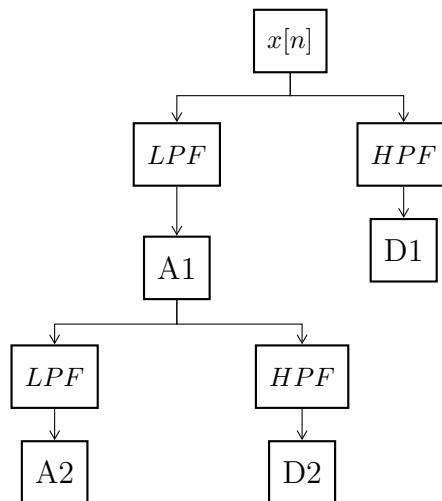


Figure 2.6: Transformation en ondelettes discrètes (DWT) comme banque de filtres

DWT donne deux séries de coefficients : **coefficients d'approximation** associé au filtre passe-bas et **coefficients de détail** associé au filtre passe-haut du DWT. En appliquant à nouveau le DWT sur les coefficients d'approximation, le niveau de décomposition suivant peut être obtenu. À chaque niveau, le signal original est sous-échantillonné d'un facteur 2, ce qui impose une limitation du nombre possible de niveaux de décomposition pour un signal donné.

#### 2.2.4 Reduction de la dimensionnalité

La réduction de la dimensionnalité se réfère au processus consistant à prendre des données à haute dimension et à trouver une bonne représentation de ces données dans une dimension inférieure tout en préservant ses caractéristiques originales. La réduction de la dimensionnalité est effectuée pour l'extraction de caractéristiques en trouvant les principales composantes ou pour la visualisation (en réduisant le nombre de dimensions à 2 ou 3). Il existe de nombreuses techniques et algorithmes de réduction de la dimensionnalité comme:

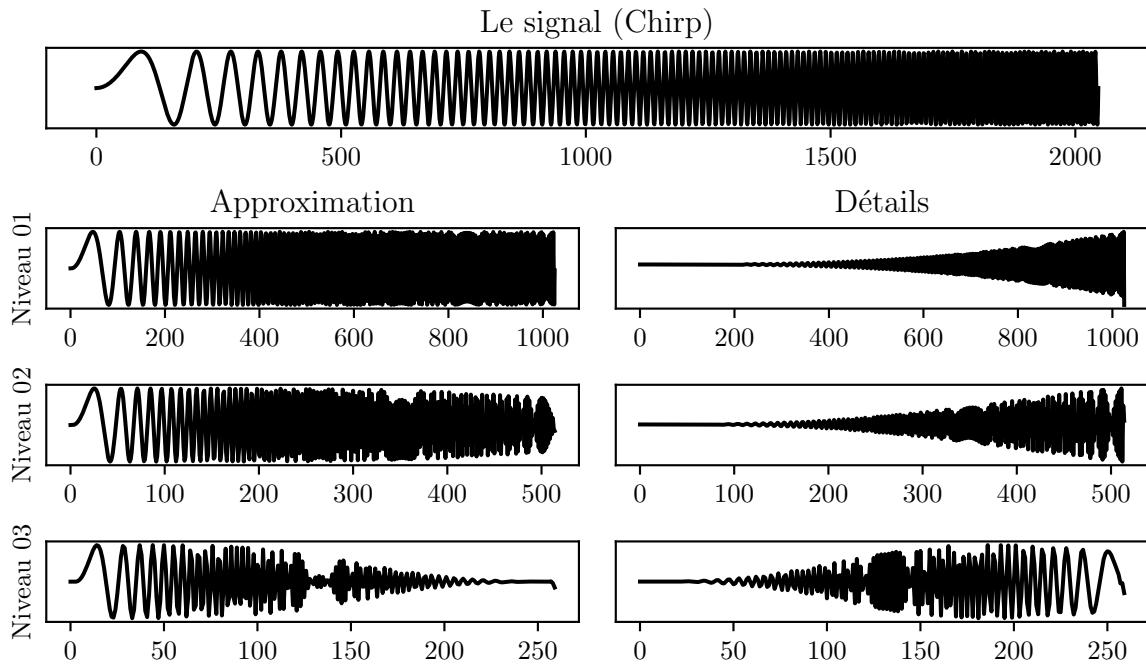


Figure 2.7: Décomposition du signal de niveau 3 à l'aide de DWT

- Principal Component Analysis (PCA)
- Autoencoders
- t-Distributed Stochastic Neighbor Embedding (t-SNE)

Il faut noter que pour les pronostics, où les données utilisées consistent en des entrées de capteurs où chaque variable a une signification physique et une interprétation directe, le fait d'effectuer une réduction de la dimensionnalité donnera les principales composantes, mais les nouvelles variables *perdront leur interprétabilité physique et devront être traitées comme des abstractions*.

## 2.3 Diagnostic

Le diagnostic se résume à un processus d'identification et de détermination de la relation entre les informations obtenues dans l'espace de mesure et les modes de défaillance des machines dans l'espace de défaillance. Le diagnostic comporte trois grandes étapes, à savoir la détection des défauts, l'isolation des défauts et l'identification des défauts. La détection des défauts est une tâche qui consiste à indiquer si un défaut a s'est déjà produite dans les machines surveillées. L'isolation des défauts consiste à trouver la composante de la défaillance et la position de la défaillance. L'identification des défauts est la dernière étape du diagnostic, qui tente de déterminer le mode et la gravité de la défaillance. Les trois étapes sont en corrélation entre eux. Cette dernière étape repose sur les résultats de la première et ne peut donc pas être réalisée individuellement [27].

## 2.4 Pronostic

Le diagnostic est l'analyse de l'événement postérieur et le pronostic est l'analyse de l'événement antérieur. Le pronostic est beaucoup plus efficace que le diagnostic pour obtenir des performances sans arrêt de production. Le diagnostic est toutefois nécessaire lorsque la prédiction des erreurs de pronostic échoue et qu'une erreur se produit [16]. Prognostics are the forecast or prediction of the future performance of a system, this prediction is based on its current state. The prediction is performed using a model, types of prognostics models are already discussed in details in section 1.3 but the focus of this current discussion is geared towards data-driven models. More concretely, what these models predict is the remaining useful life (RUL) defined in section 1.2. Prognostics models—by estimating RUL—aim at scheduling maintenance actions according to the predictions and production constraints associated with the machine in order to achieve zero unscheduled equipment downtime.

## 2.5 Décision de la maintenance

La prochaine et dernière étape de toute approche pronostique consiste à utiliser les résultats du modèle développé (c'est-à-dire les prédictions ou les estimations pour le RUL du système) afin de réaliser des actions de maintenance préventive meilleures et plus précises. Les personnes chargées de la maintenance doivent analyser les résultats du modèle avant de prendre des mesures, car tout modèle comporte une erreur inhérente associée à ses prévisions. Habituellement, les modèles basés sur des données fournissent des intervalles de confiance associés à leurs prédictions qui doivent être considérés avec soin. Il convient également de noter que ces intervalles de confiance reflètent la certitude des modèles quant à leurs prédictions sur la base des données fournies pour l'entraînement. Dans les applications réelles, un nouveau mode de défaillance imprévu peut se produire, ce qui n'était peut-être pas le cas auparavant, et n'a donc pas été prévu pour le modèle dans les données utilisées pour le construire. Ceci est plus ou moins associé à la capacité de différents types de modèles à généraliser pour de nouvelles données imprévues, mais en général, il est bien connu que tout modèle d'entraînement sera moins fiable lorsqu'il effectue des prédictions en utilisant des données que le modèle ne voyait pas avant [28]. C'est pourquoi il est essentiel de fournir des données de qualité qui reflètent différents modèles de dégradation et modes de défaillance pour que le modèle puisse faire des prédictions plus robustes et plus fiables.

## 2.6 Conclusion

L'élaboration d'un modèle de pronostic (fondé sur des données) est un processus qui nécessite de nombreuses étapes : de l'acquisition des données nécessaires à la construction du modèle à l'utilisation des prévisions du modèle pour prendre des décisions de maintenance meilleures et plus précises qui se traduisent par moins (voire zéro) de temps d'arrêt imprévu ; ceci réduit les coûts et diminue les pertes de production, ce dernier point étant le but ultime de toute cette discussion et de la littérature sur les

pronostics/la maintenance préventive en général. Cette discussion se concentre sur les modèles basés sur les données, l'un des modèles qui a prouvé sa grande capacité à apprendre des modèles non linéaires complexes dans les données sont les réseaux de neurones. Ces modèles ont une architecture différente en fonction de la structure des données. Ils seront présentés et expliqués en détail dans le prochain chapitre.

# Chapter 3

## Introduction au Réseaux de Neurones Artificiels

Les réseaux neuronaux artificiels sont des systèmes de computation capables de trouver des fonctions complexes qui relient une entrée à une sortie. ces systèmes peuvent être utilisés pour une variété de tâches comme la régression et la classification. Ils peuvent être utilisés dans le domaine de la maintenance prédictive et des pronostics pour estimer l'état de santé de l'équipement et prévoir avec une certaine incertitude sa durée de vie utile restante. Ce chapitre traite les réseaux de neurones, leur topologie, leur entraînement et leur formulation mathématique.

### 3.1 La structure de réseaux de neurones artificiels

Les réseaux de neurones artificiels sont des systèmes de computation utilisés pour trouver la correspondance entre une entrée et une sortie, ils sont constitués de plusieurs couches (couche d'entrée, couche de sortie et un nombre arbitraire de couches cachées entre l'entrée et la sortie) et chaque couche contient un certain nombre de neurones où chaque neurone de chaque couche est connecté à tous les neurones de la couche précédente et de la couche suivante (à l'exception des couches d'entrée et de sortie qui sont connectées uniquement aux couches suivante et précédente respectivement). Chaque neurone de chaque couche reçoit une entrée des neurones de la couche précédente (sous la forme d'un vecteur), multiplie le vecteur par quelques poids et somme le résultat puis applique une fonction d'activation linéaire. Chaque neurone se retrouve avec une seule valeur numérique appelée activation, qui sera transmise aux neurones de la couche suivante.

Figure 3.1 montre un réseau de neurones avec la structure suivante :

- Une couche d'entrée avec 3 entrée
- Une seule couche cachée avec 4 neurones
- Une couche de sortie avec 1 neurone

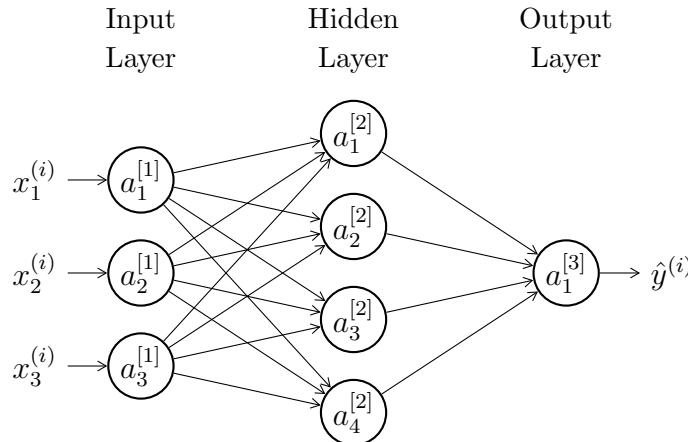


Figure 3.1: Structure d'un réseau de neurones

## 3.2 Feedforward: de l'entrée vers la sortie

L'architecture de la figure 3.1 est appelée Feedforward Neural Network, c'est une architecture acyclique où l'information circule de la première à la dernière couche sans aucune boucle interne, contrairement aux autres architectures comme les réseaux de neurones récurrents. La première couche est la couche d'entrée, elle n'effectue aucune opération et se limite à recevoir l'entrée. L'entrée est un vecteur de nombres représentant les différentes variables. Le vecteur est multiplié par une matrice de poids qui le transforme et l'envoie à la couche suivante (ou à la première couche cachée). Une fonction d'activation est appliquée aux valeurs résultant de la multiplication des valeurs de la couche précédente avec la matrice des poids, le résultat devient les valeurs de la couche suivante, ou activations (la valeur de chaque neurone est appelée activation).

Une formule générale pour passer de la couche  $l - 1$  à la couche  $l$  est donnée par l'équation 3.1:

$$a^{[l]} = g^{[l]}(W^{[l]}a^{[l-1]} + b^{[l]}) \quad (3.1)$$

Où  $g^{[l]}$  est la fonction d'activation de la couche  $l$ ,  $W^{[l]}$  est la matrice des poids qui transforme les valeurs (activations) de la couche  $l - 1$  à la couche  $l$  et  $a^{[l]}$  représente les activations de la couche  $l$ .  $b$  est la valeur du biais (ou valeur d'interception), elle est ajoutée à la multiplication entre les activations et la matrice des poids, c'est un paramètre qui peut être appris comme les poids. L'opération se répète pour chaque couche jusqu'à la couche de sortie. La première couche peut être considérée comme la couche 0 et les entrées peuvent être désignées par le vecteur  $a^{[0]}$ .

### 3.3 Fonction d'activation

Les fonctions d'activation sont appliquées au résultat de la multiplication des entrées de la couche précédente avec les poids correspondants, pour déterminer la valeur de chaque neurone. Il existe différents types de ces fonctions.

L'utilisation de la fonction d'activation non linéaire est très importante pour les réseaux de neurones, ils permettent d'apprendre la correspondance non linéaire complexe de l'entrée à la sortie. Si le réseau n'utilise pas l'activation non linéaire (par exemple, l'activation linéaire ou la fonction d'identité), alors le réseau entier (quelle que soit sa profondeur) est équivalent à un réseau avec une seule couche cachée.

Il existe une variété de fonctions d'activation qui peuvent être utilisées pour les couches cachées et de sortie. La figure 3.2 montre quelques exemples:

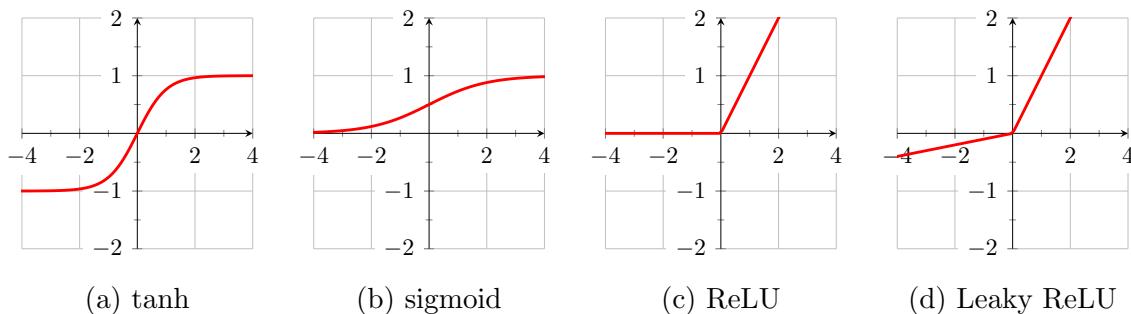


Figure 3.2: Différentes fonctions d'activation

Tableau 3.1 montre les définitions mathématiques de quelques fonctions d'activation:

Fonction d'activation	Définition mathématique
Identity (pas d'activation)	$\text{Id}(x) = x$
Sigmoid	$\sigma(x) = \frac{1}{1+e^{-x}}$
$\tanh$	$\tanh(x) = \frac{(e^x - e^{-x})}{(e^x + e^{-x})}$
Rectified Linear Unit (ReLU)	$\text{ReLU}(x) = \max(0, x)$
Leaky ReLU	$\text{LeakyReLU}(x) = \max(0.01x, x)$

Table 3.1: Définitions mathématiques de quelques fonctions d'activation

### 3.4 Entraînement du réseau

Le processus de formation d'un réseau de neurones consiste à déterminer les poids (coefficients) qui relient les neurones de chaque couche aux neurones de la couche suivante. Ce processus peut être formulé en termes plus mathématiques comme un problème d'optimisation : optimisation des coefficients du réseau pour trouver leurs valeurs qui minimisent une fonction de coût.

L'entraînement du réseau de neurones utilise des données d'entraînement (training data) qui fournissent des entrées et leurs sorties correspondantes.

### 3.4.1 Fonction de coût

La fonction de coût est la fonction utilisée pour calculer la différence entre la sortie du réseau de neurones et la sortie réelle attendue, elle quantifie la performance du réseau. Le but du processus d'entraînement est de minimiser cette fonction en utilisant Gradient Descent (voir la section suivante) pour trouver le meilleur ensemble de poids qui donne la différence la plus basse entre les données d'entraînement et la prédiction du réseau.

Il existe de nombreux types de fonctions de coût, chaque type correspondant à différentes tâches des réseaux de neurones (par exemple, régression, classification binaire, ...). La fonction de coût généralement utilisée pour les problèmes de régression est une fonction d'erreur quadratique moyenne (Equation 3.2) qui calcule la somme des distances entre les prédictions du modèle  $\hat{y}_i$  et la sortie réelle ( $y_i$ ),  $N$  le nombre de points de données disponibles pour l'entraînement :

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (3.2)$$

L'autre principale tâche des réseaux de neurones et des autres types de modèles est la classification. Il existe deux principaux types de classification : la classification binaire, ou classification de deux classes et la classification multiclasse ou classification de plusieurs classes. La première utilise la fonction de perte d'entropie croisée binaire (Equation 3.3) et la seconde utilise l'entropie croisée catégorielle.

$$BCE = -\sum_{i=1}^N \hat{y}_i \log(y_i) + (1 - \hat{y}_i) \log(1 - \hat{y}_i) \quad (3.3)$$

Le choix de la fonction d'activation pour la dernière couche du réseau est directement lié à la fonction de coût utilisée. Pour les tâches de régression, la fonction d'activation linéaire est utilisée. Pour la classification binaire, c'est la fonction sigmoïde et pour la classification multiclasse, c'est la fonction softmax.

### 3.4.2 Gradient Descent

Gradient Descent est un algorithme d'optimisation itératif utilisé pour optimiser une fonction différentiable. Gradient Descent fonctionne en calculant les gradients de la fonction objectif puis en prenant des mesures itératives dans le sens du négatif des gradients.

### 3.4.3 Convex and non-convex functions

Gradient Descent dans le contexte de réseaux de neurones est une algorithme utilisé pour optimiser (minimiser) la fonction de coût en trouvant le meilleur ensemble de poids et de biais qui donne le coût le plus bas possible. Les fonctions objectives peuvent être classées en deux types : les fonctions convexes et les fonctions non convexes. Une

fonction est dite convexe si elle a la propriété que chaque accord se trouve sur ou au-dessus de la fonction. Toute valeur de  $x$  dans l'intervalle allant de  $x = a$  à  $x = b$  peut être écrite sous la forme  $\lambda a + (1 - \lambda)b$  où  $0 \leq \lambda \leq 1$ . Le point correspondant sur l'accord est donné par  $\lambda f(a) + (1 - \lambda)f(b)$ , et la valeur correspondante de la fonction est  $f(\lambda a + (1 - \lambda)b)$  (Figure 3.3). La convexité implique alors :

$$f(\lambda a + (1 - \lambda)b) \leq \lambda f(a) + (1 - \lambda)f(b) \quad (3.4)$$

Cela équivaut à l'exigence que la dérivée seconde de la fonction soit partout positive [29]. Cette condition de convexité peut être étendue aux espaces ayant un nombre arbitraire de dimensions.

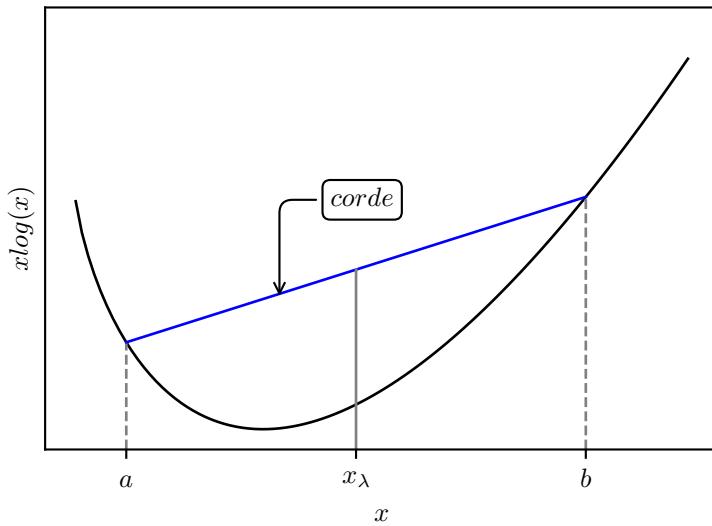


Figure 3.3: Condition de convexité

Une fonction non convexe est une fonction qui ne satisfait pas la condition de convexité. La figure 3.4 montre un exemple de fonction convexe (à gauche) et de fonction non convexe (à droite) où les paramètres de la fonction sont dans un espace à deux dimensions.

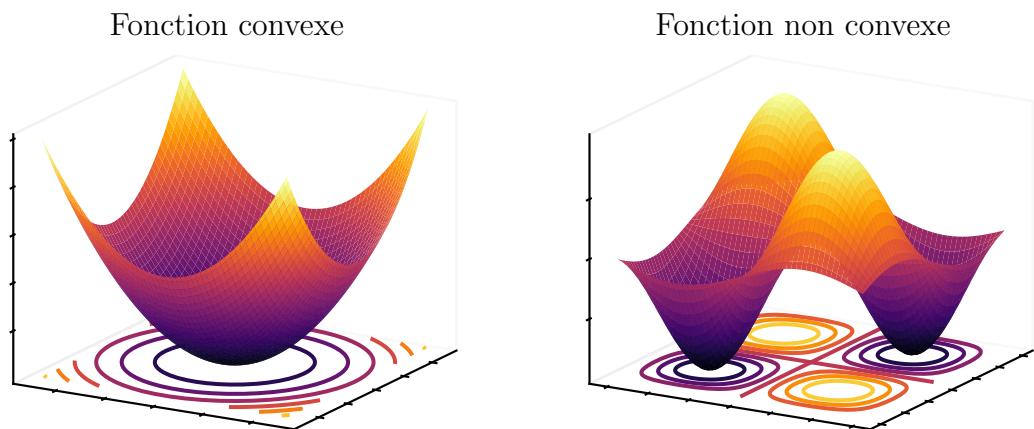


Figure 3.4: Exemple de fonction convexe (gauche) et non convexe (droite)

### 3.4.4 Les minima globaux et locaux

Le minimum global désigne la valeur la plus basse possible d'un ensemble (ou d'une fonction). Trouver le minimum global consiste à trouver l'ensemble des paramètres qui correspondent à cette valeur minimale. Lorsque la fonction est convexe, trouver le maximum global est possible et facile, des algorithmes comme le Gradient Descent convergent toujours vers le minimum global dans l'optimisation convexe. La fonction de coût de régression linéaire est un exemple de fonction de coût convexe. Pour les modèles plus complexes tels que les réseaux de neurones, la fonction de coût est très non convexe avec de nombreux minima locaux.

La figure 3.5 montre un exemple de fonction non convexe avec un minimum global et un minimum local. Le problème de l'optimisation non convexe est que l'algorithme d'optimisation peut converger vers le minimum local au lieu du minimum global. La convergence de l'algorithme est liée à l'initialisation aléatoire des poids du réseau.

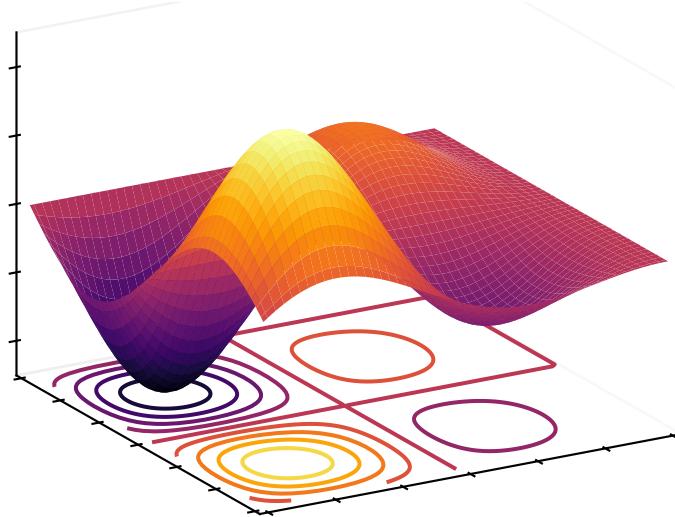


Figure 3.5: Fonction non convexe avec minima globaux et locaux

### 3.4.5 Points selle

Un point selle (saddle point) ou de minimax est un point à la surface du graphique d'une fonction où les pentes (dérivées) dans les directions orthogonales sont toutes nulles (un point critique), mais qui n'est pas un extrême local de la fonction (Figure 3.6).

### 3.4.6 Optimisation de réseaux de neurones

Dans [30], les auteurs ont montré que le fait de rester coincé dans des minima locaux mauvais n'est un problème majeur que pour les réseaux peu profonds, mais devient progressivement moins important à mesure que la taille du réseau augmente. Cela est principalement dû au fait que dans les grands réseaux, les minima locaux sont proches du minimum global, de sorte qu'ils donnent de bonnes performances similaires.

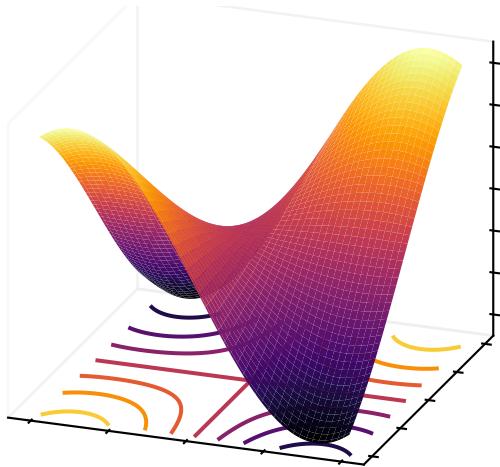


Figure 3.6: Point selle

Les hypothèses suivantes ont également été vérifiées empiriquement dans le document mentionné concernant l'apprentissage avec des réseaux de grande taille :

- Quand un réseau est grand, il n'y a pas de différence significative de performance sur les données de test parmi la plupart des minima locaux.
- La probabilité de trouver un mauvais minimum local dans les petits réseaux est plus élevée que la probabilité de les trouver dans les grands réseaux.
- Trouver le minimum global sur les données d'entraînement ne garantit pas une meilleure performance sur les données de test, mais peut plutôt entraîner une réduction de la performance à cause de overfitting.

### 3.4.7 Backpropagation

Backpropagation est l'algorithme utilisé pour calculer les gradients de la fonction de coût (la fonction objectif) d'un réseau de neurones. Comme un réseau de neurones peut être interprété comme une fonction composite, Backpropagation utilise le théorème de dérivation des fonctions composées pour trouver les gradients par rapport aux poids du réseau.

L'utilisation de Backpropagation pour l'entraînement de réseaux de neurones a été popularisée par David E. Rumelhart, Geoffrey E. Hinton et Ronald J. Williams [31], ils l'ont décrit comme une procédure qui ajuste de manière répétée les poids des connexions du réseau de manière à minimiser une mesure de la différence entre le vecteur de sortie réel du réseau et le vecteur de sortie souhaité. À la suite des ajustements de poids, des unités internes "cachées" qui ne font pas partie de l'entrée ou de la sortie en viennent à représenter des caractéristiques importantes du domaine de la tâche, et les régularités de la tâche sont saisies par les interactions de ces unités.

Figure 3.7 représente le Forward Pass et le Backward Pass, le Forward Pass calcule la sortie de réseau, le Backward pass calcule les gradients de la fonction de coût qui mesure la différence entre cette sortie et la sortie réelle souhaitée:

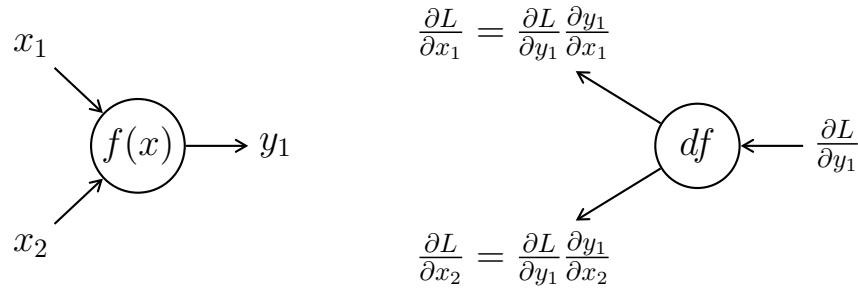


Figure 3.7: Forward (gauche) et Backward (droite) passes

## 3.5 Réseaux de neurones récurrents

Les réseaux de neurones récurrents (RNN) sont une architecture spéciale qui convient mieux à la modélisation de données séquentielles. Les RNN traitent une séquence d'entrée un élément à la fois, en maintenant dans leurs unités cachées un "vecteur d'état" qui contient implicitement des informations sur l'historique de tous les éléments passés de la séquence [32]. Figure 3.8 montre une architecture RNN. À gauche, on voit la version dépliée avec une boucle cyclique dans la couche cachée et à droite la version déroulée.  $x_1, x_2, \dots x_t$  représentent le vecteur d'entrée (séquence),  $y_1, y_2, \dots y_t$  est le vecteur de sortie (peut être une séquence de longueur égale à l'entrée, de longueur différente ou un seul élément).  $h_1, h_2, \dots h_t$  sont les neurones de la couche cachée, chaque neurone passe un vecteur de son état caché au neurone suivant.

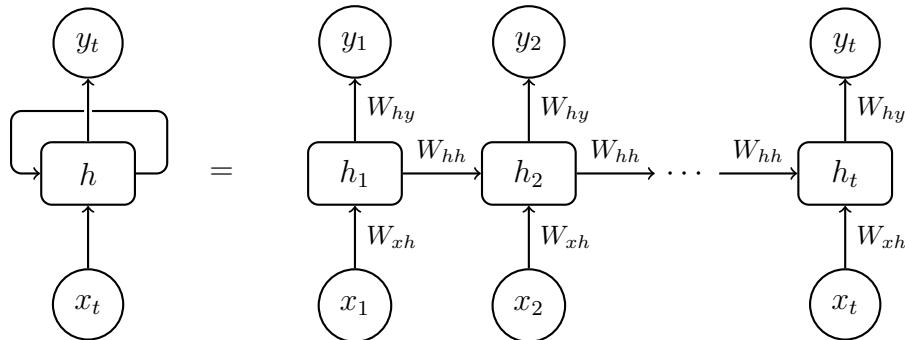


Figure 3.8: Architecture de réseaux de neurones récurrents

### 3.5.1 Long Short-Term Memory

Les RNNs simples ont des problèmes avec l'apprentissage des dépendances à long terme (c'est-à-dire faire des prédictions basées sur des prédictions faites de nombreux pas dans le passé) et disparition des gradients. Long Short-Term Memory (LSTM) a été présenté pour la première fois par Jürgen Schmidhuber et Sepp Hochreiter [33]. LSTM surmonte les problèmes liés à les RNN simples en utilisant ce que l'on appelle un état cellulaire qui permet aux réseaux LSTM d'apprendre les dépendances à long terme. Les cellules LSTM possèdent également trois types de portes (gates) qui, ensemble, contrôlent le flux d'informations à l'intérieur de la cellule :

- **Porte d'oubli** : Contrôle quelles informations sont conservées ou rejetées au

moment  $t$

- **Porte d'entrée** : Contrôle les informations à stocker dans l'état cellulaire au moment  $t$
- **Porte de sortie** : Contrôle la sortie finale de la cellule au moment  $t$

Les valeurs des portes d'entrée, d'oubli et de sortie sont calculées à l'aide des équations 3.5, 3.6 et 3.7 respectivement :

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \quad (3.5)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \quad (3.6)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) \quad (3.7)$$

$$(3.8)$$

Les portes d'entrée, d'oubli et de sortie sont calculées au temps  $t$  en utilisant des ensembles de poids et de biais ( $W_{xi}$ ,  $W_{hi}$ ,  $W_{ci}$ ,  $b_i$ ), ( $W_{xf}$ ,  $W_{hf}$ ,  $W_{cf}$ ,  $b_f$ ) et ( $W_{xo}$ ,  $W_{ho}$ ,  $W_{co}$ ,  $b_o$ ) qui contrôle comment chacun de  $x_t$ ,  $h_{t-1}$  et  $c_{t-1}$  affecte la valeur du porte respectivement. Une fonction sigmoïde est utilisée pour convertir les valeurs dans l'intervalle de 0 à 1.

À chaque étape  $t$ , un état de cellule candidate  $\tilde{c}_t$  est calculé en utilisant des poids et un terme de biais ( $W_{xc}$ ,  $W_{hc}$ ,  $b_c$ ) qui fait correspondre les valeurs de l'entrée  $x_t$  et de l'état caché précédent  $h_{t-1}$  à l'état de la cellule candidate (équation 3.9). Comme le nom l'indique,  $\tilde{c}_t$  sert de candidat pour remplacer l'état actuel de la cellule  $c_t$  au moment  $t$ . L'état de la cellule au moment  $t$  est obtenu en utilisant la porte d'oubli pour contrôler quelles informations sont conservées de l'état de la cellule précédente et la porte d'entrée pour contrôler quelles informations sont conservées de l'état de la cellule candidate selon l'équation 3.10. La sortie finale de la cellule est calculée en utilisant l'état actuel de la cellule  $c_t$  et la porte de sortie selon l'équation 3.11.

$$\tilde{c}_t = \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (3.9)$$

$$c_t = f_t c_{t-1} + i_t \tilde{c}_t \quad (3.10)$$

$$h_t = o_t \tanh(c_t) \quad (3.11)$$

Les portes contrôlent le flux d'informations à l'intérieur de la cellule LSTM, leurs valeurs vont de 0 à 1 (fonction sigmoïde) et contrôlent les informations à conserver et celles à écarter lors de la mise à jour de l'état de la cellule (portes d'oubli et d'entrée) et lors du calcul de la sortie de la cellule LSTM (porte de sortie).

La figure 3.9 illustre les différentes opérations qui se produisent dans une cellule LSTM du signle :

## 3.6 Convolutional neural networks

Les réseaux neuronaux convolutifs (Convolutional Neural Networks: CNN) ont ce qu'on appelle des couches convolutives. Une couche convulsive accepte une entrée

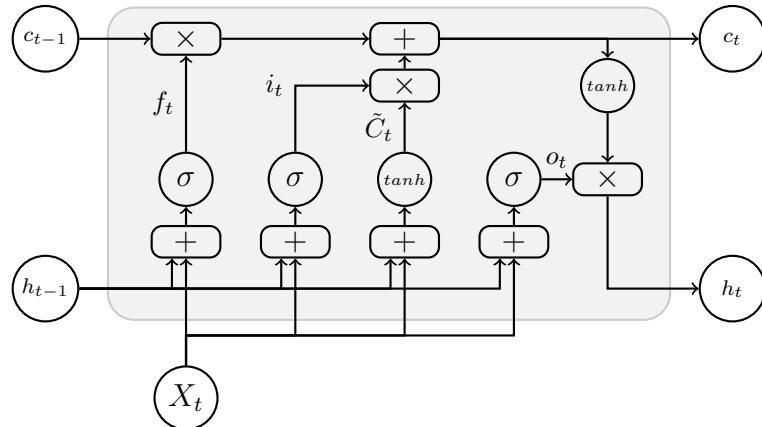


Figure 3.9: Cellule Long Short-Term Memory

(généralement des images 2D, mais elle peut aussi être 1D ou 3D) et applique une opération mathématique appelée convolution<sup>1</sup>. Les convolutions agissent comme des filtres qui extraient des caractéristiques des données brutes en multipliant récursivement ce qu'on appelle le noyau par les données d'entrée. Les différents noyaux ont des objectifs différents et peuvent extraire une grande variété de caractéristiques des données brutes. Après l'extraction de caractéristiques par la couche convective, ils servent d'entrée pour un réseau neuronal de feedforward (comme décrit dans la section 3.2) pour la classification ou la régression.

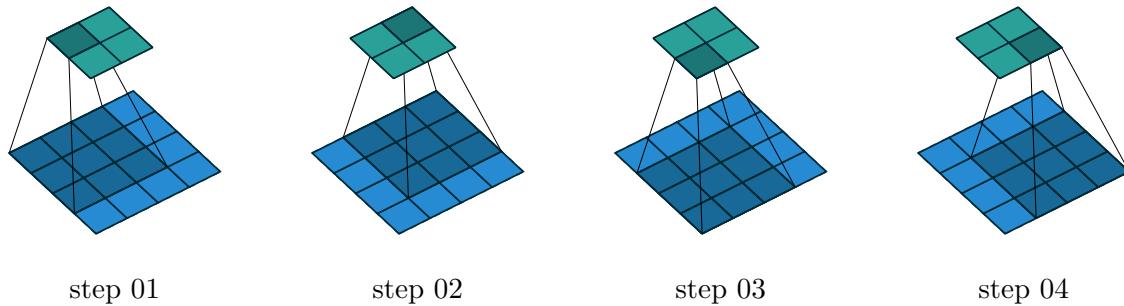


Figure 3.10: Appliquer des convolutions (en multipliant l'entrée par un noyau) de manière itérative à différentes régions (bleu foncé) de l'entrée 2D (carré bleu entier). Chaque itération donne une valeur numérique (vert foncé). Le résultat de toutes les itérations est la sortie 2D verte [3].

## 3.7 Conclusion

Les réseaux de neurones sont un outil puissant pour trouver la relation complexe entre une entrée et une sortie (e.g. les données de Condition Monitoring et la dégradation de la machine). Ils sont constitués de différentes couches de neurones interconnectées.

<sup>1</sup>Mathématiquement, l'opération qui se produit dans un CNN est appelée corrélation croisée, ce qui est un peu différent de la définition mathématique d'une convolution. Dans la littérature sur l'apprentissage automatique, ces opérations sont appelées convolutions, c'est la terminologie qui sera utilisée ici.

la connexion entre chaque 2 couches est définie par un ensemble de poids, qui sont multipliés par les valeurs de la couche précédente pour trouver les valeurs de la couche suivante, le processus est répété jusqu'à ce que la couche de sortie soit atteinte. la sortie prédictive est comparée à la sortie réelle obtenue à partir des données d'entraînement, les poids sont ajustés pour minimiser (en utilisant Backpropagation et Gradient Descent) la différence entre les prédictions et la réalité.

# Chapter 4

## Évaluation de l'État des Équipements par les Réseaux de Neurones

Ce chapitre démontre l'utilisation d'un réseau de neurones artificiels pour estimer l'état de santé d'un ensemble de turbosoufflantes à partir de l'ensemble de données C-MAPSS de la NASA et prédire la durée de vie utile restante. Il présente également différentes mesures permettant d'évaluer les performances du réseau.

### 4.1 Introduction à la base de données C-MAPSS

C-MAPSS est un outil de simulation de grands turboréacteurs commerciaux réalistes. Le logiciel est codé dans l'environnement MATLAB® et Simulink®.

Le logiciel C-MAPSS est composé de nombreux paramètres d'entrée éditables qui contrôlent la simulation. Ces entrées sont spécifiées par l'utilisateur et contrôlent de nombreux aspects de la simulation tels que le profil opérationnel, les contrôleurs en boucle fermée, les conditions environnementales, etc. [4].

La figure 4.1 est un diagramme simplifié du turboréacteur simulé montrant ses principaux éléments, comme la section de compresseur basse pression (LPC), la section de compresseur haute pression (HPC), la soufflante et la chambre de combustion. L'ensemble de données

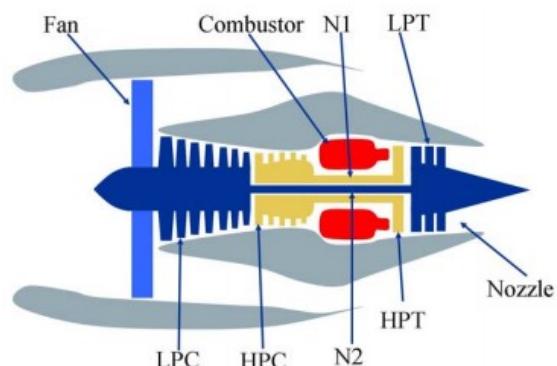


Figure 4.1: Schéma simplifié d'un turboréacteur simuler dans C-MAPSS [4]

publié par le Centre de recherche Ames de la NASA contient les données résultant de la simulation de nombreux turboréacteurs, depuis le début du fonctionnement jusqu'à la panne. L'ensemble de données a été initialement publié pour la compétition de Pronostics et gestion de la santé 2008. Le tableau 4.1 montre différentes variables, la sortie de la simulation et leurs unités, qui ont été fournies pour les participants à la compétition :

Symbol	Description	Units
<b>T2</b>	Total temperature at fan inlet	R
<b>T24</b>	Total temperature at LPC outlet	R
<b>T30</b>	Total temperature at HPC outlet	R
<b>T50</b>	Total temperature at LPT	R
<b>P2</b>	Pressure at fan inlet	psia
<b>P15</b>	Total pressure in bypass-duct	psia
<b>P30</b>	Total pressure at HPC outlet	psia
<b>Nf</b>	Physical fan speed	rpm
<b>Nc</b>	Physical core speed	rpm
<b>epr</b>	Engine pressure ratio (P50/P2)	–
<b>Ps30</b>	Static pressure at HPC outlet	psia
<b>phi</b>	Ratio of fuel flow to Ps30	pps/psi
<b>NRf</b>	Corrected fan speed	rpm
<b>NRc</b>	Corrected core speed	rpm
<b>BPR</b>	Bypass Ratio	–
<b>farB</b>	Burner fuel-air ratio	–
<b>htBleed</b>	Bleed Enthalpy	–
<b>Nf_dmd</b>	Demanded fan speed	rpm
<b>PCNfR_dmd</b>	Demanded corrected fan speed	rpm
<b>W31</b>	HPT coolant bleed	lbm/s
<b>W32</b>	LPT coolant bleed	lbm/

Table 4.1: Sortie de simulation C-MAPSS pour mesurer la réponse du système

La base de données C-MAPSS contient 4 ensembles : FD001, FD002, FD003 et FD004. Chaque ensemble a des conditions de travail et des modes de défaillance différents. Le tableau 4.2 contient les statistiques des différents ensembles :

	units number	max length	average length	min length
FD001	100	362	206.31	128
FD002	260	378	206.77	128
FD003	100	525	247.2	145
FD004	249	543	245.95	128

Table 4.2: Statistiques sur le nombre d’unités et la longueur des cycles dans la base de données C-MAPSS.

## 4.2 Visualisation de la dégradation des turboréacteurs

La visualisation des résultats de la simulation peut donner une idée de la façon dont ces variables changent pendant la durée de vie du turboréacteur. La figure 4.2 montre quatre capteurs différents de l’un des moteurs (les valeurs sont normalisées) :

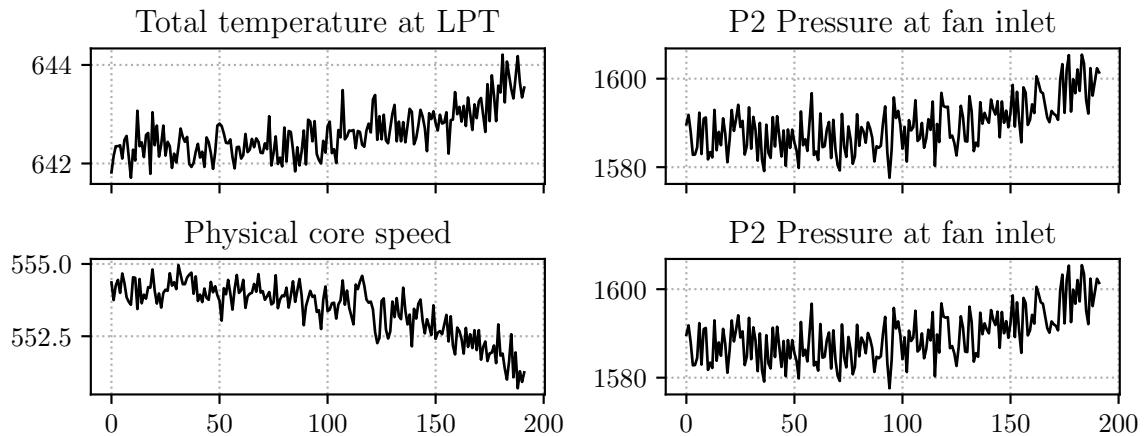


Figure 4.2: Développement de 6 capteurs de sortie d'un des turboréacteurs (normalisé)

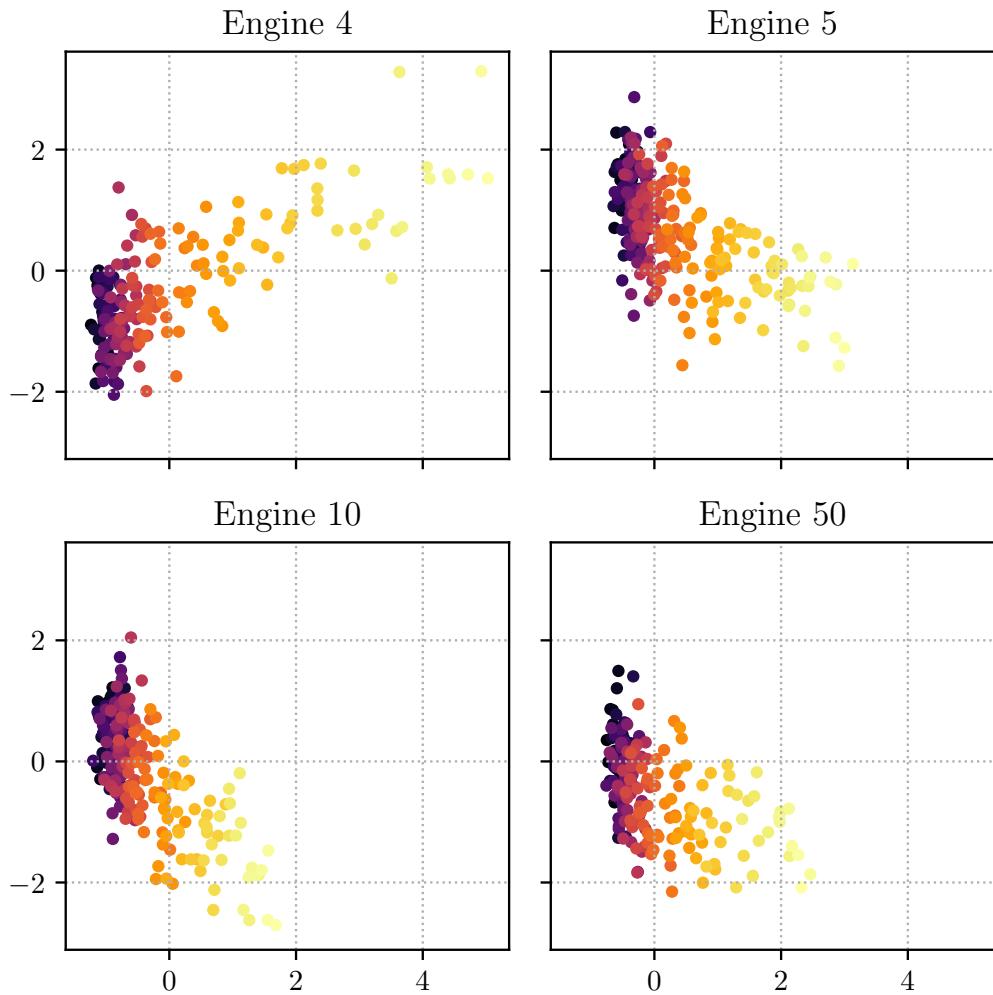


Figure 4.3: Dégradation de la santé de l'équipement (des couleurs plus claires indiquent l'avancement de la dégradation de la santé) de quatre turboréacteurs à partir de la base de données C-MAPSS

Il est évident que la sortie des capteurs suit un tendance spécifique (croissant ou décroissant) du début de l'opération jusqu'à la panne, ce qui est très utile et peut augmenter la fiabilité du modèle prédictif.

Alternativement, toutes les valeurs des capteurs peuvent être combinées et visualisées ensemble en utilisant l'analyse en composantes principales (section 2.2.4) pour révéler la tendance générale des données. Si les données de surveillance de l'état de santé de l'équipement sont directement indicatives de l'état de santé de l'équipement, la visualisation des composantes principales peut montrer des modèles de dégradation visuelle apparents.

Les valeurs des capteurs de 4 turboréacteurs différents sont combinées à l'aide de l'ACP et les deux premiers composants principaux sont représentés sur la figure 4.3.

Il existe un schéma absolument apparent de dégradation de l'état de santé dans les différentes unités, de la gauche (où les couleurs plus foncées indiquent un état de fonctionnement normal) à la droite (où les couleurs plus claires indiquent le développement de défauts).

## 4.3 classification de l'état de santé des turboréacteurs

Avant de procéder à une tâche compliquée telle que l'estimation RUL, une tâche plus simple comme la classification sanitaire des turboréacteurs peut montrer la complexité du problème. Cette section décrit l'utilisation des réseaux de neurones pour classer les états des moteurs comme étant sains ou défectueux. Les 25 premiers et derniers cycles de chaque unité sont considérés comme sains et défectueux respectivement.

Un réseau de neurones qui utilise les 24 entrées (tous les paramètres opérationnels et les capteurs) avec deux couches cachées est utilisé pour effectuer cette classification. Le tableau 4.3 résume l'architecture du réseau :

Couche (type)	Forme de la sortie	Param #
Dense1 (Dense)	(None, 8)	200
Dense2 (Dense)	(None, 4)	36
Dense3 (Dense)	(None, 1)	5
Total params: 241		
Trainable params: 241		
Non-trainable params: 0		

Table 4.3: Architecture du classificateur de l'état des unités

Le modèle est entraîné pour 200 époques avec batch size de 32 échantillons, le classificateur atteint une exactitude de **93,46%** sur les données de test. La figure 4.4 montre le processus d'entraînement du réseau. Dans le diagramme du haut, l'axe des y correspond aux pertes d'entraînement et de validation (crossentropie binaire). L'axe des y dans le diagramme du bas correspond aux précisions d'entraînement et de validation du réseau, l'axe des x partagé entre les deux diagrammes indique les époques de formation. Bien que l'exactitude de validation varie beaucoup, l'exactitude d'entraînement continue de s'améliorer à chaque époque.

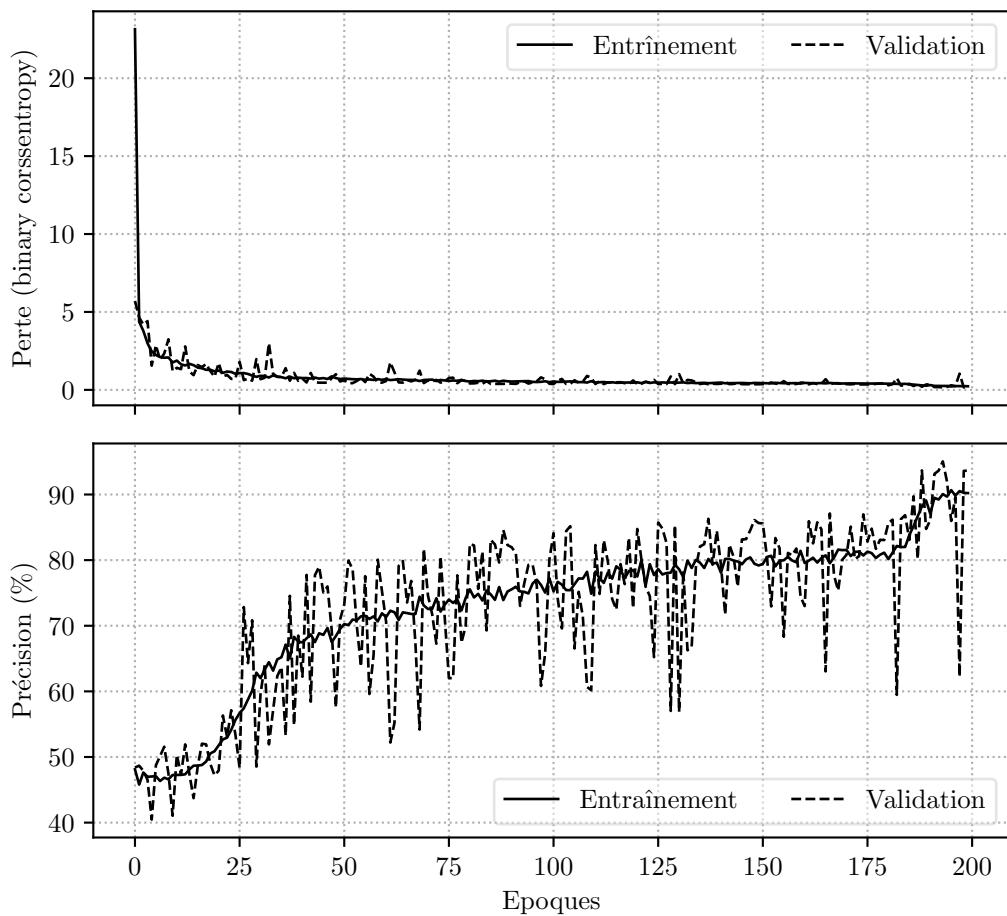


Figure 4.4: Processus d'entraînement du classificateur de l'état des unités

La figure 4.5 montre la courbe Receiver Operating Characteristics (ROC) du classificateur avec un haut Area Under Curve (AUC). D'autres mesures de classification sont présentées dans le tableau 4.4.

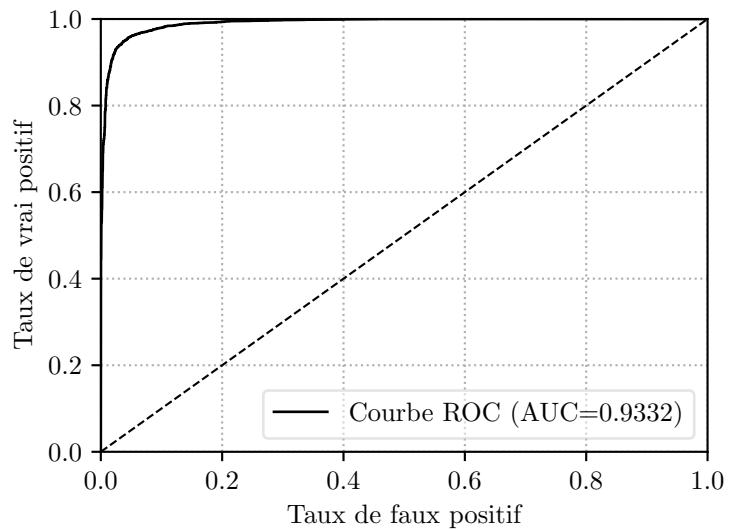


Figure 4.5: Courbe ROC du classificateur sur les données de test

Métrique	Exactitude	Précision	Rappel	F-1	ROC AUC
Valeur	93.43%	0.90	0.98	0.94	0.9332

Table 4.4: Métriques du classificateur sur les données de test

## 4.4 Prédiction de RUL

### 4.4.1 Modélisation de RUL

Afin d’entraîner un réseau de neurones pour l’estimation de Remaining Useful Life (RUL) d’une nouvelle donnée non vue provenant de base de données C-MAPSS, un RUL approprié correspondant aux données d’entraînement doit être construit.

L’expression RUL a été définie dans la section 1.2, à partir de cette définition, plusieurs approches pour construire un RUL approprié peuvent être développées pour les unités dans les données d’entraînement (où le nombre total de cycles avant la défaillance est connu). L’approche la plus simple consiste à utiliser un RUL toujours décroissant, ce qui implique que l’état de l’équipement est toujours décroissant et va vers la défaillance.

Le problème de cette approche est que le processus de dégradation n’est pas linéaire et que dans la vie réelle, les machines ne commencent pas à se dégrader dès qu’elles commencent à fonctionner. Une deuxième approche consiste à utiliser une fonction par morceaux où l’état de l’équipement est d’abord constant puis, à un point spécifique, il commence à se dégrader de manière linéaire.

Cette approche est bien meilleure que la première, mais elle n’est pas très représentative du comportement réel du processus de dégradation. Dans le contexte de ce mémoire, RUL est abordé comme une fonction polynomiale non linéaire où la dégradation est lente au début puis s’accélère vers la fin de la vie. La figure 4.6 montre les différents choix de modélisation de RUL :

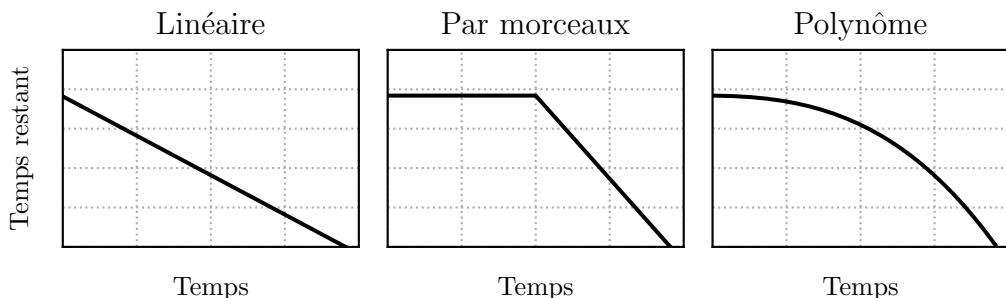


Figure 4.6: Différents modèles de RUL

Il convient de noter qu’aucune de ces approches ne peut être considérée comme représentative du processus de dégradation réel, mais le choix de la représentation la plus intuitive de RUL peut aider l’algorithme d’apprentissage à connaître les tendances de dégradation implicites dans les données.

#### 4.4.2 Prédiction de RUL par un réseau de neurones

La dernière section portait sur la classification de l'état de santé des unités comme étant sain ou défectueux. Cette section présente plutôt un réseau de neurones pour l'estimation de la RUL. Dans cette section, une architecture de réseau de neurones avec 3 couches cachées est utilisée. Comme il s'agit d'un problème de régression, l'erreur quadratique moyenne est utilisée comme fonction de perte, l'erreur absolue moyenne est utilisée comme mesure pour le modèle. Le tableau 4.5 présente les détails de l'architecture :

Couche (type)	Forme de la sortie	Param #
Dense1 (Dense)	(None, 32)	800
Dense2 (Dense)	(None, 16)	528
Dense3 (Dense)	(None, 8)	136
Dense4 (Dense)	(None, 1)	9
Total params:	1,473	
Trainable params:	1,473	
Non-trainable params:	0	

Table 4.5: Architecture d'un réseau de neurones pour la prédiction de RUL

Le réseau a été entraîné pour 300 époques, avec batch size de 128 échantillons. La figure 4.7 montre le processus de entraînement du réseau. Les figures montrent l'évolution des pertes d'entraînement et de validation (erreur quadratique moyenne) et des métriques (erreur absolue moyenne) respectivement en fonction des époques.

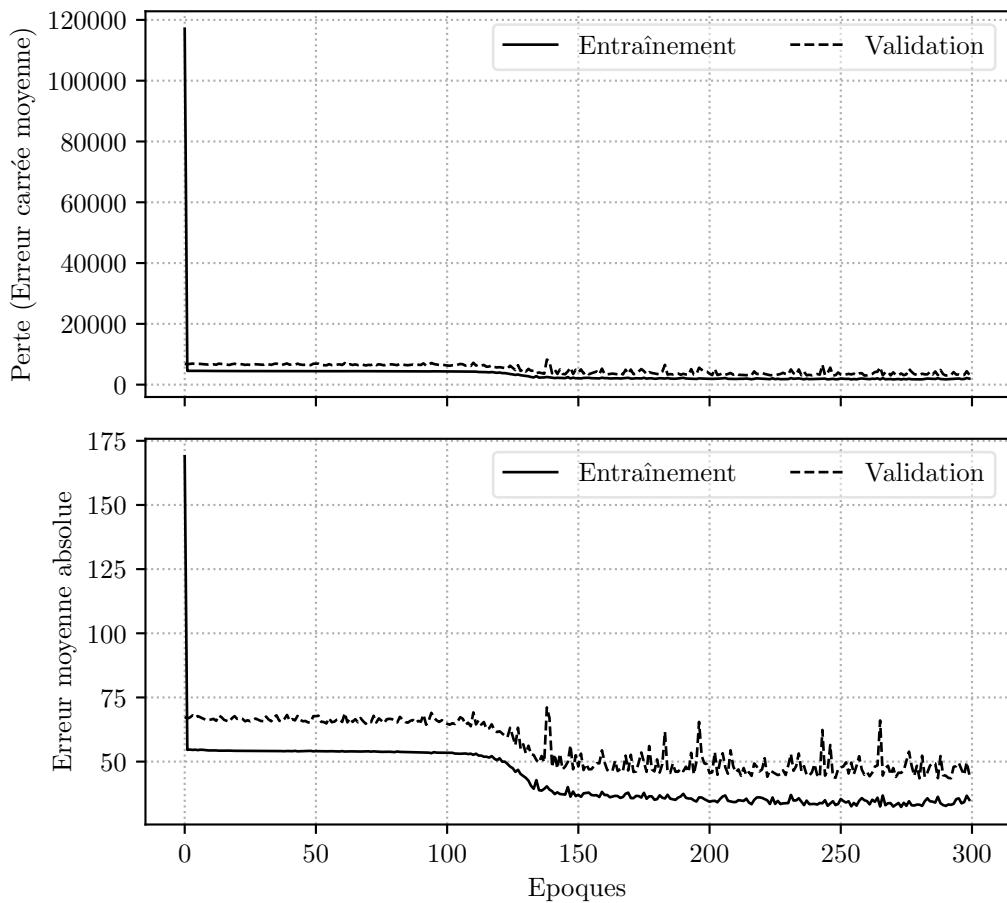


Figure 4.7: Processus d’entraînement du réseau utilis  pour pr dire RUL

Apr s l’entra nement, le mod le est ´valu  sur deux unit s de la s rie de tests. La figure 4.8 montre le RUL r  el (ligne discontinue), la pr dition du mod le  chaque cycle et un ajustement polyomial de 3 me degr  des pr dictions du mod le.

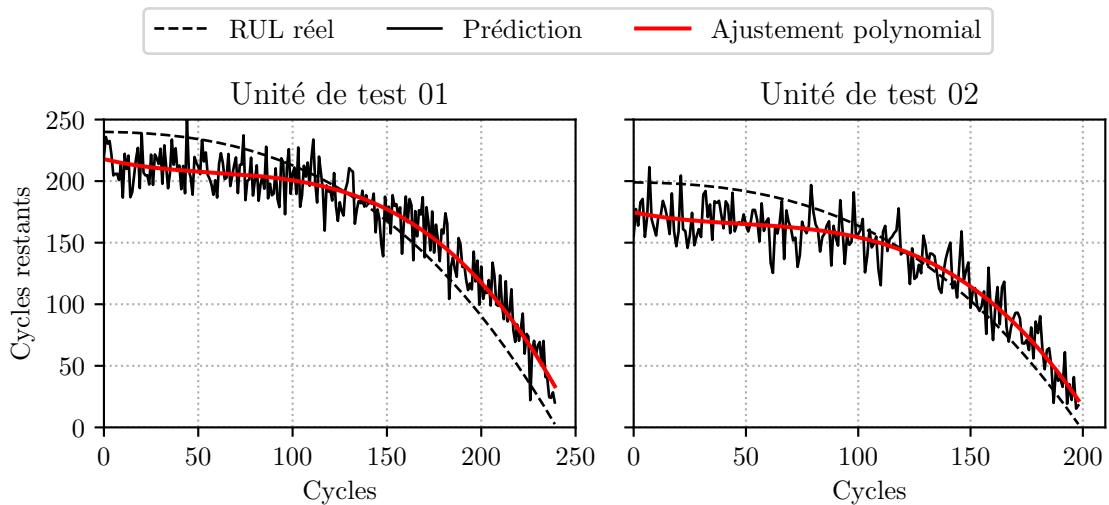


Figure 4.8: R  ultats de pr dition de RUL par un r  seau de neurones

Bien que les prévisions du modèle soient proches de la réalité, mais elles sont bruyantes et il y a beaucoup de fluctuation. Le bruit peut être réduit en lissant la sortie, par exemple en utilisant une moyenne mobile ou en adaptant les points à un polynôme. Mais la raison du bruit en premier lieu est que l’architecture entièrement connectée ne prend pas en considération les prédictions précédentes (c’est-à-dire l’architecture acyclique). Étant donné qu’à chaque instant, la RUL dépend de la RUL de la précédente, le fait de prendre en considération les étapes précédentes tout en faisant de nouvelles prédictions peut effectivement réduire le bruit et aussi améliorer la précision des prédictions. Pour y parvenir, il convient d’utiliser une architecture neurale cyclique comme les réseaux de neurones récurrents.

#### 4.4.3 Amélioration de la prédiction RUL en utilisant les réseaux LSTM

Les réseaux de neurones entièrement connectés constituent un outil puissant pour la modélisation d’un large éventail de problèmes, mais l’utilisation de cette architecture pour la prédiction de séries chronologiques, comme la prédiction RUL, peut donner un résultat très bruyant car le réseau est acyclique et chaque instant dans le temps est évalué séparément et ne tient pas compte des prédictions précédentes. Long Short-Term Memory (Section 3.5.1) est un outil puissant pour modéliser de tels problèmes et pour faire des prédictions plus fiables et moins bruyantes. Dans cette section, l’architecture de réseau de neurones entièrement connectée de la section précédente est remplacée par un réseau LSTM pour prédire RUL des turboréacteurs de la base de données C-MAPSS.

L’architecture utilisée ici est décrite dans le tableau 4.6:

Couche (type)	Forme de la sortie	Param #
LSTM1 (LSTM)	(None, 100, 100)	50000
LSTM2 (LSTM)	(None, 100, 100)	80400
LSTM3 (LSTM)	(None, 75)	52800
Dense1 (Dense)	(None, 120)	9120
Dense2 (Dense)	(None, 110)	13310
Dense3 (Dense)	(None, 100)	11100
Total params:	216,730	
Trainable params:	216,730	
Non-trainable params:	0	

Table 4.6: Architecture de réseau LSTM pour la prédiction de RUL

Il est évident que l’architecture LSTM a beaucoup plus de paramètres (216,730 paramètres) que l’architecture entièrement connectée (1,473). Cela est dû à la conception plus complexe des cellules LSTM et à leur possession de différentes portes. Il en résulte un temps d’ entraînement beaucoup plus long.

Les couches LSTM prennent un tenseur tridimensionnel de la forme (*samples, sequence length, features*) comme entrée. La longueur de la séquence a été fixée à 100 dans cette architecture, c’est pourquoi chaque échantillon d’ entraînement doit avoir une forme de (*sequence length, features*). Toutes les unités ont le même nombre de caractéristiques

mais des longueurs différentes, puisque chaque échantillon doit avoir une longueur de séquence de 100, les échantillons avec des cycles inférieurs à 100 sont complétés par -10 où les couches LSTM sont définies pour ignorer tout les instants de temps avec cette valeur.

Le réseau a été entraîné pour 50 époques avec batch size de 16 échantillons et un fractionnement de validation de 0,2 en utilisant l'algorithme d'optimization Adam. Le processus d'entraînement est visualisé sur la figure 4.9 :

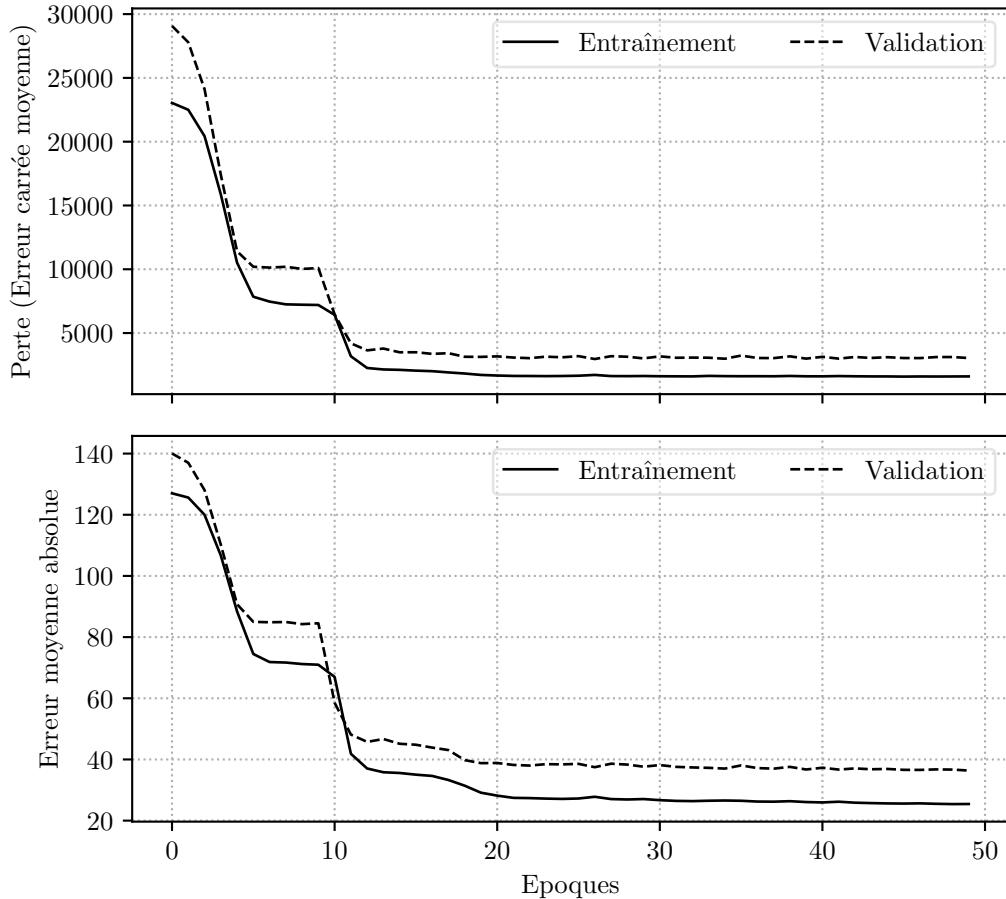


Figure 4.9: Processus d'entraînement du réseau LSTM

Le tableau 4.7 montre la perte et la métrique (c'est-à-dire l'erreur absolue moyenne) sur les données d'entraînement, validation et de test :

	Perte	Erreur absolue moyenne
Ensemble d'entraînement	1600.60	25.43
Ensemble de validation	3039.68	36.37
Ensemble de test	1379.67	23.27

Table 4.7: Résultats d'entraînement du réseau LSTM

Quatre unités différentes ont été réservées comme unités de test. Après l'entraînement,

le réseau est utilisé pour prédire RUL sur les unités de test. La figure 4.10 montre les résultats de la prédiction et la RUL réelle de deux unités différentes :

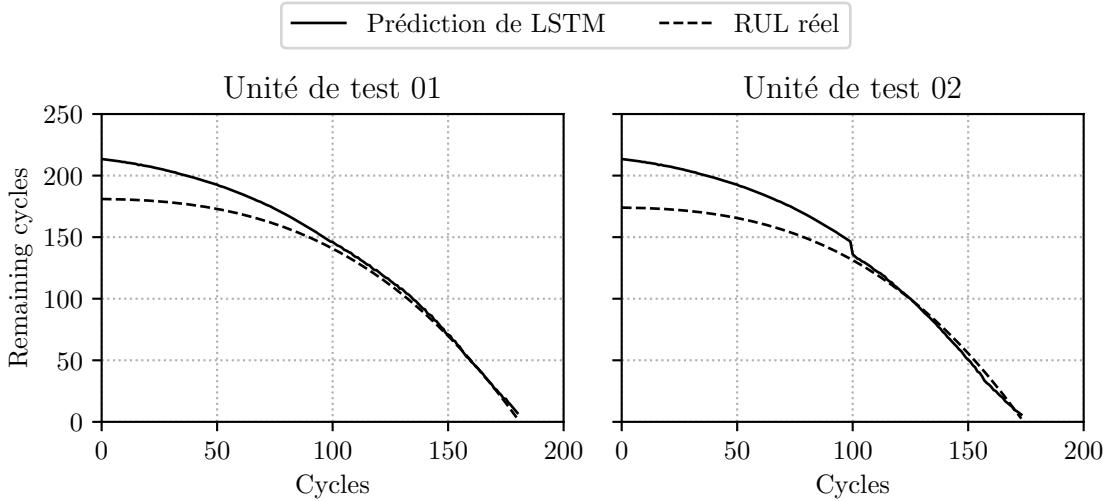


Figure 4.10: Résultats de prédiction du réseau LSTM

Il est évident que les prédictions du réseau LSTM sont bien meilleures et presque exemptes de bruit par rapport à celles faites par le réseau entièrement connecté, comme le montre la figure 4.8. La prédiction RUL est presque identique à la prédiction réelle RUL vers la fin de vie du turboréacteur.

## 4.5 Application aux équipements des chantiers pétroliers

Ce chapitre a présenté une approche de maintenance prédictive pour l'estimation des RUL basée sur les données de surveillance de l'état fournies par différents capteurs qui peuvent mesurer différentes variables physiques qui sont liées à l'état de santé de l'équipement. Les données utilisées dans les sections précédentes consistaient en une base de données de turboréacteurs C-MAPSS, mais la même approche peut être étendue à d'autres applications comme les plates-formes pétrolières.

Les chantiers pétroliers modernes contiennent un grand nombre de capteurs répartis sur presque tous les équipements critiques. Ces capteurs mesurent différentes variables physiques telles que la température, le débit, la pression.... Mais la majorité des données fournies par ces capteurs n'est pas correctement exploitée dans un cadre approprié de pronostics et de maintenance prédictive. Il existe d'énormes quantités de données historiques dans les chantiers pétroliers qui, si elles sont exploitées, peuvent améliorer considérablement les programmes de maintenance dans des applications aussi critiques où la réduction des temps d'arrêt est très importante et où l'indisponibilité des équipements peut avoir causé des pertes de production importantes.

La figure 4.11 montre différents endroits où les capteurs sont installés sur la plupart des chantiers pétroliers modernes. Le tableau 4.8 énumère les types de différents capteurs qui peuvent être utilisés pour les pronostics, les unités auxquelles ils sont attachés et les variables physiques qu'ils peuvent mesurer. Un réseau de neurones peut être entraîné sur les données historiques de la plate-forme et être utilisé pour surveiller les

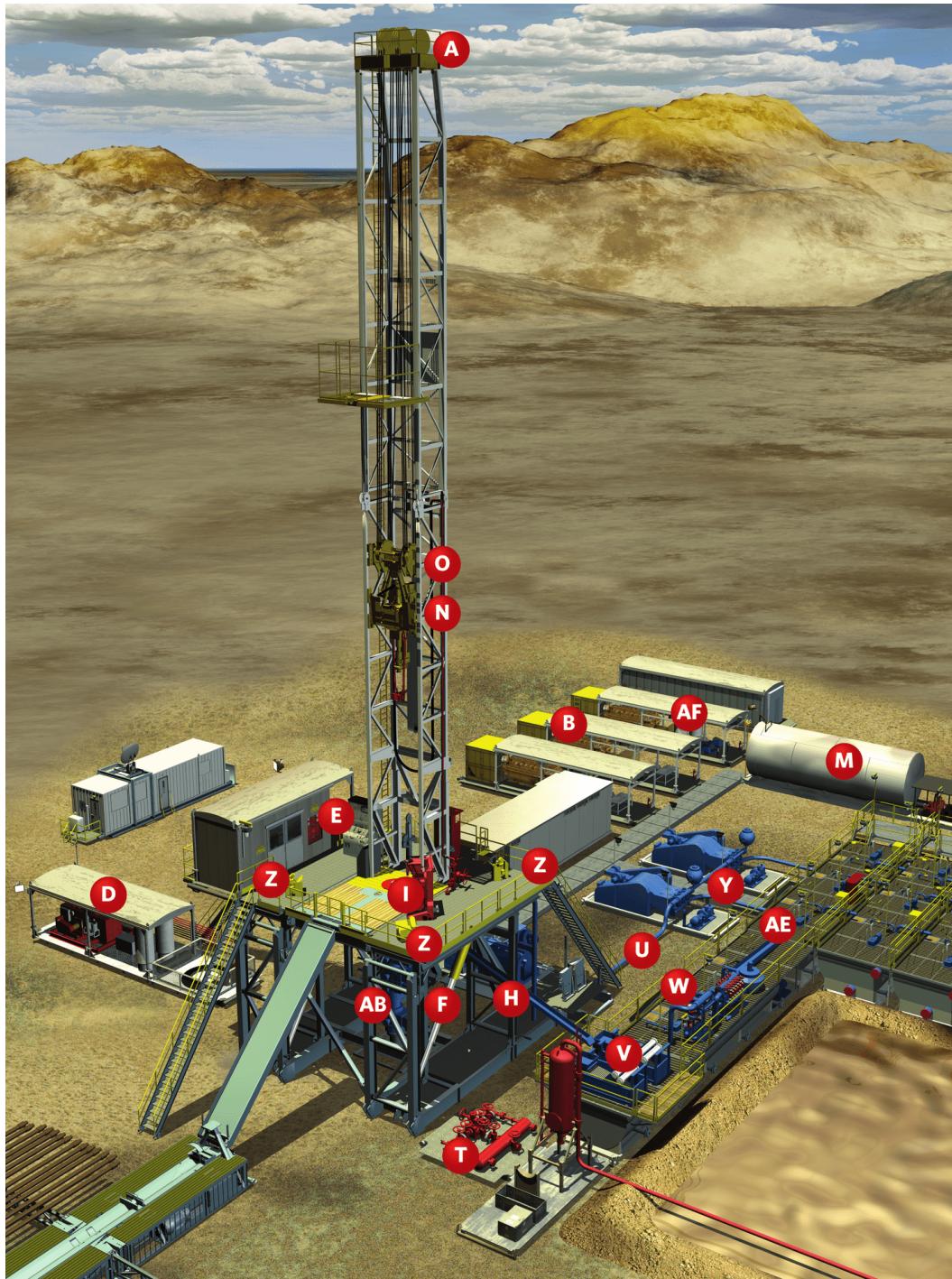


Figure 4.11: Emplacement des capteurs dans un chantier pétrolier [5]

différents équipements, leur interaction, détecter les anomalies et les signaler à temps afin de mener les actions de maintenance appropriées.

## 4.6 Conclusion

Les réseaux de neurones entièrement connectés sont un outil puissant pour quantifier l'état de santé de systèmes complexes à l'aide de données de surveillance des conditions. L'ensemble de données C-MAPSS est un exemple de système comportant de nombreuses parties en interaction qui fournissent une variété de données de surveillance de l'état de santé (par exemple, la température, la pression, le régime, ...) où la dégradation n'est pas directement liée à une composante mais est indiquée par la somme des tendances de toutes les données de surveillance. Les réseaux de neurones sont capables de saisir de telles tendances et de prédire RUL du système avant une défaillance. Les réseaux entièrement connectés ont d'abord été utilisés pour cette tâche, mais une architecture cyclique comme LSTM peut produire de meilleurs résultats avec une précision accrue et beaucoup moins de bruit. Une approche pour intégrer les réseaux de neurones dans le cadre de maintenance préventive et pronostic dans les chantiers pétroliers en exploitant les différents capteurs et les données historiques disponibles pour créer un programme de maintenance plus fiable et moderne.

Capteur	Equipment	Type de capteur	Value mesurée
A	Crown Block	Load cell	Weight on drill line via cable tension
B	Power Generation Unit	Pressure	Oil, water, and hydraulic fluid pressure
D	Accumulator Unit	Pressure	Inlet/outlet pressure with high accuracy
F	Rig Hydraulic Lift	Pressure	Hydraulic pressure, weight, force/strain, or movement, monitor raising or lowering deck for directional drilling
H	Drawworks	Load cells	Torque, load/weight/position while guiding pipe into position
M	Water/Storage Tank	Pressure	Tank liquid levels
N	Top Drive	Torque/Pressure	Monitor torque/twisting movement to ensure right amount of force is applied. Weight on drill bit. Hydraulic pressure and feed information into control system.
O	Trabeling Block	Load cells	Weight on the drill line via cable tension
R	Deadline anchor	Load cells	Tension on deadline/drilling line cable
U	Mud Return Line	Pressure	Drilling mud pressure to monitor and control mud flow
Y	Mud Pump	Pressure	Pressure and flow of mud media
AB	BlowOut Preventor	Pressure	Monitor RAM position via hydraulic volumetric or pressure behind the piston
AD	Drill Bit	Pressure	Pressure or differential pressure at high temperature and pressure ranges
AE	Fluid manifold	Pressure	Drilling fluid pressure
AF	Mud Tank/Reservoir	Pressure	Tank liquid levels

Table 4.8: Liste des capteurs dans un chantier pétrolier [5]

# Chapter 5

## Diagnostic et Pronostic des Roulements avec les Réseaux de Neurones

La surveillance des vibrations est vitale pour de nombreux systèmes industriels. Les mesures des vibrations contiennent des informations très utiles sur l'état de santé des équipements et les types de défauts. Néanmoins, obtenir des informations à partir de ces signaux dans des applications réelles se révèle être un processus complexe. Cela est principalement dû à la complexité du problème. Ce chapitre présente plusieurs approches pour le diagnostic et le pronostic des défauts de roulements à l'aide de réseaux de neurones convolutifs.

### 5.1 Etude de cas: Diagnostique des roulements

Les données utilisées dans cette section sont des données de vibrations des roulements fournies par Case Western Reserve University (CWRU). Les roulements utilisés dans le test sont des roulements à billes SKF. La figure 5.1 montre les différents composants d'un roulement à billes standard. L'essai a été réalisé là où les roulements supportent l'arbre d'un moteur de 2 chevaux dans différentes conditions de charge.

Les roulements d'essai possèdent des défauts ponctuels qui ont été introduits

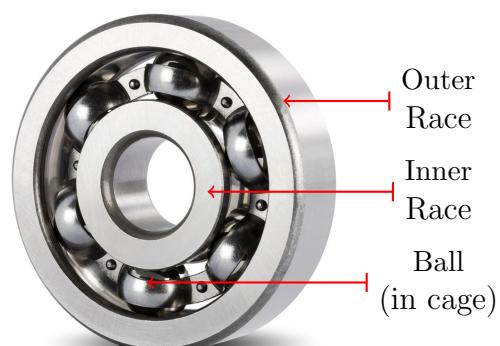


Figure 5.1: Composants d'un roulement

par électroérosion avec des diamètres de défaut de 0,18 mm, 0,36 mm, 0,53 mm, 0,71 mm et 1,02 mm. Ces défauts ont été introduits dans la bille et les chemins de roulement intérieurs et extérieurs du roulement. Des roulements SKF ont été utilisés pour les défauts de 0,18 mm, 0,36 mm et 0,53 mm, et des roulements équivalents NTN ont été utilisés pour les défauts de 0,71 mm et 1,02 mm. Le tableau 5.1 contient les dimensions des modèles de roulements SKF utilisés et les fréquences correspondantes (en multiples de RPM) associées aux différents types de défauts.

Dimension	Valeur (mm)	Défaut	Fréquence ( $\times$ RPM Hz)
Diamètre intérieur	25.00	Chemin intérieur	5.4152
Diamètre extérieur	52.00	Chemin extérieur	3.5848
Epaisseur	15.00	Cage	0.3983
Diamètre du pas	08.03	Bille	4.7135

Table 5.1: Dimensions des roulements CWRU et les fréquences de défauts

### 5.1.1 Génération de données à partir de vibration

Les données de vibration non traitées ne peuvent pas être utilisées directement pour entraîner un réseau de neurones. Ce chapitre utilise l'approche proposée dans [34] pour convertir les données de vibration non traitées en images. La figure 5.2 montre le principe de génération de données où les signaux de vibration à une dimension sont convertis en matrices à deux dimensions (images) en transformant des morceaux de longueur 4096 en matrices de  $64 \times 64$ .

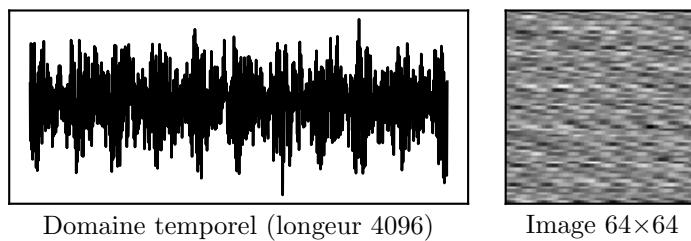


Figure 5.2: Génération de données par conversion du signal en images  $64 \times 64$

Comme mentionné précédemment, plusieurs essais ont été réalisés avec différents types de défauts de roulements (c'est-à-dire des défauts de billes, de chemins de roulement intérieurs et extérieurs) avec différents diamètres de défaut. Les signaux des différents tests sont transformés en images pour servir d'entrée à un réseau de neurones convolutif qui sera entraîné à classer les différents signaux dans les types de défauts correspondants et leurs diamètres. Les signaux sont répartis en 10 classes : trois types de défauts différents et pour chaque type de défaut, il y a trois diamètres de défaut différents, plus un signal de base normal qui appartient au roulement sain. La figure 5.3 montre quelques échantillons des signaux transformés des neuf différents types de défaut/diamètres :

Les signaux des différents types de défauts ont des longueurs variables, ce qui se traduit

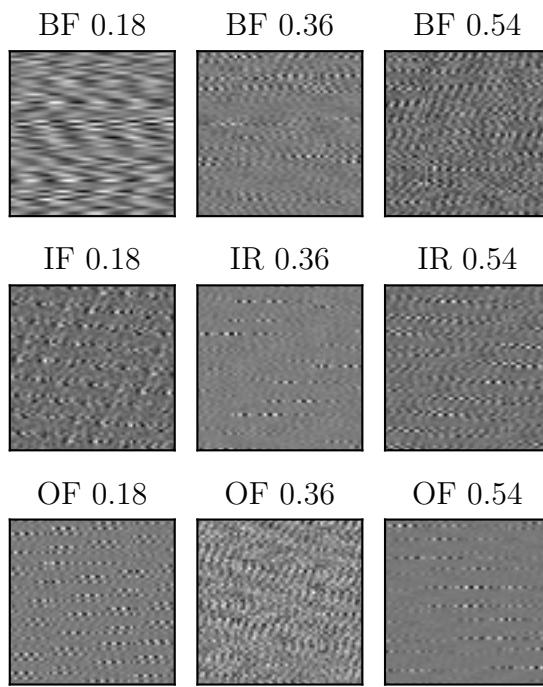


Figure 5.3: Signaux convertis de différents types de défauts

par un nombre différent d'images synthétisées par type de défaut. Le nombre d'images par type de défaut (classe) est donné par l'équation 5.1 :

$$N = \text{floor} \left( \frac{\text{signal length}}{64 \times 64} \right) \quad (5.1)$$

Le nombre d'images correspondant à chaque classe est indiqué dans le tableau 5.2:

Classe	Nombre d'échantillons
Normal bearing	295
Roller element 0.18mm	146
Roller element 0.36mm	116
Roller element 0.54mm	116
Inner race 0.18mm	295
Inner race 0.36mm	116
Inner race 0.54mm	116
Outer race 0.18mm	116
Outer race 0.36mm	116
Outer race 0.54mm	116
Total	1548

Table 5.2: Nombre d'échantillons pour chaque classe de défauts

### 5.1.2 Diagnostic des défauts de roulements à l'aide de réseaux de neurones

Après avoir généré des données en convertissant des signaux vibratoires bruts en images, ces images servent d'entrée pour un réseau de neurones convolutif.

#### Architecture de réseau

Convolutional Neural Networks (CNN) décrit un type de réseaux de neurones qui convient au traitement des images. Cette section présente l'utilisation de CNN pour classer les signaux de vibration des roulements qui ont été transformés en images dans les types de défauts correspondants. Les CNN ont été décrits en détail dans la section 3.6.

Pour effectuer cette tâche de classification, une architecture CNN est utilisée, le réseau se compose de trois couches convolutionnelles avec des couches MaxPool entre elles, suivies de trois couches Fully-Connected. Tous les détails de cette architecture sont mentionnés dans le tableau 5.3.

Couche (type)	Forme de la sortie	Param #
Conv1 (Conv2D)	(None, 1, 64 ,32)	18464
MaxPool1 (MaxPool2D)	(None, 1, 32, 32)	0
Conv2 (Conv2D)	(None, 1, 32, 64)	18496
MaxPool2 (MaxPooling2D)	(None, 1, 16, 64)	0
Conv3 (Conv2D)	(None, 1, 16, 128)	73856
MaxPool3 (MaxPooling2D)	(None, 1, 8, 128)	0
Flatten1 (Flatten)	(None, 1024)	0
Dense1 (Dense)	(None, 128)	131200
Dense2 (Dense)	(None, 64)	8256
Dense3 (Dense)	(None, 10)	650
Total params:	250,922	
Trainable params:	250,922	
Non-trainable params:	0	

Table 5.3: Architecture CNN du classificateur de défauts de roulements

#### Processus d'entraînement

La base de données contient au total 1 548 échantillons, dont 20 % sont utilisés comme test, les 80 % restants étant destinés à l'entraînement. De l'ensemble d'entraînement, 15 % sont utilisés comme fractionnement de validation pendant le processus d'entraînement. Le réseau a été entraîné pour 30 époques et batch size de 32. La figure 5.4 montre l'évolution de l'exactitude (%) et de la perte (entropie croisée catégorique) en fonction des époques d'entraînement. Le réseau converge vers la 10ème époque avec une légère augmentation de l'exactitude et une diminution de la perte par la suite.

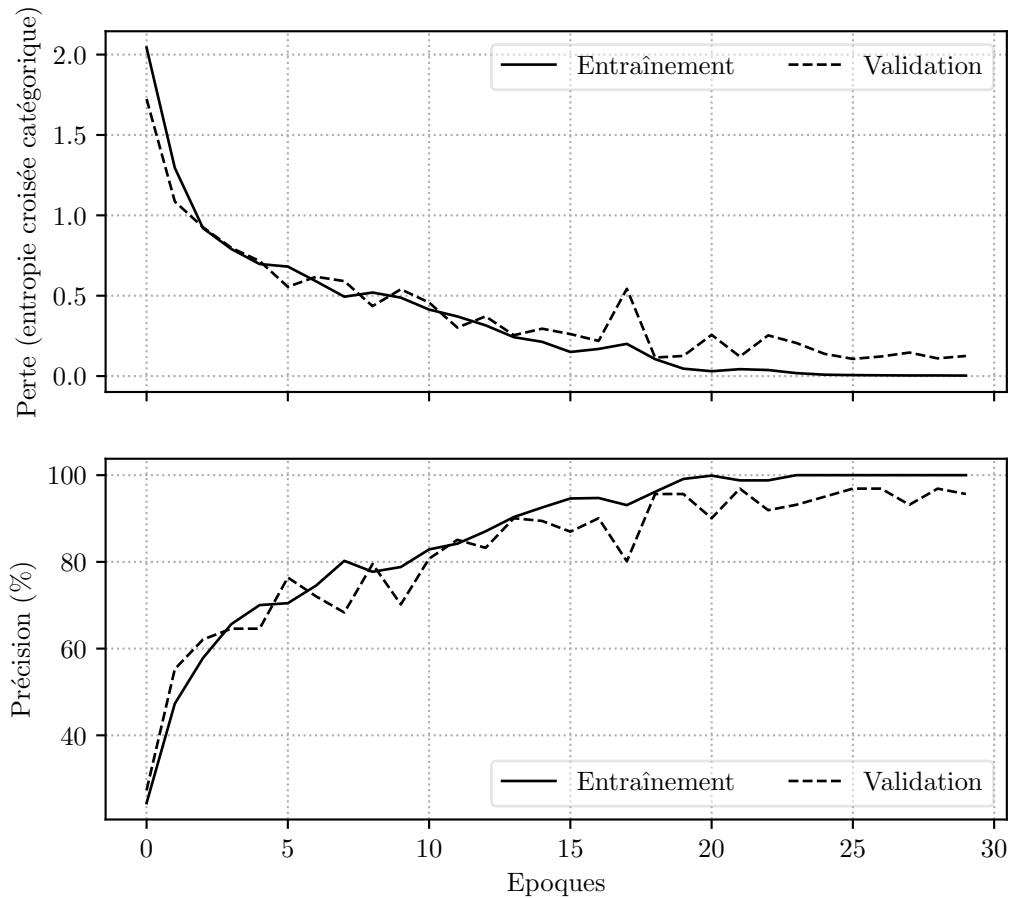


Figure 5.4: Entraînement de classificateur de défauts de roulements

### Discussion des résultats

Le modèle atteint une exactitude parfaite de 100 % sur l'ensemble d'entraînement et une exactitude de 95.65% sur l'ensemble de validation. Après l'entraînement, le modèle est évalué sur l'ensemble de test et atteint une exactitude très élevée 95.90 %. L'entraînement, la validation et la perte et l'exactitude du test sont résumées dans le tableau 5.4.

	Perte	Exactitude
Ensemble d'entraînement	0.0030	100.00%
Ensemble de validation	0.1251	95.65%
Ensemble de test	0.1074	95.90%

Table 5.4: Résultats de l'entraînement du classificateur de défauts de roulements

Pour mieux comprendre les résultats du modèle, une matrice de confusion est construite. La matrice de confusion (Figure 5.5) montre les résultats des prédictions du modèle sur l'ensemble de test où l'axe des y montre les classes réelles (réelles) dans l'ensemble de test et l'axe des x les prédictions du modèle. La diagonale de la matrice de confusion est l'endroit où les classes réelles croisent les classes prédites par le modèle

pour chaque échantillon et il est évident que le modèle atteint une classification quasi parfaite.

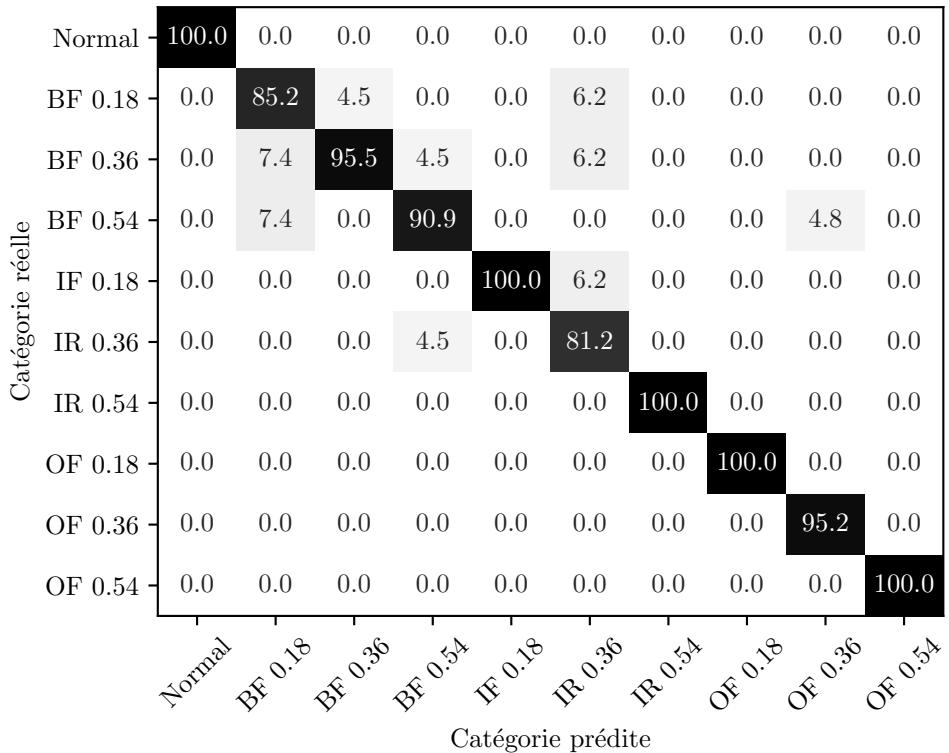


Figure 5.5: Matrice de confusion de la classification des défauts de roulements à l'aide de CNN

## 5.2 Etude de cas: Pronostique des roulements

Section 5.1 a présenté une approche pour diagnostiquer les défauts de roulements en utilisant les données de vibrations et les réseaux de neurones. Le diagnostic est un aspect important de la maintenance en général, mais il ne suffit pas à lui seul. Dans un programme de maintenance prédictive et dans le contexte de l'PHM, il est également important de surveiller l'équipement, d'estimer son état de santé et de le projeter dans le futur pour prédire l'RUL et c'est ce qui est couvert dans cette section. C'est ce qui est traité dans cette section. La base de données des roulements FEMTO est utilisée.

### 5.2.1 Introduction au base de données FEMTO

La base de données FEMTO Bearings a été fournie par l'institut FEMTO-ST (Besançon - France) dans le cadre du PHM 2012 Data Challenge. La base de données est constituée de mesures de vibrations et de températures provenant d'expériences de vieillissement accéléré (run to failure) de roulements qui ont été réalisées à l'aide de la plate-forme PRONOSTIA.

PRONOSTIA (Figure 5.6) est une plateforme expérimentale utilisée pour les tests de vieillissement accéléré des roulements. PRONOSTIA permet de réaliser ces tests en

quelques heures. Les données fournies par la plate-forme correspondent à des roulements naturellement dégradés plutôt qu'à des roulements présentant des défauts artificiels (par exemple, les données des roulements CWRU).

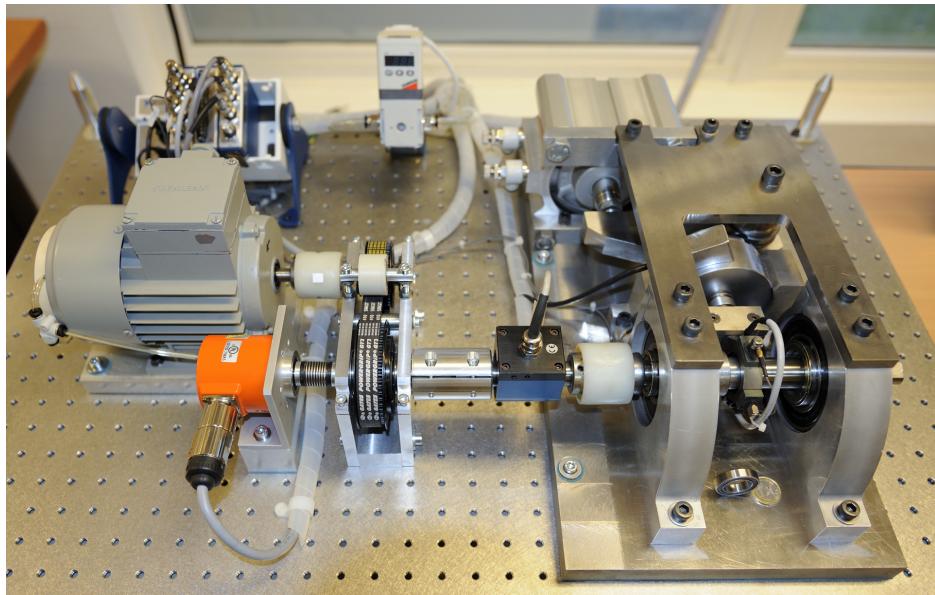


Figure 5.6: Plateforme PRONOSTIA [6]

La plateforme PRONOSTIA se compose de trois parties principales :

- **Partie tournante** : Moteur asynchrone d'une puissance de 250W qui transmet un mouvement de rotation par l'intermédiaire d'un réducteur qui permet une vitesse de rotation jusqu'à 2830 tr/min. La vitesse et le sens de rotation du moteur sont contrôlés par l'utilisateur par le biais d'une interface homme-machine (IHM).
- **Pièce de charge** : Qui est utilisé pour appliquer une force radiale afin d'induire des défauts dans les roulements. En appliquant une charge qui dépasse la charge dynamique maximale du roulement (4000 N), cette pièce permet de réduire drastiquement la durée de vie du roulement et de réaliser le test de vieillissement accéléré. La charge est générée à l'aide d'un actionneur de force qui consiste en un vérin pneumatique alimenté par une pression délivrée par un régulateur électropneumatique numérique.
- **Pièce de mesure** : Les conditions de fonctionnement comprennent trois variables : a) la force radiale appliquée, b) la vitesse de l'arbre et c) le couple infligé aux roulements, tous trois échantillonnés à une fréquence de 100Hz. La santé des roulements est mesurée par deux variables : a) les vibrations (horizontales et verticales) qui sont mesurées par deux accéléromètres placés sur la bague extérieure du roulement et b) la température mesurée par une sonde RTD (Resistance Temperature Detector) placée à proximité de la bague extérieure. Les vibrations sont échantillonnées à 26,5 kHz et la température à 10 Hz. Les mesures de vibrations consistent en des instantanés de 2560 échantillons prélevés toutes les 10 minutes, tandis que les mesures de température sont de 600 échantillons prélevés chaque minute.

Les roulements dans les bases de données sont regroupés sous trois conditions de fonctionnement :

- Premières conditions de fonctionnement : vitesse de rotation de 1800 tr/min et charge de 4000N.
- Deuxième conditions de fonctionnement : vitesse de rotation de 1650 tr/min et charge de 4200N.
- Troisième condition de fonctionnement : vitesse de rotation de 1500 tr/min et charge de 5000N.

Le tableau 5.5 montre différentes conditions de fonctionnement et leurs roulements correspondants dans les bases de données d'apprentissage et de test :

Datasets	Operation conditions		
	Conditions 1	Conditions 2	Conditions 3
Learning set	Bearing1_1	Bearing2_1	Bearing3_1
	Bearing1_2	Bearing2_2	Bearing3_2
Testing set	Bearing1_3	Bearing2_3	
	Bearing1_4	Bearing2_4	
	Bearing1_5	Bearing2_5	Bearing3_3
	Bearing1_6	Bearing2_6	
	Bearing1_7	Bearing2_7	

Table 5.5: Différentes conditions de fonctionnement et roulements correspondants dans la base de données FEMTO

### 5.2.2 Génération de données à partir des scaleogrammes

Il est important de détecter la dégradation des roulements avant que la défaillance réelle ne se produise. Dans cette section, une architecture de réseau de neurones convolutifs sera utilisée pour classer les roulements sains et les roulements défectueux à l'aide des scaleogrammes provenant des données de vibration de la base de données de roulements FEMTO.

Les figures 5.7 et 5.8 montrent les différents scaleogrammes des données de vibrations horizontales et verticales du roulement1\_1 (respectivement) prises à différentes étapes de la vie du roulement.

Les scalogrammes peuvent fournir des informations importantes sur l'état de santé du roulement, mais les différents roulements peuvent avoir des profils de dégradation différents. C'est pourquoi une base de données de scalogrammes est construite à partir des données de vibrations. Les 80 premiers enregistrements de vibrations de chaque roulement sont considérés comme représentant un roulement en bonne santé, tandis que les 80 derniers enregistrements de vibrations sont considérés comme défectueux. Ceci est considéré comme une tâche de classification binaire, où un réseau de neurones convolutifs sera utilisé pour extraire automatiquement les caractéristiques des bases de

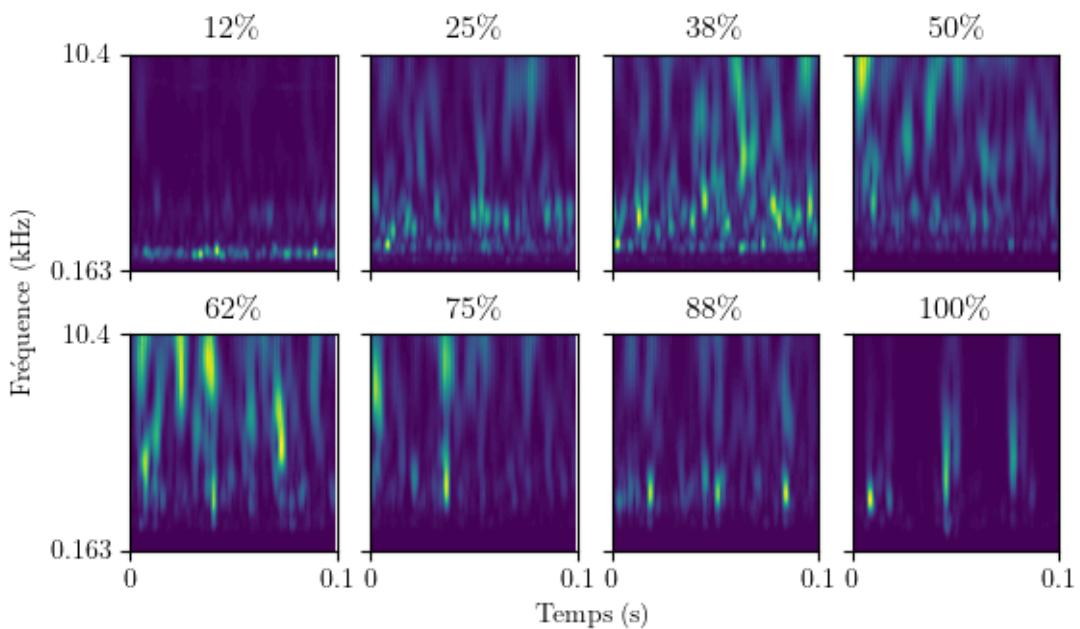


Figure 5.7: Scalogrammes des différentes phases de la vie du roulement 1\_1 (vibrations horizontales)

données des scaleogrammes (données de vibrations horizontales et verticales) et classifier l'état de santé du roulement. Pour éviter la complexité et le coût de l'entraînement d'un grand modèle, les scaleogrammes sont redimensionnés à une forme de  $(128 \times 128)$ , de sorte que l'entrée du réseau est de forme  $(2 \times 128 \times 128)$  où les deux canaux correspondent aux scaleogrammes de vibrations horizontales et verticales. Ces données représentent les roulements de la première condition de fonctionnement (du Roulement1\_1 au Roulement1\_7).

### 5.2.3 Détection de la défaillance des roulements à l'aide de réseaux de neurones convolutifs

#### Architecture de réseau

Pour cette tâche, un réseau de neurones convolutionnel (CNN) avec 3 couches convolutionnelles, 3 couches de maxpooling et 3 couches entièrement connectées est utilisé. L'architecture complète du réseau avec le nombre de paramètres dans le réseau est représentée dans le tableau 5.6.

#### Processus d'entraînement

La base de données générée contenait un total de 1120 échantillons, 716 échantillons ont été utilisés pour l'entraînement, 180 pour la validation et 224 pour le test. Le réseau a été entraîné pour 50 époques avec des lots de 64 échantillons. La figure 5.9 montre le processus d'entraînement.

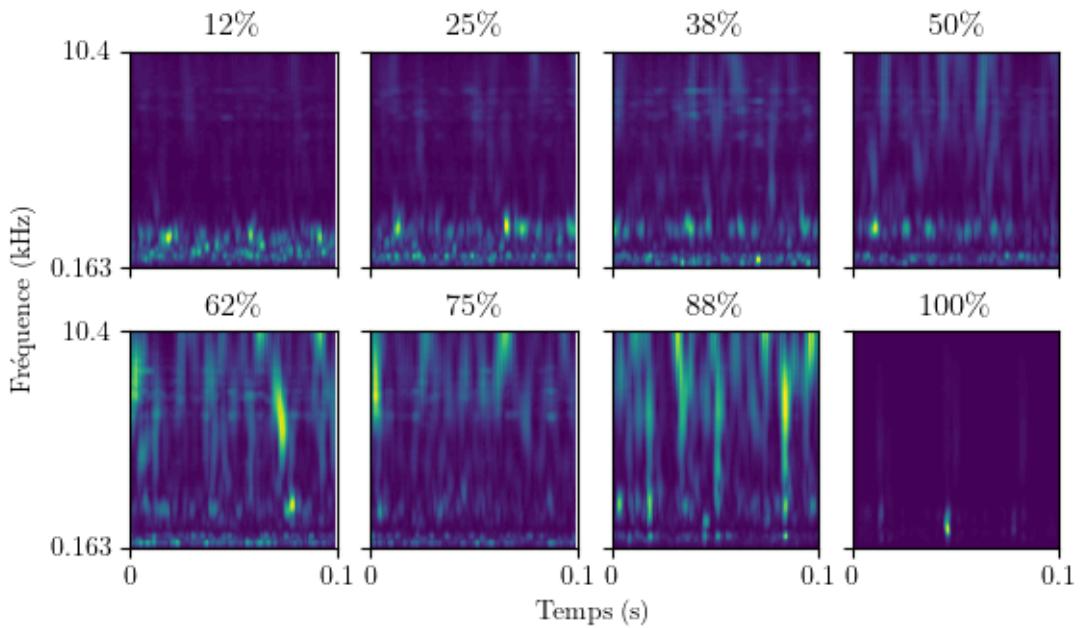


Figure 5.8: Scalogrammes des différentes phases de la vie du roulement1\_1 (vibrations verticales)

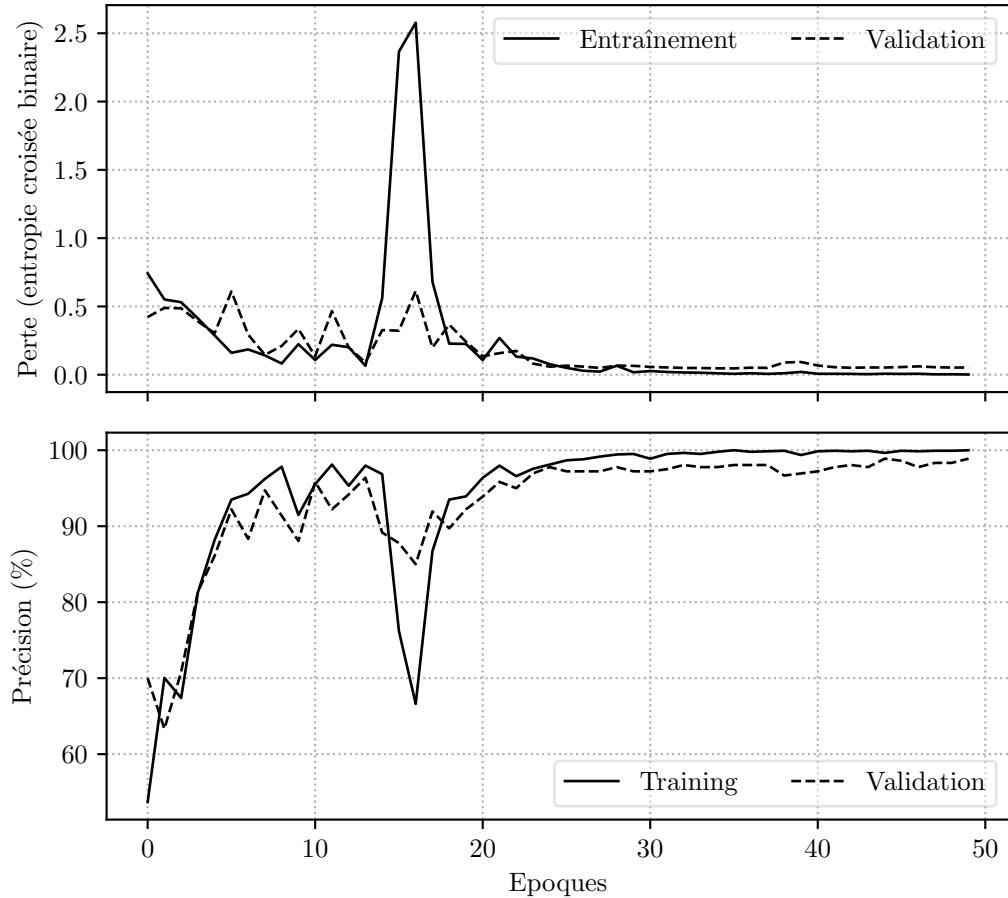


Figure 5.9: Entraînement du classificateur de l'état de santé des roulements

Couche (type)	Forme de la sortie	Param #
Conv1 (Conv2D)	(None, 2, 128, 16)	18448
MaxPool1 (MaxPooling2D)	(None, 1, 64, 16)	0
Conv2 (Conv2D)	(None, 1, 64, 32)	4640
MaxPool2 (MaxPooling2D)	(None, 1, 32, 64)	0
Conv3 (Conv2D)	(None, 1, 32, 128)	18496
MaxPool3 (MaxPooling2D)	(None, 1, 16, 128)	0
Flatten1 (Flatten)	(None, 2048)	0
Dense1 (Dense)	(None, 1024)	524800
Dropout1 (Dropout)	(None, 512)	0
Dense2 (Dense)	(None, 64)	32832
Dropout2 (Dropout)	(None, 64)	0
Dense3 (Dense)	(None, 2)	66
Total params:	559,346	
Trainable params:	599,346	
Non-trainable params:	0	

Table 5.6: Architecture du classificateur de l'état de santé de roulements

## Discussion des résultats

Le réseau a atteint une exactitude d'entraînement parfaite de 100 % et une exactitude de validation et de test similaire d'environ 98 %. Le tableau 5.7 montre la perte et l'exactitude des données sur l'entraînement, la validation et les tests. Le tableau 5.8 montre les indicateurs supplémentaires de précision, de rappel et de score F-1.

	Perte	Exactitude
Ensemble d'entraînement	0.0017	100.00%
Ensemble de validation	0.0528	98.89%
Ensemble de test	0.0527	98.44%

Table 5.7: Résultats de l'entraînement du classificateur de l'état de santé de roulements

Métrique	Exactitude	Précision	Rappel	F-1	ROC AUC
Valeur	98.44%	0.9896	0.9694	0.9794	0.8348

Table 5.8: Indicateurs supplémentaires pour la performance du réseau

La figure 5.10 montre la courbe ROC du classificateur :

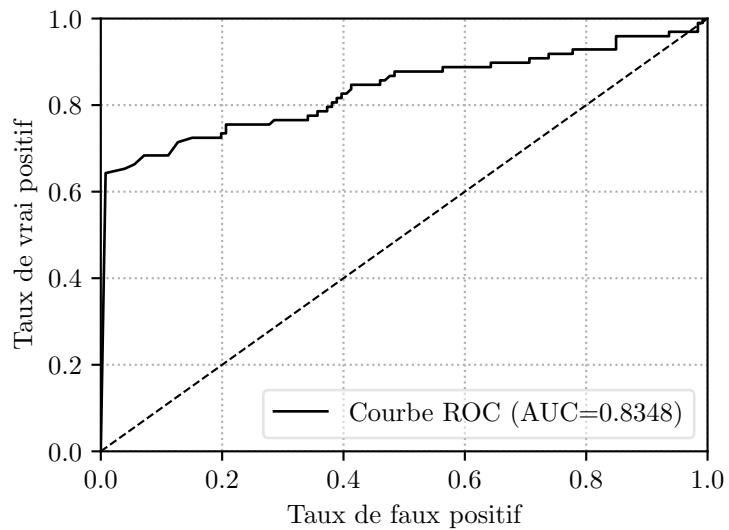


Figure 5.10: Courbe ROC du classificateur de l'état de santé des roulements sur l'ensemble de test

### 5.2.4 Le besoin de nouvelles caractéristiques de pronostic

L'entrée du modèle de pronostic est essentielle à la précision de ses prédictions [35] et puisque c'est le cas, une étape de sélection des caractéristiques doit être effectuée lors du développement du modèle, qui vise à sélectionner un ensemble de caractéristiques appropriées qui peuvent rendre la prédiction plus précise [36]. C'est pourquoi il est important d'utiliser des mesures concrètes qui peuvent quantifier la qualité des caractéristiques pronostiques sélectionnées et aider à les comparer directement et à sélectionner les plus appropriées.

#### Trendabilité et monotonie

Dans [35], les auteurs ont proposé un ensemble de mesures qui peuvent être utilisées pour quantifier la qualité des caractéristiques de pronostic et comparer directement leur aptitude à être utilisées dans des modèles prédictifs, deux de ces mesures sont **monotonie** et **tendabilité**. Selon leur définition, la monotonie fait référence à la nature de la variable, qu'elle soit croissante ou décroissante, et elle a une valeur comprise entre 0 et 1, une variable qui est toujours croissante ou décroissante ayant une monotonie de 1. La monotonie est définie par l'équation 5.2 :

$$M = \frac{\text{no. of } \frac{d}{dx} > 0}{n - 1} - \frac{\text{no. of } \frac{d}{dx} < 0}{n - 1} \quad (5.2)$$

La tendabilité, en revanche, a été définie comme le fait de savoir si une caractéristique donnée suit la même tendance (à la hausse ou à la baisse) dans une population de systèmes. Une caractéristique qui est toujours à la baisse ou à la hausse a une plus grande tendance qu'une autre qui ne suit pas une tendance spécifique dans différents cas. En conséquence, la tendance est définie par l'équation 5.3 :

$$t_i = \frac{\text{no. of } \frac{d}{dx} > 0}{n-1} + \frac{\text{no of } \frac{d^2}{dx^2} > 0}{n-2} \quad (5.3)$$

$$Trendability = 1 - \text{std}(t_i)$$

### 5.2.5 Caractéristiques trigonométriques et descripteurs cumulatifs

Dans [9], les auteurs ont proposé une nouvelle procédure d'extraction des caractéristiques des données de vibration basée sur une transformée en ondelettes discrètes (Section 2.2.3) et des fonctions trigonométriques. D'abord, le signal de vibration brut est décomposé en utilisant une transformée en ondelettes discrètes avec une ondelette db4 et un 4ème niveau de décomposition, puis les coefficients de décomposition sont mis à l'échelle en utilisant des fonctions trigonométriques (par exemple asinh, atan), enfin l'écart-type des coefficients mis à l'échelle est calculé pour obtenir la valeur finale. Selon les auteurs, ces fonctions sont moins sensibles au bruit et à la variabilité des signaux de vibration bruts.

Le tableau 5.9 montre la définition mathématique des différentes caractéristiques classiques et trigonométriques utilisées dans le litearture :

Trigonometric features	Formula
Standard deviation of asinh	$\sigma \left( \log \left[ x_j + \sqrt{(x_j^2 + 1)} \right] \right)$
Standard deviation of atan	$\sigma \left( \frac{i}{2} \log \left( \frac{i+x_j}{i-x_j} \right) \right)$
Classic features	Formula
Entropy	$E(x) = \sum_j E(x_j)$
Energy	$e = \sum_{j=0}^n x_j^2$
Root mean square	$RMS = \sqrt{\frac{1}{n} (x_1^2 + \dots + x_n^2)}$
Skewness	$\frac{\sum_{j=1}^n (x_j - \bar{X})^3}{(n-1)\sigma^3}$
Kurtosis	$\frac{\sum_{j=1}^n (x_j - \bar{X})^4}{(n-1)\sigma^4}$
Upper bound	$\max(x) + \frac{1}{2} \frac{\max(X) - \min(X)}{n-1}$

Table 5.9: Caractéristiques trigonométriques et classiques de pronostic [9]

La figure 5.11 montre deux caractéristiques trigonométriques (asinh et atan) pour le palier 1\_1 à partir de la base de données FEMTO. Elles ont été filtrées à l'aide du filtre Savitsky–Golay pour réduire le bruit et la variabilité.

Même si les caractéristiques trigonométriques sont moins sensibles au bruit que les caractéristiques classiques (par exemple: rms, skewness, kurtosis, etc.), ce n'est pas toujours le cas. ), ce n'est pas toujours le cas. Les auteurs ont également proposé l'utilisation de descripteurs cumulatifs (c'est-à-dire la somme cumulative) afin d'assurer la monotonie des caractéristiques. La caractéristique cumulative associée à chaque caractéristique est calculée selon l'équation 5.4 :

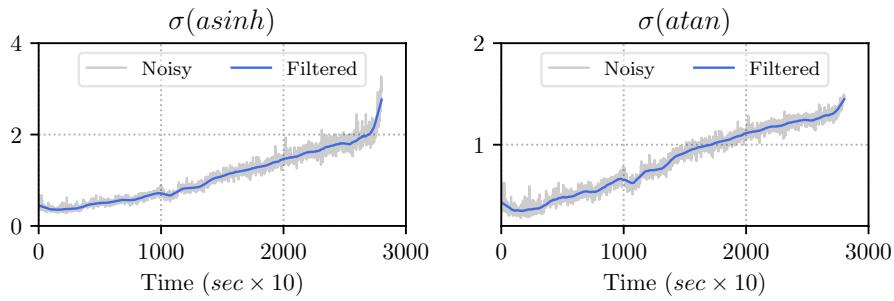


Figure 5.11: Caractéristiques trigonométriques du roulement1\_1

$$Cf_{nk} = \frac{\sum_{i=1}^n f_{ik}}{\sqrt{abs(\sum_{i=1}^n f_{ik})}} \quad (5.4)$$

Dans la figure 5.12, les caractéristiques classiques et trigonométriques cumulées sont comparées à leurs homologues non cumulées. Même avec une simple inspection visuelle, il est évident que l'utilisation de descripteurs cumulatifs améliore considérablement la qualité des caractéristiques de pronostic en réduisant le bruit et les fluctuations.

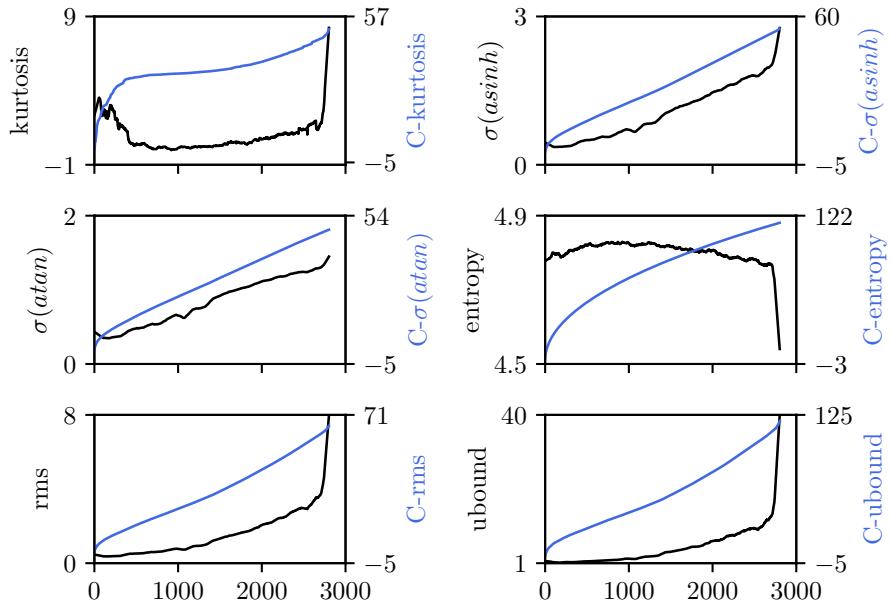


Figure 5.12: Caractéristiques cumulatives classiques et trigonométriques

Pour quantifier l'amélioration causée par l'utilisation de descripteurs cumulatifs sur les caractéristiques de pronostic, la monotonie et la tendabilité de chaque caractéristique et de sa contrepartie cumulative sont calculées et résumées dans les tableaux 5.10 et 5.11 respectivement.

Il est très évident que la monotonie des caractéristiques a augmenté de manière significative pour les caractéristiques cumulatives par rapport aux caractéristiques non cumulatives. Mais pour la tendabilité, la différence était moins importante. La figure

Feature	Monotonicity	Cumulative Feature	Monotonicity
$\sigma(\text{atan})$	0.486	C- $\sigma(\text{atan})$	1
$\sigma(\text{asinh})$	0.481	C- $\sigma(\text{asinh})$	1
kurtosis	0.059	C-kurtosis	0.998
entropy	0.035	C-entropy	1
rms	0.481	C-rms	1
ubound	0.287	C-ubound	1

Table 5.10: Différence de monotonicité entre les caractéristiques trigonométriques et classiques et leurs descripteurs cumulés

Feature	Trendability	Cumulative Feature	Trendability
$\sigma(\text{atan})$	0.987	C- $\sigma(\text{atan})$	0.993
$\sigma(\text{asinh})$	0.989	C- $\sigma(\text{asinh})$	0.995
kurtosis	0.985	C-kurtosis	0.890
entropy	0.994	C-entropy	0.976
rms	0.988	C-rms	0.993
ubound	0.990	C-ubound	0.996

Table 5.11: Différence de trendabilité entre les caractéristiques trigonométriques et classiques et leurs descripteurs cumulés

Figure 5.13 montre la monotonicité (axe x) et la trendabilité (axe y) des deux types de caractéristiques :

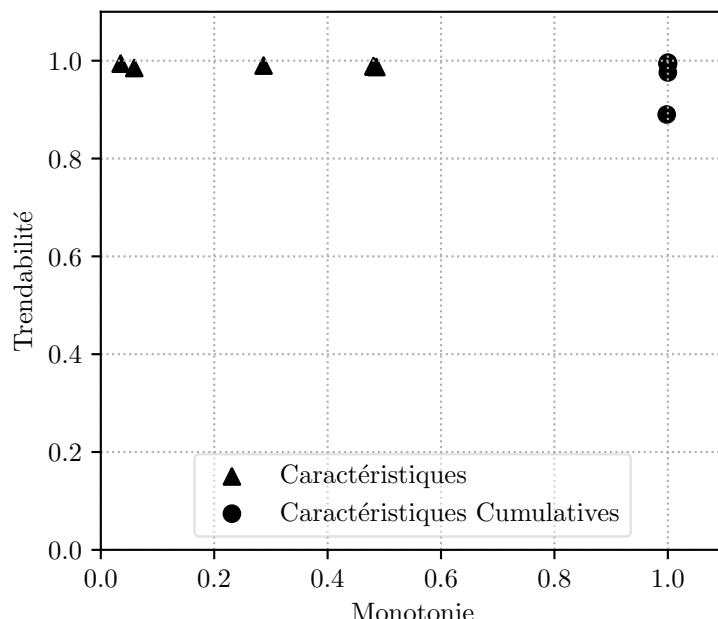


Figure 5.13: Features fitness

## 5.3 Application aux équipements des chantiers pétroliers

### 5.3.1 Top Drive

Dans les platesformes de forage pétrolier, le Top Drive est un dispositif qui sert à faire tourner le train de sonde et à remplacer la table de rotation classique et la tige carrée. Le principal avantage de Top Drive par rapport aux solutions conventionnelles est la possibilité de forer en utilisant trois joints de triges au lieu d'un seul, ce qui réduit considérablement le temps de forage. Top Drives aide également les foreurs à minimiser le coût et la fréquence des incidents liés aux tiges coincées [37].

Figure 5.14 montre le Top Drive Bentec 500-HT.

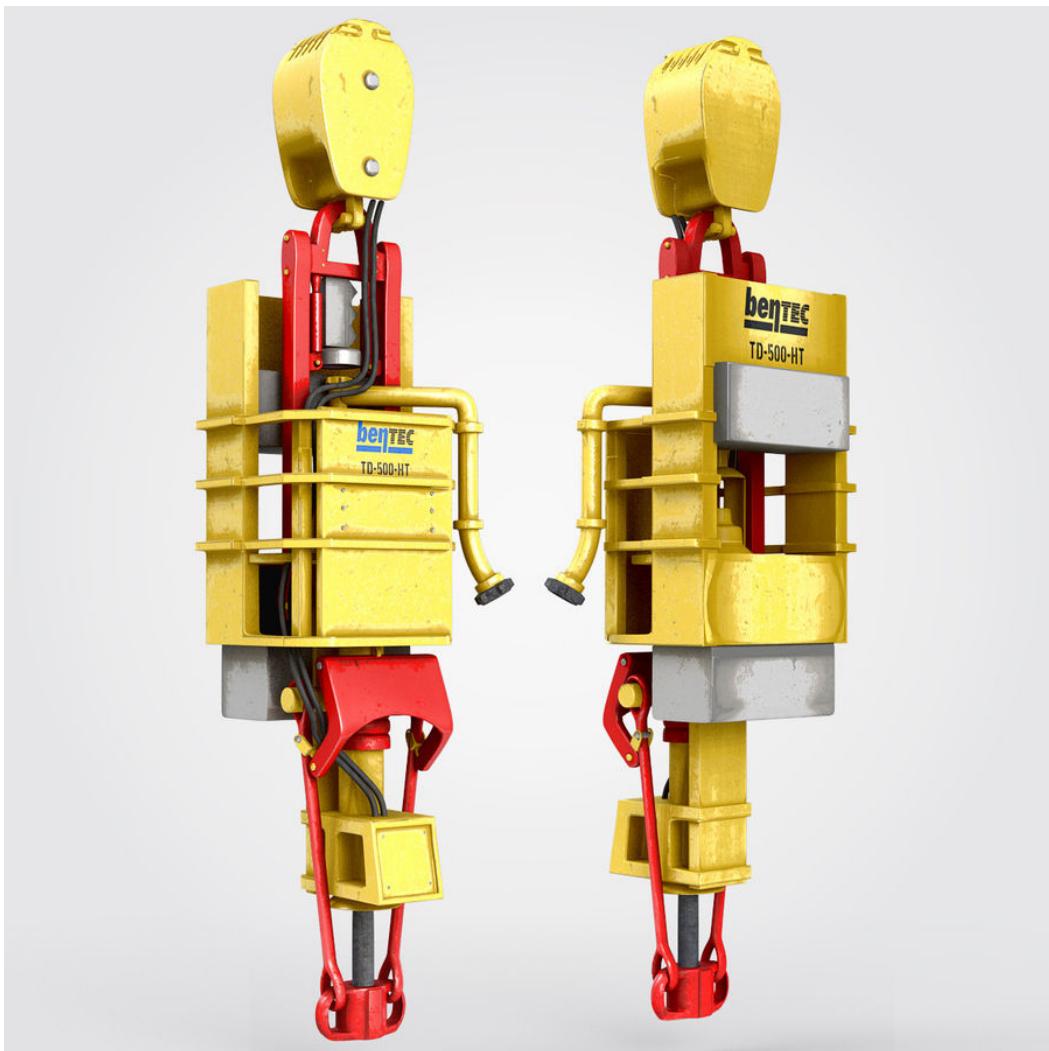


Figure 5.14: Top Drive Bentec 500-HT

From an engineering perspective, Top Drives are much more complicated systems than conventional rotary table and kelly, thus they need more rigorous maintenance programs to ensure their availability considering their major role in drilling operations. Every year the oil and gas technology is pushed harder, to drill in harder conditions and to drill deeper and more challenging oil wells, this poses stricter requirements on used equipment: to be able under and sustain harsher conditions while maximizing its

availability which is of great importance in the field, for example Top Drives downtime cost can reach 1m\$ per day and cause significant delay to drilling operations [38].

In [39] the authors reported an industry survey that was conducted to assess the impression of different groups of people (operators, contractors, rig builders...) on the usage of Top Drives. The survey showed an average 60% satisfaction level across all the groups in the survey, which is an indicative that Top Drives aren't really meeting the expectations in the industry. They were also asked which are the features they would like to see in Top Drives, the most wanted things are less downtime and better ability to detect failures before the breakdown of the system. This means that a more reliable preventive maintenance and prognostics approach is the most wanted improvement to these systems in the field.

D'un point de vue technique, les Top Drives sont des systèmes beaucoup plus compliqués que les tables de rotation conventionnelles et les tiges carrées, ils ont donc besoin de programmes de maintenance plus rigoureux pour assurer leur disponibilité compte tenu de leur rôle majeur dans les opérations de forage. Chaque année, la technologie du pétrole et du gaz est poussée plus loin, pour forer dans des conditions plus difficiles et pour forer des puits de pétrole plus profonds et plus difficiles, ce qui impose des exigences plus strictes aux équipements utilisés : être capable de supporter des conditions plus difficiles tout en maximisant sa disponibilité, ce qui est d'une grande importance sur le terrain, par exemple le coût du temps d'arrêt des Top Drives peut atteindre 1m\$ par jour et causer des retards importants dans les opérations de forage [38].

Dans [39], les auteurs ont rapporté un sondage réalisé dans l'industrie pour évaluer l'impression de différents groupes de personnes (opérateurs, entrepreneurs, constructeurs des appareils de forage...) sur l'utilisation de Top Drives. Le sondage a montré un niveau de satisfaction moyen de 60 % pour tous les groupes interrogés, ce qui indique que les Top Drives ne répondent pas vraiment aux attentes de l'industrie. On leur a également demandé quelles étaient les caractéristiques qu'ils aimeraient voir dans Top Drives, les choses les plus souhaitées étant une réduction des temps d'arrêt et une meilleure capacité à détecter les pannes avant que le système ne tombe en panne. Cela signifie qu'une maintenance préventive et une approche pronostique plus fiables sont les améliorations les plus souhaitées de ces systèmes sur le terrain.

### **5.3.2 Les composants de Top Drive**

Les Top Drives sont constitués de nombreux sous-ensembles qui sont présentés dans la figure 5.15<sup>1</sup>.

---

<sup>1</sup>Toutes les figures suivantes des composants Top Drive sont extraites du bulletin technique de Bentec pour les Top Drives TD-500-HT et TD-350-HT

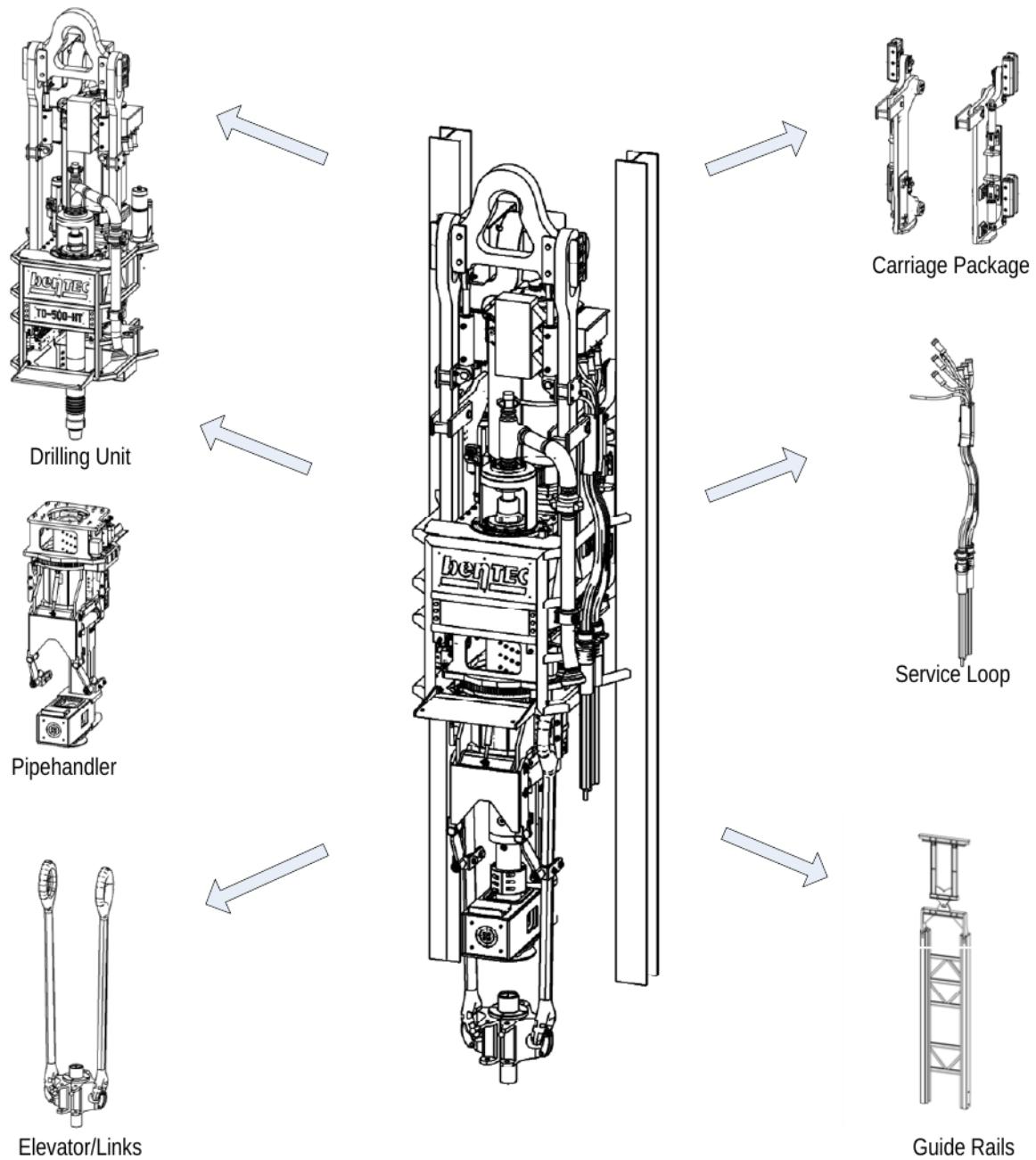


Figure 5.15: Sous-ensembles du Top Drive Bentec 500-HT

The Top Drive subassembly of interest here is the drilling unit (Figure 5.16) which is responsible for generating the rotation motion and transferring it to the drillstring. Drilling unit is composed of protection frame, hydraulic unit, mudd supply, hanger assembly and—the most important for the current discussion—a drive.

Le sous-ensemble Top Drive qui nous intéresse ici est l'unité de forage (Figure 5.16) qui est chargée de générer le mouvement de rotation et de le transférer au train de sonde. L'unité de forage est composée d'un cadre de protection, d'une unité hydraulique, d'une alimentation en boue, d'un ensemble de suspension et—le plus important pour la discussion actuelle—d'un entraînement.

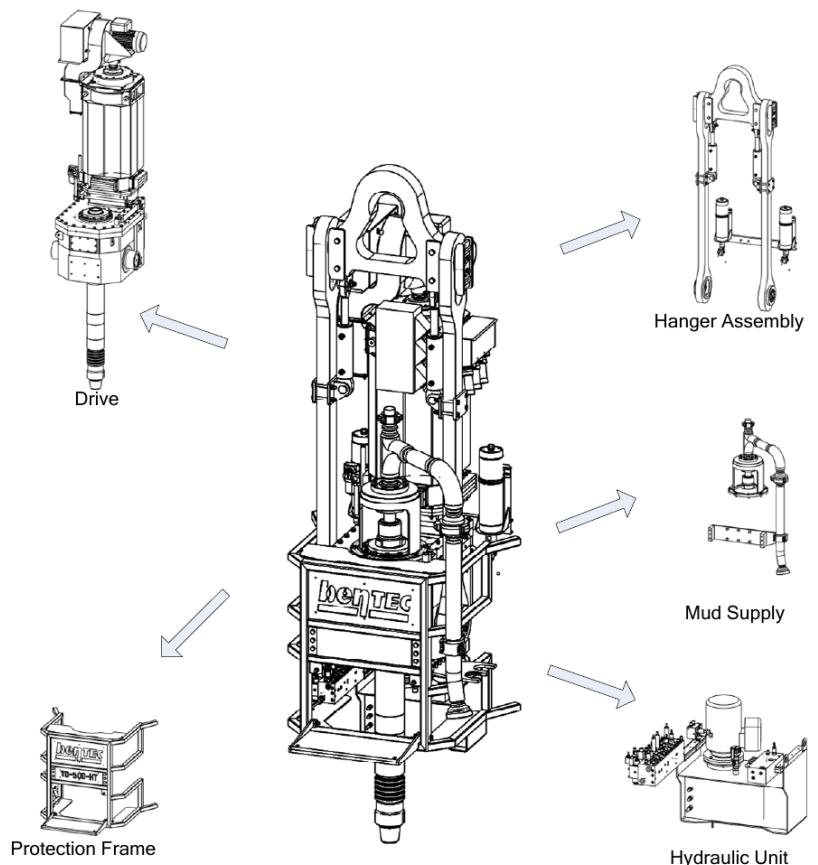


Figure 5.16: Unité de forage du Top Drive Bentec 500-HT

L'entraînement lui-même est composé de :

- Engine cooling system
- Brake
- AC Motor
- Gearbox (Figure 5.17(a))
- Mainshaft (Figure 5.17b)

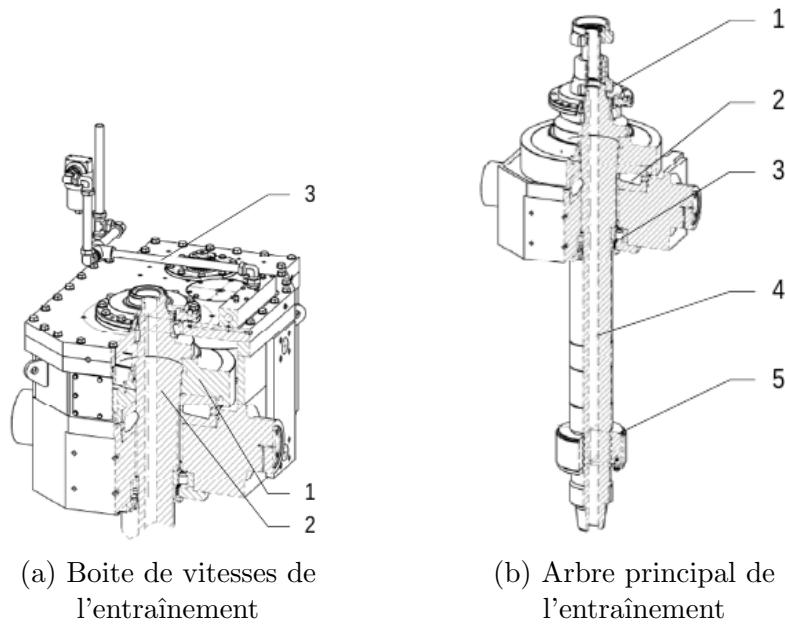


Figure 5.17: Les composants principaux de l' entraînement

La boîte de vitesses comporte deux étages avec un rapport de transmission de 14:1. La vitesse du moteur est réduite deux fois puis transférée à l'engrenage (1) de la boîte de vitesses. La transmission est située sur le palier de butée. La lubrification de la boîte d'engrenages est effectuée par une librification combinée par barbotage et pression (3). L'arbre principal (4) est alimenté par la transmission de la boîte de vitesses et se trouve dans la boîte de vitesses avec le palier de butée (2) et il est conduit par le palier inférieur (3). Un tuyau de lavage (1) est connecté pour exécuter le fluide de forage. L'arbre principal contient également le collier de charge (5) qui porte l'adaptateur de liaison à la tige de forage.

La figure 5.18 montre un exemple de palier conique habituellement utilisé dans les entraînements Top Drives. La figure 5.19 montre le palier dans un entraînement Top Drive.

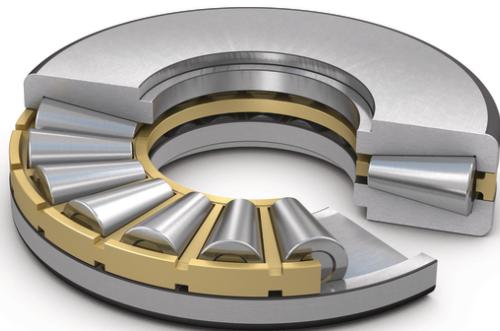


Figure 5.18: Palier de butée à rouleaux coniques SKF [7]

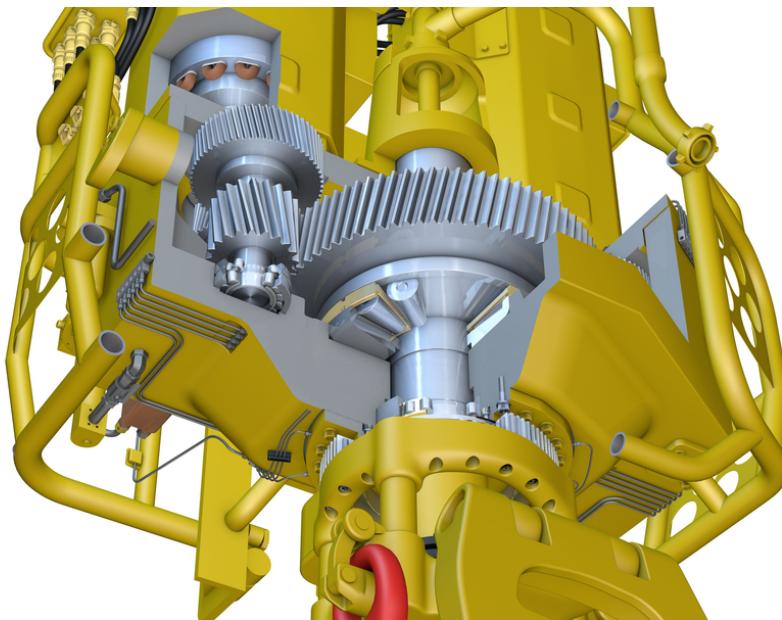


Figure 5.19: Position des roulements dans le Top Drive [8]

### 5.3.3 Proposed approach for Top Drive monitoring using neural networks

Les Top Drives sont déjà équipés de nombreux capteurs pour surveiller leur état comme des capteurs de température, de pression et de débit, mais l'information que ces capteurs ne sont utilisés que de manière simplifiée par les opérateurs [39]. Afin d'améliorer les programmes de maintenance et d'adopter des approches de maintenance préventive et de pronostic, une approche plus sophistiquée et méthodologique du traitement des données fournies par ces capteurs doit être utilisée, ainsi que l'introduction de nouveaux capteurs de vibrations qui sont les principaux capteurs de surveillance utilisés pour les roulements et les engrenages, les éléments présents dans le Top Drive et qui travaillent sous des charges élevées et variables.

La figure 5.20 est une représentation schématique simplifiée des procédures proposées pour employer des capteurs supplémentaires (points rouges) pour les vibrations dans l' entraînement afin de surveiller différents éléments comme les engrenages et les roulements. Il faudrait installer suffisamment de capteurs pour capter les vibrations dans les trois axes : les vibrations verticales, horizontales et axiales. Lorsque le Top Drive est en service, les données recueillies par ces capteurs peuvent être utilisées pour développer une architecture de réseau de neurones afin de détecter la dégradation des roulements (Section 5.2) ou—si suffisamment de données historiques sont disponibles—classer les différents modèles de dégradation (Section 5.1). La détection correcte des dégradations ou même la classification des modèles de dégradation peut être utilisée pour planifier et préparer les actions de maintenance appropriées et les exécuter en temps voulu avant qu'une dégradation grave ne se produise, ce qui peut entraîner un temps improductif important et des coûts de maintenance corrective élevés.

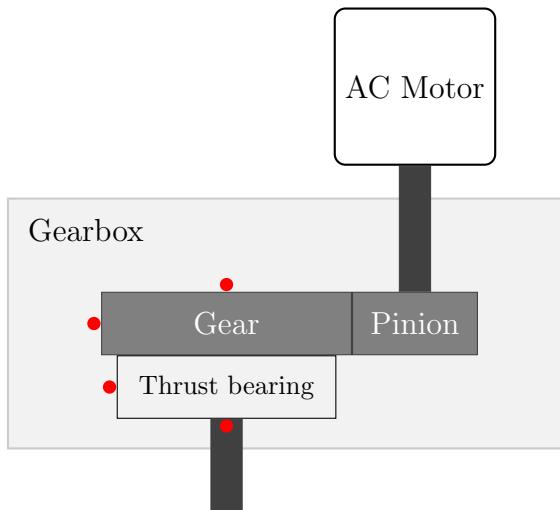


Figure 5.20: Approche proposée pour la surveillance du Top Drive à l'aide de réseaux de neurones

## 5.4 Conclusion

Ce chapitre présente deux approches différentes pour le diagnostic et le pronostic des roulements à partir des données de vibrations. Les réseaux de neurones présentés ont permis d'obtenir une très grande exactitude de test. Ensuite, une méthodologie a été présentée pour intégrer ces modèles neuronaux dans les équipements critiques des champs pétrolifères (c'est-à-dire les entraînements Top Drives), ce qui peut avoir un grand impact sur les programmes de maintenance traditionnels existants et contribuer à réduire les temps improductifs et les coûts élevés résultant de la panne de ces équipements importants.

# Conclusion

La maintenance est un processus important dans toutes les installations industrielles qui garantit la disponibilité des équipements et réduit les temps d'arrêt. Nous pouvons distinguer différents types d'actions de maintenance, chaque type ayant un domaine d'application spécifique. Pour les équipements critiques dont le coût de réparation est élevé et dont les temps d'arrêt peuvent entraîner des pertes importantes, la maintenance prédictive a été la politique de maintenance privilégiée pendant de nombreuses décennies. Cependant, avec l'avancée technologique des deux dernières décennies où la puissance de calcul est devenue moins chère et les capteurs omniprésents, l'industrie a commencé à s'orienter vers ce que l'on appelle la maintenance prédictive et les pronostics, qui visent à fournir des prévisions sur le comportement futur du système et les pannes potentielles. Il existe de nombreuses approches de la maintenance prédictive et des pronostics. Cette thèse s'est concentrée sur l'utilisation de méthodes basées sur les données, principalement des techniques d'apprentissage machine et d'apprentissage profond et des données de surveillance des équipements pour développer des modèles prédictifs capables d'apprendre les modèles de dégradation d'un équipement spécifique et de les utiliser pour prédire les performances d'un nouvel équipement. Le chapitre 4 a utilisé la base de données C-MAPSS de la NASA pour illustrer l'utilisation des réseaux de neurones artificiels (tant les réseaux de neurones entièrement connectés que les réseaux LSTM) pour modéliser la dégradation des équipements à l'aide de données provenant de capteurs qui surveillent et mesurent différentes variables physiques. Les modèles développés ont montré une grande performance dans la prévision de la durée de vie utile restante des équipements (RUL). Plus tard, le chapitre a présenté l'adoption potentielle de cette approche dans les chantiers pétroliers sur l'équipement pétrolier.

Le chapitre 5 s'est concentré sur une application plus spécialisée des réseaux de neurones : la surveillance des roulements en utilisant les données de vibration pour le diagnostic et le pronostic. Ce chapitre a utilisé des techniques de traitement de données plus sophistiquées pour extraire les caractéristiques appropriées des données vibratoires brutes. Deux réseaux de neurones convolutionnels différents (CNN) ont été utilisés, le premier pour diagnostiquer les roulements et classifier le type de défaut à partir des données de vibration, et le second pour utiliser des techniques de traitement du signal et de transformation en ondelettes continues afin de construire une base de données de scalogrammes à partir des données de vibration qui a ensuite été utilisée comme entrée dans le réseau de neurones pour classifier les roulements sains et les roulements défectueux. Le chapitre a également présenté de nouvelles techniques d'extraction de caractéristiques de la littérature en utilisant des caractéristiques trigonométriques et

des descripteurs cumulatifs ainsi que des mesures utilisées pour quantifier l'adéquation de ces caractéristiques aux pronostics. Enfin, une procédure a été décrite pour adopter les solutions présentées dans ce chapitre aux entraînements de pointe dans les chantiers pétroliers.

# Bibliography

- [1] K. Javed, R. Gouriveau, and N. Zerhouni, “State of the art and taxonomy of prognostics approaches, trends of prognostics applications and open issues towards maturity at different technology readiness levels,” *Mechanical Systems and Signal Processing*, vol. 94, pp. 214–236, 09 2017.
- [2] F. Mangili, *Development of advanced computational methods for prognostics and health management in energy components and systems*. PhD thesis, Politecnico di Milano, Mar. 2013.
- [3] V. Dumoulin and F. Visin, “A guide to convolution arithmetic for deep learning,” *ArXiv e-prints*, mar 2016.
- [4] A. Saxena, K. Goebel, D. Simon, and N. Eklund, “Damage propagation modeling for aircraft engine run-to-failure simulation,” in *2008 International Conference on Prognostics and Health Management*, IEEE, oct 2008.
- [5] Honeywell, “Sensors and switches in oil rig applications.” Online.
- [6] P. Nectoux, R. Gouriveau, K. Medjaher, E. Ramasso, B. Chebel-Morello, N. Zerhouni, and C. Varnier, “Pronostia: An experimental platform for bearings accelerated degradation tests,” pp. 1–8, 06 2012.
- [7] SKF, “Tapered roller thrust bearings.” Online.
- [8] SKF, “Skf helps develop 1000-ton top drive and consolidate oem supply chain.” Online.
- [9] K. Javed, R. Gouriveau, N. Zerhouni, and P. Nectoux, “A feature extraction procedure based on trigonometric functions and cumulative descriptors to enhance prognostics modeling,” pp. 1–7, 06 2013.
- [10] ISO/IEC 2382:2015, “Information technology — vocabulary,” standard, International Organization for Standardization, Geneva, CH, 2015.
- [11] S. Kadry, ed., *Diagnostics and Prognostics of Engineering Systems*. IGI Global, 2013.
- [12] J.-H. Shin and H.-B. Jun, “On condition based maintenance policy,” *Journal of Computational Design and Engineering*, vol. 2, pp. 119–127, apr 2015.

## BIBLIOGRAPHY

---

- [13] T. Tinga and R. Loendersloot, “Aligning phm, shm and cbm by understanding the physical system failure behaviour,” in *European Conference of the Prognostics and Health Management Society 2014*, 07 2014.
- [14] E. R. Brown, N. N. McCollom, E.-E. Moore, and A. Hess, “Prognostics and health management a data-driven approach to supporting the f-35 lightning II,” in *2007 IEEE Aerospace Conference*, IEEE, 2007.
- [15] K. L. Tsui, N. Chen, Q. Zhou, Y. Hai, and W. Wang, “Prognostics and health management: A review on data driven approaches,” *Mathematical Problems in Engineering*, vol. 2015, pp. 1–17, 2015.
- [16] A. K. Jardine, D. Lin, and D. Banjevic, “A review on machinery diagnostics and prognostics implementing condition-based maintenance,” *Mechanical Systems and Signal Processing*, vol. 20, pp. 1483–1510, oct 2006.
- [17] M. I. M. Samuel Telford and I. Howard, “Condition based maintenance (cbm) in the oil and gas industry: An overview of methods and techniques,” in *Proceedings of the 2011 International Conference on Industrial Engineering and Operations Management*, 2011.
- [18] L. Eriksen, “The impact of digital on unplanned downtime.” GE Digital, Oct. 2016.
- [19] NASA Prognostics Center of Excellence data repository. <https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/>.
- [20] A. Hess, J. S. Stecki, and S. D. Rudov-Clark, “The maintenance aware design environment: Development of an aerospace phm software tool,” 2008.
- [21] A. Cubillo, S. Perinpanayagam, and M. Esperon-Miguez, “A review of physics-based models in prognostics: Application to gears and bearings of rotating machinery,” *Advances in Mechanical Engineering*, vol. 8, aug 2016.
- [22] Y. Lei, N. Li, L. Guo, N. Li, T. Yan, and J. Lin, “Machinery health prognostics: A systematic review from data acquisition to RUL prediction,” *Mechanical Systems and Signal Processing*, vol. 104, pp. 799–834, may 2018.
- [23] X.-S. Si, W. Wang, C.-H. Hu, and D.-H. Zhou, “Remaining useful life estimation – a review on the statistical data driven approaches,” *European Journal of Operational Research*, vol. 213, pp. 1–14, aug 2011.
- [24] R. Priemer, *Introductory Signal Processing*. World Scientific Publishing Co Pte Ltd, 1990.
- [25] Y. Lei, *Intelligent Fault Diagnosis and Remaining Useful Life Prediction of Rotating Machinery*, ch. Introduction and Background, p. 8. Elsevier - Health Sciences Division, 2016.
- [26] Y. Lei, *Intelligent Fault Diagnosis and Remaining Useful Life Prediction of Rotating Machinery*, ch. Introduction and Background, p. 9. Elsevier - Health Sciences Division, 2016.

## BIBLIOGRAPHY

---

- [27] Y. Lei, *Intelligent Fault Diagnosis and Remaining Useful Life Prediction of Rotating Machinery*, ch. Introduction and Background, p. 10. Elsevier - Health Sciences Division, 2016.
- [28] Y. Chung, P. J. Haas, E. Upfal, and T. Kraska, “Unknown examples & machine learning model generalization,” *CoRR*, vol. abs/1808.08294, 2018.
- [29] C. M. Bishop, *Pattern recognition and machine learning*. New York: Springer, 2006. Christopher M. Bishop., Includes bibliographical references and index.
- [30] A. Choromanska, M. Henaff, M. Mathieu, G. B. Arous, and Y. LeCun, “The loss surfaces of multilayer networks,”
- [31] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning representations by back-propagating errors,” *Nature*, vol. 323, pp. 533–536, oct 1986.
- [32] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, pp. 436–444, may 2015.
- [33] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, pp. 1735–1780, nov 1997.
- [34] L. Wen, X. Li, L. Gao, and Y. Zhang, “A new convolutional neural network-based data-driven fault diagnosis method,” *IEEE Transactions on Industrial Electronics*, vol. 65, pp. 5990–5998, jul 2018.
- [35] J. Coble and J. Hines, “Identifying optimal prognostic parameters from data: A genetic algorithms approach,” *Annual Conference of the Prognostics and Health Management Society*, 01 2009.
- [36] K. Javed, R. Gouriveau, R. Zemouri, and N. Zerhouni, “Features selection procedure for prognostics: An approach based on predictability,” vol. 8, 08 2012.
- [37] Schlumberger, “Top drive.” Online Oilfield Glossary.
- [38] SKF, “Strength and reliability for the next generation of top drives.” Online.
- [39] P. Pournazari, D. Adams, P. Ashok, E. van Oort, and K. Holliday, “Real-time health monitoring of top drives for offshore operations using physics based models and new sensor technology,” in *SPE Deepwater Drilling and Completions Conference*, Society of Petroleum Engineers, 2016.