

الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne Démocratique et Populaire
وزارة التعليم العالي والبحث العلمي
Ministère de l'enseignement supérieur et de la recherche scientifique

Université Mohamed Khider – Biskra
Faculté des Sciences et de la technologie
Département : Génie Electrique
Réf :



جامعة محمد خيضر بسكرة
كلية العلوم والتكنولوجيا
قسم: الهندسة الكهربائية
المرجع:

Thèse présentée en vue de l'obtention
Du diplôme de
Doctorat Sciences en Génie Electrique
Spécialité (Option) : Automatique

Contribution au pronostic de défaut dans les systèmes complexes par les techniques intelligentes

Présentée par :
DRICI Djalel

Soutenue publiquement le : 25/06/2023

Devant le jury composé de

Nadjiba TERKI	Professeur	Président	Université de Biskra
Yahia KOURD	Professeur	Encadreur	Université de Souk-Ahras
Mostefa Mohamed TOUBA	MCA	Co-Encadreur	Université de Biskra
Mohamed SAHRAOUI	Professeur	Examineur	Université de Biskra
Khaled KHELIL	Professeur	Examineur	Université de Souk-Ahras
Abdelghani CHELIHI	MCA	Examineur	Université de Constantine 1

Année universitaire 2022/2023

Sommaire

Remerciements	i
Abstract	ii
Résumé	iii
ملخص	iv
Acronymes	v
Liste des Figures	vii
Liste des tableaux	ix
Introduction générale	01
Chapitre I. État de l’art sur le diagnostic et le pronostic	
I.1 Introduction	04
I.2 Forme de maintenance	05
I.2.1 La maintenance corrective	06
I.2.2 La maintenance préventive	07
I.3 La maintenance conditionnelle et prévisionnelle CBM	08
I.4 Préparation des données	11
I.4.1 Prétraitement des données	11
I.5 Diagnostic de défaut	17
I.5.1 Classification des méthodes de diagnostic	20
I.5.2 Critère de Choix de la méthode de diagnostic	30
I.6 Pronostic de défaut	30
I.6.1 Les approches de pronostic	31
I.7 Conclusion	39
Chapitre II. Détection et classification par SVM des défauts de moteur à induction en utilisant WPA	
II.1 Introduction	40
II.2 Modèle mathématique de la machine	41
II.2.1 le cas sain	41
II.2.2 le cas d’un défaut	44
II.3 Détection des défauts et classification	44
✚ Générations des caractéristiques	45
II.3.1 La transformées des ondelettes pour l’extraction des caractéristiques	45
✚ Conceptions du classificateur	54
II.4 Machines à vecteurs de support (SVM)	54
II.4.1. Le problème de discrimination binaire	56
II.5 Résultats et discussions	63
II.6 Conclusion	71
Chapitre III. Approche proposée pour l’estimation du RUL en utilisant en utilisant l’apprentissage profond	

III.1 Introduction	72
III.2 Généralité	74
III.3 Apprentissage profond	74
III.3.1 Les réseaux de neurones artificiels	74
III.3.2 Modèles du Deep Learning	79
III.4 LSTM et l'estimation de RUL	85
III.4.1 Modèle LSTM	85
III.4.2 Apprentissage du LSTM	86
III.5 La base de données et le prétraitement :	87
III.5.1 La base de données	87
III.5.2 Augmentation des données comme contribution	92
III.5.3 Prétraitement des données	94
III.6 Évaluation du modèle	94
III.7 Paramètres du modèle LSTM construit	95
III.8 Résultats et discussion	96
III.8.1 Performance du pronostic	96
III.9 Conclusion	97
Conclusion générale & Perspectives	98
Références bibliographiques	100

Remerciements

Je dois tout d'abord remercier notre dieu **ALLAH**, pour m'avoir donné la santé, la force et le courage pour accomplir ce modeste travail.

En guise de reconnaissance et de gratitude, mes sincères remerciements et ma profonde reconnaissance sont adressés à toutes les personnes, qui ont aidé à l'accomplissement de ce mémoire, notamment :

Monsieur **Yahia KOURD**, Professeur à l'Université de Souk Ahras pour m'avoir fait l'honneur par son encadrement, pour sa très grande gentillesse et pour le soutien et l'aide qu'il m'a apporté lors de mes travaux de recherche, pour avoir suivi la réalisation de ce travail, et pour votre aide précieuse dans l'élaboration de ce document. Chaleureux remerciements.

Monsieur **Mohamed Mostefa TOUBA**, Maitre de conférences « A » à l'Université de Biskra, pour l'honneur qu'il m'a fait par son Co-encadrement de ma thèse, recevez mes remerciements sincères pour vos conseils éclairés, Soyez assuré de ma profonde gratitude et mes sincères remerciements.

Monsieur **Dimitri LEFEBVRE**, Professeur à l'Université du havre, pour l'honneur qu'il m'a fait diriger de ma thèse, recevez mes remerciements sincères pour vos conseils éclairés, Soyez assuré de ma profonde gratitude et mes sincères remerciements.

Madame **Nadjiba TERKI**, Professeur à l'université de Biskra, pour l'intérêt qu'elle a accordée à ce travail en acceptant de le juger et de présider le jury.

Monsieur **Mohamed SAHRAOUI**, Professeur à l'université de Biskra, pour m'avoir honoré de faire partie du jury et de juger ce travail, et qu'il trouve ici l'expression de ma gratitude.

Monsieur **Khaled KHELIL**, Professeur à l'université de Souk Ahras, pour m'avoir honoré de faire partie du jury et de juger ce travail, et qu'il trouve ici l'expression de ma gratitude.

Monsieur **Abdelghani CHELIHI**, maitre de conférences « A » à l'université de Constantine 1, trouve ici l'expression de mes vifs remerciements pour avoir bien accepté de juger ce travail.

Je tiens également à remercier mon père, Je ne pourrais vous remercier assez, pour sa patience et son soutien, je vous dois tout.

Je dédie ce modeste travail et ma profonde gratitude à l'âme à ma mère, qui a je dois la réussite, pour l'éducation qu'elle m'a prodiguée. Merci beaucoup à toute ma famille mes frères et sœurs.

Je veux remercier également ma femme. Elle a fait preuve de patience et de soutien durant ma thèse. Elle m'a toujours encouragé et épaulé dans les moments difficiles. A mes fleurs, mes enfants, dont je puise la force.

Un grand merci à tous mes amis et mes collègues, merci pour tous les bons moments passés ensemble.

Je remercie également, et sans exception, l'ensemble du personnel du Département Génie Electrique de l'université de Biskra.

DRICI Djalel

ABSTRACT

This work introduces a novel approach that implements data-driven methods for fault prognosis. These methods rely on the availability of data that describes the degradation process. However, when there is a shortage of data, accurately predicting states using deep models, which necessitate a significant amount of training data, becomes challenging. To address the issue of data scarcity in predicting the Remaining Useful Life (RUL) using Long Short-Term Memory (LSTM), a type of recurrent neural network, we propose the utilization of a simple data augmentation strategy.

The obtained results are shown that the incorporation of this data augmentation, strategy can substantially enhance the performance of RUL prediction using LSTM techniques. The approach presented in this work is validated using data from the NASA Commercial Modular Aero-Propulsion System Simulation (C-MAPPS) dataset.

KEYWORDS: RUL, Fault Prognosis, Fault diagnosis, Deep Learning, Prediction, LSTM

RESUME

Nous avons présenté une nouvelle approche basée sur l'utilisation d'une méthode guidée par les données pour le pronostic des défauts. Cette méthode requiert des données décrivant le processus de dégradation. Lorsque les données sont insuffisantes, la prédiction des états devient difficile avec les modèles profonds de type mémoire à long terme (LSTM), qui nécessitent une quantité importante de données d'apprentissage. Pour résoudre ce problème de rareté des données dans la prédiction de la durée de vie restante (RUL), nous proposons d'adopter une stratégie d'augmentation des données.

Les résultats obtenus sont démontrent que l'application d'une stratégie d'augmentation des données, peut améliorer les performances de prédiction de la RUL en utilisant les techniques LSTM.

Nous avons validé cette approche en utilisant les données de la NASA Commercial Modular Aero-Propulsion System Simulation (C-MAPPS).

Mots Clés : RUL, Pronostic de Défaut, Diagnostic de Défaut, Apprentissage Profond, Prédiction, LSTM.

ملخص

يقدم هذا العمل منهجاً جديداً، يستعمل أساليب توجيه البيانات للتشخيص المستقبلي للأعطال. تعتمد هذه الأساليب على توفر البيانات التي تصف عملية التدهور. مع العلم عندما يكون هناك نقص في البيانات، يصبح من الصعب التنبؤ بالحالات بدقة باستخدام النماذج العميقة، والتي تتطلب كمية كبيرة من بيانات التدريب. لمعالجة مشكلة نقص البيانات هذه في توقع الحياة المفيدة المتبقية (RUL) وباستخدام الذاكرة القصيرة الطويلة (LSTM: تقنية من الشبكة العصبية المتكررة)، نقترح استخدام استراتيجية لزيادة البيانات.

تُظهر النتائج المحصل عليها أن دمج هذه الاستراتيجية لزيادة البيانات يمكن أن يعزز بشكل كبير أداء توقع RUL باستخدام تقنيات LSTM. يتم تحقيق النهج المقدم في هذا العمل باستخدام البيانات من مجموعة بيانات NASA Commercial Modular Aero-Propulsion System Simulation (C-MAPPS).

الكلمات المفتاحية: الحياة المفيدة المتبقية، تنبأ الخطأ، تشخيص الخطأ، التعلم العميق، الذاكرة القصيرة الطويلة.

Acronymes

ACP : Analyse en Composante Principales
AFNOR : Association Française de NORmalisation
ALE : Adaptive Line Enhancer
ALM : Accelerated Life Model
AMDEC : Analyse des Modes de Défaillance, de leurs Effets, et de leur Criticité
ANC : Adaptive Noise Cancellation
AR : Auto-Regressive
ARMA: Auto-Regressive Moving Average.
BD : Blind Deconvolution
CBM : Condition Based Maintenance.
C-MAPSS: Commercial Modular Aero-Propulsion System Simulation
CNN: Convolutional Neural Network
CWT: Continuous Wavelet Transform
DL: Deep Learning
DWNN: Dynamic Wavelet Neural Networks
DWT : Discrete Wavelet Transform
EDF : Electricité de France
FDI : Fault Detection and Isolation
FFT : Fast Fourier Transform
HHT : Hilbert Huang Transform
IA : Intelligence Artificiel
BIAT: Bayesian Inference for Aging Treatment
IHM : Interface Homme-Machine
LF : Logique Floue
LSTM : Long Short Term Memory
LSTMAAD : LSTM Avec Augmentation des Données
LSTMSAD : LSTM Sans Augmentation des Données
MAPE : Mean Absolute Percentage Error
MCO : Maintien en Condition Opérationnelle
MIMOSA: Machinery Information Management Open Standards Alliance
ML: Machine Learning
MLP: Multilayer Perceptron
OSA-CBM: Open System Architecture for Condition Based Maintenance
PHM: Prognostics and Heath Management
PHM: Proportional Hazard Model
RBF: Base Radiale Neurale
RMSE: Root-Mean-Squared Error
RNA : Réseaux de Neurones Artificiels
RNN : Recurrent Neural Network
RUL : Remaining Useful Life
SANC: Self-Adaptive Noise Cancellation

SCADA: Supervisory Control and Data Acquisition

STFT: Short-Time Fourier Transform

SVR: Support Vector Regression

TSA: Time Synchronous Averaging

WPA: Wavelet Packet Analysis

Liste des figures

Chapitre I. État de l'art sur le diagnostic et le pronostic

Figure I.1: L'architecture de PHM	04
Figure I. 2 : Méthodes de maintenance	06
Figure I. 3 : Les couches de l'architecture OSA-CBM	10
Figure I.4: Chronologie de progression de défaillance	19
Figure I.5-a : Détection, diagnostic, pronostic et ordonnancement dans les systèmes de maintenance intelligente	19
Figure I.5-b : Détection, diagnostic et pronostic les aspects phénoménologiques	20
Figure I.6: Classification générale des méthodes de diagnostic	21
Figure I.7: Le principe de l'approche à Base de Modèles	22
Figure I.8: Principe de l'approche FDI à base de résidus	23
Figure I.9: Principe de l'approche logique basée sur la cohérence	25
Figure I.10: Classification des Approches de Pronostic	32
Figure I.11: Architecture d'un réseau de neurone à trois couches	36

Chapitre II. Détection et classification par SVM des défauts de moteur à induction en utilisant WPA

Figure II.1: Topologie du circuit des mailles du rotor	41
Figure II.2: Circuit de mailles de rotor dans le cas d'une barre de rotor	44
Figure II.3: la nature hiérarchique de l'analyse multi résolution	49
Figure II.4 : Reconstruction unidimensionnelle par ondelettes	49
Figure II.5 : Décomposition en cascade par ondelettes	50
Figure II.6: Décompositions en ondelettes par paquets d'ordre 3	52
Figure II.7: Répartition des nœuds dans un arbre de décomposition	53
Figure II.8: Exemples de quatre problèmes de discrimination binaire impliquant la séparation des points bleus des croix rouges. Les frontières de décision sont représentées en bleu	56
Figure II.9: Les deux notions de marges de la discrimination linéaire séparable en deux dimensions.	57
Figure II.10: Une illustration du concept de vecteur support dans le contexte de problèmes linéairement séparables. Les vecteurs de support pour les deux classes sont entourés d'un cercle rouge.	60
Figure II.11: Illustration du concept d'écart (à gauche) et de la fonction de coût charnière (à droite). Dans cet exemple, tous les écarts sont nuls, à l'exception d'un point bleu mal classé. L'écart mesure la distance entre le point et la marge numérique de l'hyperplan de séparation	62
Figure II.12: étapes pour la détection de défauts à l'aide d'ondelettes par paquets et SVM	64
Figure II.13 : Modèle réduit de machine triphasé à induction à cage d'écureuil sous Matlab Simulink	65
Figure II.14 : Courant statorique machine saine sans charge	65
Figure II.15 : Courant statorique machine saine avec charge	65
Figure II.16 : Courant statorique cassure d'une barre	65

FigureII.17 : Courant statorique cassure de deux barres	65
FigureII.18 : Courant statorique cassure de trois barres	66
FigureII.19 : Variation d'énergie dans les bandes de fréquence au 9eme niveau (db44)	
□ Machine saine sans charge x Machine saine avec charge ◇ Machine avec cassure de barre ' Machine avec cassure de deux barres *Machine avec cassure de trois barres	68

Chapitre III. Approche proposée pour l'estimation du RUL en utilisant en utilisant l'apprentissage profond

Figure III.1: Principes de dégradation et de pronostic	73
Figure III.2: la différence entre l'intelligence artificielle, machine Learning et le Deep Learning	74
Figure III.3: Neurone artificiel simple avec une seule sortie	75
Figure III.4: Réseau de neurones à un seul niveau	75
Figure III.5: Le réseau de neurones multi-niveaux	76
Figure III.6: Illustration d'un perceptron multicouche	78
Figure III.7: Cellule RNN à travers le temps	80
Figure III.8: Types de séquences d'entrée pour un réseau de neurones récurrent	81
Figure III.9: Cellules LSTM	81
Figure III.10: Chaîne de cellules LSTM	82
Figure III.11: état de la cellule	82
Figure III.12: Une porte LSTM	82
Figure III.13: porte d'oubli (Forget gate)	83
Figure III.14: Mise à jour de la cellule LSTM	83
Figure III.15: Mise à jour de l'ancien état de cellule	84
Figure III.16: la sortie de réseau LSTM	84
Figure III.17: Modèle LSTM pour l'estimation RUL	86
Figure III.18: Schéma des modules du turboréacteur.	87
Figure III.19: lectures des 21 capteurs de base de données CMAPSS	91
Figure III.20 : Données de capteur normalisées du moteur 177 dans FD002	94
Figure III.21 Estimation RUL pour quatre moteurs aléatoires	96

Liste des tableaux

Chapitre I. État de l'art sur le diagnostic et le pronostic

Tableau I.1 : Quelques caractéristiques du domaine temporel	14
Tableau I.2 : quelques caractéristiques du domaine fréquentielles	15

Chapitre II. Détection et classification par SVM des défauts de moteur à induction en utilisant WPA

Tableau II.1 le résumé de deux scénarios possibles	61
Tableau II.2: Bandes de fréquences résultantes de la décomposition en ondelettes par paquets	66
Tableau II.3: Bandes de fréquences résultantes de la décomposition en ondelettes par paquets dans la bande [0-312.5Hz]	67
Tableau II.4 : l'énergie des 16 premières bandes de 9 ^{ième} niveau (ondelettes par paquets db44) pour les cinq modes de fonctionnement de la machine	67
Tableau II.5 : Description de la base de données	69
Tableau II.6 : Résultat de classification	70

Chapitre III. Approche proposée pour l'estimation du RUL en utilisant en utilisant l'apprentissage profond

Tableau III.1 : les fonctions d'activation les plus couramment utilisées	77
Tableau III.2 : Liste et significations physiques du jeu de données C-MAPSS.	88
Tableau III.3 : Description détaillée du jeu de données C-MAPSS	91
Tableau III.4 Description détaillée de données CMAPSS après l'augmentation	93
Tableau III.5 Comparaison de RMSE et SMAPE sur le jeu de données C-MAPSS pour les modèles de prédiction RUL.	97

Introduction générale

Dans la médecine, des analyses médicales sont effectuées, basées sur les mesures de certains paramètres liés aux conditions de santé, ont pour objectif de détecter des anomalies, et ensuite de diagnostiquer des maladies et de prévoir leur évolution. Par analogie, les procédures techniques de gestion de la santé sont utilisées pour capturer l'état fonctionnel des équipements industriels à partir d'enregistrements historiques de paramètres mesurables des systèmes ou de sous-systèmes, comme indiqué par le travail de Qiu, S. et al. (2023) [1].

Une gestion réussie des événements anormaux nécessite la détection rapide des conditions anormales, le diagnostic et le pronostic sont des éléments de base de la procédure de correction du défaut de l'équipement et le contrôle des conditions de sécurité, comme l'explique Jarrell, D. B. et al. (2004) [2]. Du point de vue de la sécurité, la connaissance de l'état des équipements (diagnostics) et la prédiction de leur évolution future (pronostics) permettent un fonctionnement plus sûr et plus fiable, avec un minimum d'arrêts non programmés de la production, et une maintenance économique, cela est souligné par le travail de Jardine, A. K. et al. (2006) [3]. Mais le problème est qu'il devient de plus en plus difficile de compter uniquement sur des opérateurs humains pour faire face aux événements anormaux et aux urgences. En particulier, les tâches de diagnostic et de pronostic dans une usine complexe sont rendues assez difficiles par la variété des occurrences de défaillance d'équipement et des réponses de processus associées, et par le grand nombre de variables de processus surveillées (de l'ordre d'un millier dans les usines de processus modernes) qui conduisent à une surcharge d'informations. Cela peut donner lieu à un accident catastrophique avec des impacts importants sur l'économie, la sécurité et l'environnement (l'accident de la raffinerie pétrochimique du Koweït en 2000 a entraîné des dommages estimés à 100 millions de dollars, comme l'a démontré Venkatasubramanian, V. et al. (2003) [4]). D'autre part, les accidents mineurs se produisent relativement fréquemment et peuvent entraîner de nombreux accidents de travail et maladies professionnelles, et selon les études menées par Nimmo, I. (1995) [5] et Laser, M. (2000) [6], les coûts pertinents pour l'industrie, aux États-Unis et au Royaume-Uni, sont estimés à environ 20 milliards de dollars par an. Cela explique le grand intérêt et l'attention portés au développement de méthodes et de procédures efficaces pour la gestion des événements anormaux, et la création de méthodes adéquates et automatisées pour le PHM des équipements industriels, en soutien aux opérateurs humains.

L'état de l'équipement est généralement surveillé à intervalles réguliers et une fois que la lecture du signal surveillé dépasse un niveau de seuil bien déterminé au préalable par le constructeur, un avertissement est déclenché et des actions de maintenance peuvent être planifiées en fonction de la prédiction de l'évolution future du processus de dégradation. L'intervalle de surveillance influence le coût global et les performances de l'équipement : un intervalle plus court peut augmenter le coût de la surveillance, tandis qu'un intervalle plus long augmente le risque de défaillance. Mais aussi, le système de surveillance doit être fiable afin d'éviter les fausses alarmes, qui peuvent entraîner une décision qui doit être prise à chaque fois qu'une alarme est signalée ; dans l'autre côté, ignorer une alarme peut entraîner de graves conséquences en cas d'accidents avec dommages et pertes de biens. Alors la première possibilité consiste, en premier temps, à effectuer une enquête plus approfondie sur l'alarme, sans arrêter l'équipement ; une autre possibilité consiste à arrêter l'équipement pour une révision. Dans la première possibilité, une fausse alerte entraînerait un surcoût dû au temps et à la main-d'œuvre

nécessaires pour faire le diagnostic ; la deuxième possibilité pourrait entraîner des pertes plus importantes, où la perte de production et les coûts de main-d'œuvre se produisent simultanément.

La planification dynamique de la maintenance conditionnelle est envisagée comme une solution, mais elle constitue une tâche complexe et difficile, qui nécessite la prédiction de l'évolution des variables surveillées représentant l'état de l'équipement. Lors de la détection de précurseurs de défaillance, le pronostic devient une tâche fondamentale ; il s'agit de prévoir la fiabilité ou la probabilité de défaillance de l'équipement à l'avenir, et la durée de vie résiduelle (RUL), c'est-à-dire la durée de vie pendant laquelle l'équipement continuera à remplir sa fonction conformément aux spécifications de conception. Ce processus de prédiction / prévision / extrapolation doit tenir compte de l'évaluation de l'état actuel et des conditions opérationnelles futures attendues. Alors la tâche de pronostic est l'intégration intelligente des informations et des données disponibles dans des modèles précis résolus par des algorithmes de calcul efficaces. Dans cette partie, l'accent est mis sur les pronostics. En effet, le développement de systèmes de pronostic peut s'appuyer sur des informations et des données assez différentes sur le comportement passé, présent et futur de l'équipement ; il peut y avoir des situations dans lesquelles des données statistiques suffisantes et pertinentes sur les défaillances de l'équipement sont disponibles, d'autres dans lesquelles le comportement de l'équipement est connu de manière suffisamment précise pour permettre la construction d'un modèle suffisamment précis. La base de données C-MAPSS est largement utilisée dans ce domaine. Cependant, afin d'améliorer davantage les résultats, nous utilisons également un processus d'augmentation des données. Cette technique consiste à générer de nouvelles données synthétiques en manipulant et en transformant les données d'origine. Elle vise à augmenter la diversité et la quantité des données disponibles pour l'entraînement du modèle.

L'ajout du processus d'augmentation des données vise à améliorer la capacité du modèle à généraliser et à capturer les différentes variations et schémas de dégradation des moteurs à double flux. En conséquence, nous anticipons une amélioration significative de la précision de la régression et des résultats de prédiction de la RUL.

L'organigramme de la thèse est schématisé comme suit :

L'introduction aborde tout d'abord une vue d'ensemble succincte et étendue du domaine de recherche. Cette introduction décrit également l'importance, les objectifs et les innovations de cette recherche. Elle montre comment retracer les objectifs de recherche à partir des problèmes identifiés. L'originalité et les principales contributions de ce travail sont aussi présentées.

Le chapitre I a pour objet de situer et limiter le cadre de notre travail. Il présente une revue détaillée de la littérature des techniques de supervision actuelles de diagnostic et de pronostic.

Le chapitre II décrit les étapes principales de la détection et la classification SVM des défauts de moteur à induction en utilisant WPA. Nous avons essayé de déterminer le vecteur des attributs optimal adapté au diagnostic en utilisant les ondelettes par paquets et la classification SVM. Enfin, les performances de différents classifieurs sont présentées, validées et comparées sur les données du modèle de machine à induction sous MATLAB® /Simulink.

Le chapitre III est réservé à l'explication d'une nouvelle approche pour l'estimation du RUL en utilisant le Deep Learning, avec un éclaircissement de la contribution qui est une

méthode d'augmentation des données de (C-MAPSS NASA), pour améliorer la précision et la robustesse de la régression. Les performances de pronostic sont présentées, validées et comparées sur des données avec et sans augmentation. Enfin, la thèse se conclut par une conclusion générale résumant les contributions de nos travaux de recherche présentés dans cette thèse et décrivant les orientations futures des travaux.

Chapitre I.

*État de l'art sur le diagnostic et le
pronostic*

I.1 Introduction :

La sécurité de fonctionnement des systèmes industriels et la garantie de leur disponibilité à moindre coût est l'un des enjeux les plus importants de cette industrie. La disponibilité correspond à la durée pendant laquelle un système industriel est capable d'accomplir sa tâche. Cela dépend de l'efficacité de la stratégie de maintenance mise en place. En conséquence, la maintenance est une partie croissante de l'industrie d'aujourd'hui. Ces dernières années, sa politique est passée d'une maintenance purement corrective à une maintenance prenant en compte la santé du système, c'est-à-dire prédiser ses défaillances en prédisant sa durée de vie restante. La mise en œuvre de la nouvelle politique de maintenance nécessite une approche scientifique, notamment différentes tâches telles que la surveillance, la modélisation, l'analyse et l'aide à la décision. Ces tâches font partie du concept PHM (Prognostics and Health Management) ou bien Pronostic et gestion de la santé pouvant se traduire par la prédiction et la gestion de la santé des systèmes industriels.

Le PHM est une discipline dynamique, intéressé par l'étude des mécanismes de défaillance des systèmes pour mieux gérer l'utilisation des informations sur l'état du système. Le PHM fait référence à un domaine qui se concentre sur les principales tâches associées à la détection de la dégradation des systèmes ou des sous-systèmes, au diagnostic des pannes et à la prédiction des pannes. Cela a été souligné dans les travaux de Zio (2013) [7].

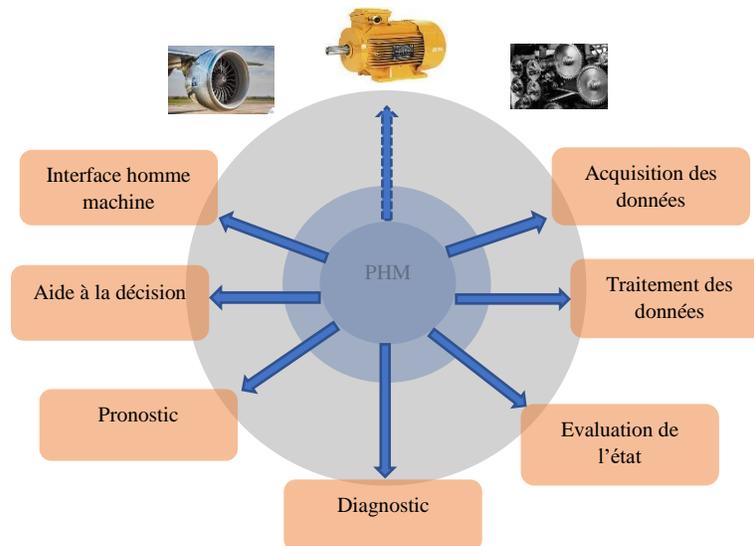


Figure I.1: L'architecture de PHM

Le processus PHM se compose de sept modules, comme la montre la figure I.1 :

1. Acquisition de données : y compris la mesure de grandeurs physiques telles que la tension, le courant et la température etc., avec l'utilisation de capteurs, de logiciels et d'observation humaine. Ces données sont obtenues grâce à un système d'acquisition qui collecte, prétraite les données pour les envoyer à d'autres modules et les stocker dans une base de données fiable et sécurisée.
2. Traitement du signal/des données : Analyse et interprétation des signaux afin d'extraire des informations caractérisant le comportement du système dans les domaines temporel et/ou fréquentiel.

3. Évaluation de l'état actuel : sera obtenu à partir de caractéristiques du système et permettra l'utilisation du comportement nominal pour détecter différentes anomalies possibles.
4. Diagnostic : localiser et identifier la cause de l'anomalie ou de la panne.
5. Pronostic : la durée de vie avant défaillance est prédite en fonction de l'état actuel du système et des résultats de la détection et/ou du diagnostic.
6. Aide à la décision : concernant les stratégies de maintenance mises en œuvre pour maintenir le système en bon état. Le module se base sur toutes les informations obtenues (état actuel du système, RUL, connaissances contextuelles, etc.).
7. Interface Homme-Machine (IHM) : Fournit un moyen de présenter et de stocker des informations utiles sous différentes formes.

I.2 Forme de maintenance :

Il existe différentes formes de la maintenance ainsi que leur évolution conceptuelle jusqu'à présent.

Il faut d'abord commencer par une définition standard de la maintenance, selon la norme NF EN 13306 (2010), la maintenance peut être définie comme l'ensemble de toutes les actions techniques, administratives et de management durant le cycle de vie d'un bien, destinées à le maintenir ou à le rétablir dans un état dans lequel il peut accomplir la fonction requise. », elle comprend ainsi un ensemble d'actions de dépannage, de réparation, de contrôle et de vérification des équipements, matériels, et doit contribuer à l'amélioration des processus industriels. Dans la vision traditionnelle, la fonction maintenance permet de garantir les caractéristiques de sûreté de fonctionnement des matériels, notamment la disponibilité. Elle vise donc globalement à appréhender les phénomènes de défaillance et à agir en conséquence, afin d'assurer que le système (le bien) soit à même de remplir la fonction pour laquelle il a été conçu (maintien en condition opérationnelle MCO), cela a été souligné par Gouriveau, R (2013) [8].

L'analyse des différentes formes de maintenance repose sur 4 concepts :

- Les événements qui sont à l'origine de l'action : référence à un échancier, la subordination à un type d'événement (auto diagnostic, information d'un capteur, mesure d'une usure, etc.), l'apparition d'une défaillance.
- Les méthodes de maintenance : maintenance préventive systématique, maintenance préventive conditionnelle, maintenance corrective.
- Les opérations de maintenance : inspection, contrôle, dépannage, réparation, etc.
- Les activités connexes : maintenance d'amélioration, rénovation, reconstruction, modernisation, travaux neufs, sécurité, etc.

Le diagramme suivant synthétise selon la norme NF X 60-000 les différentes méthodes de maintenance.

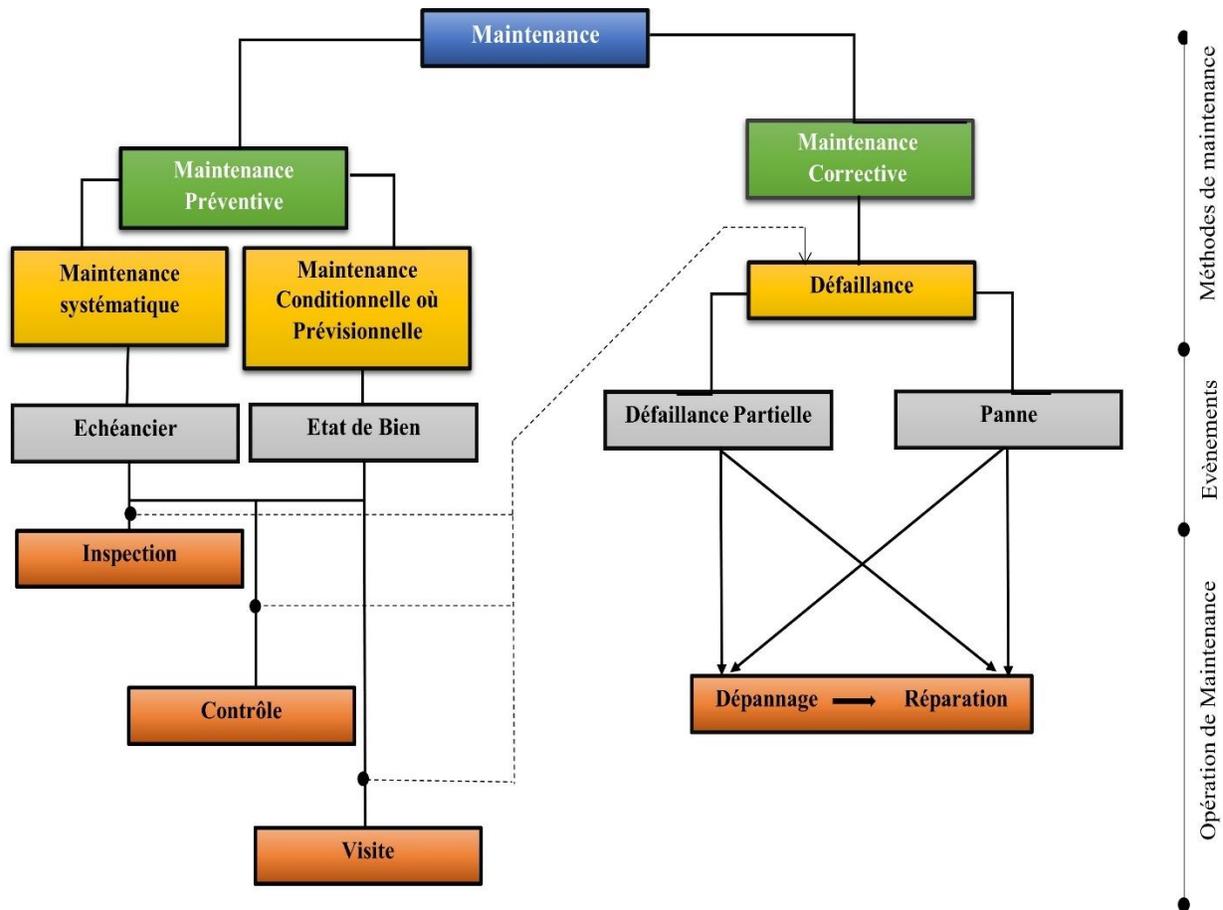


Figure I. 2 : Méthodes de maintenance

Lors du choix d'une méthode de maintenance, il est nécessaire de comprendre les objectifs de la direction de l'entreprise, des directions politique de maintenance, mais aussi de comprendre le fonctionnement et les caractéristiques de l'équipement, le comportement de l'équipement en fonctionnement, les conditions d'exploitation, l'application de chaque méthode, les coûts de maintenance et le coût des pertes de production.

Notons ici que la maintenance se décompose en deux grandes parties, selon le schéma ci-dessus

I.2.1 La maintenance corrective :

La maintenance corrective, parfois appelée curative (terme non normalisé), vise à redonner aux équipements la qualité perdue nécessaire à leur utilisation. Des défauts, dysfonctionnements ou dommages divers qui peuvent entraîner une indisponibilité des équipements concernés et/ou une diminution de la quantité et/ou de la qualité des prestations fournies nécessitant une maintenance corrective.

Elle se décompose en deux types :

- *La maintenance palliative* : qui comprend les interventions de type dépannage
- *La maintenance curative* : qui comprend les interventions de type réparations.

Les différentes phases d'une intervention corrective sont les suivantes :

- Test : comparaison des réponses d'un dispositif à une sollicitation avec celle d'un dispositif de référence.
- Détection : action de déceler par une surveillance accrue, l'apparition d'une défaillance.
- Localisation : action de localiser les éléments par lesquels la défaillance se manifeste.
- Diagnostic : identification de la cause de la défaillance à l'aide d'un raisonnement logique.
- Dépannage, réparation.

Et pour rendre semblables la terminologie en matière de maintenance selon la norme NF X 60-010, on présente les définitions suivantes :

Défaillance : Altération ou cessation de l'aptitude d'une entité à accomplir une fonction requise. Après défaillance d'une entité, celle-ci est en état de panne, donc la défaillance est un passage d'une entité d'un état de fonctionnement normal à un état de fonctionnement anormal ou de panne., on distingue deux types de défaillance :

1. *Défaillance partielle* : altération ou dégradation de l'aptitude d'un bien à accomplir la fonction requise.
2. *Défaillance complète* : cessation de l'aptitude d'un bien à accomplir la fonction requise.

Défaut : un défaut représente un écart entre une caractéristique réelle d'une entité et la caractéristique voulue, cet écart dépassant des limites d'acceptabilité. Un défaut dans le système n'affecte pas en général l'aptitude du système à accomplir une fonction requise. Par conséquent, on peut constater que le défaut ne conduit pas toujours à une défaillance. De ce fait, le défaut est vu comme une opinion sur le bon fonctionnement. Par contre une défaillance conduit systématiquement à un défaut, et d'autre manière un défaut peut générer une défaillance du système.

Panne : est définie par l'inaptitude d'une entité à accomplir une fonction requise. Une panne est la conséquence d'un défaut qui est donc toujours associée à une défaillance.

I.2.2 La maintenance préventive :

Maintenance effectuée dans l'intention de réduire la probabilité de défaillance d'un bien ou d'un service rendu. On peut subdiviser la maintenance préventive en trois types:

1. *La maintenance préventive systématique* : c'est une maintenance obéissant à un échéancier établi en fonction du temps et du nombre d'unités d'exploitation. Mais selon la définition de la « (normes AFNOR X 60 010 et 60 011 » : c'est celle exécutée à des intervalles de temps préétablis ou selon un nombre défini d'unités d'usage mais sans contrôle préalable de l'état de l'équipement ».
2. *La maintenance préventive conditionnelle* : d'après la définition Afnor, il s'agit d'une forme de maintenance préventive basée sur une surveillance de fonctionnement du bien et/ou des paramètres significatifs de ce fonctionnement, et intégrant les actions qui en découlent. La maintenance conditionnelle assure une surveillance continue des équipements en service, avec des décisions d'intervenir lorsqu'il existe une évidence expérimentale de défauts imminents ou de seuils de dégradation prédéterminés. Cela implique certains types de pannes, de pannes arrivantes progressivement ou par dérive. L'étude de la dérive dans le cadre d'interventions de maintenance préventive permet de détecter des seuils d'alarme en technologie mécanique et électronique.

3. *La maintenance préventive prévisionnelle* : C'est une maintenance qui commence par le suivi de l'état du matériel et procède à des analyses périodiques pour déterminer l'évolution de la dégradation du matériel et la période d'intervention. La « maintenance prédictive » est à éviter car en maintenance on prévoit plutôt qu'on ne prédit).

Définition selon les normes européennes : "Maintenance conditionnelle basée sur des prédictions extrapolées d'analyse et d'évaluation des paramètres importants de la dégradation des actifs"

I.3 La maintenance conditionnelle et prévisionnelle CBM

L'objectif de la maintenance prévisionnelle conditionnelle (CBM) comme le souligne Cochetoux, P. (2010) [9] est généralement d'identifier les défaillances avant qu'elles ne deviennent graves et critique. Selon Bengtsson, M et al. (2004) [10] ; Garga, A. K et al. (2001) [11], cela permet d'avoir une vision et une planification plus précises. Il est basé sur une analyse des tendances des grandeurs caractéristiques du système et des indicateurs de performance par rapport aux limites fixées par le fabricant, dans le but de détecter, d'analyser et de corriger les problèmes dans le processus, cette approche, comme le souligne Azam, M et al. (2002) [12], est mise en place. Il est nécessaire de mesurer les performances à l'aide d'une instrumentation ou de ses composants (tels que des capteurs) adaptés au procédé. Le traitement de ces informations est également utilisé pour le diagnostic et le pronostic en estimant la durée de vie restante (RUL). Le CBM nécessite moins de temps que les interventions de maintenance corrective classique et permet une utilisation prolongée du système tout au long de son cycle de vie, ce qui entraîne moins d'interventions et une meilleure planification de celles-ci. Alors. D'après Muller, A et al. (2005) [13], il existe quatre approches architecturales des stratégies de maintenance. Prenant en charge un ensemble de processus dans leur couche de traitement de l'information pour prendre en charge les décisions de maintenance :

Watch Dog (2003) [14], PEDS (Prognostic Enhancements to Diagnosis Systems) (2002) [15], SIMP (Integrated Predictive Maintenance Systems) (2005) [13], OSA-CBM (Open System Architecture for Condition Based Maintenance) [16].

1. *Méthode Watch Dog*: Djurdjanovic, D et al. (2003) [14] a proposé le système Watch Dog qui combine des composants physiques et logiciels, rassemblant ainsi toutes les connaissances nécessaires (modèles, méthodes, etc.). Le procédé permet d'obtenir des indicateurs de santé représentatifs par la collecte et le traitement de données du système, et effectue simultanément un diagnostic et une évaluation pronostic pour évaluer l'état de santé du système. Sa particularité est qu'elle restitue plutôt des modèles statistiques de défaillance plutôt qu'un diagnostic ainsi que des modèles de pronostic issus des données. Cependant, les mesures des données et la fonction d'aide à la décision sur le système sont réalisées en dehors de Watch Dog.
2. *Méthode PEDS* : Cette méthode proposée par Byington, C. S et al. (2002) [15] peut combiner deux fonctions de pronostic, l'une dite verticale, qui est basée sur des statistiques de données de défaillance, et l'autre dite horizontale et consécutive à la détection d'un défaut au moyen d'une fonction de diagnostic.
3. *Méthode SIMP* : Développée par Muller A et al. (2005) [13], cette méthode est une stratégie visant à intégrer globalement le pronostic et la maintenance prévisionnelle dans le système

d'information d'une entreprise et ses différents processus internes et externes. La méthode SIMP met en œuvre une série de trois sous-processus séquentiels : surveillance, le pronostic et aide à la décision.

4. Méthode OSA-CBM : Introduite par Lebold, M et al. (2001) [16], l'architecture illustrée à la Figure I.3 est divisée en 7 couches, créant une succession linéaire de sous-processus nécessaires pour effectuer cette maintenance. Le pronostic suit un processus " Health assessment " qui définit l'état du système. Bien que cela ne soit pas explicitement indiqué, ce processus intègre des diagnostics pour comprendre les schémas de défaillance/dégradation courants. Le pronostic est en amont du processus « Decision support », qui permet de sélectionner les opérations de maintenance à programmer pour remettre le système dans un état et des performances donnés. L'architecture précédente a été décrite en termes de couches fonctionnelles, couvrant le domaine de l'acquisition de données jusqu'à la prise de décision, identifiant les principales données échangées entre sous-processus, et modélisées par les travaux de l'Institut Machinery Information Management Open Standards Alliance (MIMOSA), qui fournit une architecture standardisée pour un système de maintenance conditionnelle CBM appelée (OSA-CBM) (Open System Architecture for Condition Based Maintenance).

Les fonctions générales des couches sont spécifiées ci-dessous :

- Couche 1 : Couche d'acquisition de données (Data Acquisition layer) c'est la couche d'acquisition (fonction capteur) qui convertit les grandeurs physiques en signaux électriques.
- Couche 2 : Couche de manipulation des données (Data Manipulation layer), c'est la couche qui traite les signaux de la couche d'acquisition pour l'extraction des grandeurs représentatives de la santé du système.
- Couche 3 : Couche de surveillance de l'évolution des données ou Data Evolution Monitoring (Condition Monitoring/State Detection Layer) c'est la couche qui extrait les données des couches précédentes pour les comparer aux valeurs limites. Cette couche génère des alertes lorsque les limites sont franchies.
- Couche 4 : Couche de Diagnostic (Health Assessment layer), cette couche reçoit des données de la couche de surveillance pour déterminer si la santé du système, du sous-système ou du composant surveillé est dégradée. Elle génère un diagnostic d'une ou plusieurs conditions de panne associées à un niveau de confiance. Les diagnostics doivent prendre en compte les tendances de l'historique de santé, de la charge, de l'état de fonctionnement et de l'historique de maintenance.
- Couche 5 : Couche pronostic (Pronostics layer), reçoit les données de toutes les couches précédentes. Son principal objectif est de projeter l'état de santé actuel du système dans le futur en fournissant des informations sur la durée de vie restante du système.
- Couche 6 : Couche d'aide à la décision (Decision Support layer), c'est la couche qui fournit des recommandations et des alternatives pour maintenir le système en bon état. Pour ce faire, elle utilise les données des couches diagnostic et pronostic. Les

recommandations peuvent être des opérations de maintenance planifiées ou des modifications de la configuration de fonctionnement du système.

- Couche 7 : Couche Présentation (Presentation/GUI layer), c'est la couche qui assure l'interface entre le système et un ou plusieurs opérateurs, notamment pour présenter les informations produites par ces derniers.



Figure I. 3 : Les couches de l'architecture OSA-CBM

Par rapport aux stratégies traditionnelles de la maintenance, la maintenance conditionnelle (CBM) offre un avantage majeur pour la minimisation des instances des défaillances des équipements, une réduction dans les activités planifiées de maintenance et une maximisation de la disponibilité des équipements. Dans ce contexte, des détails sur les différentes étapes de cette stratégie sont présentés. En partant de l'étape d'acquisition des données et leur conditionnement vers l'étape de diagnostic, et de pronostic comme support à la décision tout en passant par les techniques existantes de traitements de signal considérées comme un pas fondamental pour l'extraction et la sélection des caractéristiques de diagnostic et de pronostic pertinentes.

I.4 Préparation des données

I.4.1 Prétraitement des données

En général, dans toute application basée sur les données, et en particulier pour la maintenance prédictive, selon Zhang, L et al. (2010) [17], le prétraitement des données est essentiel pour rendre les données significatives et utilisables.

Nous considérerons tout traitement effectué sur les données avant l'apprentissage d'un modèle comme un prétraitement. On peut citer le nettoyage des données, le filtrage du bruit, la normalisation et la sélection de caractéristiques etc.

On peut donc considérer le prétraitement comme une étape formée de plusieurs étapes, chacune ayant un but particulier, dont l'ordre peut parfois être interverti mais dans lequel la propriété commutative n'est en général non rempli.

Les données impliquées dans chaque problème lié à la maintenance prédictive ont des propriétés spécifiques. Par exemple, les données relatives à la détection de défauts ont tendance à être fortement déséquilibrées car les informations concernant les situations défectueuses sont beaucoup moins fréquentes que celles concernant les situations sans défaut.

En général, les propriétés des données doivent être prises en compte lors du choix d'une stratégie de prétraitement, par exemple, tous les ensembles de données pour les problèmes de prédiction de la durée de vie restante (RUL), ne sont pas identiques. Les propriétés de chaque ensemble de données doivent être déterminées. De plus, parfois, avec les mêmes propriétés, un schéma de prétraitement fonctionne pour un problème et pas pour un autre.

Dans les systèmes hautement dynamiques, des stratégies de prétraitement appropriées sont à la base de la construction d'un modèle précis et robuste.

Le reste de cette partie établit une taxonomie, fournit des descriptions brèves mais au-delà d'une simple citation de plusieurs techniques pour chacune des étapes de prétraitement pour la surveillance du système dans la maintenance prédictive.

Nous définissons le prétraitement comme l'ensemble des actions effectuées sur les données brutes avant une performance de modélisation ultérieure. Dans le but d'améliorer les capacités de modélisation, l'amélioration peut être comprise de plusieurs manières, telles que l'augmentation de la précision, l'augmentation de la robustesse, le raccourcissement du temps de calcul, la diminution des besoins en mémoire et/ou en puissance de calcul, ou la réduction des coûts.

I.4.1.1 Taxonomie

La classification implique six étapes de prétraitement, que nous examinerons ensuite en détail :

- 1. Nettoyage des données :** L'hypothèse de données complètes, fiables et sans bruit est utilisée dans la plupart des stratégies basées sur les données. Cependant, les données du monde réel sont rarement aussi pures que les données idéales, ce qui nécessite le développement de solutions pour faire face aux valeurs aberrantes et au bruit. De plus, en raison de la nature des données ou en raison d'un manque de stratégie d'acquisition de données adéquate, des fonctionnalités redondantes ou non pertinentes pourraient être considérées dans l'ensemble de données, qui pourraient être traitées à la fois dans l'étape de nettoyage des données ou plus tard dans l'étape d'ingénierie des caractéristiques.

L'une des techniques appliquées lors de l'étape de nettoyage des données est le filtrage passe-bas, car le bruit domine le signal dans les hautes fréquences. La fréquence de coupure doit être au moins le double de la fréquence utile du signal. Pour filtrer les signaux acquis, une technique de "moyenne synchronisée dans le temps" (TSA) étudiée par Jardine, A. K et al. (2006) [3] ; Heng, A et al. [19], a été utilisée pour moyenniser le signal brut sur plusieurs évolutions afin de supprimer le bruit aléatoire. Il est exprimé par la formule suivante :

$$\bar{s}(t) = \frac{1}{N} \sum_{n=0}^{N-1} s(t + nT), \quad 0 \leq t \leq T \quad (\text{I.1})$$

Où $s(t)$ est le signal, T est la période de calcul de la moyenne et N est le nombre d'échantillons moyen. L'auteur McFadden, P. D (1987) [20] a démontré que la suppression de bruit périodique de fréquence connue peut optimiser en choisissant un nombre approprié de moyenne.

L'annulation adaptative du bruit (ANC) est un autre algorithme efficace proposé par Widrow, B et al. (1975) [21]. Constaté qu'il sépare les signaux de vibrations des éléments roulants des roulements du reste des signaux vibratoires pour aider à diagnostiquer les défauts.

Techniques « Adaptive Line Enhancer » (ALE) ont été proposées par Shiroishi, J et al. (1999) [22] et "Self-Adaptive Noise Cancellation" (SANC) proposé par Antoni, J et al. (2004) [23] pour réduire le nombre de capteurs en remplaçant le bruit de référence par une version décalée du signal mesuré. Cependant, SANC présente certains inconvénients dans les applications pratiques, telles que l'ajustement des paramètres lors de la phase d'adaptation.

Pour surmonter ce problème, Antoni, J et al. (2004) [24] ont proposé une estimation du gain en fréquence pour l'annulation du bruit dans le domaine fréquentiel.

Qiu, H et al. (2003) [25] ont proposé un filtre à ondelettes pour améliorer les caractéristiques faibles pour l'identification des défauts.

Karimi, M et al. (2006) [26] ont appliqué la technique de "Blind Deconvolution" (BD), pour leur étude pour récupérer le signal souhaité sans connaissance à priori.

2. Normalisation des données : Les données provenant d'une variété de sources hétérogènes sont facilement collectées, faisant de l'ensemble de données réel une collection d'ensembles de données obtenus de différentes manières par différentes parties du système. Cette fusion d'ensembles de données, appelée intégration de données, n'est pas considérée par de nombreux auteurs, comme faisant partie du prétraitement, mais plutôt comme faisant partie de la collecte de données. Certains algorithmes sont très sensibles à diverses échelles et plages de variables, ce qui peut entraîner une dégradation des performances si les données ne sont pas homogénéisées.

Il existe diverses méthodes qui peuvent être utilisées pour la normalisation des ensembles de données, notamment la normalisation min-max, la normalisation du z-score, decimal scaling, standardized moment, etc. Les deux techniques de normalisation les plus populaires et largement utilisées, qui sont min-max. et normalisation du z-score.

• **Normalisation Min-max :**

Dans la normalisation min-max, telle que décrite par Han, J et al. (2006) [27], les caractéristiques sont normalisées dans la plage [0,1] à l'aide de l'équation suivante :

$$\hat{v} = \frac{v - \min_A}{\max_A - \min_A} \quad (\text{I.2})$$

Ici, \min_A et \max_A sont les valeurs minimale et maximale de la caractéristique A . La valeur d'origine et la valeur normalisée de l'attribut A sont représentées par v et \hat{v} respectivement. On peut observer à partir de l'équation ci-dessus que les valeurs maximales et minimales des caractéristiques sont mappées à 1 et 0, respectivement.

• **Normalisation z -score :**

La valeur normalisée de l'attribut pour la normalisation du z-score peut être calculée à l'aide de l'équation suivante, comme indiqué par Han, J et al. (2006) [27] :

$$z = \frac{v - \mu_A}{\sigma_A} \quad (I.3)$$

Ici, μ_A et σ_A sont la moyenne et l'écart type de l'attribut A respectivement. Les valeurs de caractéristique d'origine et normalisées sont représentées par v et z , respectivement. Après la normalisation, la moyenne et l'écart type de toutes les valeurs des caractéristiques deviennent 0 et 1, respectivement.

3. **Transformation des données :** Bien que les étapes précédentes et certaines des étapes postérieures impliquent effectivement des transformations des données, nous réservons ce nom aux transformations des données au moyen de certaines fonctions. Par exemple si nous effectuons la maintenance prédictive de certaines machines industrielles en utilisant des informations sur la composition chimique des déchets résiduels par des données spectroscopiques (un processus connu sous le nom d'étalonnage multivarié chimiométrique), nous pouvons utiliser la loi de Beer-Lambert pour déterminer la relation entre la composition chimique et les données spectroscopiques d'absorbance.

La polyvalence des données utilisées dans la maintenance prédictive ouvre de nombreuses possibilités de transformation. Il existe deux branches principales de la transformation des données pour la maintenance prédictive, que nous identifions comme la transformation statistique et le traitement du signal.

- **Transformation statistique :** Les transformations statistiques sont inspirées des transformations historiquement utilisées dans l'inférence statistique comme l'explique Johnson, N et al. (1995) [28]. L'utilisation de l'un ou l'autre type dépend de l'application et du type de données.

Dans les statistiques, les transformations de données sont appliquées lorsque certaines informations à priori sont disponibles. Certains des plus célèbres sont la transformation *logit* de la régression logistique, liée aux réseaux de neurones et aux deep Learning ; transformation racine carrée « square root transformation » à partir d'une régression quadratique ; et transformation réciproque « reciprocal transformation ».

- **Traitement du signal :** Nous nous concentrerons sur les données de forme d'onde, car la plupart des données de maintenance prédictive sont basées sur ce type. Les données de forme d'onde peuvent être observées à partir de trois domaines liés : le domaine temporel, le domaine fréquentiel, et le domaine temps-fréquence.

➤ **L'analyse du domaine temporel :** est l'analyse des données de forme d'onde, qui sont, d'un point de vue mathématique une séquence chronologique, on peut citer la variance, la moyenne et Kurtosis.....etc. Malheureusement, l'approche dans le domaine temporel seul

est souvent incapable d'identifier le composant défectueux et donc insuffisant pour diagnostiquer la plupart des problèmes du procédé.

Tableau I.1 : Quelques caractéristiques du domaine temporel

N°	Caractéristiques	Description mathématique
01	Moyenne	$T_m = \frac{1}{n} \sum_{i=1}^n x_i$
02	Racine Carrée Moyenne	$T_{rms} = \left[\frac{1}{n} \sum_{i=1}^n x_i^2 \right]^{\frac{1}{2}}$
03	Root	$T_r = \left[\frac{1}{n} \sum_{i=1}^n x_i ^{\frac{1}{2}} \right]^2$
04	Std	$T_{sd} = \left[\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \right]^{\frac{1}{2}}$
05	Skewness	$T_{sk} = \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{(n-1)T_{sd}^3}$
06	Kurtosis	$T_{ku} = \frac{\sum_{i=1}^n (x_i - \bar{x})^4}{(n-1)T_{sd}^4}$
07	Facteur de forme	$T_{sf} = \frac{T_{rms}}{\bar{x}}$
08	Facteur de crête	$T_{cf} = \frac{x_{max}}{x_{rms}}$
09	Facteur d'impulsion	$T_{if} = \frac{x_{max}}{\bar{x}}$
10	Clearance factor	$T_{clf} = \frac{x_{max}}{x_r}$
11	Skewness factor	$T_{skf} = \frac{T_{sk}}{T_{rms}^3}$
12	Kurtosis factor	$T_{kuf} = \frac{T_{ku}}{T_{rms}^4}$

Ici x_i est une série chronologique de signaux pour $i = 1, 2, \dots, n$, et n est le nombre de points de données, $x_{max} = \max x_i$ et $\bar{x} = \frac{1}{n} \sum_{i=1}^n |x_i|$ est la valeur moyenne absolue.

➤ **L'analyse dans le domaine fréquentiel** : est en mesure, de surmonter les inconvénients de l'analyse dans le domaine temporel.

Bracewell, R.N (2000) [29] a souligné qu'il existe plusieurs méthodes pour transformer les signaux du domaine temporel au domaine fréquentiel. La transformée de Fourier, proposée par Weaver, H. J. (1983) [30], est mise en évidence dans ce domaine. Tan, L et al. (2018) [31] ont souligné l'importance de la transformée de Laplace et de la transformée en Z (dite version discrète de la transformée de Laplace), car ce sont les plus pertinentes. Il existe des algorithmes efficaces pour les calculer ainsi que leurs inverses. Par exemple, l'algorithme de la transformée de Fourier rapide (FFT) développé par Duhamel, P et al. (1990) [32] ; Johnson, S et al. (2006)

[33] ; Welch, P (1967) [34], est une méthode efficace pour calculer la transformée de Fourier d'un signal.

Tableau I.2 : quelques caractéristiques du domaine fréquentielles

N°	Caractéristiques	Description mathématique
01	Fréquence centrale (FC)	$\frac{\int_0^{+\infty} f s(f) df}{\int_0^{+\infty} s(f)}$
02	Fréquence Quadratique Moyenne (MSF)	$\frac{\int_0^{+\infty} f^2 s(f) df}{\int_0^{+\infty} s(f)}$
03	Racine de fréquence Quadratique Moyenne (RMSF)	\sqrt{MSF}
04	Variance de fréquence (VF)	$\frac{\int_0^{+\infty} (f - FC)^2 s(f) df}{\int_0^{+\infty} s(f)}$
05	Racine de Variance de Fréquence (RVF)	\sqrt{VF}

Où $s(f)$ est le spectre de puissance du signal. FC , MSF et $RMSF$ montrent le changement de position des principales fréquences. Le VF et RVF décrivent la convergence de la puissance du spectre.

➤ **L'analyse dans le domaine temps-fréquence**

Malgré son immense succès, la technique FFT a plusieurs défauts, on peut citer le manque de localisation temporelle. En effet, l'analyse de Fourier permet de connaître les différentes fréquences excitées dans un signal, c'est à dire son spectre, mais ne permet pas de savoir à quels instants ces fréquences ont été émises. Cette analyse donne une information globale et non locale, car les fonctions d'analyse utilisées sont des sinusoides qui oscillent indéfiniment sans s'amortir. Cette perte de localité n'est pas un inconvénient pour analyser des signaux stationnaires, mais le devient pour des signaux non stationnaires.

Pour surmonter cet effet, il est possible d'utiliser la transformée de Fourier à court terme (STFT) en utilisant une fenêtre temporelle de largeur fixe, tel que suggéré par Covell, M et al. (1991) [35] ; Tan, L et al. (2013) [36].

Néanmoins il y a un autre problème avec STFT, venant du fait qu'une bonne résolution dans un domaine implique une mauvaise résolution dans l'autre. Cela nous oblige à choisir la largeur de la fenêtre pour qu'il y ait un compromis fin entre la résolution dans les deux domaines. Une autre solution consiste à employer une transformation en ondelettes (domaine temps-fréquence), ce qui nous donne le même effet que d'avoir des résolutions dynamiques en temps et en fréquence. Il existe des transformations en ondelettes continues et discrètes telles que les transformations de Daubechies, I (1988) [37] ; Mallat, S. G (1989) [38], ces dernières étant plus efficaces en termes de calcul.

Certaines nouvelles approches plus sophistiquées ont été développées par la suite. La transformation de Hilbert Huang qui a été proposée par Huang, N. E et al.(1998) [39], une méthode en deux étapes consistant en (1) la décomposition de mode empirique, c'est-à-dire la décomposition du signal en un nombre fini de fonctions de mode intrinsèques, et (2) la transformée de Hilbert des fonctions de mode intrinsèques. Le fait que ces fonctions soient orthogonales d'après Sharpley, R. C et al (2006) [40], implique qu'elles peuvent être comprises comme ayant une signification physique, donc applicables en maintenance prédictive comme le soulignent également Wu, T. Y et al. (2012) [41].

Finalement, la distribution de Wigner Ville qui a été proposé par Cohen, L (1995) [42] a été adaptée par Ville, J (1948) [43] à partir des travaux de Wigner dans le domaine de la mécanique quantique. Il s'agit d'une transformation intégrale quadratique sous la forme d'une transformée de Fourier bidimensionnelle d'une fonction d'autocorrélation temps-fréquence liée à la fois au temps et à la fréquence. Ce n'est pas une méthode basée sur la fenêtre et elle offre la meilleure résolution. Néanmoins, lorsqu'un signal est une composition de deux signaux, effectivement, lors de l'analyse de signaux, il peut y avoir des termes croisés qui introduisent des distorsions et interfèrent avec les résultats. Cela peut rendre l'interprétation et l'analyse plus complexes. Kadambe, S et al. (1992) [44] ont également abordé cette problématique et propose des solutions pour atténuer ces effets indésirables. Sinon l'étude des différences dans les termes croisés peut être utilisée dans des problèmes de maintenance prédictive, comme le suggèrent Wang, C et al. (2008) [45].

- 4. Traitement des valeurs manquantes :** On doit considérer le problème du nettoyage de données, il s'agit d'identifier les données aberrantes, les individus atypiques et de traiter les données manquantes

Les données aberrantes peuvent prendre plusieurs formes :

- Données catégorielles
- Données positives
- Valeurs extrêmes de probabilité très faible

Dans le cas de données issues d'une distribution continue multivariée, on peut repérer les données aberrantes par projection sur un sous espace (ACP, ACP non linéaire) ou éventuellement par classification, on doit alors considérer les données correspondantes comme des données manquantes.

Il est parfois difficile de savoir si une donnée est aberrante ou atypique, en statistique on parle de valeur manquante lorsqu'on n'a pas d'observation pour une variable ,le problème de la gestion des données manquantes est un sujet très vaste, les données manquantes ne peuvent pas être ignorées lors d'une analyse statistique , mais selon leur proportion et leur type ,des solution différentes vont être choisies, on pourra soit retirer les variables présentant des données manquantes ou imputer des valeurs aux données manquantes ou encore développer des méthodes qui permettent de mener les analyses en présence de données manquantes. On distingue deux types de données manquantes :

- ✓ Données manquantes complètement aléatoirement : les données sont complètement aléatoirement si la probabilité qu'une observation soit manquante ne dépend pas des mesures observées ou non observées.

- ✓ Données manquantes non aléatoirement : Ce cas correspond à un mécanisme de non réponse non-ignorable ; ça signifie que :
 - Même en tenant compte de toute l'informations observées, les raisons pour lesquelles des données sont manquantes dépendent des données manquantes elles-mêmes.
 - Pour obtenir des estimations valides, un modèle joint des données complètes sachant le mécanisme de réponse est nécessaire.

Lorsque nous avons des données manquantes, et pour traiter ce problème, la solution est comme suite :

- Obtenir les réponses
- Imputation des données manquantes
- Développement des méthodes adaptées

Les méthodes d'imputation les plus simple sont

- Imputation par la moyenne : la méthode consiste à remplacer les données manquantes par leur moyenne ou leur médiane.
- Imputation par tirage conditionnel : cette méthode améliore l'idée de l'imputation par la moyenne en réalisant de l'imputation par tirage conditionnel, le principe est d'utiliser l'information apportée par les variables renseignées.
- Imputation par analyse factorielle : considérons le cas de données issues de variables quantitatives, l'analyse en composantes factorielles permet de reconstruire des données par projection dans un espace de dimension réduite, cette caractéristique peut être exploitée pour remplacer des données manquantes

5. Ingénierie des caractéristiques : Il n'y a pas de définition standard pour l'ingénierie des caractéristiques. Il fait référence à l'utilisation d'une ou plusieurs des techniques suivantes : sélection de caractéristiques (sélection des caractéristiques les plus importantes sur la base d'un ensemble de critères de qualité), extraction de caractéristiques (création de nouvelles caractéristiques à partir de certaines ou de toutes les caractéristiques d'origine) et la discrétisation. (Transformation d'éléments continus en éléments discrets)

6. Traitement des données déséquilibré : Si notre problème de maintenance prédictive est supervisé de manière à ce que certains types d'échantillons soient extrêmement rares par rapport aux autres (classe minoritaire), alors nous sommes confrontés à un problème d'apprentissage déséquilibré. Il y a deux manières logiques de procéder : (1) équilibrer d'une manière ou d'une autre les données, et (2) compenser en donnant d'une manière ou d'une autre plus d'importance aux échantillons de la classe minoritaire. La première est liée aux techniques d'échantillonnage et la seconde aux techniques de pondération.

I.5 Diagnostic de défaut

Les définitions du terme diagnostic selon la communauté CBM, sont de plus en plus acceptées comme l'a souligné Vachtsevanos, G. J et al. (2006) [46].

- Le diagnostic de défaut concerne la détection, l'isolement et l'identification, d'une condition de défaut imminente, ou naissante dans un système à l'aide d'un raisonnement logique fondé sur un ensemble d'informations provenant d'une inspection, d'un contrôle ou d'un test. Le

terme défaut implique que le système sous observation est encore opérationnel, mais ne peut pas continuer à fonctionner indéfiniment sans intervention de maintenance.

- Le diagnostic de défaillance concerne la détection, l'isolement et l'identification d'un système qui a cessé de fonctionner.

D'après Butler, S (2012) [47], les termes "détection, isolement et identification de défauts" ont généralement les significations suivantes :

Détection : Elle consiste à classer la défaillance du système dans l'une des catégories opérationnelles en fonction des conditions observées ; normale ou anormale.

Localisation : La localisation ou l'isolement d'une défaillance, consiste à tracer les symptômes pour retrouver tous les éléments défaillants. Ce problème est difficile à résoudre. En fait, les défaillances causées par des défauts peuvent être déterminées. En revanche, le problème inverse est plus difficile à résoudre, Car la panne peut être causée par un ou plusieurs défauts.

Identification : Il s'agit d'identifier la source d'une situation anormale. Ces causes peuvent être internes ou externes au système.

Une fois le type de défaut déterminé, et selon les connaissances disponibles sur le système, selon Ribot, P. (2009) [48], les effets d'un défaut sur les équipements d'un système peuvent parfois se propager et avoir des conséquences qui peuvent être prédites.

Pour comprendre la portée du diagnostic et du pronostic considéré, la figure I.4 illustre la chronologie de la progression des défaillances pour les composants typiques du système selon G. Vachtsevanos, G et al. (2006) [46]. Au début de la durée de vie du composant, celui-ci est considéré en bon état, Après une période de fonctionnement, une condition de défaillance initiale se produit. Par conséquent, la gravité de la défaillance augmente jusqu'à ce que le composant finisse par tomber en panne. Si le système est autorisé à continuer à fonctionner, il est susceptible d'endommager davantage d'autres composants mineurs ou le système. Selon Vachtsevanos, G et al. (2006) [46], le domaine du diagnostic se produit généralement au moment de la défaillance du composant, ou sur l'intervalle entre la défaillance du composant et la défaillance éventuelle de l'ensemble du système. Cependant, si des conditions de défaillance peuvent être détectées au stade de bourgeonnement, Les actions de maintenance peuvent être retardées jusqu'à ce que le défaut progresse vers un état plus avancé (critique), mais toujours dans la défaillance. L'intervalle de temps entre la détection d'une condition de défaut naissant et l'apparition du défaut, c'est le pronostic, L'intervalle suffisant entre la détection de défaillance naissant et défaillance du système, d'après Butler, S (2006) [47], cette compréhension de la progression des défaillances permet de prévoir et de planifier à l'avance une variété de fonctions opérationnelles et des travaux de maintenance corrective. Le positionnement relatif des processus "détection", "diagnostic" et "pronostic" est représenté sur la figure I.5-a.

- La détection vise à identifier le mode de fonctionnement, l'état du système.
- Les diagnostics permettent d'isoler et d'identifier les défauts lorsqu'ils surviennent des composants qui ont cessé de fonctionner (propagation arrière : des effets vers les causes),
- Le pronostic vise à prédire l'état futur du système (propagation avant : des causes vers les effets).

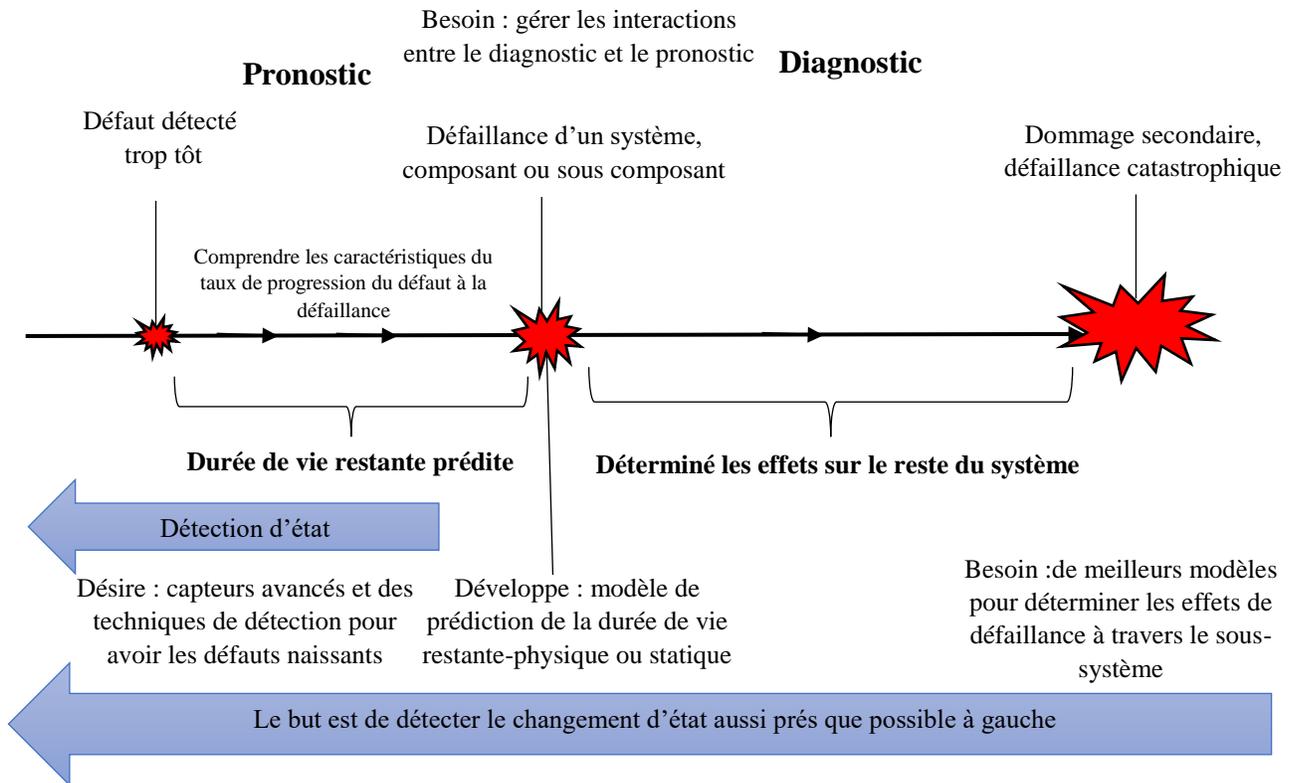


Figure I.4: Chronologie de progression de défaillance d'après Huang, N. E et al. (1998) [39]

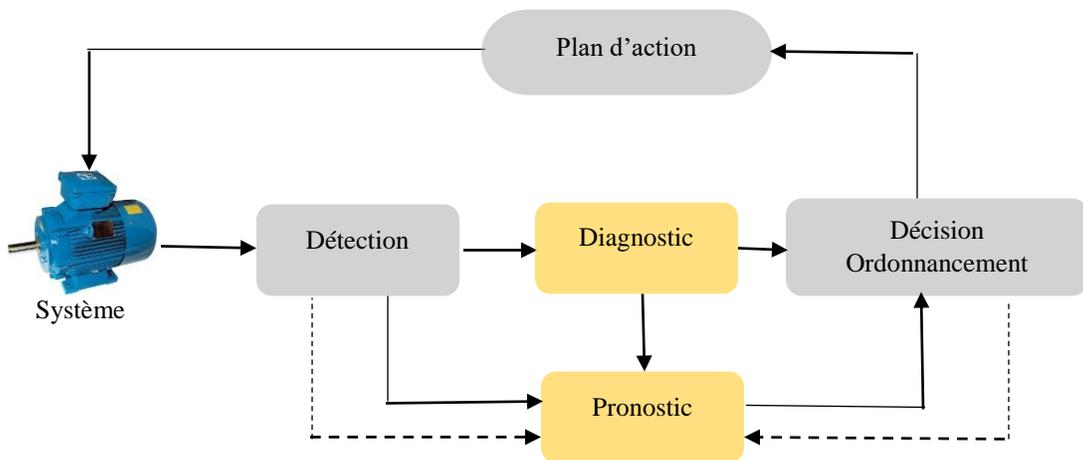


Figure I.5-a : Détection, diagnostic, pronostic et ordonnancement dans les systèmes de maintenance intelligente

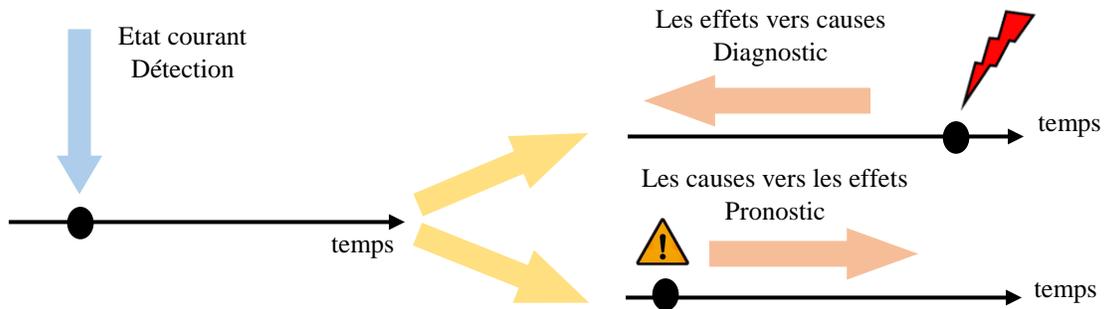


Figure I.5-b : Détection, diagnostic et pronostic les aspects phénoménologiques

I.5.1 Classification des méthodes de diagnostic :

Le développement de la technologie dans le domaine industriel a produit une variété des approches et méthodes de diagnostic, dépendant du type de système (thermique, mécanique, équipements électroniques, etc.), selon Mous, M. D (2006) [49]. Plusieurs classifications ont été proposées dans la littérature selon le type de traitement ou analyse de données. Dans cette partie, nous présenterons une classification récente inspirée des travaux de Patrick, LYONNET et al. (2012) [50] ; Zwingelstein, G (1995) [51] ; Mous, M. D. (2006) [49]. L'objectif de classification est de déterminer la méthode la plus appropriée pour la résolution de problèmes de détection et de diagnostic, ainsi que de positionner la classe de méthodes de diagnostic qui nous intéresse parmi les différentes méthodes de la littérature. Nous proposons une classification des méthodes de diagnostic en deux grandes familles

1. Méthodes sans modèle mathématique qui ne nécessitent pas une connaissance accrue du système physique, mais l'utilisation de connaissances superficielles.
2. Les méthodes basées sur des modèles qui nécessitent la connaissance approfondie du système physique.

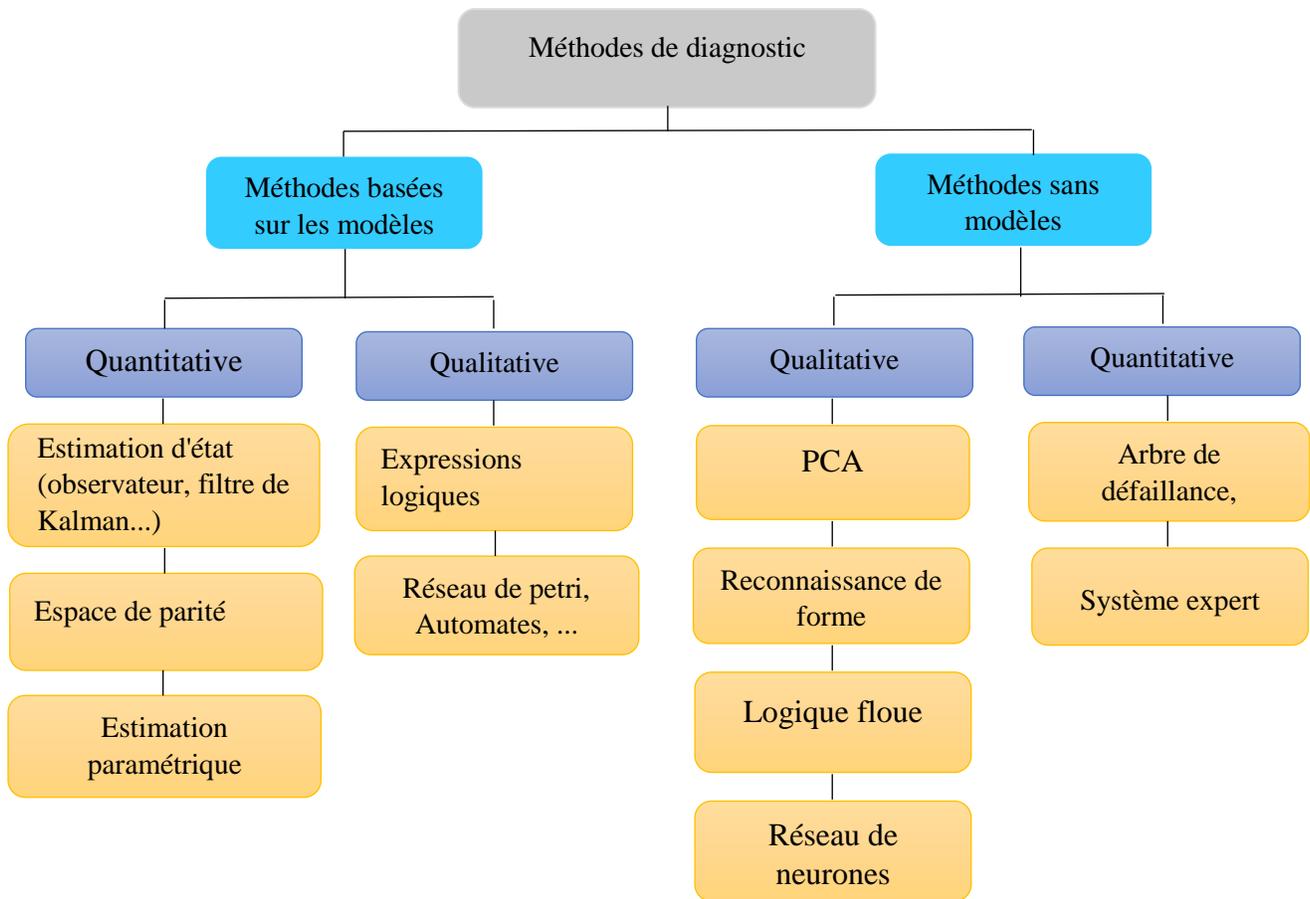


Figure I.6 : Classification générale des méthodes de diagnostic.

I.5.1.1 Les méthodes basées sur les modèles :

L'utilisation des modèles pour le diagnostic remonte au début des années 1970. Comme de nombreuses études ont été proposées par Willsky, A. S. (1976) [52] ; Chow, E. Y. E. Y et al. (1984) [53] ; Basseville, M. (1988) [54] ; Patton, R. J. (1997) [55] ; Evsukoff, A et al. (1997) [56] ; Isermann, R et al. (1997) [57] et Balle, P et al. (1998) [58]. Une étude complète des méthodes basées sur des modèles peut être trouvée dans l'étude de Frank, P. M. (1996) [59] ou dans des ouvrages comme le livre de Patan, K. (2008) [60], et le livre de Chiang, L. H et al. (2000) [61]. La structure générale de la plupart de ces méthodes repose sur l'idée de redondance analytique selon Chow, E. Y. E. Y. (1984) [53]. Le principe des méthodes basées sur les modèles est d'identifier l'écart entre le système réel et son modèle. Ces méthodes reposent sur des schémas comportementaux explicites du système diagnostiqué.

Les méthodes de diagnostic basées sur des modèles reposent sur une connaissance physique approfondie du système à diagnostiquer. Un système est représenté sous la forme d'un ou plusieurs modèles qui décrivent la structure du système et son comportement nominal ou son comportement en présence de défauts. Les méthodes de diagnostic sont basées sur la comparaison du comportement réel observé sur le système physique avec le comportement prédit à l'aide du modèle. La détection des incohérences peut tirer des conclusions sur les défaillances qui se sont produites

dans le système. Le modèle de dysfonctionnement (modèles de défauts) peut localiser les défauts et éventuellement les identifier.

LA figure I.7 illustre les concepts de base d'une approche de diagnostic typique basée sur un modèle.

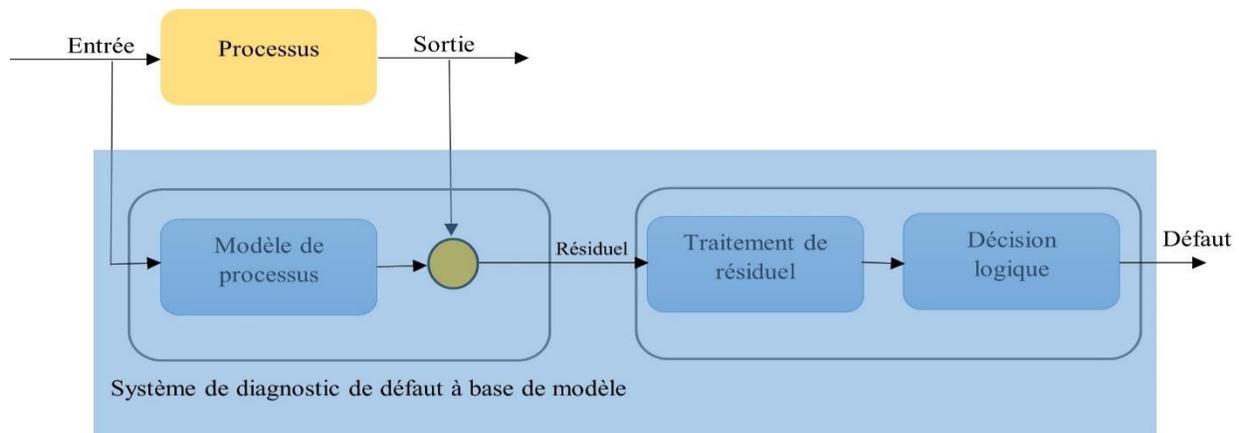


Figure I.7 : Le principe de l'approche à Base de Modèles

Deux approches principales peuvent être distinguées dans les méthodes de diagnostic basées sur les modèles.

- L'approche FDI "Fault Detection and Isolation" de la communauté de l'automatisation utilise des modèles quantitatifs pour décrire les modèles de comportement des systèmes. Les équations différentielles algébriques peuvent exprimer un comportement continu de systèmes avec une certaine précision numérique.
- La méthode DX est basée sur la théorie logique du diagnostic, trouve son origine dans la communauté de l'intelligence artificielle. Ils utilisent des modèles qualitatifs qui peuvent représenter efficacement les interactions entre composants ou systèmes.

I.5.1.1.1 Approches FDI :

L'approche FDI repose sur une compréhension approfondie du fonctionnement du système. Elle utilise des modèles quantitatifs de référence qui peuvent être dérivés de lois physiques fondamentales selon Frank, P. M. (1996) [62]. Ces modèles analytiques numériques décrivent le comportement normal du système à travers un ensemble d'équations différentielles d'après Gertler, J. (1998) [63]. Le diagnostic consiste à générer des indicateurs indiquant s'il existe un dysfonctionnement du système, ces indicateurs sont appelés résidus. Pour obtenir ces résidus, le modèle de référence se restreint aux entrées et sorties observables (mesurées par des capteurs sur le système réel). Les valeurs non nulles des résidus sont interprétées comme des modifications anormales (ou des écarts inacceptables) des propriétés ou des paramètres caractéristiques du système modélisé.

D'après Frank, P. M. (1987) [64] ; Staroswiecki, M et al. (1990) [65], les équations différentielles utilisées pour décrire le comportement d'un système peuvent être linéaires ou non linéaires, en temps discret ou en temps continu. Selon Narasimhan, S et al. (2007) [66] ; Coquempot, V et al. (2005) [67], le système de modélisation peut également être un système hybride. En raison des erreurs de

modélisation et du bruit, les résidus ne sont jamais vraiment nuls même s'il n'y a pas de défauts dans le système. La décision de détection d'un défaut nécessite d'évaluer les expressions des résidus obtenues en utilisant des mesures du système pour déterminer si les différences qui leur sont associées sont significatives. Pour l'évaluation des résidus, la reconnaissance des formes, la logique floue, selon Basseville, M et al. (1993) [68], les techniques de seuillage sont utilisées, ainsi que les techniques de la théorie statistique de la décision. La figure I.8 illustre le principe de l'approche FDI basée sur le calcul des résidus.

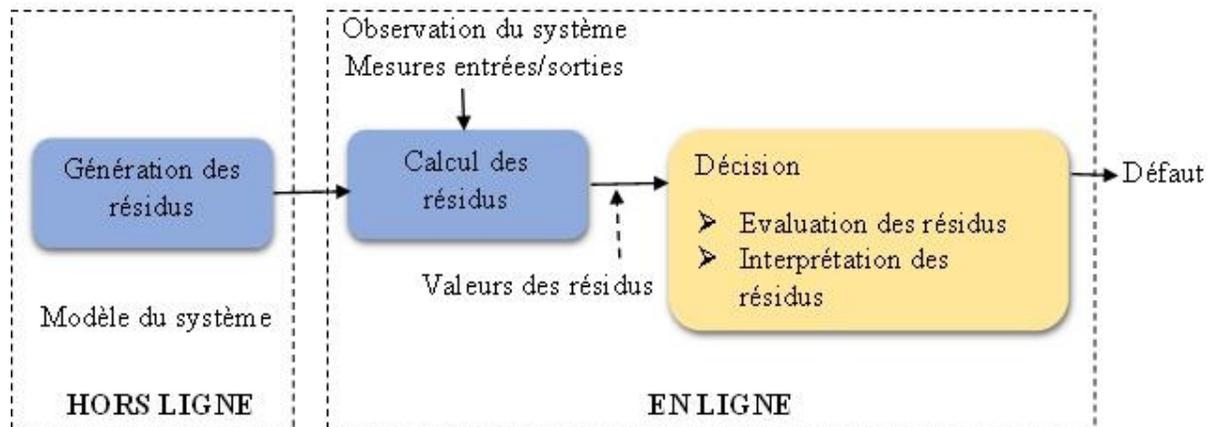


Figure I.8 : Principe de l'approche FDI à base de résidus

Pour obtenir des expressions analytiques des résidus, plusieurs techniques peuvent être utilisées : relations de parité, observateurs, filtres de détection. D'autres méthodes s'intéressent davantage à l'estimation des paramètres de détection de défauts. Certains articles de revue, comme celui de Isermann, R. (1997) [69], décrivent ces différentes méthodes utilisées dans la méthode FDI

a) Espace de parité :

Les méthodes d'espace de parité reposent sur la génération de résidus selon Staroswiecki, M et al. (2001) [70]. Le comportement normal (nominal) du système est caractérisé par des valeurs nulles pour les résidus. Ces mesures sont alors cohérentes avec les propriétés de l'espace de parité. La conception de l'espace de parité est basée sur le développement d'expressions analytiques, la méthode repose sur la détermination de la redondance analytique entre les entrées et les sorties du système, comme le mentionnent Patton, R. J et al. (1991) [71] ; Maquin, D et al. (1997) [72]. Les relations de parité utilisent la redondance directe à travers des relations algébriques qui relient différentes grandeurs du modèle de système. Les relations de redondance analytique sont obtenues en éliminant les variables d'état non observables du modèle sur l'espace de parité. Nous obtenons le modèle entrée/sortie du système, où le vecteur d'état x est inconnu et n'apparaît pas selon Gertler, J. (1998) [63]. Les équations ainsi obtenues ne font intervenir que des variables d'entrée/sortie mesurables, donc connues. Les relations de redondance analysées sont ensuite directement utilisées pour construire des indicateurs de défaut pour le diagnostic. Staroswiecki, M et al. (1991) [73] ; Desai, M et al. (1981) [74], sont des références reconnues pour les applications industrielles de l'espace de parité, pour la détection de relations statiques. Schwarte, A et al. (2003) [75] ont développé des applications industrielles sur les systèmes dynamiques dans le domaine automobile.

Les techniques d'observateurs peuvent compléter la méthode de l'espace de parité, lorsque certaines sorties ne peuvent pas être directement mesurées.

b) Estimation d'état : observateurs :

Les observateurs sont des systèmes dynamiques qui peuvent être utilisés pour la détection et l'isolation des défauts dans les systèmes linéaires ou non linéaires. Les méthodes d'observateurs impliquent la reconstruction des sorties non mesurables du système à partir d'un modèle analytique et d'un ensemble d'observations partielles (entrées/sorties) du système par l'estimation de l'état du système. Les observateurs sont utilisés pour estimer la valeur des variables d'état. A partir de ces valeurs reconstruites et du modèle dynamique, des valeurs de sortie sont calculées. On obtient un vecteur de résidus, représentant la différence entre les valeurs de sortie mesurées et estimées. Dans le cas des systèmes linéaires, Staroswiecki, M et al. (1991) [73] ; Magni, J. F et al. (1994) [76] sont des références qui ont travaillé sur la détection de défaut à l'aide d'observateurs. Frank, P. M. (1987) [64] ; Hammouri, H et al. (1999) [77] ont proposé une solution pour l'observation des systèmes non linéaires.

c) Estimation paramétrique :

L'estimation des paramètres permet d'analyser l'effet des défauts sur les paramètres structurels du modèle de système. C'est donc une façon naturelle de détecter et isoler les défauts multiplicatifs des paramètres du modèle dynamique du système. L'estimation des paramètres utilise des méthodes analytiques pour calculer les valeurs des paramètres du modèle dont la structure est connue. Ces paramètres de structure sont généralement constants.

Pour estimer ces paramètres, des techniques de filtrage sont utilisées. Par exemple, les filtres de Kalman estiment de manière récursive les paramètres structurels des modèles linéaires en tant que représentations d'état à partir des mesures bruitées.

Pour les systèmes non linéaires mais localement linéarisables, des filtres de Kalman étendus peuvent être utilisés, et pour les systèmes non linéaires et non linéarisables, des filtres à particules (techniques analogiques et statistiques) sont appliqués.

La détection des défauts est réalisée en comparant les paramètres estimés avec les paramètres nominaux caractérisant le comportement normal du système. L'idée principale de cette méthode est de minimiser la distance entre la valeur réelle et la valeur estimée du paramètre. Ainsi, les méthodes d'estimation sont caractérisées par la définition de distances (maximum de vraisemblance, moindres carrés, etc.). Dans l'article de Gertler, J (1995) [78], la méthode d'estimation des paramètres est comparée à la méthode de l'espace de parité. L'équivalence entre les techniques utilisant l'espace de parité, Patton, R. J et al. (1991) [71] ont établi l'importance de l'observateur et de l'estimation des paramètres.

d) Localisation et identification des défauts :

Pour localiser les défauts, différents types de résidus sont générés : des résidus structurés (propriétés booléennes) et des résidus directionnels (propriétés géométriques). Les résidus structurés sont conçus de sorte que chaque résidu soit sensible à un sous-ensemble de défauts connus et insensible aux autres. Lorsqu'une défaillance connue se produit, certains résidus auront une valeur nulle ou proche de zéro, tandis que d'autres ne seront pas nuls. L'ensemble des différentes valeurs

résiduelles représente la signature du défaut. Toutes les signatures pouvant apparaître pour divers défauts connus dans un système est appelé une matrice de signature.

Le deuxième type de résidu qui permet la localisation des défauts est le résidu directionnel. Ils sont représentés sous forme de vecteurs de résidus, selon L'orientation spécifique de l'espace résiduel, en fonction du type de défaut présent, la connaissance de diverses défaillances ou modes de défaillance du système représenté par le modèle permet l'identification des défauts comme l'a souligné Ribot, P. (2009) [48].

I.5.1.1.2 Approche DX :

La méthode DX est une méthode qualitative basée sur la cohérence issue du domaine de l'intelligence artificielle. Les techniques de diagnostic de cohérence consistent à comparer le comportement réel du système observé avec son comportement attendu, tel qu'il peut être prédit par un modèle de bon comportement. Elle est basée sur la théorie de diagnostic logique introduite par Reiter, R. (1987) [79] dans sa publication. Le but du diagnostic logique est d'identifier les dispositifs (ou composants) du système dont le fonctionnement anormal peut expliquer l'incohérence détectée entre le comportement prédit, et les observations du système selon De Kleer, J et al. (1987) [80]. Il s'agit d'un diagnostic adductif, où le but est de trouver toutes les raisons qui expliquent les observations. Le principe de cette approche logique basée sur la cohérence est illustré à la figure I.9.

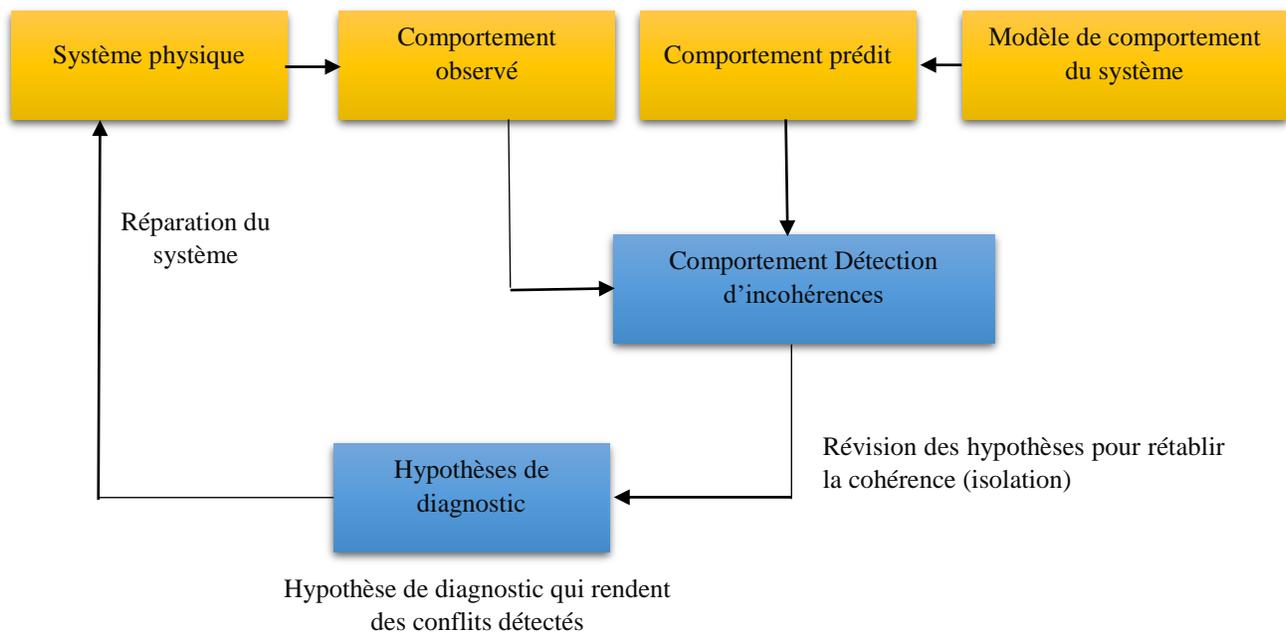


Figure I.9 : Principe de l'approche logique basée sur la cohérence

Le principal avantage d'une approche de diagnostic basée sur des modèles est la capacité de détecter des défauts non-anticipés car les modèles utilisés sont toujours basés sur la physique de défaillance. Cela contraste avec les approches basées sur les données, qui nécessitent en fait des exemples historiques pour chaque condition de défaut détectée. Cependant, dans des situations pratiques, il est impossible d'appliquer des méthodes de diagnostic basées sur des modèles mathématiques précis car le processus est compliqué.

I.5.1.2 Les méthodes sans modèles :

Pour les applications industrielles, la conception d'un modèle mathématique est difficile ou impossible à obtenir, en raison des nombreuses reconfigurations impliquées dans le processus de production ou de la complexité des phénomènes impliqués. Dans ce cas, nous utilisons des méthodes qui ne nécessitent pas une connaissance détaillée du processus.

Deux classes de ce type d'approches sont possibles selon Methnani, S. (2012) [81] :

- Méthodes quantitatives appelées aussi méthodes basées sur la connaissance du système
- Méthodes qualitatives basées sur le traitement des données.

I.5.1.2.1 Méthodes qualitatives :

Les méthodes qualitatives consistent à exploiter une base de connaissances symbolique et nécessitent l'existence d'un large éventail de données historiques correspondant aux différents modes d'installation. On peut citer :

a) Analyse en composantes principales (ACP) : l'analyse en composantes principales (ACP) est une technique statistique multivariée, capable de comprimer les données et de réduire leur taille. Il peut être considéré comme une technique de projection orthogonale linéaire qui projette les cas représentés dans un espace multidimensionnel de dimension n (n est le nombre de variables observées) dans un sous-espace de dimension $< n$, maximisant la variance de la projections (ou minimisation de l'erreur quadratique). Selon les travaux de Jolliffe, I. T. [82] (1990) ; Pessel, N et al. (2007) [83], la méthode basée sur la théorie de diagnostic logique a été appliquée avec succès dans des études de diagnostic.

b) Reconnaissance de formes : le but d'une méthode de reconnaissance de formes est la classification automatique d'un objet suivant sa ressemblance par rapport à un objet de référence. Dans un problème de diagnostic, une classe est formée par les ensembles d'observations caractérisant une situation ou une opération de processus : par exemple, la classe C1 peut être liée au fonctionnement normal du procédé, la classe C2 pour le mode dégradé et la classe C3 pour le mode de défaillance. Le diagnostic consiste à associer de nouvelles observations à une classe. Le problème de diagnostic équivaut à la recherche de frontières entre les classes qui minimisent les erreurs de classification. Le calcul de la distance (distance euclidienne) entre les classes peut être choisi comme critère de décision pour affecter une forme à une classe et déterminer quelle confiance est affectée par la décision.

c) Analyse spectrale : Dans certaines conditions normales, certaines mesures ont un spectre de fréquence typique ; tout écart des caractéristiques fréquentielles d'un signal est lié à une anomalie. Cette méthode est utile pour l'analyse de signaux présentant des oscillations sur de longues périodes (débits, pressions...), parmi les travaux qui utilisent l'analyse spectrale on cite les travaux de Basseville, M. [84] (1988) ; M. Basseville (1997) [85].

d) Réseaux de neurones artificiels (RNA) : RNA est un système informatique composé d'un certain nombre de nœuds élémentaires interconnectés qui traitent dynamiquement les informations qui arrivent des signaux externes. En général, l'utilisation de l'RNA se fait en deux phases.

1. La synthèse du réseau comprend plusieurs étapes : le choix du type de réseau, le type de neurones, le nombre de couches et la méthode d'apprentissage. L'apprentissage permet alors, à l'optimisation d'un critère, elle consiste en la recherche de paramètres de poids et peut se faire en deux façons : supervisé (le réseau utilise les données d'entrée et de sortie du système à modéliser) et non supervisé (seules les données d'entrée du système sont fournies)
2. Lorsque les résultats d'apprentissage obtenus par l'RNA sont satisfaisants, le réseau peut être utilisé pour la généralisation. Il s'agit de la deuxième phase où de nouveaux exemples qui n'ont pas été utilisés pour l'apprentissage sont présentés à RNA pour évaluer également sa capacité à prédire le comportement du système modélisé, leur faible sensibilité au bruit de mesure, leur capacité de résoudre des problèmes non linéaires et multivariés, stocker de manière compacte des connaissances, apprentissage en ligne et en temps réel, sont des propriétés qui rendent l'utilisation des RNA commun et reconnaissable, leur utilisation peut alors se faire à trois niveaux :
 - En tant que le modèle d'un système à observer l'état normal et générer un résidu d'erreur entre les observations et les prédictions.
 - Comme système d'évaluation des résidus pour le diagnostic.
 - Comme un système de détection en une étape (comme classificateur), ou en deux étapes (pour la génération de résidus et le diagnostic).

Plusieurs travaux ont été réalisés dans la littérature pour développer des méthodes de diagnostic des processus, basées sur les différentes informations disponibles pour décrire le comportement des systèmes. Assaf, E. M et al. (1996) [86] ; Barton, J. A et al. (1984) [87] ont présenté les résultats donnés par le modèle neuronal pour les éventuelles situations d'emballement thermique de processus fortement exothermiques. L'objectif est d'établir un algorithme d'alarme fiable pour la détection et la prévention précoce de cette situation. Ainsi, les résultats montrent que le modèle neuronal NARMAX est représentatif du comportement dynamique du processus chimique non linéaire. Gustin, J. L. (1996) [88] dans son article a classé les causes d'emballement thermique où la perte de la capacité de refroidissement est la cause la plus importante de l'apparition de l'emballement thermique.

Le diagnostic de défaut avec RNA à base des fonctions radiale et une couche cachée peut être appliqué pour développer une représentation non linéaire du réacteur de polymérisation basée sur une structure répétitive selon Yu, D. L et al. (1999) [89]. Dans la première étude de cas d'un réacteur discontinu, un défaut de capteur de température et un encrassement de la chemise ont été étudiés et classés avec succès à l'aide du modèle de processus neuronaux et du classificateur de la fonction de base radiale neurale (RBF). La capacité des RNA à extraire des informations directement à partir des données disponibles sur le processus peut-être la raison la plus importante d'appliquer cette approche dans l'industrie. Ils sont utiles dans des problèmes tels que l'analyse de données de capteurs, ce qui dépasse le cadre des techniques classiques du système expert algorithmique.

Dans les processus industriels, les RNA ont été appliqués pour la détection et le diagnostic des défauts. Par exemple, Cubillos, F et al. (1998) [90] ont décrit un système hybride adaptatif basé sur des connaissances a priori et des réseaux de neurones pour modéliser les stratégies de contrôle de processus et les paramètres incertains dans un réacteur continu fortement non linéaire. Power, Y et al. (2004) [91] ont décrit un cadre de supervision en deux étapes de diagnostic de défaut utilisant

des réseaux de neurones. Sur la base de ce cadre, un système de détection des pannes a été mis en place pour identifier l'emplacement exact des défauts et les diagnostiquer. Les RNA ont fait des développements rapides dans le diagnostic des défauts des processus. La détection des défauts d'un système par des techniques d'intelligence artificielle a été étudiée par Venkatasubramanian, V et al. (1990) [92] ; Hoskins, J et al. (1988) [93] ; Watanabe, K et al. (1994) [94].

Comme indiqué précédemment, l'RNA est utilisé pour le diagnostic des défauts, leur faible sensibilité au bruit de mesure, leur capacité à résoudre des problèmes non linéaires et multivariés, à stocker de manière compacte les connaissances, à apprendre en ligne et en temps réel, sont en effet de nombreuses propriétés qui les rendent attractifs pour cet usage.

e) Logique floue (LF) : est une technologie d'intelligence artificielle. Elle a été formalisée par Lotfi Zadeh en 1965 et est utilisée dans l'automatisation (freinage ABS), la robotique (reconnaissance de forme), la gestion du trafic routier (feux rouges), le contrôle du trafic aérien, l'environnement (météorologie, climatologie, sismologie, analyse du cycle de vie), la médecine (aide au diagnostic), assurance (sélection et prévention des risques), etc.

L'idée de l'approche floue est de construire un dispositif, appelé système d'inférence floue capable d'imiter les décisions d'un opérateur humain à partir de règles verbales traduisant sa connaissance sur le processus.

Trouver une relation mathématique entre une panne et ses symptômes est souvent difficile. Cependant, sur la base de leur expérience, les opérateurs humains sont en mesure de déterminer le composant défectueux à l'origine des symptômes observés. Ce type de connaissances peut être exprimée à l'aide de règles de la forme : SI condition ALORS conclusion, où la partie condition comprend les symptômes observés et la partie conclusion comprend le composant défectueux. Ainsi le problème de diagnostic est considéré comme un problème de classification. Le vecteur des symptômes du classifieur, élaboré à partir des valeurs mesurées sur le système, peut être vu comme une forme, il s'agit de se ranger parmi toutes les formes correspondant à un fonctionnement normal ou non.

I.5.1.2.2 Méthodes quantitatives :

Des méthodes quantitatives ou une base de connaissances sont mises en œuvre lorsque la majorité des mesures ne sont pas disponibles et lorsque la construction du modèle est difficile. Elles peuvent être utilisées pour identifier les causes des défaillances d'un procédé industriel. Il s'agit d'analyses fonctionnelles et structurelles basées sur l'expérience et les connaissances de l'opérateur. Parmi les méthodes les plus connues nous citons :

a) Arbre de défaillance : apparu pour la première fois par Villemeur, A (1988) [95], est l'un des outils majeurs d'analyse des risques technologiques.

L'objectif de cette approche est d'identifier les différentes combinaisons possibles d'événements qui provoquent la création d'un événement indésirable unique. La représentation graphique est composée d'une arborescence permettant un traitement à la fois qualitatif et quantitatif. L'arbre de défaillance est constitué de plusieurs couches dont la racine correspond à l'événement indésirable. Les niveaux sont ordonnés successivement de telle sorte que chaque événement est généré à partir des événements de niveau inférieur via des opérateurs logiques (ET, OU). La décomposition s'arrête

aux niveaux des événements élémentaires, caractérisés par le fait qu'ils sont indépendants et non décomposables en éléments plus simples.

b) AMDEC : La méthode AMDEC est un outil de qualité pour l'analyse préventive afin d'identifier et de traiter les causes potentielles de défauts et de défaillances avant qu'elles ne surviennent. Du fait de la mutualisation des informations et des données, la méthode AMDEC est une méthode de travail rigoureuse et très efficace. Il se déroule en petits groupes et permet à chaque participant d'apporter sa propre expérience et ses connaissances. Il existe 3 types d'AMDEC. En effet, l'approche AMDEC peut être mise en œuvre au niveau du produit, du processus ou des moyens.

- L'approche AMDEC produit est mise en œuvre pour vérifier que le produit développé répond aux exigences du client.
- L'AMDEC processus est conçue pour vérifier la fiabilité du processus de fabrication.
- L'AMDEC moyenne permet de vérifier la fiabilité des équipements.

Le principe de l'approche AMDEC est d'identifier toutes les causes potentielles pour chaque mode de défaillance. Ensuite, il faut évaluer la criticité du mode de défaillance, la criticité est obtenue grâce à une triple cotation :

1. Gravité (G) ou sévérité de l'effet du défaut ou de la défaillance,
2. Occurrence (O) ou fréquence d'apparition de la cause,
3. La Détection (D) : la probabilité que la cause ne soit pas détectée.

Par conséquent, l'indice de criticité (C) est obtenu en multipliant les trois notes :

$$\text{Criticité} = \text{Gravité} \times \text{Occurrence} \times \text{Détection}$$

Plus la criticité est élevée, plus la défaillance considérée est importante. Lorsque la valeur critique dépasse l'ensemble des limites prédéfinies par le groupe, ce dernier recherche les améliorations possibles pour la ramener à des niveaux acceptables par des ajustements sur la gravité, occurrence et la détection.

c) Système expert (le raisonnement par cas) : Le raisonnement par cas modélise l'expertise et les capacités de raisonnement de spécialistes qualifiés dans le domaine émergent. Ce raisonnement est qualifié pour résoudre un problème en s'appuyant sur des expériences passées. Les connaissances sont stockées de manière appropriée. Un cas est une connaissance contextuelle, représentant une expérience qui peut être utilisée pour atteindre les objectifs du moteur de raisonnement. Ainsi, un cas peut être vu comme une situation avérée dans le passé, associée au résultat d'une action pertinente. Le raisonnement à partir des cas est un raisonnement par analogie. Les attributs d'une situation servent d'indice dans le cas de la bibliothèque pour en tirer le meilleur, selon certains critères de similarité, et ainsi déterminer la solution.

Systèmes experts basés sur l'utilisation :

- Une base de connaissances qui contient l'expertise du spécialiste décrite sous forme de règles dont la structure est la suivante : SI <condition> ALORS <conclusion>.
- Une base factuelle qui contient les informations de base nécessaires pour établir un diagnostic.
- Un moteur d'inférence qui imite le processus de raisonnement de l'expert.

La difficulté ici est de bien définir les cas, c'est-à-dire d'identifier ceux utiles et nécessaires à la description d'une situation. Leur détermination pour les systèmes dynamiques est loin d'être évidente selon Farreny, H. (1989) [96].

I.5.2 Critère de Choix de la méthode de diagnostic

Le choix de la méthode de diagnostic appropriée pour un système/sous-système dépend de deux critères principaux, la disponibilité des données et la complexité. Si le système n'est pas complexe et peut être modélisé, une approche "basée sur un modèle" peut être utilisée, et le diagnostic devient une quantification interne. Sinon, si le système devient très difficile à modéliser, seules les données de l'historique de fonctionnement sont valables, et la modélisation mathématique de la relation (cause - effet) est la seule solution valable, et le type de diagnostic devient qualitatif externe. Le système d'évaluation de la santé mettra également l'accent sur la recherche de Forme (études qualitatives) et des signes importants entraînés par des mesures de capteurs (mesures effectuées en dehors du système (externe)) pour évaluer la santé du système.

Et lorsque la majorité des mesures ne sont pas disponibles et lorsque la construction du modèle est difficile, les méthodes quantitatives sans modèle sont les plus favorisées

I.6 Pronostic de défaut :

Le mot pronostic vient du mot grec prognostikos (connaissance préalable). Il combine pro (avant) et gnosis (un savoir). Hippocrate a utilisé le mot pronostic, tout comme nous le faisons aujourd'hui, pour signifier une prédiction de l'évolution d'une maladie. Dans le domaine de la gestion de la santé des systèmes d'ingénierie, le pronostic est considéré comme une science et souvent appelé pronostic.

Dans le marché hautement concurrentiel d'aujourd'hui, les industries s'efforcent de minimiser leurs coûts d'investissement et d'exploitation en essayant d'utiliser le cycle de vie complet de leurs équipements ou composants sans sacrifier la sécurité humaine, de production ou environnementale. La maintenance conditionnelle (maintenance prédictive) est très utile pour prévoir les pannes d'équipement et éviter les activités de maintenance inutiles.

Les pronostics sont la capacité d'évaluer la santé actuelle d'une pièce et de prédire à l'avenir la santé d'une pièce pour un horizon de temps fixe ou de prédire le temps jusqu'à la défaillance.

Être capable d'effectuer un pronostic fiable est la clé du PHM/CBM. Les pronostics sont essentiels au système, sous-système ou au composant pour améliorer la sécurité, planifier les missions, planifier les coûts de maintenance et les temps d'arrêt.

On observe souvent dans la pratique que la vie d'un équipement de production peut être divisée en deux étapes. La première étape est appelée étape de fonctionnement normal où aucun écart significatif par rapport à l'état de fonctionnement normal n'est observé. La deuxième étape est appelée le délai de période de défaillance, car un défaut peut être initié et évoluer progressivement vers une véritable panne, c'est-à-dire que l'équipement est dans une phase défectueuse mais fonctionne encore pendant cette phase. Avec l'aide de la surveillance de l'état, les défauts cachés déjà présents dans l'équipement peuvent être détectés, mais à des fins de planification de la maintenance, la prédiction du point d'initiation de la deuxième étape et, plus important encore, la durée de vie résiduelle par la suite est importante selon Wang, W. (2007) [97].

Un système de pronostic fiable est très utile pour prédire la tendance de propagation des défauts dans un équipement ou un composant et pour fournir une alarme avant qu'un défaut n'atteigne des niveaux critiques. Un système de pronostic en ligne peut également être utilisé pour améliorer la fiabilité du diagnostic de défaillance de l'équipement en vérifiant de manière adaptative les résultats du diagnostic et en modifiant la base de connaissances (règles). Ces aspects ont été abordés dans les travaux de recherche de Pourahmadi, M. (2001) [98] ; Wang, W et al.(2001) [99] ; Wang, W. (2007) [100].

La maintenance prédictive est devenue une stratégie efficace pour les industries modernes, pour analyser les tendances des équipements ou performance des composants par rapport aux limites techniques connues dans le but de détecter, d'analyser et de corriger les problèmes avant qu'une défaillance ne se produise. Des pronostics plus avancés sont axés sur la surveillance et l'évaluation de la dégradation des performances.

Afin que les pannes puissent être prédites et pour atteindre l'objectif du pronostic, trois étapes cruciales sont nécessaires. Premièrement, le défaut ou l'anomalie doit pouvoir être détecté à un stade précoce. Deuxièmement, les performances de l'équipement ou des composants doivent être évaluées de manière rigoureuse et suivies en permanence. Enfin, la durée de vie utile restante (ou durée de vie utile résiduelle) et le mode de défaillance possible du système ou du composant doivent être prédits efficacement. L'estimation de la durée de vie utile restante est la plus importante dans ces trois étapes, car la durée de vie utile restante sert directement les variables de décision des pronostics.

Cependant, il existe des défis pour prédire efficacement la durée de vie utile restante des composants. L'un des défis de la prédiction de la durée de vie est de trouver comment mettre en place un indicateur de dégradation approprié basé sur un signal.

Généralement, les caractéristiques temporelles, telles que la moyenne, la valeur efficace, le kurtosis ou le facteur de crête des signaux, et les caractéristiques de fréquence telles que la moyenne des amplitudes de la fréquence défectueuse et de ses harmoniques dans le temps, comme des indices de dégradation, cependant, ces indices ont soit une faible sensibilité aux défaut naissant ou ne correspondent pas, sous des tests très accélérés. Des travaux de recherche ont montré que chaque caractéristique n'est efficace que pour un certain défaut. Une bonne méthode d'évaluation des performances devrait profiter des informations mutuelles provenant de plusieurs caractéristiques pour l'évaluation de la dégradation du système.

I.6.1 Approches de pronostic :

Lebold, M et al. (2001) [101] ont proposé la première classification des méthodes pronostics. Dans leur article, les auteurs proposent trois niveaux de classification pyramidale (voir figure I.10) : les méthodes basées sur l'expérience, les méthodes basées sur les données et les méthodes basées sur les modèles. Afin de différencier ces trois types de méthodes, les critères suivants doivent être pris en compte : Le coût et la complexité de la mise en œuvre, la précision des résultats obtenus et l'applicabilité de la méthode. D'autres travaux ont proposé de nouveaux critères de classification des méthodes qui peuvent être en fonction du résultat attendu ou de la nature des informations utilisées

pour le pronostic. Par conséquent, les auteurs McKee, K et al. (2014) [18] ont défini une classification basée sur le résultat attendu du pronostic. Ils distinguent les travaux qui cherchent à obtenir une RUL et ceux qui cherchent à obtenir une probabilité de défaillance à une date donnée.

Cependant, les auteurs ont trouvé que la seconde classe contient très peu de travaux de recherche et divise la première classe en trois sous-catégories "Méthodes statistiques", "Méthodes intelligentes" et "Méthodes basées sur des modèles".

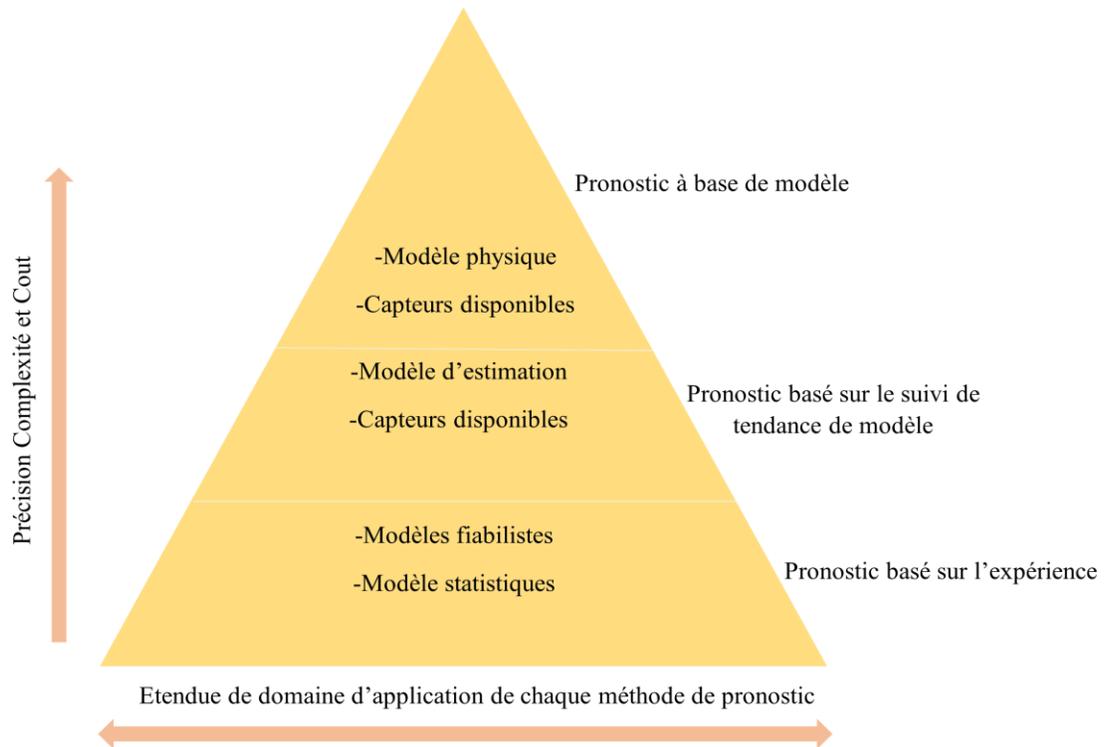


Figure I.10 : Classification des Approches de Pronostic

Par conséquent, nous choisissons la classification basée sur les méthodes suivantes :

- Les méthodes basées sur des modèles physiques
- Les méthodes guidées par les données
- Les méthodes basées sur l'expérience.

I.6.1 .1 Approches basées sur un modèle physique :

La mise en œuvre de cette méthode est généralement basée sur la représentation mathématique du mécanisme de dégradation. Cela signifie une connaissance spécifique des défauts et du mode de fonctionnement du système analysé. Les relations causales dérivées des lois de la physique sont utilisées pour modéliser les interactions entre les entités du système. Par conséquent, *les résidus* sont utilisés comme un outil mathématique. Ces résidus sont le résultat de contrôles de cohérence entre les mesures détectées par le système réel et la sortie de modèle mathématique. De

grands résidus impliquent des défauts, et de petits résidus impliquent des perturbations normales telles que du bruit ou des erreurs de modélisation. Des techniques statistiques sont souvent utilisées pour définir des seuils permettant de détecter la présence de défauts selon J. Yan et al. [102]. Le principe de la prédiction basée sur des modèles physiques est de déterminer le degré de dégradation actuel du système (dû aux processus de surveillance et de diagnostic) et en fonction de la sollicitation du système, une courbe d'évolution de la dégradation est utilisée pour estimer le temps restant avant la défaillance.

Les principaux avantages de cette approche selon Yan, J et al. (2004) [102] sont :

- Flexibilité : Si une des propriétés du système ou de la dégradation change, le modèle peut être réajusté pour prendre en compte cette modification,
- Interprétabilité : est la capacité à intégrer la compréhension physique du système dans le modèle. La connaissance de la structure du modèle permet de relier la variation d'indicateurs à une modification de paramètre.

Les principaux inconvénients du pronostic basé sur des modèles physiques sont :

- Le développement de modèles coûte très cher, cela nécessite un haut niveau de qualification et une certaine expérience.
- Difficulté à construire un modèle global pour les systèmes complexes : la complexité de la modélisation des interactions entre les différents mécanismes, les difficultés de calcul liées à la résolution de systèmes d'équations différentielles (explosion des temps de calcul pour les simulations Monte Carlo).
- Il peut ne pas être possible de généraliser une approche basée sur un modèle physique orientée « composant » à un niveau « système »,
- Il est nécessaire de comprendre le mécanisme de dégradation et les facteurs qui l'influencent.

Les techniques utilisées pour cette approche reposent sur les outils suivants : espace de parité, observateur (filtre de Kalman, etc.) et estimation de paramètres.

a) Espace de parité :

Un espace de parité est une relation analytiquement redondante représentée par une équation dans laquelle toutes les variables sont connues. La génération de cette relation permet de générer des résidus. En l'absence de défaut, les résidus sont statistiquement nuls. Lorsqu'un défaut survient, son amplitude change considérablement.

b) Observateur

Utilisez l'estimation d'état pour générer des résidus, y compris la reconstruction d'état ou, plus généralement, la sortie du processus à l'aide d'observateurs, et à utiliser l'erreur d'estimation comme résidu. Cette méthode s'est beaucoup développée car elle donne lieu à la conception de générateurs de résidus flexibles.

e) Estimation paramétrique

L'approche d'estimation des paramètres suppose que les effets des défaillances se reflètent dans les paramètres, et pas seulement dans les variables du système physique comme le fait l'observateur. Le principe de la méthode consiste à estimer en continu les paramètres du procédé en utilisant les mesures d'entrées/sorties et en évaluant la distance qui les sépare des valeurs de référence de l'état normal du procédé. L'estimation des paramètres a l'avantage de fournir des informations sur l'importance des déviations. Cependant, l'un des principaux inconvénients de cette approche est qu'elle nécessite un système physique excité en permanence. Ainsi, cela pose des problèmes d'ordre pratique dans le cas de procédés dangereux, coûteux ou exécutés en mode stationnaire.

I.6.1.2 Approches de pronostics guidées par les données :

L'approche basée sur les données découle directement des données d'exploitation du système (par exemple, étalonnage, données colorimétriques, Données spectrales, puissance, vibration et acoustique, température, pression, débris dans l'huile, courant et tension...). Dans de nombreuses applications, les données d'entrée/sortie mesurées sont la principale source d'informations sur le comportement de dégradation du système. Ce type d'approche vient donc apprendre les phénomènes de dégradation et capturer les relations subtiles entre les données, même si ces relations sont inconnues ou difficiles à décrire. Elles sont appuyées sur l'historique des données (données d'apprentissage) et donner des prédictions basées sur les données de surveillance conditionnelle (CM).

Une approche basée sur les données repose sur les hypothèses suivantes : Avant un événement de défaillance du système, les caractéristiques statistiques des données sont relativement uniformes. La mise en œuvre de l'approche « data-driven » est relativement simple, car elle ne nécessite pas de connaissances formelles sur le mécanisme de dégradation. Le principe est qu'une même cause aura toujours le même effet. Aussi, Cette approche peut être adaptée à tout type d'application avec suffisamment de données en quantité et en qualité. D'autre part, c'est aussi sa principale limite : l'efficacité des méthodes pronostiques basées sur les données dépend fortement de l'existence de telles données opérationnelles. En outre, le processus de surveillance détermine également l'exactitude du pronostic grâce à sa capacité à fournir une estimation fiable et précise de l'état de santé actuel du système selon Yan, J et al. (2004) [102].

Dans ce qui suit, les approches basées sur les données sont divisées en deux catégories :

- Méthodes statistiques (analyse de séries temporelles, etc.)
- Méthodes par l'intelligence Artificielle (IA).

a) Méthodes statistiques :

Les techniques statistiques les plus populaires sont les techniques de lissage (telles que le lissage exponentiel) et les techniques d'analyse de séries temporelles telles que les modèles de régression, y compris les modèles AR (Auto-Regressive) et ARMA (Auto-Regressive Moving Average models). Ces techniques sont également classées dans l'analyse des tendances. Elles ont l'avantage d'être faciles à programmer. Cependant, elles supposent que la dégradation du système est

monotone et ne peut pas intégrer les changements des conditions d'utilisation du système : l'évolution de la dégradation doit être Semblable à ce qui a été appris.

➤ **Approches par analyse des séries temporelles :**

Si une quantité suffisante de données temporelles est disponible, des techniques d'analyse de séries temporelles sont souvent utilisées pour déterminer l'état du système à un moment donné dans le futur. Ces techniques reposent fortement sur des données antérieures pour prédire les performances futures.

✓ **Techniques de régression :**

Une situation courante en séries temporelles est d'avoir à sa disposition deux ensembles de données de taille n , $\{y_1, y_2, \dots, y_n\}$ et $\{x_1, x_2, \dots, x_n\}$ obtenus expérimentalement ou mesurés. Le problème de la régression consiste à rechercher d'une relation pouvant éventuellement exister entre les x et les y , un modèle général de régression linéaire est donné par :

$$Y_i = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \dots + \beta_{p-1} X_{i,p-1} + \varepsilon_i \quad i = 1, \dots, n \quad (I.4)$$

Avec les hypothèses relatives du modèle

1. La distribution de l'erreur ε est indépendante de X ou X est fixe,
2. L'erreur est centrée et de variance constante (homoscédasticité) :

$$\forall i = 1, \dots, n \quad E(\varepsilon_i) = 0, \quad Var(\varepsilon_i) = \sigma^2, \quad \varepsilon \sim N(0, \sigma^2)$$

3. Les constants $\beta_0, \beta_1, \beta_2, \dots, \beta_{p-1}$ sont les paramètres à estimer.

L'estimation des paramètres $\beta_0, \beta_1, \beta_2, \dots, \beta_{p-1}$, est obtenue en maximisant la vraisemblance, sous l'hypothèse que les erreurs sont gaussiennes, ou encore par minimisation de l'erreur quadratique entre l'observations et le modèle (moindres carrés). Les deux approches conduisent aux mêmes estimations tandis que le maximum de vraisemblance induit de meilleures propriétés des estimateurs. Pour une séquence d'observations $\{(x_i, y_i) \mid i = 1 \dots, n\}$, le critère des moindres carrés s'écrit :

$$\min \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_{i,1} - \beta_2 X_{i,2} - \dots - \beta_{p-1} X_{i,p-1})^2 \quad (I.5)$$

Le maximum de vraisemblance peut également être utilisé pour estimer $\beta_0, \beta_1, \beta_2, \dots, \beta_{p-1}$, si la distribution de probabilité du terme d'erreur est connue.

Neter, J et al. (1996) [103] ont proposé un modèle de régression logistique pour calculer la probabilité de défaillance pour une condition donnée des variables, et un modèle de série temporelle autorégressif moyenne mobile (ARMA) pour étendre les conditions de prédiction de défaillance possibles. Hush, D. R et al. (1993) [104] ont fourni plus de détails sur les modèles de régression linéaires et non linéaires, et les dangers de l'extrapolation au-delà des données observées.

b) Approches à base d'intelligence artificielle (IA)

Les réseaux de neurones artificiels ou réseaux de neurones (RN) sont une tentative de modéliser le cerveau par l'interconnexion dense d'un large ensemble d'éléments de traitement simples. Les détails du traitement et du développement des réseaux neuronaux peuvent être trouvés dans plusieurs articles et livres.

L'avantage des RNA est leur capacité d'apprendre utilisée l'historique, puis essayer de prédire l'état future quand un ensemble de données présent à l'entrée. Ces techniques sont fréquemment utilisées dans les méthodes de pronostic actuelles.

L'une des méthodes d'apprentissage machine utilisée pour le pronostic consiste à utiliser des réseaux de neurones artificiels pour modéliser des systèmes. Un réseau de neurones artificiels est un modèle (généralement non linéaire) qui établit un ensemble de relations fonctionnelles interconnectées entre l'entrée du stimulus et la sortie souhaitée, où les paramètres de la relation fonctionnelle doivent être réglés pour des performances optimales. La fonction de l'RNA est proche de celle des vrais neurones dans le cerveau humain. Chaque neurone a une dendrite, qui est la voie d'entrée, un soma qui traite l'entrée et un axone qui transporte la sortie. Un réseau de neurones artificiels est constitué d'un ensemble de ces neurones artificiels. La structure de l'RNA à trois couches est représentée sur la figure I.11. Ce réseau a généralement une couche d'entrée, une ou plusieurs couches cachées et une couche de sortie. Chaque couche sauf la couche d'entrée est constituée de nombreux neurones. Attribuer des poids à chaque entrée (de la couche d'entrée aux neurones de la couche cachée). Compte tenu du résultat de la fonction de transfert de la couche cachée, appliquer un autre ensemble de poids en entrée de la couche de sortie. Le résultat final de la fonction de transfert de la couche de sortie forme l'ensemble de sortie de l'RNA.

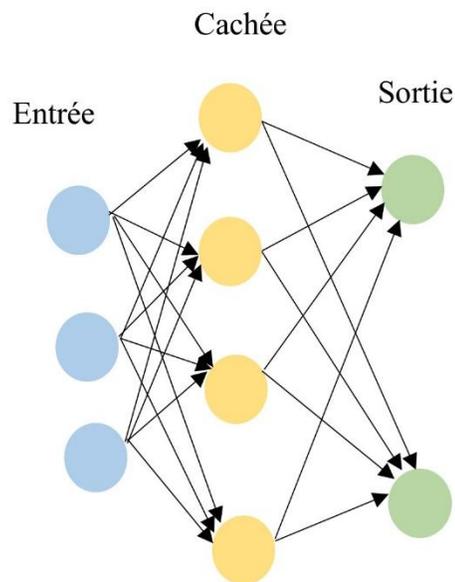


Figure I.11 : Architecture d'un réseau de neurone à trois couches

L'article de Li, C. J et al. (1995) [105] fournissent des introductions accessibles et informatives sur deux des réseaux de neurones les plus populaires : le perceptron multicouche (MLP) et le réseau de neurones à fonction de base radiale (RBF)

Les réseaux de neurones se sont avérés utiles dans une variété de domaines : détection, classification, approximation de fonctions multidimensionnelles et modélisation, prédiction. Ils sont idéaux pour développer des transformations non linéaires pour mapper les données d'entrée aux sorties.

Ainsi, ils peuvent être utilisés pour la classification ainsi que pour la prédiction. Les RNA sont « formés » en présentant des exemples de paires de données d'entrée/sortie. Pour la plupart des applications, les données de sortie sont « étiquetées » en fonction de la réponse correcte de la classe ou de la fonction. Pendant l'apprentissage, les paramètres du réseau neuronal sont ajustés jusqu'à ce que les performances de classification du réseau neuronal atteignent un niveau acceptable.

Deux types de RNA sont largement utilisés qui sont les MLP et RBF.

MLP est avéré excellent pour l'approximation de fonctions non linéaires et pour résoudre des problèmes de classification. Ils se sont révélés être des classificateurs bayésiens. Cependant, les MLP ne sont pas capables d'effectuer une détection de nouveauté : un MLP ne peut pas détecter lorsque des données pour lesquelles il n'a pas été formé sont présentes aux entrées.

Le RBF est également avéré être un excellent classificateur et approximateur de fonction. Pour l'implémentation standard du RBF, tous les éléments de traitement de la couche intermédiaire appliquent une fonction gaussienne multidimensionnelle aux données d'entrée.

La couche de sortie est la somme pondérée des sorties de la fonction de base. Le RBF peut être considéré comme un classificateur du plus proche voisin.

Les techniques de réseau de neurones artificiels pour l'estimation de la durée de vie utile restante (RUL) ont été utilisées par de nombreux chercheurs. Vachtsevanos, G et al. (2001) [106] ont examiné l'utilité des réseaux de neurones à rétropropagation pour la prédiction des dommages dus à la fatigue. Leur méthode a été utilisée pour réduire le temps de calcul requis par les méthodes traditionnelles pour résoudre numériquement des équations différentielles non linéaires. De plus, l'auteur indique qu'en plus de la prédiction de la durée de vie restante, ces réseaux de neurones pourraient potentiellement être utilisés pour l'analyse en temps réel des modèles de dommages par fatigue et d'autres types de modèles de défaillance.

Un réseau neuronal à ondelettes dynamiques (DWNN) a été utilisé pour transformer les données des capteurs en une évolution temporelle des défauts afin de prédire la durée de vie restante des roulements d'après Yam, R. C. M et al. (2001) [107].

L'auteurs de l'article Gustin, J. L. (1996) [88] a utilisé un RNA pour effectuer une prédiction en un pas sur les données conditionnelles (CM) du réducteur. Il a prouvé que les RNA peuvent apprendre La tendance des données temporelles non linéaires peut être adaptée pour la prédiction de séries temporelles selon Heng, A et al. (2009) [19].

Les auteurs Antoni, J et al. (2004) [23] utilisent des informations d'incertitude dans l'estimation de la durée de vie utile restante via les réseaux de neurones bayésiens et l'hamiltonien Monte Carlo.

Une approche de filtrage de réseau neuronal pour l'estimation de la durée de vie utile restante basée sur la similarité, interprétée par Wang, P et al. (2001) [108].

Les techniques de logique floue sont une autre technique d'IA qui sont aussi populaire pour la prédiction (pronostic). La logique floue fournit un langage (avec une syntaxe et une sémantique locale) dans lequel des connaissances qualitatives sur le problème à résoudre peuvent être traduites. En particulier, la logique floue permet de modéliser des systèmes dynamiques à l'aide de variables linguistiques. Ces variables prennent des valeurs floues caractérisées par des phrases et des fonctions d'appartenance. La signification des variables linguistiques peut être interprétée comme des contraintes élastiques sur leurs valeurs. Ces contraintes sont propagées par des opérations d'inférence floue. Le mécanisme d'inférence résultant a de fortes propriétés d'interpolation, ce qui rend la logique floue remarquablement robuste aux changements des paramètres du système et aux perturbations.

Lorsqu'elle est appliquée à la prédiction (pronostic), la logique floue est souvent utilisée en conjonction avec des méthodes d'apprentissage machine pour faire face à une partie de l'incertitude à laquelle sont confrontées toutes les estimations de pronostic. Watson, M et al. (2004) [109]; Byington, C. S et al. (2004) [110] ont utilisé des techniques de logique floue pour évaluer de manière fiable l'état de santé système. Pour estimer automatiquement l'état de santé, ils ont utilisé cette technique pour représenter la gravité ou le degré de dégénérescence. Les auteurs Wang, W. Q et al. (2004) [111] ont présenté un article majeur sur le pronostic Neuro-Flous (NF). Tous les deux évaluent les techniques RNN et Neuro-Fuzzy (NF) à l'aide des données de test de référence des taches solaires et des données de test en ligne des réducteurs. Dans le cas des taches solaires, NF sans interpolation n'est pas aussi précis que RNN. Mais NF avec interpolation produit des résultats plus précis que RNN et fournit des résultats similaires dans environ dix pour cent (10%) des époques d'apprentissage. Dans le test en ligne, NF appréhende plus rapidement le comportement dynamique du système, montré supérieur à RNN.

Aussi Giantomassi, A et al. (2011) [112] ont proposé un Modèle de Markov caché pour l'estimation de la santé et le pronostic des moteurs à double flux.

Une autre approche courante en apprentissage automatique est l'utilisation d'algorithmes de détection d'anomalies, également connus sous les noms d'algorithmes de détection de nouveauté ou de détection de valeurs aberrantes. Les travaux de Bock, J. R et al. (2006) [113]; Clifton, D et al. (2006) [114] se concentrent sur cette approche spécifique.

I.6.1.3 Approches basées sur l'expérience :

Les méthodes de pronostic basées sur l'expérience s'appuient sur la modélisation stochastique des phénomènes de dégradation (lois de fiabilité, processus de Markov ou non-Markov). Les modèles basés sur des expériences similaires passées sont également utilisés dans certaines approches. Les travaux de Yan, J et al. (2004) [102] explorent cette idée.

Les méthodes basées sur l'expérience sont principalement dérivées des méthodes traditionnelles de modélisation et de traitement de la fiabilité. En ce sens, le pronostic est analogue à l'étude de la fiabilité prévisionnelle, dont le but est d'identifier des paramètres distribués aléatoirement et décrivant des phénomènes de dégradation ou de défaillance (loi de Poisson, loi exponentielle, loi de Weibull, loi log-normale). De ce groupe, la loi de Weibull reste la plus répandue selon Coquempot, V et al. (2005) [67]. Les méthodes traditionnelles de pronostic basées sur la fiabilité peuvent être divisées en deux catégories : la fiabilité basée sur les défaillances et la

fiabilité basée sur la dégradation. La fiabilité basée sur les défaillances est utilisée pour estimer la distribution de la durée de vie et ses paramètres lorsque des données suffisantes, complètes et/ou censurées existent. S'il existe une connaissance préalable de la distribution de la durée de vie de composants similaires, Une distribution de vie est généralement supposée suivre la même distribution de composants similaires.

Contrairement à la fiabilité basée sur les défaillances, la fiabilité basée sur la dégradation se concentre sur l'utilisation d'une mesure de la dégradation des composants, plutôt que sur les données de défaillance, pour évaluer la durée de vie restante d'un composant. La dégradation est également connue sous le nom de dommages cumulatifs selon Patton, R. J. (1997) [55]. Aussi, utiliser un modèle de fiabilité évolutif de type ALM (Accelerated Life Model), ou un modèle à risques proportionnels PHM (Proportional Hazard Model), ou mettre en œuvre une approche bayésienne pour mettre à jour les paramètres de la loi de dégradation, Les informations disponibles peuvent également représenter une solution. La simulation Monte Carlo permet de combiner divers phénomènes aléatoires, mais elle se confronte au problème d'explosion du temps de simulation selon Yan, J et al. (2004) [102].

Les modèles (PHM) sont couramment utilisés pour la prédiction des défaillances et l'analyse de la fiabilité. Le PHM suppose que le risque varie proportionnellement aux covariables et que la constante de proportionnalité est toujours la même. Parmi les travaux récents dans le domaine du pronostic basé sur l'expérience, on peut citer les suivants. Les auteurs Evsukoff, A et al. (1997) [56] ont proposé le logiciel EXAKT pour optimiser le remplacement des équipements critiques (turbines, vannes, moteurs, etc.).

Balle, P et al. (1998) [58] ont proposé la suite du logiciel IBTV « Bayesian Inference for Aging Treatment » d'EDF, qui intègre une approche pronostique originale dans laquelle les stratégies de maintenance reposent sur le déclenchement d'interventions dès la détection du vieillissement d'un Composants (par exemple vannes, relais, mécanique). La méthode est basée sur une approche bayésienne qui modélise a priori le comportement de défaillance (composition loi exponentielle/Weibull). De plus, l'efficacité de l'action envisager la maintenance est prise en compte suivant l'avis d'un expert.

Banjevic, D et al. (2006) [115] ont discuté de l'estimation RUL pour les processus de Markov de défaillance temporel qui incluent un modèle PHM commun et des propriétés de Markov pour l'évolution Covariables comme cas particulier.

I.7 Conclusion

Les méthodes de diagnostic/pronostic ci-dessus ne constituent pas un état de l'art exhaustif des méthodes existantes. Cependant, les techniques décrites sont les plus connues et les plus couramment utilisées. Le choix de la méthode de diagnostic dépend de la connaissance du système et de la présence de capteurs ou de modèles capables de suivre l'état réel du composant/sous-système/système.

Pour cela, nous limitons la portée des travaux de la section diagnostics (chapitre II) à l'amélioration du taux de classification des défauts du moteur à induction, tels que les défauts de rupture de barre rotorique. À cet égard, nous proposons une approche basée sur les ondelettes par paquet pour l'extraction du vecteur caractéristiques et la classification par la technique SVM.

Chapitre II.

Détection et classification par SVM des défauts de moteur à induction en utilisant WPA

II.1 Introduction

Dans de nombreux secteurs industriels, le suivi des installations est un enjeu majeur pour assurer une compétitivité optimale des outils de production. A ce titre, leur productivité est une préoccupation quotidienne des dirigeants d'entreprise. La compétitivité de ce secteur passe nécessairement par la maîtrise et la disponibilité des outils de production, l'amélioration de leur fonctionnement, la sécurité des utilisateurs, et la réduction des coûts de maintenance. Cela nécessite l'optimisation de la fiabilité, la maintenabilité, et la sécurité des systèmes industriels. L'efficacité du dispositif de détection de la dégradation opérationnelle du système est utile réaliser des gains de productivité élevés ; ce dispositif, qui fait partie intégrante d'installation, permet la détection et la localisation d'une éventuelle défaillance du matériel.

Les systèmes technologiques peuvent évoluer dans divers modes de fonctionnement. Ces modes ne sont pas forcément tous connus, en effet, si certains de ces modes correspondent à un fonctionnement normal, d'autres modes ont des caractéristiques qui n'interviennent qu'en cas de défaillance et ne sont donc généralement pas tolérés. Pour la surveillance automatique d'équipements industriels, il est nécessaire de développer des systèmes de diagnostic avec certaines capacités adaptées aux situations permettant une identification fiable du mode de fonctionnement dans lequel le système fonctionne. Une démarche de diagnostic peut être définie comme un processus en trois étapes : détection d'un mode de fonctionnement, identification, et localisation.

La rupture des barres du rotor dans les moteurs à induction à cage d'écureuil fait depuis longtemps l'objet de nombreuses études scientifiques. Le but de ce chapitre est d'améliorer la précision de la classification des défauts du modèle de moteur à induction adapté aux études par simulation numérique de défauts tels que les défauts cassure des barres rotoriques.

En générale, on cherche à développer une technique de surveillance de l'état de fonctionnement du moteur à induction par la mesure des paramètres mesurables (courant, tension, vitesse, etc...). On propose une méthode d'extraction de caractéristiques de défaut d'un moteur à induction à cage d'écureuil basée sur le calcul d'énergie des sous-bandes du signal après une décomposition par la transformée en ondelettes par paquets du courant statorique. Les ondelettes sont un outil d'analyse temps-fréquence avec une largeur de la fenêtre d'analyse automatiquement ajustée pour diverses composantes de fréquence ; la capacité de localisation temporelle et l'analyse multi-résolution rendent les ondelettes très attrayantes pour la détection et le diagnostic des défauts.

L'idée principale de la méthode proposée est d'utiliser la décomposition en ondelettes par paquets pour décomposer le signal de courant statorique dans l'espace-temps-fréquence et d'utiliser le résultat pour calculer l'énergie des seize premières sous-bandes, et faire la sélection des énergies des sous-bandes les plus pertinentes pour les utiliser afin d'améliorer le taux de classification des défauts. On a proposé l'utilisation de classification par la technique Machine à vecteurs de support : SVM (Support Vector Machine) multi-classe. Les résultats de la simulation et l'interprétation, avec une conclusion, sont présentés.

II.2 Modèle mathématique de la machine

II.2.1 le cas sain

Le modèle mathématique global de la machine à induction à cage d'écuréuil est développé et simulé via le logiciel MATLAB® /Simulink.

En considérant que la machine à induction étudiée a un enroulement triphasé concentrique dans le circuit du stator avec N_r barres, le rotor est ce qui peut être décrit comme N_r boucles de rotor identiques et équidistantes, selon Bossio, G et al. (2006) [116]. Chaque boucle comprend deux barres de rotor ainsi que les pièces de connexion de l'extrémité, comme illustré dans la figure II.1.

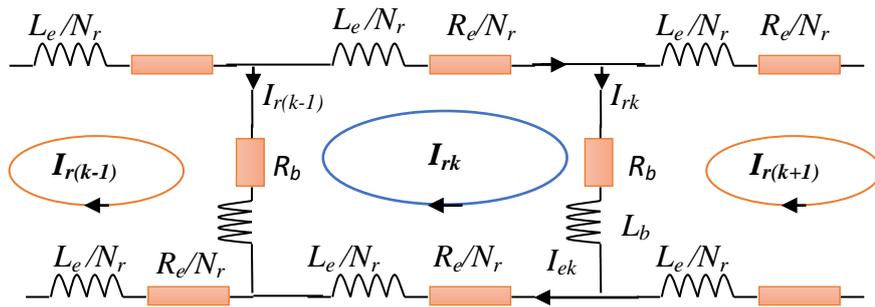


Figure II.1. Topologie du circuit des mailles du rotor

Le modèle mathématique global de la machine à induction à cage d'écuréuil est représenté par la matrice suivante, comme décrit par Razik, H. (2002) [117] :

$$[V] = [R][I] \frac{d}{dt} ([L][I]) \quad (\text{II.1})$$

Où, les matrices de tension et de courant sont respectivement données par l'équation II.2 et l'équation II.3 :

$$[V] = \begin{bmatrix} [V_s] \\ [V_r] \end{bmatrix} \quad (\text{II.2})$$

$$[I] = \begin{bmatrix} [I_s] \\ [I_r] \end{bmatrix} \quad (\text{II.3})$$

Avec, le système des différentes matrices de vecteurs de stator et rotor des tensions et courants se présente comme suit :

$$\begin{aligned} [V_s] &= [V_{s1} \quad V_{s2} \quad V_{s3}]^t \\ [I_s] &= [I_{s1} \quad I_{s2} \quad I_{s3}]^t \\ [V_r] &= [0 \quad 0 \quad 0 \quad \dots \quad 0]_{1 \times N_r+1}^t \\ [I_r] &= [I_{r1} \quad I_{r2} \quad I_{r3} \quad \dots \quad I_{rk} \quad \dots \quad I_{rN_r} \quad I_e]_{1 \times N_r+1}^t \end{aligned} \quad (\text{II.4})$$

La résistance globale est exprimée par la matrice suivante :

$$[R] = \begin{bmatrix} [R_s]_{3 \times 3} & \vdots & [0]_{3 \times N_r} & \vdots & [0]_{3 \times 1} \\ \dots & \dots & \dots & \dots & \dots \\ [0]_{N_r \times 3} & \vdots & [R_r]_{N_r \times N_r} & \vdots & -\frac{R_e}{N_r} [1]_{N_r \times 1} \\ \dots & \dots & \dots & \dots & \dots \\ [0]_{1 \times 3} & \vdots & -\frac{R_e}{N_r} [1]_{1 \times N_r} & \vdots & R_e \end{bmatrix} \quad (II.5)$$

Où, la matrice des résistances du stator et du rotor sont données respectivement par l'équation (II.6) et l'équation (II.7) :

$$[R_s]_{3 \times 3} = \begin{bmatrix} R_s & 0 & 0 \\ 0 & R_s & 0 \\ 0 & 0 & R_s \end{bmatrix} \quad (II.6)$$

$$[R_r]_{N_r \times N_r} = \begin{bmatrix} R_{b_0} + R_{b_{(N_r+1)}} + \frac{2R_e}{N_r} & -R_{b_0} & 0 & \dots & \dots & 0 & -R_{b_{(N_r+1)}} \\ 0 & \dots & -R_{b_{(N_r-1)}} & R_{b_k} + R_{b_{(N_r-1)}} + \frac{2R_e}{N_r} & -R_{b_k} & 0 & \dots & 0 \\ \dots & \dots \\ -R_{b_{(N_r-1)}} & 0 & \dots & 0 & -R_{b_{(N_r-2)}} & R_{b_{(N_r-1)}} + R_{b_{(N_r-1)}} + \frac{2R_e}{N_r} & \dots & \dots \end{bmatrix} \quad (II.7)$$

La matrice d'inductance mutuelle entre les phases du stator et le maillage du rotor peut s'écrire comme suit :

$$[M_{sr}]_{N_r \times 1} = \begin{bmatrix} \dots & -M_{sr} \cos(\theta + ka) & \dots \\ \dots & -M_{sr} \cos\left(\theta + ka - \frac{2\pi}{3}\right) & \dots \\ \dots & -M_{sr} \cos\left(\theta + ka - \frac{4\pi}{3}\right) & \dots \end{bmatrix} \quad (II.8)$$

Où, k est compris entre 0 et N_r-1

La matrice des inductances des phases statoriques exprimée par la relation (II.9) est de l'ordre (m, m), où m=3 :

$$[L_s]_{3 \times 3} = \begin{bmatrix} L_{sp} & M_s & M_s \\ M_s & L_{sp} & M_s \\ M_s & M_s & L_{sp} \end{bmatrix} \quad (II.9)$$

La matrice des inductances du rotor N_r barres s'exprime par la relation suivante :

$$[L_R]_{N_r \times N_r} = \begin{bmatrix} L_{rp} + 2L_b + \frac{2L_e}{N_r} & M_{rr} - L_b & M_{rr} & M_{rr} & \dots & M_{rr} - L_b \\ M_{rr} - L_b & L_{rp} + 2L_b + \frac{2L_e}{N_r} & M_{rr} - L_b & M_{rr} & \dots & M_{rr} \\ M_{rr} & M_{rr} - L_b & L_{rp} + 2L_b + \frac{2L_e}{N_r} & M_{rr} - L_b & M_{rr} & \dots \\ M_{rr} & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ M_{rr} - L_b & M_{rr} & M_{rr} & \dots & M_{rr} - L_b & L_{rp} + 2L_b + \frac{2L_e}{N_r} \end{bmatrix} \quad (\text{II.10})$$

L'équation mécanique du mouvement s'écrit sous la forme suivante :

$$J \frac{d\omega_r}{dt} = C_{em} - C_r \quad (\text{II.11})$$

$$\frac{d\theta_r}{dt} = \omega_r \quad (\text{II.12})$$

Avec :

θ_r : angle qui définit la position du rotor.

ω_r : vitesse de rotation du rotor.

C_r : couple résistant.

C_{em} : couple électromagnétique

II.2.2 le cas d'un défaut

La modélisation d'une barre de rotor cassée peut-être modélisée en augmentant sa valeur de résistance, de sorte que le courant transitoire qui la traverse est dévolu à zéro, la nouvelle structure du circuit de mailles de rotor dans une barre de rotor cassée sera comme illustré à la figure II.2.

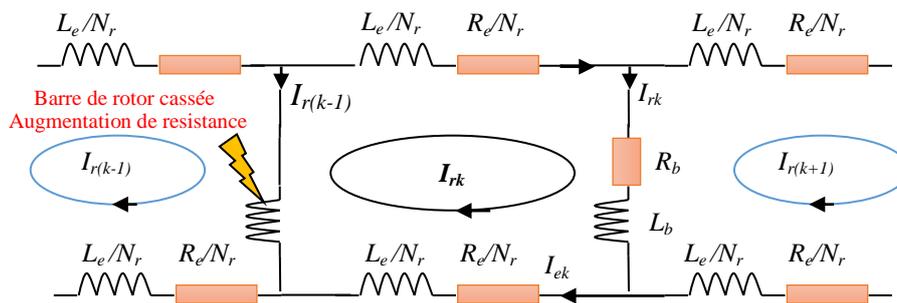


Figure II.2 Circuit de mailles de rotor dans le cas d'une barre de rotor

Ceci est introduit dans la matrice des résistances par l'addition de la matrice de la résistance rotorique $[R_r]$ avec la matrice de défaut $[R_d]$.

$$[R_d] = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ \vdots & 0 & \dots & \dots & 0 \\ \vdots & R_{k,k} & R_{k,k+1} & 0 & \vdots \\ 0 & R_{k+1,k} & R_{k+1,k+1} & \vdots & \vdots \\ 0 & \dots & \dots & 0 & 0 \end{bmatrix} \quad (\text{II.13})$$

En cas de barre cassée, la valeur de résistance de cette barre est multipliée par un facteur M avec :

$$R_{k,k} = R_{k+1,k+1} = (M + 1).R_b + 2.R_e \quad (\text{II.14})$$

$$R_{k+1,k} = R_{k,k+1} = (-M).R_b \quad (\text{II.15})$$

Par conséquent, la matrice de résistance de la cage du rotor, compte tenu du défaut des barres cassées, est définie comme suit :

$$[R_{rd}] = [R_r] + [R_d] \quad (\text{II.16})$$

II.3 Détection des défauts et classification :

Le diagnostic de défauts présente un domaine d'application approprié pour les méthodes de classification, car son objectif principal est d'obtenir une cartographie optimale de l'état actuel du système surveillé entre un ensemble d'états système prédéfinis, tous les états du système comprend un état normal et un état défectueux. Par conséquent, l'utilisation de techniques de classification est un choix naturel lors de la conception d'un système de diagnostic de défauts. Les principales questions qui découlent de la réponse à la tâche de classification sont :

- Comment générer les caractéristiques ?
- Quel est le nombre optimal des caractéristiques à utiliser ?
- Comment se fait la conception du classificateur ? Nous divisons notre travail en trois parties décrites dans les paragraphes suivants :

✚ Générations des caractéristiques

II.3.1 La transformées des ondelettes pour l'extraction des caractéristiques :

II.3.1.1 Introduction

Ces dernières années, les problèmes de défaillance dans les machines sont devenus plus importants et préoccupants dans les applications industrielles. La volonté d'améliorer la fiabilité des systèmes d'entraînement industriels a conduit à des activités de recherche et de développement concertées dans plusieurs pays pour évaluer les causes et les conséquences de diverses conditions de défaillance. En particulier, des travaux de recherche sont actuellement axés sur les défauts des barres de rotor et sur le développement de techniques de diagnostic.

D'après Deleroi, W (1982) [118], la majorité des pannes de rotor sont causées par une combinaison de diverses contraintes. Plusieurs travaux publiés sont liés aux cassures de barres de rotor dans les moteurs à induction à cage d'écureuil et sont basés sur l'analyse de signature du courant de la machine, dont nous citons notamment Vieira, M. (1995) [119]. De cette manière, la transformée de Fourier rapide est utilisée pour montrer l'effet de ce défaut sur le spectre de fréquence du signal. Les barres de rotor cassées dans la machine à induction à cage d'écureuil entraînent la présence de deux bandes latérales à la fréquence $(1 \pm 2s) f$ autour de la composante fréquentielle principale dans le spectre du courant statorique, comme mentionné par Ho, S. L et

al. (1993) [120]. L'estimation de la résistance du rotor est une autre approche, qui est rapportée par Castro, J. D et al. (2000) [121]. Bien que la transformée de Fourier soit une méthode efficace, elle est utile pour le traitement des signaux stationnaires seulement.

Si l'on considère le cas des machines asynchrones, la non-stationnarité des signaux est une propriété très courante mais difficile à maîtriser. Certains processus fonctionnent sous des couples de charge qui varient dans le temps. D'après Henao, H al. (2003) [122], les techniques de traitement temps-fréquence et temps-échelle sont bien adaptées à ces situations. La transformée de Fourier n'est pas suffisante pour représenter ce type de signaux. Au cours des dernières années, le développement de méthodes statistiques de surveillance des signaux a fourni un outil efficace pour gérer les signaux non stationnaires. Ensuite, la transformée temps-fréquence fournit le meilleur cadre mathématique pour analyser les signaux non stationnaires.

Parmi les transformations temps-fréquence, on cite les ondelettes. La transformée en ondelettes est une méthode efficace pour l'analyse des signaux non stationnaires et utilise une nouvelle description de la décomposition spectrale via le concept de mise à l'échelle. Une autre méthode de détection de cassure de barres du rotor utilise la transformée en ondelettes et calcule l'énergie des bandes fréquentielles du signal, comme rapporté par Abbaszadeh, K et al. (2001) [123].

II.3.1.2 La théorie des ondelettes :

Dans la méthode de Fourier, les "fonctions de bases" sont totalement concentrées en fréquence (impulsions de Dirac) et totalement réparties dans le temps (sinusoïdes non amorties s'étendant de $-\infty$ à $+\infty$) C'est une autre façon d'expliquer que le passage dans l'espace de Fourier donne le maximum d'informations sur la répartition des fréquences mais perd entièrement les informations relatives au temps.

Dans la méthode de la transformée de Fourier à fenêtre glissante, les informations temps-fréquences restent couplées par un compromis, limité par la relation d'incertitude, sur la localisation à la fois en temps et en fréquence. Cela met en évidence les avantages de cette méthode par rapport à celle de Fourier.

Cependant, la méthode de la transformée de Fourier à fenêtre glissante présente l'inconvénient majeur d'avoir une fenêtre de longueur fixe. Il n'est donc pas possible d'analyser simultanément des phénomènes dont les échelles de temps sont différentes. De plus, le calcul de la STFT (Short-time Fourier transform) doit s'effectuer pour chaque changement de taille de la fenêtre, d'où une charge de calcul importante.

Une autre méthode d'analyse, qui ne privilégie aucune échelle particulière mais qui généralise à toutes les échelles, l'analyse locale des fréquences obtenues par la STFT, est donc nécessaire.

La technique des ondelettes propose une analyse très fine des signaux et permet de détecter la non-stationnarité dans les signaux où cette particularité n'est pas disponible dans les techniques classiques telles que l'analyse de Fourier et l'analyse de Fourier à fenêtre glissante, etc.

Cette transformation est née de la convergence des travaux théoriques déjà anciens, notamment ceux de Haar (1910), de Littlewood et Paley (1930), de Zygmund (1930), de Gabor (1940), puis vers 1960 de Calderon, et des idées récentes mises en avant-propos ; pour le traitement numérique de certains signaux par Morlet (le premier à avoir proposé le nom d'ondelettes, 1982), ou pour le développement d'outils mathématiques utilisés en physique théorique par Grossmann (1983).

Dès lors, de nombreux chercheurs ont apporté des bases mathématiques solides en faisant apparaître la notion de base orthogonale (Meyer, 1985), d'analyse multirésolution (Mallat, 1989), et d'ondelettes à support compact (Daubechies, 1988). Pour obtenir plus de renseignements sur l'histoire des ondelettes, le lecteur peut se référer aux ouvrages [124] et [125].

Les ondelettes, famille de fonctions déduites d'une même fonction, appelée ondelette mère, par opérations de translations et dilatations, ont trouvé, de par la puissance de leur théorie, des applications dans de nombreux domaines aussi variés que les mathématiques (Daubechies, I. (1991) [126] a apporté sa contribution à l'analyse ; Derin, H et al. (1987) [127] ont également apporté leur contribution dans le domaine des probabilités ; Muzy, J. F et al. (1994) [128] dans le domaine des fractales.), le traitement du signal (S. Mallat et al. (1992) [129] ont apporté leur contribution dans le domaine de la compression ; Bijaoui, A. (1991) [130] dans le domaine de l'astronomie ; Bouchereau, E. B. (1997) [131], dans le domaine de la sismique.), la physique (Burt, P. J et al. (1983) [132], se sont intéressés à la mécanique quantique, tandis que Farge, M et al. (1996) [133], se sont consacrés à l'étude de la turbulence.). En effet, cet outil permet la représentation de fonctions de L^2 , dans une base bien localisée en espace et en fréquence, offrant les avantages de l'analyse de Fourier et s'affranchissant des inconvénients du manque de localisation de cette dernière.

II.3.1.3 Transformée en ondelettes continue :

A partir d'une fonction de base ψ , appelée ondelette mère, on construit une famille de fonctions analysantes :

$$\psi_{a,b} = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) \quad a > 0 \quad (\text{II.17})$$

Où le paramètre a est le facteur d'échelle (dilatation), et b est le paramètre de translation.

On définit alors les coefficients de la transformée en ondelettes d'un signal $s(t)$, comme étant les produits scalaires :

$$C_s(a, b) := \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} s(t) \psi\left(\frac{t-b}{a}\right) dt \quad (\text{II.18})$$

Ces coefficients mesurent, en un certain sens, les fluctuations du signal $s(t)$ autour du point $t = b$, à l'échelle fournie par a , d'après l'équation (II.17), en diminuant a , le support de $\psi_{a,b}$ réduit en temps et donc couvre une plage en fréquence plus grande et vice versa, donc $\frac{1}{a}$ est proportionnel à une fréquence.

Pour un facteur d'échelle assez grand, la représentation des coefficients d'ondelettes en fonction de b , la position donne une représentation de "la forme générale de la fonction". Par contre un facteur d'échelle faible correspond à une représentation des singularités. La fonctionnalité de ce "microscope" s'avère extrêmement utile lorsqu'il s'agit d'étudier la régularité d'une fonction, comme le soulignent Mallat, S et al. (1992) [129]. Il est possible de démontrer que si la fonction d'analyse (l'ondelette) est choisie de manière appropriée, la transformation en ondelettes est réversible, comme l'a mis en évidence dans leur article Bijaoui, A. (1991) [130]. Le signal $s(t)$ peut être reconstruit après double intégration suivant le facteur d'échelle a et le paramètre de translation b :

$$s(t) = \frac{1}{c_\psi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{1}{a^2} C_s(a, b) \psi_{a,b}(t) da db \quad (\text{II.19})$$

II.3.1.4 Transformée en ondelettes discrète

L'intérêt de ce paragraphe est de présenter très brièvement la transformée en ondelettes discrète pour montrer la relation avec la transformée en ondelettes continue.

Les valeurs discrètes du facteur d'échelle a et du paramètre de translation b seront considérées sous la forme : $a = 2^{-j}$ et $b = k2^{-j}$, avec ces valeurs de a et b , l'équation (II.18) devient :

$$C_s(2^{-j}, k2^{-j}) = 2^{\frac{j}{2}} \int_{-\infty}^{+\infty} s(t) \psi(2^j t - k) dt \quad (\text{II.20})$$

Si la fonction $s(t)$ est discrétisée, en supposant une période d'échantillonnage égale à 1 , pour des raisons de simplicité, l'équation (II.20) s'écrit alors :

$$C_s(2^{-j}, k2^{-j}) = 2^{\frac{j}{2}} \sum_n s(n) \psi(2^j n - k) \quad (\text{II.21})$$

Pour calculer la transformée en ondelettes d'une fonction en différents points dans le plan temps-échelle, il n'est pas nécessaire de connaître les valeurs de la fonction sur l'ensemble de l'axe des temps. Il suffit simplement de connaître les valeurs de la fonction aux temps où l'ondelette est différente de zéro.

D'après l'équation (II.21), la transformée en ondelettes discrète d'une fonction translatée dans le temps est différente de la transformée en ondelettes discrète de la fonction originale. Pour l'expliquer, supposons :

$$s_m(t) = s(t - t_m) \quad (\text{II.22})$$

La transformée en ondelettes de $s_m(t)$ s'écrit :

$$C_s(2^{-j}, k2^{-j}) = 2^{\frac{j}{2}} \int_{-\infty}^{+\infty} s_m(t) \psi(2^j t - k) dt \quad (\text{II.23})$$

$$\begin{aligned} &\approx 2^{\frac{j}{2}} \sum_n s(n - m) \psi(2^j n - k) \\ &= 2^{\frac{j}{2}} \sum_n s(n) \psi[2^j n - (k - m2^j)] \\ &\approx C_s(2^{-j}, (k - m2^j)2^{-j}) \end{aligned} \quad (\text{II.24})$$

II.3.1.5 Analyse multi résolution et algorithmes

L'idée de l'analyse multirésolution a été développée par Bouchereau, E. B. (1997) [131] ; Mallat, S. G. (1989) [134] et Burt, P. J et al. (1983) [132] ; Farge, M et al. (1996) [133]. Dans l'ouvrage de Mallat, S et al. (1992) [135], l'analyse multirésolution est présentée comme une théorie qui définit des opérateurs linéaires permettant d'analyser un signal à différentes échelles. Nous pourrions dire que la construction d'une multirésolution permet de regarder un signal de "très près" ou de "très loin". Ce Zoom est effectué à l'aide d'une fonction d'échelle, qui se dilate à travers les échelles. Le signal projeté sur cette fonction donne une représentation de notre signal d'origine à l'échelle supérieure. Cette représentation (coefficients de projection) provoque un zoom arrière de notre signal d'origine, d'où le terme approximation. Afin de reconstruire notre signal, à partir des coefficients d'approximation, nous devons également projeter notre signal originel sur un espace perpendiculaire (conservation de toute l'information). La fonction générant ce deuxième espace vectoriel sera une ondelette. Finalement, le signal est projeté sur une fonction d'échelle créant une approximation du signal et sur une ondelette pour récupérer

l'information perdue lors de la première projection. Cette deuxième projection contient les détails du signal d'origine.

Les bases $\varphi_{j,k}(t)$ et $\psi_{j,k}(t)$ sont respectivement nommées fonctions d'échelles et ondelettes.

a. L'analyse multi résolution

Une analyse multi résolution de $L^2(R)$ est, par définition, une suite croissante V_j , avec $j \in R$, de sous espaces vectoriels fermés de $L^2(R)$ ayant les propriétés suivantes :

- $V_j \subset V_{j+1}$; pour tous j , V_j est un sous espace de V_{j+1}
- $s(t) \in V_j \Leftrightarrow s(2t) \in V_{j+1}$; V_{j+1} Est l'image de V_j par une dilatation d'un facteur 2.
- $s(t) \in V_j \Leftrightarrow s(t - 2^{-j}k) \in V_j, \forall k \in R$; V_j est invariant par translation de 2^{-j} .
- $\cup_{j \in Z} V_j = L^2(R)$; la réunion des V_j est dense dans L^2 .
- $\cap_{j \in Z} V_j = \{0\} = \lim_{j \rightarrow \infty} V_j$; l'intersection de V_j est réduite à 0 dans L^2 .

Il existe une fonction d'échelle φ et V_0 telle que la suite $\{\varphi(t - k), k \in R$, soit une base de Riesz pour V_0

Les sous espaces W_j sont définis comme étant les compléments de V_j dans V_{j+1}

$$V_j \oplus W_j = V_{j+1} \tag{II.25}$$

$$V_j \cap W_j = \{0\} \tag{II.26}$$

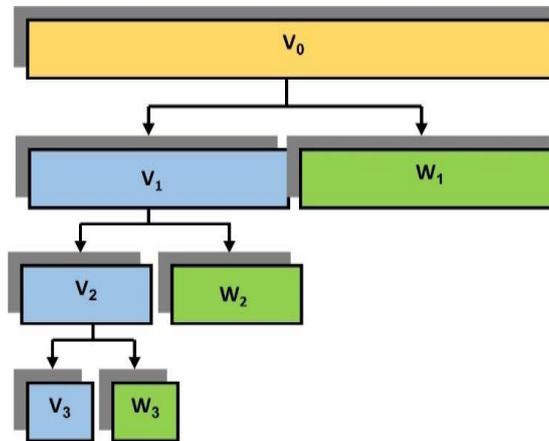


Figure II.3: la nature hiérarchique de l'analyse multirésolution.

b. Algorithmes

Deux algorithmes principaux ont été mis en évidence, l'algorithme à trous d'après Marroquin, J. L. (1985) [136] et l'algorithme de Mallat selon Burt, P. J. et al. (1983) [132] ; Farge, M et al. (1996) [133] ; Meyer, Y et al. (1992) [137]. Le premier concerne des analyses multirésolution non orthogonales, le second est pratiquement le seul utilisé dans le cas des analyses multirésolution orthogonales et biorthogonales.

➤ **Algorithme de décomposition**

L'algorithme de décomposition (analyse) est très employé en traitement du signal. Il est aussi bien utilisé pour la compression de données.

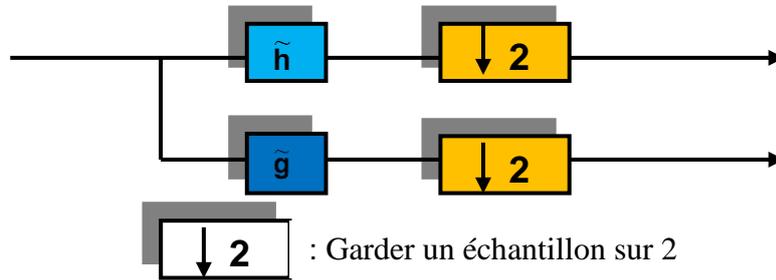


Figure II.4 : Reconstruction unidimensionnelle par ondelettes.

➤ **Algorithme de reconstruction**

Il existe une unique transformation en ondelettes discrète inverse (synthèse), telle que le signal original peut être reconstruit parfaitement à partir de ses composantes à différentes échelles. L'algorithme de reconstruction peut être schématisé de la façon suivante :

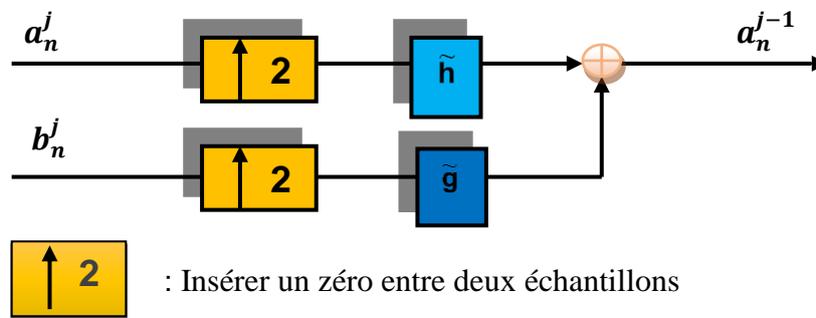


Figure II.5 : Décomposition en cascade par ondelettes.

II.3.1.6 Différentes familles d'ondelettes

Les premières ondelettes qui sont nées des travaux de Mallat, S. G. (1989) [134] et Burt, P. J et al. (1983) [132], sont les ondelettes orthogonales. Ils existent un certain nombre de familles d'ondelettes orthogonales couramment utilisées. Les plus connues sont sans doute les ondelettes de Daubechies selon Daubechies, I. (1992) [138].

Les familles d'ondelettes non orthogonales ont reçu l'appellation d'ondelettes biorthogonales. En effet, le préfixe "bi" fait référence à l'utilisation de deux bases d'ondelettes, une pour la décomposition (la base duale) et une pour la reconstruction, comme décrit par Cohen, A. (1992). [139] ; Cohen, A et al. (1992) [140]. Le fait que la fonction d'échelle et la fonction ondelette soient biorthogonales ne signifie pas nécessairement que l'analyse multirésolution ne soit pas

orthogonale. Selon Chui, C. K. (1992) [141], lorsque les fonctions d'échelle et les ondelettes biorthogonales génèrent une analyse multirésolution orthogonale, on les qualifie de semi-orthogonales.

➤ **Ondelette de Haar :**

Un premier exemple d'ondelette orthogonale est l'ondelette de Haar. La fonction d'échelle dans ce cas est $\varphi_{haar}(t) = X_{[0,1]}(t)$ et l'ondelette correspondante :

$$\psi_{haar}(t) = X_{[0, \frac{1}{2}]}(t) - X_{[\frac{1}{2}, 1]}(t) \quad (\text{II.27})$$

Dans la pratique, l'ondelette de Haar n'est pas beaucoup utilisée en raison de ses discontinuités.

➤ **Ondelette de Shannon :**

L'ondelette de Shannon donnée par la relation suivante :

$$\psi_{SH} = \frac{\sin 2\pi t - \sin \pi t}{\pi t} \quad (\text{II.28})$$

Dont la fonction d'échelle est $\varphi_{SH} = \text{sinc}(\pi t)$, n'est pas beaucoup utilisée en raison de sa très faible décroissance à l'infini. En effet, φ_{SH} est très mal localisé en temps ($\Delta \varphi_{SH} = \infty$). La raison de cette mauvaise localisation en temps vient du fait que dans le domaine fréquentiel, $\hat{\varphi}_{SH}(w)$ comporte deux discontinuités à $-\pi$ et π . En conséquence, dans le domaine temporel, la fonction décroît en $\frac{1}{t}$ et donc $\Delta \varphi_{SH} = \infty$

➤ **Ondelette de Meyer :**

Bouchereau, E. B. (1997) [131] ; Meyer, Y. (1990) [142] ont construit une fonction d'échelle de telle sorte que sa transformée de Fourier soit lisse aux endroits de discontinuités de $\hat{\varphi}_{SH}(w)$ (voir ondelette de Shannon). En temps, cela se traduit par une décroissance plus rapide à l'infini par rapport à Shannon. La fonction d'échelle et l'ondelette sont symétriques respectivement par rapport à 0 et -1/2. Les ondelettes de Meyer sont des ondelettes indéfiniment dérivables, de support infini. Leur implémentation se fait plutôt dans le domaine fréquentiel.

➤ **Ondelette de Daubechies :**

Les ondelettes de Daubechies introduites par Chui, C. K. (1992) [141] sont probablement les plus utilisées en ce qui concerne les ondelettes orthogonales. Elles sont à support compact (les filtres H et G ont une réponse impulsionnelle finie donc les filtres h_l et g_l sont finis). Ces ondelettes seront notées dbN , où db est le symbole donné pour Daubechies, et N est le nombre de moments nuls de l'ondelette. Les ondelettes de Daubechies sont supportées sur un intervalle de longueur $2N-1$. Ces ondelettes présentent l'inconvénient de ne pas être symétriques ou antisymétriques, excepté quand $N = 1$ ce qui correspond à l'ondelette de Haar.

➤ **Symlets :**

Pour obtenir une ondelette symétrique ou antisymétrique, le filtre h_l doit être symétrique ou antisymétrique par rapport au centre de son support $\hat{h}_1(w)$ (à une phase linéaire). Les symlets sont des ondelettes de Daubechies construites de telle sorte que la phase de $\hat{h}_1(w)$ soit la plus linéaire possible.

➤ **Coiflets :**

Pour une application en analyse numérique monsieur Coifman a demandé à monsieur Cohen, A et al. (1992) [140] de construire une famille d'ondelettes avec N moments nuls et un support de taille minimum, et dont la fonction d'échelle vérifie :

$$\int_{-\infty}^{+\infty} \varphi(t) dt = 1 \quad \int_{-\infty}^{+\infty} t^k \varphi(t) dt = 0 \quad \text{pour } 1 \leq k \leq N \quad (\text{II.29})$$

II.3.1.7 Critères de Choix de l'ondelette analysante

L'analyse de la sélection des ondelettes analytiques (ondelettes mères) sera basée sur des tests de plusieurs ondelettes analytiques. Cela se fera en étudiant l'erreur qui existe entre le signal d'origine (état d'origine) et le signal synthétisé (signal reconstruit). En ce sens, le paramètre d'erreur caractérise la reconstruction (ou synthèse). L'expression suivante, selon Cohen, A et al. (1992) [140], représente l'erreur qui sera calculée à chaque itération :

$$\frac{E_o}{r} = \frac{\sum_{i=1}^N |S_{oi} - S_{ri}|}{N} \quad (\text{II.30})$$

S_o : signal original ; S_{oi} : Echantillon i de S_o

S_r : signal de synthèse ; S_{ri} : échantillon i de S_r

II.3.1.8 La transformée en ondelettes par paquets :

La transformée en ondelettes par paquets est une généralisation de la décomposition en ondelettes qui offre une analyse très fine du signal original.

L'ondelette par paquets peut être utilisée pour de nombreuses applications telles que le diagnostic. Pour l'analyse du signal en ondelettes, le signal est décomposé en détail et approximation. L'approximation, on est ensuite coupé dans la même approximation et détail du deuxième niveau et le processus est répété pour la décomposition de n niveau, il y a $(n + 1)$ les moyens possibles pour se décomposer ou de coder le signal.

En utilisant l'analyse d'ondelettes par paquets, il est possible de décomposer à la fois les détails et les approximations d'un signal. On obtient ainsi 2^{2n-1} différentes décompositions, comme l'a mentionné Bazine, S. (2009) [143].

L'arbre de décomposition de l'ondelette par paquets est représenté sur la figure II.6

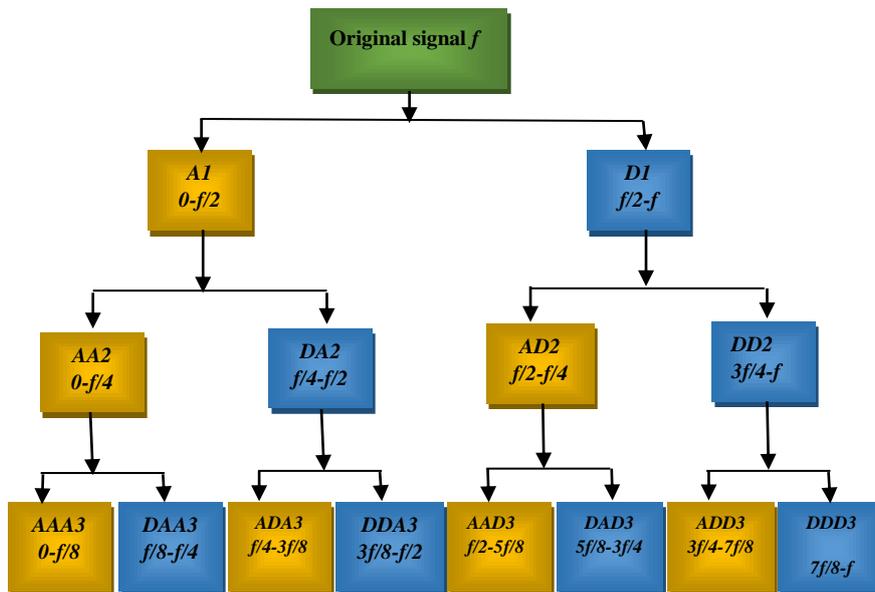


Figure II.6 Décompositions en ondelettes par paquets d'ordre 3

Une ondelette par paquets décompose un signal d'origine, stationnaire ou non stationnaire en bandes de fréquences indépendantes, aucune information redondante dans les bandes de fréquence décomposées. La méthode basée sur la multi-résolution est une approche d'analyse efficace et peut être proposée comme une méthode de diagnostic des défauts, comme souligné par Bouzida, A. (2008) [144]. La transformée en ondelettes est la génération de la transformée en ondelettes, en définissant les deux fonctions suivantes :

$$W_0(k) = \varphi(k) \quad (\text{II.31})$$

$$W_1(k) = \psi(k) \quad (\text{II.32})$$

Tel que $\varphi(n)$ et $\psi(k)$ sont la fonction d'échelle et l'ondelette mère respectivement, la décomposition d'ordre m donne la fonction $W_m(n)$ qu'est exprimée par :

$$W_{2m}(k) = 2 \sum_{n=0}^{2^m-1} h(n)W_m(2k - n) \quad (\text{II.33})$$

$$W_{2m+1}(k) = 2 \sum_{n=0}^{2^m-1} g(n)W_m(2k - n) \quad (\text{II.34})$$

$$W_{j,m,n}(k) = 2^{\frac{j}{2}} W_m(2^{-j}k - n) \quad (\text{II.35})$$

Tels que j : paramètre d'échelle, n : paramètre de localisation temporel. La différence entre la décomposition en ondelettes ordinaire et la décomposition en ondelettes par paquets est que ces derniers décomposent les détails et les approximations en composants plus fins, ce qui donne un arbre de paquets d'ondelettes.

Un exemple d'arbre de décomposition en ondelettes par paquets à trois niveaux est illustré à la Figure II.7. Chaque nœud dans l'arbre de décomposition en ondelettes par paquets sont classées par un couple d'entiers (j, k) , où j est le niveau correspondant à la décomposition et k est l'ordre de la position du nœud à un emplacement particulier noter. Dans chaque niveau j , il y a 2^j nœuds d'ordre $k = 0, 1, \dots, 2^j - 1$. Un vecteur de coefficients d'ondelettes par paquets C_{jA} correspondant à chaque nœud (j, k) . La reconstruction du signal est basée sur la somme des coefficients d'approximation et les détails à chaque niveau par la relation suivante.

$$j = 1, f(n) = A1(n) + D1(n)$$

$$j = 2, f(n) = AA2(n) + DA2(n) + AD2(n) + DD2(n)$$

$$j = 3, f(n) = AAA3(n) + DAA3(n) + ADA3(n) + DDA3(n) + AAD3(n) + DAD3(n) + ADD3(n) + DDD3(n) \quad (\text{II.36})$$

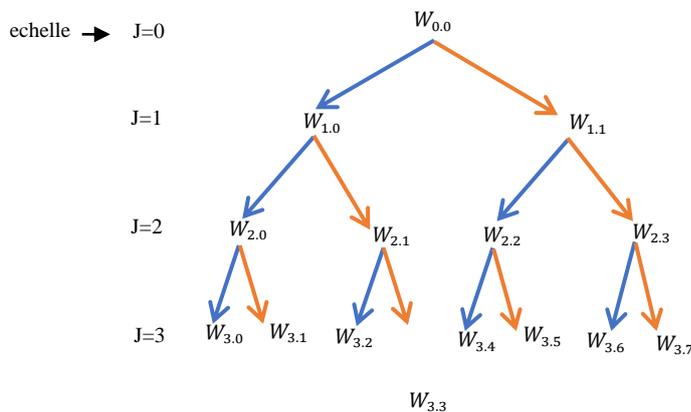


Figure II.7 Répartition des nœuds dans un arbre de décomposition

En ondelettes par paquet où les nœuds sont étiquetés par leur niveau de décomposition j et l'ordre du nœud k

II.3.1.8.1 Le vecteur caractéristique :

Le diagnostic de défaillance est basé sur l'observation et la comparaison entre les niveaux de décomposition du signal qui contiennent l'information de défaut pour des systèmes à diagnostiquer.

Lorsqu'il y a une défaillance, les informations de défaut de signal est inclus dans chaque bande de fréquence résultant de la décomposition et en calculant l'énergie associée, d'après Ye, Z et al. (2003) [145], il est possible de construire un outil de diagnostic très efficace, à l'aide de cet indice, nous pouvons explorer la distribution "d'énergie" du signal mesuré pour déterminer si le système est sain ou non.

La valeur propre d'énergie de chaque bande de fréquence est définie par :

$$E_j = \sum_{k=1}^{k=n} D_{j,k}^2(n) \quad (\text{II.37})$$

Où j est le niveau de décomposition.

Où D_j est l'amplitude en chaque point discret du coefficient d'ondelettes du signal dans la bande de fréquence correspondante, basé Sur la valeur propre de l'énergie, le vecteur est donné par :

$$T = \left[\frac{E_0}{E}, \frac{E_1}{E}, \dots, \dots, \frac{E_{2^m-1}}{E} \right] \quad (\text{II.38})$$

Tel que :

$$E = \sum_{j=0}^{2^m-1} |E_j|^2 \quad (\text{II.39})$$

Les valeurs propres d'énergie des niveaux de décomposition contiennent les signaux d'information du système à diagnostiquer, le tracé de ces valeurs peut être utilisé pour diagnostiquer un défaut, il est également possible d'identifier le degré de défaut, la déviation d'une certaine valeur indique appropriée la gravité du défaut.

II.3.1.8.2 Sélection du niveau de décomposition

Le choix optimal du nombre de niveaux de décomposition (n_{L_s}) dépend de la fréquence d'échantillonnage f_e du signal analysé, comme discuté dans les approches proposées par Antonino-Daviu et al. (2006) [146] ; Cvetkovic, Z et al. (1995) [147].

D'après Chow, T et al. (2004) [148], le nombre minimum de niveaux de décomposition nécessaire pour obtenir un signal d'approximation (A_{n_f}) de sorte que la limite supérieure de sa bande de fréquence associée soit inférieure à la fréquence fondamentale, est décrit par la condition suivante :

$$2^{-(n_{L_s}+1)} f_s < f \quad (\text{II.40})$$

A partir de cette condition, le niveau de décomposition du signal est l'entier n_{L_s} donné par :

$$n_{L_s} = \text{int} \left(\frac{\log\left(\frac{f_e}{f_s}\right)}{\log(2)} \right) \quad (\text{II.41})$$

Avec f_e : la fréquence d'échantillonnage

f_s : la fréquence du réseau

Pour cette approche, une décomposition supplémentaire de ce signal doit être effectuée afin que la bande de fréquence $[0 - f]$ soit décomposée en plusieurs bandes. Habituellement, deux

niveaux de décomposition supplémentaires (c'est-à-dire $n_{Ls} + 2$) seraient suffisants pour l'analyse.

L'équation (II.41) devient :

$$n_{Ls} + 2 = \text{int} \left(\frac{\log\left(\frac{f_e}{f_s}\right)}{\log(2)} \right) + 2 \quad (\text{II.42})$$

Conceptions du classificateur

II.4 Machines à vecteurs de support (SVM) :

Les SVM, souvent traduits par le nom de séparateurs à vaste marges (SVM), sont une classe d'algorithmes d'apprentissage définis à l'origine pour la discrimination, Hartert, L. (2010) [149] se réfère à la prédiction de variables qualitatives binaires. Elles sont ensuite généralisées à la prédiction de variables quantitatives. Dans le cas de la distinction de variables dichotomiques, elles reposent sur une recherche du meilleur hyperplan marginal qui classe ou sépare correctement les données lorsque cela est possible, tout en restant le plus éloigné possible de toutes les observations. Le principe est donc de trouver un classifieur ou une fonction discriminante dont la capacité de généralisation (qualité de prédiction) est la plus grande.

Cette approche découle directement des travaux de Cortes, C et al. (1995) [150] sur la théorie de l'apprentissage. Il se concentre sur les propriétés de généralisation (ou de prédiction) du modèle en contrôlant la complexité du modèle. La dimension de Vapnik-Chervonenkis est un indicateur du pouvoir de séparation d'une famille de fonctions associée à un modèle, qui contrôle la qualité de ses prédictions. Les SVM sont spécifiquement conçues pour incorporer une estimation avec un contrôle de la complexité, où le nombre de paramètres est lié au nombre de vecteurs supports. Une autre idée fondamentale de Vapnik dans ce développement était d'éviter de substituer l'objectif initial, c'est-à-dire le problème discriminant, par des problèmes qui finissent par devenir plus complexes, tels que l'estimation non paramétrique des densités de lois multidimensionnelles dans l'analyse discriminante.

Le principe fondamental des SVM consiste à transformer le problème de discrimination en un problème linéaire de recherche de l'hyperplan optimal. Deux idées ou astuces peuvent contribuer à atteindre cet objectif :

- La première consiste à définir un hyperplan comme solution à un problème d'optimisation sous contraintes dont la fonction objective est représentée uniquement par des produits scalaires entre vecteurs, où le nombre de contraintes "actives" ou vecteurs supports contrôle la complexité du modèle.
- Un moyen de trouver des surfaces de séparation non linéaires obtenus par l'introduction d'une fonction noyau (*Kernel*) dans le produit scalaire pour induire implicitement une transformation non linéaire des données vers un espace intermédiaire de plus grande dimension (espace des caractéristiques), D'où le nom commun de la machine du noyau ou *kernel machine*. Au niveau théorique, la fonction noyau définit un espace de Hilbert, dit autorépliatif et isométrique par une transformation non linéaire de l'espace initial, et dans lequel le problème linéaire est résolu.

Cet outil est largement utilisé dans de nombreux types d'applications et s'est avéré être un concurrent sérieux des algorithmes les plus performants (agrégation de modèles). Introduisez le noyau, en particulier pour Etant donné un problème, donnez-lui beaucoup de flexibilité pour s'adapter à des situations très diverses (reconnaissance de formes, séquences de génomes,

caractères, détection de spam, diagnostics, etc.). Il faut noter qu'algorithmiquement Régulièrement, ces algorithmes sont plus pénalisés par le nombre d'observations, c'est-à-dire le nombre de vecteurs supports potentiels, que par le nombre de variables. Cependant, une version performante de l'algorithme peut prendre en compte de grandes bases de données avec un temps de calcul acceptable.

Un ouvrage de référence sur ce sujet est celui de Schölkopf, B et al. (2002) [151] ; Cornuejols, A et al. (2002) [152].

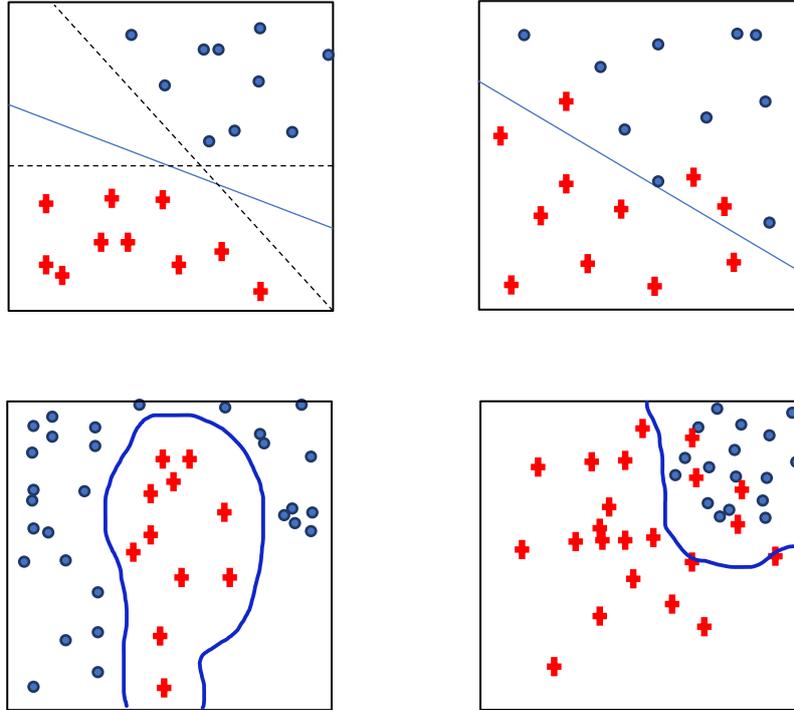


Figure II.8 Exemples de quatre problèmes de discrimination binaire impliquant la séparation des points bleus des croix rouges. Les frontières de décision sont représentées en bleu.

II.4.1. Le problème de discrimination binaire

Le problème abordé est la discrimination binaire. Le but est de trouver un moyen de construire une fonction de décision qui associe chaque observation à sa classe. On va s'occuper de ça dans un cadre probabiliste spécifique, Supposons que les formes à discriminer sont des vecteurs $x \in \mathbb{R}^p$. Le cadre probabiliste du problème repose sur l'hypothèse qu'il existe une loi inconnue $\mathbb{P}(x, y)$ sur $(\mathbb{R}^p, \{-1, 1\})$ [149]. Les problèmes discriminants visent à construire des estimateurs de fonctions de décision idéales $D : \mathbb{R}^p \rightarrow \{-1, 1\}$, minimisant la probabilité d'erreur $\mathbb{P}(D(x) \neq y|x)$ pour toutes les observations x . Pour construire cet estimateur, nous supposons qu'il existe un échantillon $\{(x_i, y_i), i = 1, n\}$ (également appelé ensemble d'apprentissage). Variables indépendantes et identiquement distribuées i.i.d de la loi parente $\mathbb{P}(x, y)$ est inconnue.

II.4.1.1 la discrimination linéaire :

Le problème de discrimination peut être dit linéairement séparable lorsqu'il existe une fonction de décision linéaire (également appelée séparateur linéaire) de la forme $D(x) = \text{signe}(f(x))$ où $f(x) = vx + a$, $v \in \mathbb{R}^p$ et $a \in \mathbb{R}$, classent correctement toutes les observations

dans l'ensemble d'apprentissage ($D(x_i) = y_i, i \in [1, n]$). La fonction f est appelée fonction caractéristique. Il s'agit d'un problème spécial qui semble très spécifique, mais il fournit une introduction pédagogique aux grands principes de SVM : marges, programmation quadratique, vecteurs support, formules duales et matrices de gram. Nous généraliserons ensuite au cas des observations inséparables et non linéaires en introduisant des variables d'écart et des noyaux. Ces différents types de problèmes sont présentés dans la figure II.8.

Pour toute fonction de décision, ainsi que pour les fonctions de décision linéaires, une frontière de décision peut être associée :

$$\Delta(v, a) = \{x \in \mathbb{R}^p | v^T x + a = 0\} \tag{II.43}$$

Comme pour les fonctions de décision linéaires, cette frontière de décision est définie comme un terme multiplicatif pré, c'est-à-dire que le couple (v, a) le même que le nombre généré par $(kv, ka) \forall k \in \mathbb{R}$. Cela est lié à la définition de l'hyperplan affine lié à la fonction caractéristique. Pour garantir l'unicité de la solution, un hyperplan standard (par exemple $\|v\| = 1$) ou un hyperplan canonique par rapport à un point x (par exemple $v^T x + a = 1$) peut être considéré.

II.4.1.2 La marge d'un classifieur :

Pour un échantillon donné, deux marges peuvent être associées au même classifieur linéaire : géométrique et numérique. La marge géométrique m d'un échantillon linéairement séparable est donnée par la distance minimale d'un point de l'échantillon à la frontière de décision. La marge numérique μ est donnée par la valeur minimale de la fonction de décision.

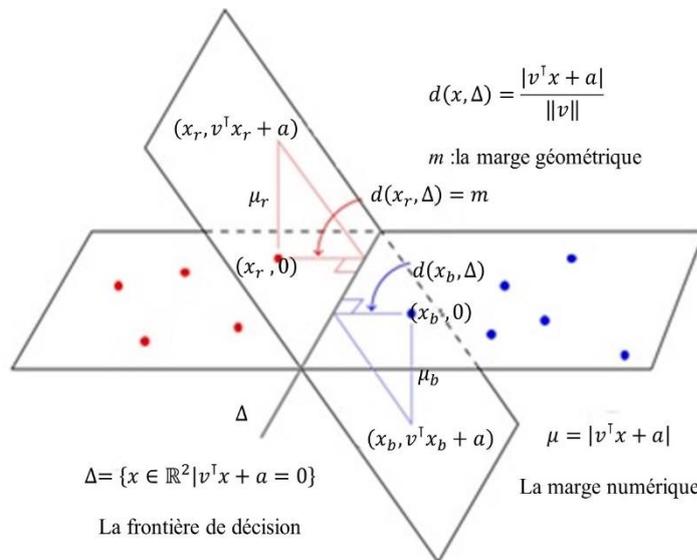


Figure II.9 Les deux notions de marges de la discrimination linéaire séparable en deux dimensions.

À un point de l'échantillon, leur définition mathématique est :

Marge géométrique :

$$m = \min_{i \in [1, n]} \text{dist}(x_i, \Delta(v, a))$$

(II.44)

Marge numérique :

$$\mu = \min_{i \in [1, n]} |v^\top x_i + a| \quad (\text{II.45})$$

La figure II.9 illustre ces deux concepts de marge pour un exemple en deux dimensions. On voit que pour une frontière de décision donnée, la marge géométrique m est fixe, tandis que la marge numérique μ dépend de la "pente" hyperplan de décision (donc $\|v\|$). En effet, pour une observation donnée, les deux marges forment les côtés adjacents d'un triangle rectangle dont l'hypoténuse est définie par la fonction caractéristique $v^\top x_i + a$.

II.4.1.3 Maximisation de la marge d'un classifieur :

Lorsque les observations sont linéairement séparables, comme la montre la figure II.8. Dans le cas général, un échantillon séparable par une infinité de frontières de décision linéaires. Le concept de marges fournit un critère de sélection dans toutes ces solutions, reconnaissant que maximiser les marges signifie également maximiser la confiance et donc minimiser la probabilité d'erreur associée au classifieur. Nous aborderons les problèmes suivants :

$$\max_{v, a} \min_{i \in [1, n]} \text{dist}(x_i, \Delta(v, a)) \quad (\text{II.46})$$

Marge : m

En introduisant explicitement les marges comme variable, ce problème est réécrit comme un problème d'optimisation sous contrainte :

$$\left\{ \begin{array}{l} \max_{v, a} \quad m \\ \text{avec } \min_{i \in [1, n]} \frac{|v^\top x_i + a|}{\|v\|} \geq m \end{array} \right. \quad (\text{II.47})$$

C'est un problème mal posé, si (v, a) est une solution, alors $(kv, ka) \forall 0 < k$ l'est aussi, Une façon de contourner cette difficulté est de changer les variables : $w = \frac{v}{m\|v\|}$ et $b = \frac{a}{m\|v\|}$ puis de réécrire le problème :

$$\left\{ \begin{array}{l} \max_{w, b} \quad \frac{1}{\|w\|} \\ \text{avec } y_i(w^\top x_i + b) \geq 1 \quad : i = 1, n \end{array} \right. \quad (\text{II.48})$$

Parce que $\|w\| = \frac{1}{m}$ Cela revient à fixer la marge numérique (celui de norme minimale) du classificateur recherché à 1. Alors la formulation "classique" pour SVM est obtenue en minimisant $\|w\|^2$ au lieu de maximiser l'inverse de la norme, Cela donne le problème suivant, qui admet la même solution que le précédent.

Alors soit $\{(x_i, y_i) : i = 1, n\}$ un ensemble de formes vectorielles étiquetées avec $x_i \in \mathbb{R}^p$ et $y_i \in \{1, -1\}$, les séparateurs à vaste marge linéaires (SVM) sont des discriminateurs linéaires de la forme : $D(x) = \text{signe}(w^\top x + b)$ où $w \in \mathbb{R}^p$ et $b \in \mathbb{R}$, Les travaux de Chendeb, M. (2006) [153] sont basés sur la résolution des problèmes suivants :

$$\left\{ \begin{array}{l} \min_{w, b} \quad \frac{1}{2} \|w\|^2 \\ \text{avec } y_i(w^\top x_i + b) \geq 1 \quad i = 1, n \end{array} \right. \quad (\text{II.49})$$

Ce problème d'optimisation sous contraintes est un programme quadratique de la forme :

$$\begin{cases} \min_z \frac{1}{2} z^T A z - d^T z \\ \text{avec } B z \leq e \end{cases} \quad (\text{II.50})$$

Où $z = (w, b)^T \in \mathbb{R}^{p+1}$, $d = (0, \dots, 0)^T \in \mathbb{R}^{p+1}$, $A = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}$, I est la matrice identité de \mathbb{R}^p , $B = -[\text{diag}(y)X, y]$, $e = -(0, \dots, 0)^T \in \mathbb{R}^n$, $y \in \mathbb{R}^n$ le vecteur des signes des observations et X une matrice de taille $n \times d$ des observations dont la $i^{\text{ème}}$ ligne est le vecteur x_i^T . Ce problème est convexe car la matrice A est semi-définie positive. Par conséquent, il admet une solution unique (qui existe parce que le problème est linéairement séparable par hypothèse) et les conditions d'optimalité nécessaires du premier ordre sont également suffisantes. Ce problème (appelé problème primal) admet une formule duale équivalente, qui est aussi un programme quadratique. La résolution du problème SVM sur des données linéairement séparables peut se faire directement (à partir de la formulation originale), par exemple en utilisant des méthodes stochastiques de type Gauss-Seidel, une méthode d'ensemble actif, un algorithme de point intérieur, de Newton avec région de confiance ou type gradient conjugué. Cependant, il est intéressant de passer par le dual de ce problème :

- Le problème dual est un programme quadratique de taille n (égal au nombre d'observations) qui peut s'avérer plus facile à résoudre que le problème original,
- La formulation duale fait apparaître la matrice de Gram XX^T , ce qui permet d'introduire la non-linéarité par le noyau dans le cas général (non-linéarité).

Pour trouver cette formulation duale, nous allons maintenant expliquer les conditions optimales du Lagrangien et de Karush, Kuhn et Tucker pour le problème. Ces conditions vont nous permettre d'introduire la notion importante de vecteurs supports.

II.4.1.4 Conditions d'optimalité et vecteurs supports :

Pour tenir compte des conditions d'optimalité nécessaires du premier ordre, l'approche classique consiste à interpréter le lagrangien d'un problème d'optimisation sous contrainte. Dans le cas de SVM, il s'écrit comme ceci :

$$\mathcal{L}(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^n \alpha_i (y_i (w^T x_i + b) - 1) \quad (\text{II.51})$$

Où $\alpha_i \geq 0$ est le multiplicateur lagrangien associé à la contrainte. Les conditions d'optimalité de Karush, Kuhn et Tucker pour la programmation quadratique associée à SVM permettent de caractériser la solution du problème primal (w^*, b^*) et le multiplicateur lagrangien α^* associé par le système d'équations suivant :

Stationnarité :

$$w^* - \sum_{i=1}^n \alpha_i^* y_i x_i = 0 \quad (\text{II.52})$$

$$\sum_{i=1}^n \alpha_i^* y_i = 0 \quad (\text{II.53})$$

Complémentarité :

$$\alpha_i^* (y_i (w^{*T} x_i + b^*) - 1) = 0 \quad i = 1, \dots, n \quad (\text{II.54})$$

Admissibilité primale :

$$y_i (w^{*T} x_i + b^*) \geq 1 \quad i = 1, \dots, n \quad (\text{II.55})$$

Admissibilité duale :

$$\alpha_i^* \geq 0 \quad i = 1, \dots, n \quad (\text{II.56})$$

Les conditions de complémentarité permettent de définir l'ensemble A des indices des

contraintes actives (ou saturées) à l'optimum dont les multiplicateurs de Lagrange $\alpha_i^* > 0$ sont strictement positifs :

$$A = \{i \in [1, n] | y_i(w^{*\top} x_i + b^*) = 1\} \quad (\text{II.57})$$

Pour les autres contraintes, la condition de complémentarité signifie que leur multiplicateur lagrangien est égal à 0, et l'observation associée vérifie strictement l'inégalité

$$\forall j \notin A, y_j(w^{*\top} x_j + b^*) > 1 \quad (\text{II.58})$$

La solution (w, b, α_A) Vérifiez le système linéaire suivant :

$$\begin{cases} w - X_A^\top D_y \alpha_A = 0 \\ -D_y X_A w - b y_A = -e_A \\ -y_A \alpha_A = 0 \end{cases} \quad (\text{II.59})$$

Où $D_y = \text{diag}(y_A)$, $\alpha_A = \alpha(A)$, $y_A = y(A)$ et $X_A = X(A; :)$. Le premier sous-système (qui est la condition de stationnarité pour w) est spécial Très important car il précise que la solution au problème s'écrit ainsi :

$$w = \sum_{i \in A} \alpha_i y_i x_i \quad (\text{II.60})$$

Dans le cas optimal, le vecteur w est une combinaison linéaire des observations x_i associées aux contraintes actives $i \in A$, où $|w^{*\top} x_i + b^*| = 1$. Ces observations sont appelées vecteurs supports car elles supportent l'hyperplan discriminant car leurs marges numériques sont égales à 1. Les autres données n'interviennent pas dans le calcul, et leur marge numérique est strictement supérieure à 1. La figure II.10 montre que le vecteur de support définit la frontière de décision. Ensuite, selon que l'on considère des variables primales ou duales, la fonction de décision peut s'écrire sous deux formes différentes :

$$D(x) = \text{sign}(f(x)) \quad (\text{II.61})$$

Avec :

$$f(x) = \sum_{j=1}^d w_j x_j + b = \sum_{i \in A} \alpha_i y_i (x^\top x_i) + b \quad (\text{II.62})$$

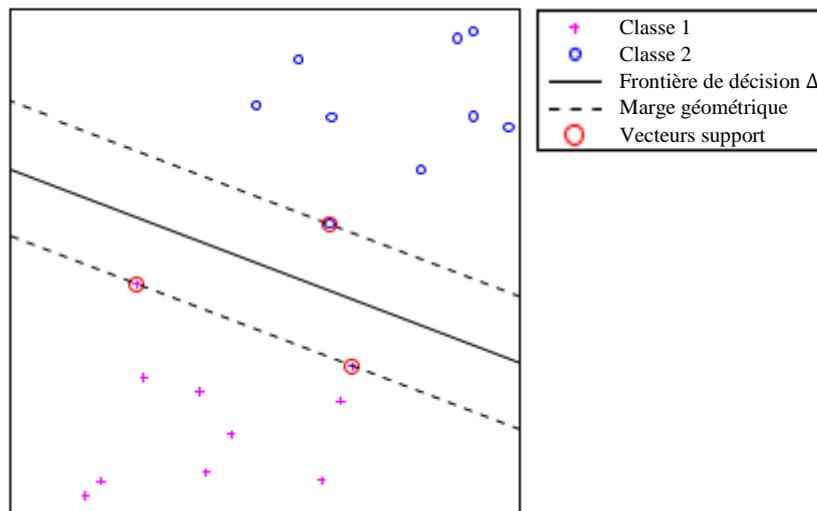


Figure II.10 Une illustration du concept de vecteur support dans le contexte de problèmes linéairement séparables. Les vecteurs de support pour les deux classes sont entourés d'un cercle rouge.

Tableau II.1 le résumé de deux scénarios possibles.

Observation inutile	Observation importante
Contrainte non saturée	Vecteur support
$i \notin A$	Contrainte active
$\alpha_i = 0$	$i \in A$
$y_i(w^{*\top}x_i + b^*) > 1$	$\alpha_i > 0$
	$y_i(w^{*\top}x_i + b^*) = 1$

Résoudre ce problème équivaut à rechercher dans les exemples disponibles ceux qui seront des vecteurs supports, ce qui nous donne l'ensemble A . Une fois cet ensemble connu, le problème peut être résolu en éliminant w en utilisant le premier sous-système, le problème devient :

$$\begin{cases} D_y X_A X_A^\top D_y \alpha_A + b y_A = e_A \\ y_A^\top \alpha_A = 0 \end{cases} \quad (\text{II.63})$$

Le cardinal de A est égal au nombre d'inconnues ($p + 1$) et la solution est :

$$b = \frac{y_A^\top (D_y X_A X_A^\top D_y)^{-1} e_A}{y_A^\top (D_y X_A X_A^\top D_y)^{-1} y_A} \quad (\text{II.64})$$

et

$$\alpha_A = (D_y X_A X_A^\top D_y)^{-1} (e - b y_A) \quad (\text{II.65})$$

Ceci permet d'utiliser le fait que la matrice $D_y X_A X_A^\top D_y$ est définie positive, l'algorithme pertinent est :

Algorithme : calcul de la solution des SVM connaissant A

$$Xa = \text{diag}(y(A)) * X(A, :)$$

$$U = \text{chol}(Xa * Xa')$$

$$a = U \setminus (U' \setminus e(A))$$

$$c = U \setminus (U' \setminus y(A))$$

$$b = (y(A)' * a) \setminus (y(A)' * c)$$

$$\text{alpha} = U \setminus (U' \setminus (e(A)' - b * y(A)))$$

$$w = Xa' * \text{alpha}$$

II.4.1.5 Le cas des données non séparables :

Dans le cas où les données ne sont pas linéairement séparables, la même approche que dans le cas séparable peut être utilisée, mais au prix de l'introduction d'une variable d'écart, l'idée est de modéliser l'erreur potentielle par une variable d'écart positive ξ_i associée à chaque observation $(x_i, y_i), i \in [1, n]$. La variable d'écart est nulle lorsque le point satisfait la contrainte marginale $y_i(w^{*\top}x_i + b^*) \geq 1$. Nous avons donc les deux cas suivants :

Pas d'erreur :

$$y_i(w^\top x_i + b) \geq 1 \Rightarrow \xi_i = 0 \quad (\text{II.66})$$

Erreur :

$$y_i(w^\top x_i + b) < 1 \Rightarrow \xi_i = 1 - y_i(w^\top x_i + b) > 0 \quad (\text{II.67})$$

Cette définition est associée à une fonction de coût appelée "coût charnière", son graphe illustré à la figure II.11 :

$$\xi_i = \max (0.1 - y_i(w^\top x_i + b)) \quad (II.68)$$

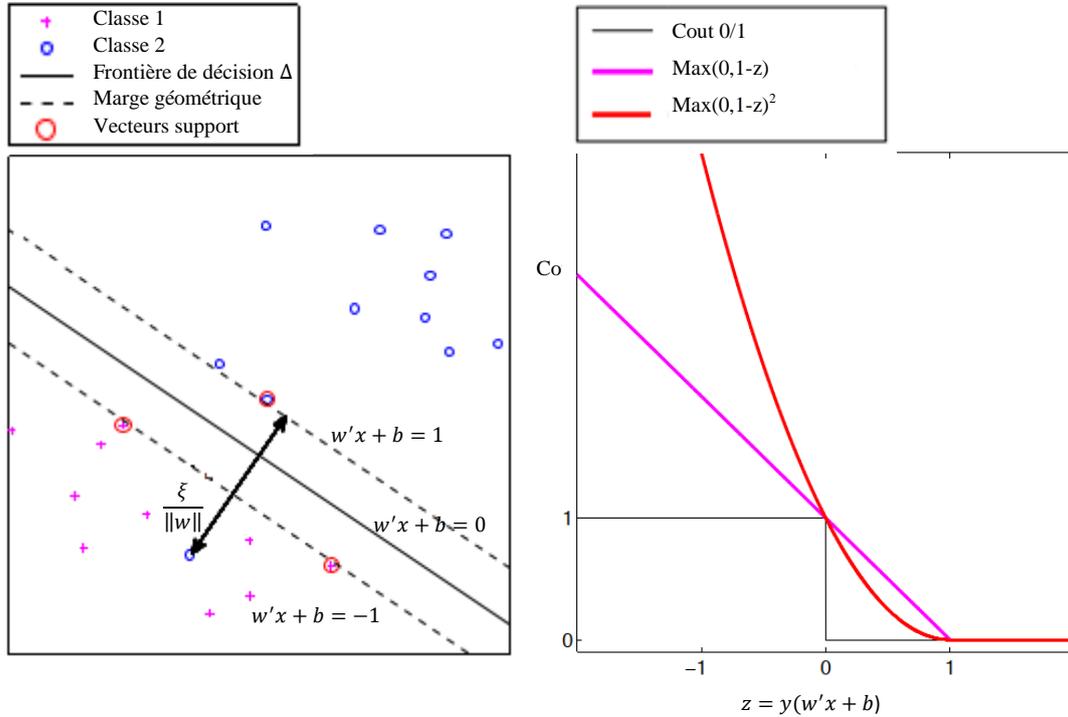


Figure II.11 Illustration du concept d'écart (à gauche) et de la fonction de coût charnière (à droite).

Dans cet exemple, tous les écarts sont nuls, à l'exception d'un point bleu mal classé. L'écart mesure la distance entre le point et la marge numérique de l'hyperplan de séparation.

II.4.1.6 Évaluation des performances des systèmes de reconnaissance de formes :

Quelle que soit la méthode utilisée pour mettre en œuvre le processus de diagnostic de reconnaissance de formes, la robustesse du classificateur choisi doit être démontrée. Ainsi que les attributs qui caractérisent chaque observation. En fait, les erreurs de diagnostic peuvent être dues à une mauvaise performance de la procédure de décision. De plus, la performance de la classification dépend (surtout dans le diagnostic) de la pertinence du vecteur forme (attribut), autrement dit la signature du système analysé. Par conséquent, Il est souvent utile d'appliquer une phase de test afin d'estimer l'erreur de classification réelle, c'est-à-dire le taux d'erreur de classement. Cela a été souligné par Zhang, N. (2010) [154] ; Casimir, R. (2003) [155].

Dans notre cas, nous tirons un échantillon X de taille N de l'ensemble de données disponible $X = \{x_1, x_2, x_3, \dots, x_N\}$ dont les classes d'appartenance sont connu. Une partie X_{ap} peut-être extraite de cet échantillon et réservée à l'apprentissage, et l'autre partie X_t est constituée de vecteurs réservés aux tests. Ces deux groupes sont disjoints. L'ensemble X_{ap} constituant la base d'apprentissage est utilisée pour construire des règles de décision, et l'ensemble de test X_t

est utilisé pour évaluer la performance des règles de décision. Pour évaluer ces performances, deux méthodes peuvent être utilisées pour sélectionner X_{ap} et X_t .

X_t Contient un seul vecteur à la fois, et X_{ap} contient tous les $(N - 1)$ autres vecteurs tel que $X_{ap} = \{X_1, X_2, \dots, X_{k-1}, X_{k+1} \dots X_N\}$, et $X_t = X_k$. Dans ce cas la règle de décision est appliquée pour chacune des $N - 1$ partitions. Les taux d'erreur de classification pour chaque observation X_t sont moyennés afin d'évaluer le taux d'erreur de la procédure de classification. C'est la méthode du leave-one-out. L'inconvénient de cette méthode est son coût calculatoire même si elle fournit une bonne estimation du taux d'erreur.

X_t Consiste en un nombre non négligeable d'observations qui ne font pas partie de l'ensemble d'apprentissage. La méthode peut être exécutée à plusieurs reprises en sélectionnant au hasard les données X_{ap} et X_t . Le taux d'erreur de classification pour chaque observation X_t est moyenné pour évaluer le taux d'erreur de la procédure de classification. Pour faciliter le calcul, un seul tirage peut être utilisé. Cependant, le taux d'erreur ou taux de classement reste proche de la valeur obtenue avec quelques tirages. C'est Pour cela, nous avons adopté cette dernière méthode. Le critère d'évaluation de la performance des règles de décision choisies pour le problème posé est défini par le taux d'observations bien classées (taux de classification) :

$$T_B(\%) = \frac{\text{Nombre d'observation bien classées}}{N_{\text{total}}} \quad (\text{II.69})$$

Ou le taux d'observations mal classées (taux d'erreur) :

$$T_M(\%) = \frac{\text{Nombre d'observation mal classées}}{N_{\text{total}}} \quad (\text{II.70})$$

Si le taux de classement est proche de 100%, alors on peut estimer que non seulement le vecteur en amont, le vecteur forme caractérise bien l'ensemble d'apprentissage, mais aussi la règle de décision est adaptée. Dans le même temps, il est souhaitable d'avoir une valeur très faible de la probabilité de mauvaise classification pendant la phase de test, afin de minimiser les erreurs de diagnostic pour les nouvelles observations. Le taux d'erreur ou de classification peut être déduit de la matrice de confusion notée C :

$$C = \begin{bmatrix} C_{11} & \dots & C_{1k} \\ \vdots & \ddots & \vdots \\ C_{k1} & \dots & C_{kk} \end{bmatrix} \quad (\text{II.71})$$

L'élément $C_{i,j}$ de la matrice de confusion C représente le nombre d'observations qui appartenait à l'origine à la classe C_i et qui ont été affectées à la classe C_j par la règle de décision. La somme des éléments hors diagonale est égale au nombre d'observations mal classées, et la somme des éléments diagonaux nous donne le taux de bonne classification. En plus du taux d'erreur ou de classification, cette matrice peut savoir quelles sont les classes confondre et des observations mal classées.

II.5 Résultats et discussions :

Dans cette étape, notre objectif est de construire une base de données de signaux à partir de notre modèle. Ces signaux correspondent aux courants statoriques de la machine asynchrone

et représentent cinq modes de fonctionnement différents : fonctionnement sain avec et sans charge, avec une barre rotorique cassée, et avec deux ou trois barres rotoriques cassées.

Après avoir généré, extrait et sélectionné les vecteurs caractéristiques dans l'étape précédente, nous obtenons une base de données. Cette base de données est ensuite divisée en deux sous-ensembles : un pour l'apprentissage et un pour les tests. Nous passons maintenant à la phase de classification, où nous utilisons les observations du sous-ensemble d'apprentissage pour définir les classes. L'objectif de la classification est de construire une règle de décision, qui sera ensuite utilisée dans l'étape de test ou de décision pour attribuer une observation inconnue à l'une des classes définies lors de la phase d'apprentissage.

Dans notre cas, nous utilisons des SVM (Support Vector Machines) non linéaires pour réaliser la classification. Ces SVM sont des algorithmes d'apprentissage supervisé puissants qui nous permettent de bien séparer les différentes classes dans notre base de données de signaux.

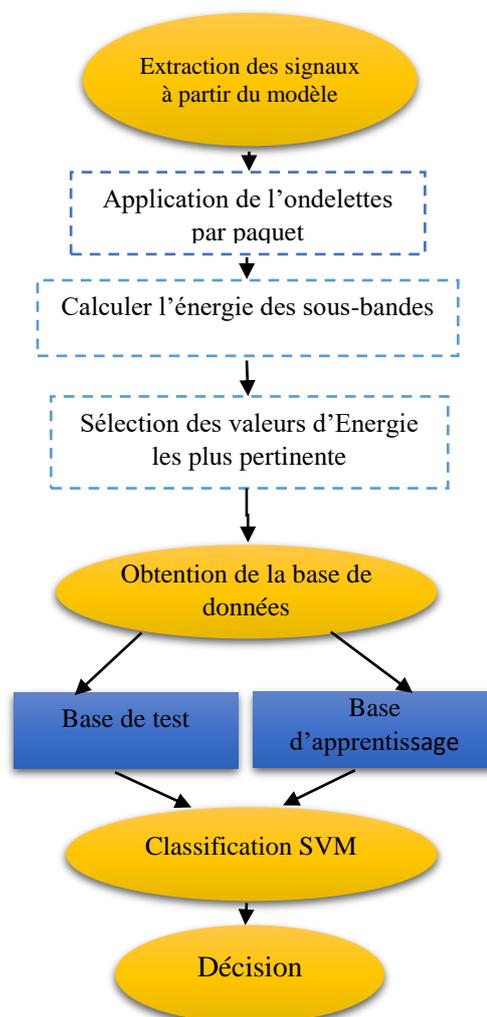


Figure II.12 : étapes pour la détection de défauts à l'aide d'ondelettes par paquets et SVM

Dans cette étude, nous utilisons un modèle de machine triphasé à induction à cage d'écureuil à l'aide du logiciel MATLAB® /Simulink, avec une puissance nominale 3 kW et les caractéristiques suivantes :

- Tension nominale 220V
- Courant nominal 5A

- Vitesse nominale 314 tr/min
- le couple résistant 3,5 n.m

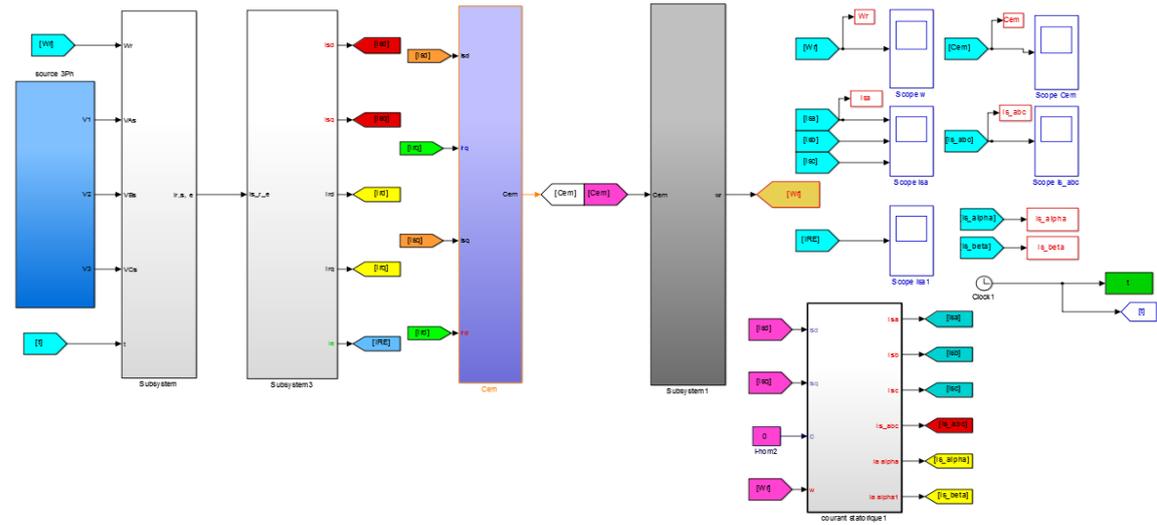


Figure II.13 : Modèle de la machine triphasé à induction à cage d'écureuil sous Matlab Simulink

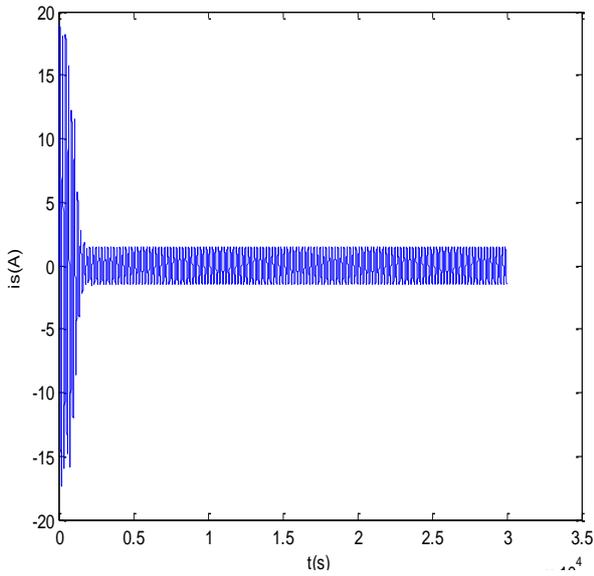


Figure II.14 Courant statorique machine saine sans charge

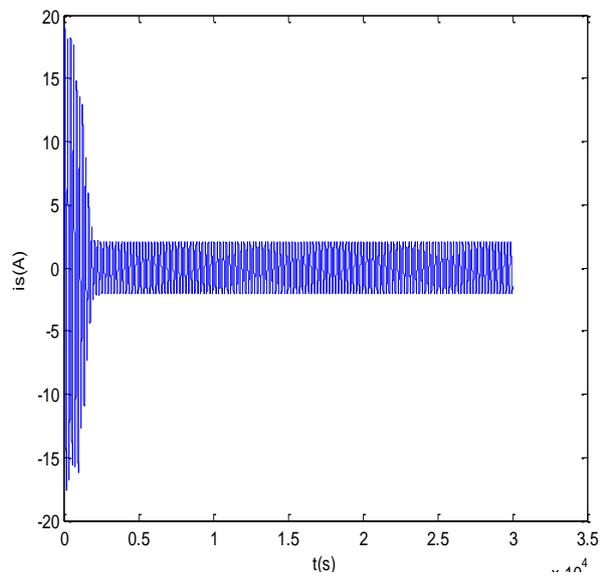


Figure II.15 Courant statorique machine saine avec charge

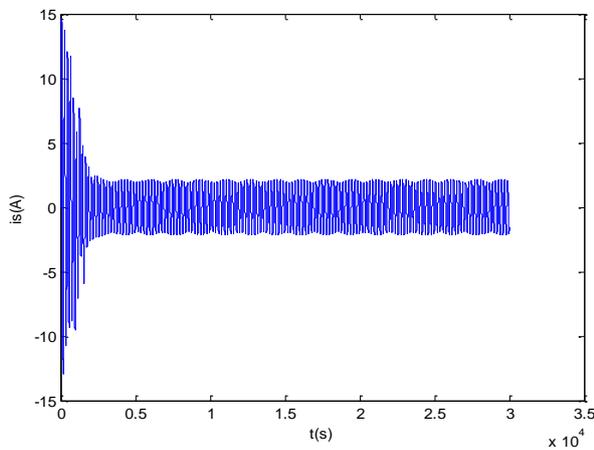


Figure II.16 Courant statorique cassure d'une barre

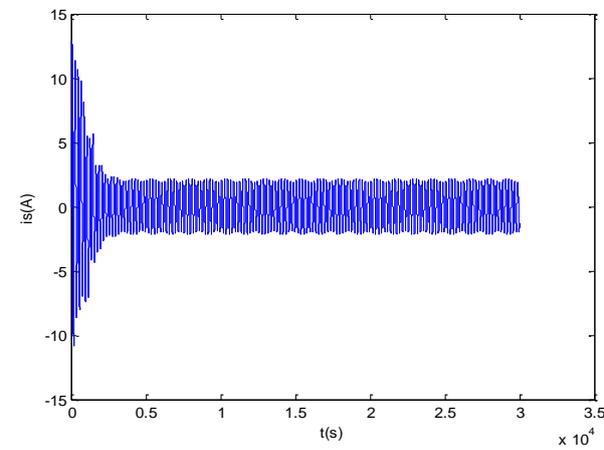


Figure II.17 Courant statorique cassure de deux barres

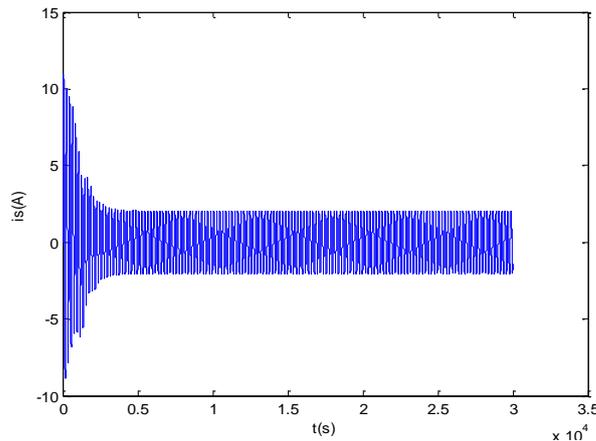


Figure II.18 Courant statorique cassure de trois barres

La décomposition en ondelettes par paquets du courant statorique est réalisée à l'aide de l'ondelette mère Daubechies44 (db44), le niveau de décomposition nécessaire est calculé d'après la relation (II.42) avec f_s la fréquence du réseau, f_e : fréquence d'échantillonnage.

Connaissant $f_s = 50 \text{ Hz}$ et $f_e = 10 \text{ KHz}$, on peut calculer le nombre de décompositions appropriés qu'est égales à :

$$n_{Ls} + 2 = \text{int} \left(\frac{\log(10^4/50)}{\log(2)} \right) + 2 = \text{int}(7.6439) + 2 = 9 \text{ niveau}$$

Tableau II.2: Bandes de fréquences résultantes de la décomposition en ondelettes par paquets.

0-10000					
0-5000					5000-10000
0-2500			2500-5000		
0-1250		1250-2500		2500-3750	3750-5000
0-625	625-1250	1250-1875	1875-2500		
0-312.5	312.5-625				

Dans les machines à induction à cage d'écureuil fonctionnant sur un réseau de 50 Hz, la plage de fréquence efficace est inférieure à 400 Hz, selon le tableau (II.3). Nous ne nous intéressons qu'à la bande de fréquence (0-312.5Hz).

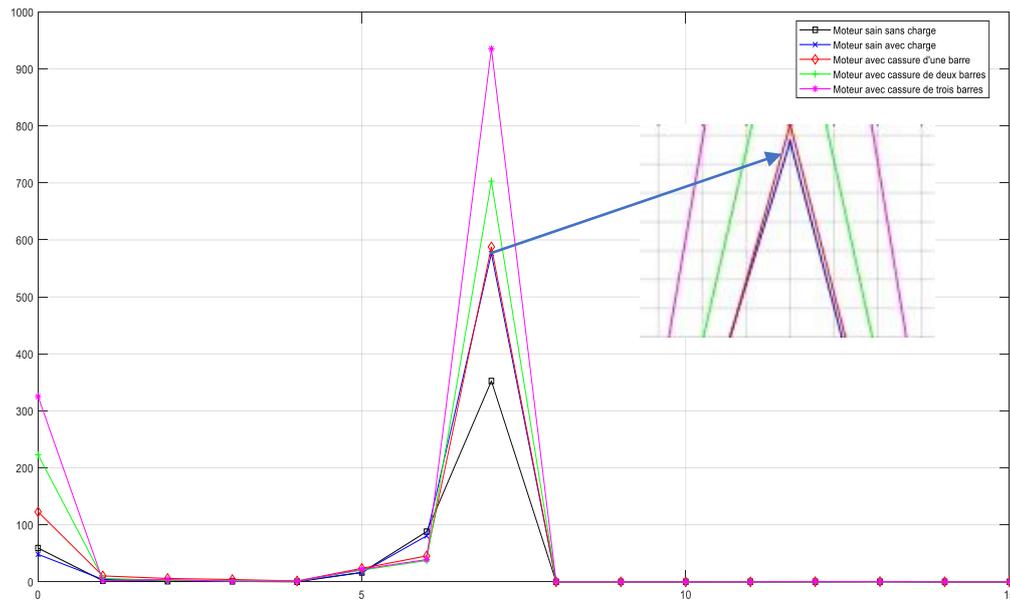


Figure II.19 : Variation d'énergie dans les bandes de fréquence au 9^{ème} niveau (db44)

□ Machine saine sans charge x Machine saine avec charge

◇ Machine avec cassure de barre ' Machine avec cassure de deux barres

* Machine avec cassure de trois barres

En se référant à la figure II.19, il est observé que l'énergie stockée dans les 16 bandes de fréquence obtenues par la décomposition au niveau 9, en particulier l'énergie de la bande de fréquence (7.9), joue un rôle crucial dans la distinction des différents types de défauts. Dans le cas d'une rupture de barre, il est constaté que l'énergie emmagasinée dans les 16 bandes de fréquence est supérieure à celle de la machine en bon état fonctionnant sous charge.

De même, en cas de rupture de deux barres, l'énergie stockée est strictement supérieure à celle en cas de rupture d'une seule barre, indiquant ainsi que l'augmentation de l'énergie stockée est proportionnelle au nombre de barres cassées. Par conséquent, l'énergie dans la bande (7.9) s'avère être le meilleur moyen pour détecter les ruptures des barres du rotor, mais il n'est pas suffisant à lui seul. Des informations supplémentaires doivent être ajoutées.

Il est remarqué que la majeure partie de l'énergie est concentrée dans les bandes ((0.9), (5.9), (6.9), (7.9)), tandis que l'énergie des autres bandes est négligeable.

Ainsi, le travail consiste à construire un vecteur caractéristique en combinant la bande (7.9) avec le reste des bandes, c'est-à-dire :

Combinaison de 1 (C1) : (7.9) tout seul

Combinaison de 2 (C2) : ((0.9), (7.9)) ; ((5.9), (7.9)) ; ((6.9), (7.9))

Combinaison de 3 (C3) : ((0.9), (5.9), (7.9)) ; ((0.9), (6.9), (7.9)) ; ((5.9), (6.9), (7.9)) ;

Combinaison de 4 (C4) : ((0.9), (5.9), (6.9), (7.9)).

Nous avons constaté que chaque vecteur décrivant un mode de fonctionnement de la machine asynchrone est représenté par une combinaison possible. Nous avons obtenu 100 échantillons pour chaque mode, ce qui équivaut à une matrice de dimensions (100 x combinaisons possibles) pour l'apprentissage et (90 x combinaisons possibles) pour le test. Le tableau ci-dessous fournit une description de la base de données.

Tableau II.5 : Description de la base de données

Les combinaisons possibles	Taille de base de données Apprentissage		Taille de base de données Test	
	Taille par mode	Taille totale	Taille par mode	Taille totale
Combinaison (C1)	100x1	500x1	90x1	450x1
Combinaison (C2)	100x2	500x2	90x2	450x2
Combinaison (C3)	100x3	500x3	90x3	450x3
Combinaison (C4)	100x4	500x4	90x4	450x4

Tableau II.6 : Résultat de classification

Classifieur		C1	C2			C3			C4
		(7.9)	((0.9), (7.9))	((5.9), (7.9))	((6.9), (7.9))	((0.9), (5.9), (7.9))	((0.9), (6.9), (7.9))	((5.9), (6.9), (7.9))	((0.9), (5.9), (6.9), (7.9))
SVM	Temps d'apprentissage (s)	1.6695	1.6825	1.6985	1.6895	1.8300	1.9400	1.9795	2.1700
	Temps de teste (s)	1.7425	1.8620	1.8595	1.8605	1.9435	2.0155	2.1007	2.3425
	Matrice de confusion	90 0 0 0 0 0 80 10 0 0 0 0 90 0 0 0 4 4 79 3 0 0 0 0 90	90 0 0 0 0 0 82 8 0 0 0 0 90 0 0 0 0 7 83 0 0 0 0 0 90	90 0 0 0 0 0 84 2 2 2 2 0 88 0 0 0 0 2 87 1 0 0 0 0 90	90 0 0 0 0 0 84 4 2 0 0 2 88 0 0 0 2 3 85 0 0 0 0 2 88	90 0 0 0 0 0 86 4 0 0 0 0 90 0 0 0 0 0 89 1 0 0 1 1 88	90 0 0 0 0 0 86 2 2 0 0 0 89 1 0 0 0 0 90 0 0 0 0 0 90	90 0 0 0 0 0 86 4 0 0 0 0 88 2 0 0 0 0 88 2 0 0 0 0 90	90 0 0 0 0 0 88 2 0 0 0 1 89 0 0 0 0 0 90 0 0 0 0 0 90
	CCR	95.33%	96.66%	97.75%	96.66%	98.77%	98.88	98.22	99.33
RNA	Temps d'apprentissage (s)	20.010	25.220	25.150	25.150	28.180	28.220	28.102	30.090
	Temps de teste (s)	21.093	25.200	25.250	25.203	27.407	27.812	27.941	30.065
	CCR	83.55%	84.66%	84.35%	82.64%	86.02%	86.05%	86.02%	88.92%

Dans le Tableau II.6, nous présentons quatre combinaisons de vecteurs d'attributs. La première combinaison, C1, se compose des énergies des sous-bandes (7.9). La deuxième combinaison, C2, inclut les vecteurs ((0.9), (7.9)), ((5.9), (7.9)), et ((6.9), (7.9)). La troisième combinaison contient les vecteurs ((0.9), (5.9), (7.9)), ((0.9), (6.9), (7.9)), et ((5.9), (6.9), (7.9)). Enfin, la dernière combinaison englobe l'ensemble ((0.9), (5.9), (6.9), (7.9)).

Les résultats de classification démontrent l'efficacité de notre contribution en termes de taux de classification, particulièrement en ajoutant des informations supplémentaires à la bande centrale (7.9).

La matrice de confusion présente les différentes classes de modes de fonctionnement de la machine. Chaque colonne correspond à une classe, allant de la première colonne représentant le moteur sain sans charge, jusqu'à la cinquième colonne représentant le moteur avec une cassure de trois barres.

Il est évident que l'utilisation de la transformation en ondelettes par paquet db44, combinée avec notre méthode de sélection du vecteur d'attributs et un classifieur SVM, permet d'obtenir d'excellents résultats pour la classification des défauts de cassure des barres de la machine asynchrone.

II.6 Conclusion

Au cours de ce chapitre, notre objectif principal a été d'identifier le vecteur d'attributs optimal adapté au diagnostic et à la classification des défauts rotoriques de la machine asynchrone. Pour cela, nous avons utilisé les machines à vecteurs de support (SVM), qui sont des algorithmes d'apprentissage couramment utilisés en apprentissage automatique pour résoudre des problèmes de classification.

Nous avons réalisé une comparaison approfondie entre plusieurs vecteurs d'attributs afin de déterminer ceux qui offriraient une bonne classification et, par conséquent, un bon diagnostic en cas de présence d'un défaut. Notre objectif était de trouver les attributs les plus pertinents et informatifs pour détecter et classer avec précision les défauts de la machine asynchrone.

En utilisant les SVM et en évaluant les performances des différents vecteurs d'attributs, nous avons cherché à identifier la combinaison optimale permettant d'obtenir une classification précise et fiable des défauts. Notre approche visait à proposer un modèle de diagnostic efficace basé sur les caractéristiques extraites de la machine asynchrone.

Chapitre III.

*Approche proposée pour l'estimation du
RUL en utilisant en utilisant
l'apprentissage profond*

III.1 Introduction :

Par rapport à la maintenance standard, le pronostic et la gestion de la santé (PHM) ou la maintenance prédictive peuvent fournir une méthode améliorée. Le PHM peut améliorer la fiabilité et la disponibilité des systèmes industriels tout en réduisant les pannes imprévues et les coûts de maintenance.

La durée de vie utile restante (RUL) est un terme utilisé dans les pronostics pour décrire le temps qui s'écoule en moyenne jusqu'à ce qu'une panne se produise. Le RUL est calculé sur la base de l'évolution des performances ou des conditions du système dans le temps, ce qui est utilisé pour prédire la dégradation et estimer le RUL (voir figure III.1).

Les méthodes de prédiction de RUL existantes peuvent être classées en trois classes, selon la littérature : les travaux de Li, L. L., Zhang et al. (2019) [156], se concentrent sur les méthodes basées sur un modèle ; les travaux de Wang, Y et al. (2014) [157], portent sur les méthodes basées sur les données ; et les travaux de Liao, L, et al.(2016) [158], se concentrent sur le développement et l'application de méthodes hybrides.

En pratique, une variété de facteurs influence le processus de développement des défaillances, ce qui entraîne un grand nombre de paramètres incertains dans la modélisation des défaillances. Pour prédire le RUL, les approches basées sur les données effectuent une analyse approfondie et une exploration des signaux de surveillance collectés du processus de dégradation du système. Elles sont moins difficiles à mettre en œuvre que les deux autres types de techniques.

Les signaux de séries chronologiques sont largement utilisés dans la prédiction de la durée de vie restante (RUL) basée sur les données, car ils offrent une description précise de la détérioration de la santé. Les travaux de Ramasso, E et al. (2012) [159], confirment cette observation. En utilisant des données de séries chronologiques, de nombreuses techniques de prédiction de RUL ont donné de bons résultats pour les pronostics, tels que les travaux de Lei, Y et al (2018) [160] ; Liu, J et al. (2016) [161]. Les réseaux de neurones artificiels (RNA) sont une méthode couramment utilisée pour estimer la durée de vie restante (RUL) des moteurs d'avion en raison de leur capacité à approximer efficacement les fonctions non linéaires. Le travail de Hornik, K. (1991) [162], met en évidence cette capacité des RNA à modéliser et à prédire des relations complexes. Les chercheurs Peel, L. (2008) [163] ; Tamilselvan, P et al. (2013) [164] ; Zhang, X et al.(2015) [165] ; Zhao, Z et al.(2017) [166] ; Ompusunggu, A et al.(2015) [167] ont utilisé la méthode de prédiction de la durée de vie restante (RUL) pour les embrayages de transmission automatique basée sur un filtre de Kalman (KF). Li, X et al.(2018) [168] ont proposé une série temporelle variable dans les signaux de surveillance acquis pour construire un réseau de neurones à convolution pour la prédiction de RUL. Khelif, R et al. (2016) [169] ont utilisé une méthode d'estimation RUL directe basée sur la régression du vecteur de support (SVR) qui évite l'estimation de l'état de dégradation et la définition du seuil de défaillance. Wu, J et al. (2017) [170] ont proposé l'utilisation de la machine d'apprentissage extrême (Extreme Learning Machine, ELM) pour la prédiction de RUL des batteries lithium-ion, qui a été optimisée à l'aide de la technique GSO. Morando, S et al. (2014) [171] ont utilisé des données de séries chronologiques pour construire un modèle de réseau d'état d'écho (Echo State Network, ESN) afin de prédire la durée de vie restante (RUL) des piles à combustible. Ces méthodes de la prédiction de RUL, qui cartographient le lien entre les signaux de surveillance et les valeurs RUL, ignorent la signification des signaux de séries chronologiques à différents moments qui reflètent des micro-changements dans l'état de santé.

En tant que solution du problème, un réseau neuronal récurrent (RNN) peut extraire des informations vitales utiles à partir de données précédemment traitées à travers des pas de temps et les intégrer dans l'état actuel de la cellule pour modéliser des données séquentielles selon les Liu, H et al. (2018) [172]. Dans son article Chandra, R. (2015) [173], a utilisé un réseau de neurones récurrents (RNN) avec une approche de convolution coopérative compétitive pour prédire des séries chronologiques chaotiques. Malhi, A et al. (2004) [174] ont proposé une approche de prédiction à long terme basée sur le RNN pour la surveillance de la santé des machines. Les RNN traditionnels, en revanche, Les travaux de Lukoševičius, M et al. (2009) [175] ; Wu, Y et al.(2018) [176], soulignent que les réseaux de neurones récurrents (RNN) peuvent être limités dans leur application en raison du problème de la disparition ou de l'explosion des gradients pendant l'apprentissage.

Heureusement, une nouvelle structure RNN appelée mémoire longue à court terme (LSTM) a été développée pour résoudre ce problème et est déjà utilisée pour prédire les données de séries chronologiques. LSTM a démontré une excellente capacité à apprendre des données robustes et sensibles en introduisant un ensemble de neurones de mémoire. Likhitha, D et al. (2019) [177] proposent l'utilisation de l'intelligence artificielle (IA) et de l'apprentissage en profondeur à l'aide de réseaux de neurones à mémoire longue à court terme (LSTM) pour prédire le RUL d'un moteur d'avion.

Des études réalisées par Li, H., Huang et al.(2020) [178], ont démontré que l'utilisation de prédicteurs basés sur plusieurs capteurs améliore la précision et la fiabilité des prédictions.

Nous suggérons dans ce chapitre un modèle (LSTM) pour l'estimation de RUL pour les moteurs à double flux et testons ses performances sur l'ensemble de données C-MAPSS.

La principale contribution de ce chapitre est résumée comme suit,

Tout d'abord, un modèle LSTM pour la prédiction de RUL est construit, avec quelques explorations pour améliorer la structure et les paramètres du réseau. Deuxièmement, la méthode proposée dans les travaux de recherche consiste en un processus d'augmentation de données visant à améliorer la précision et la robustesse de la régression. Cette approche a donné des résultats supérieurs à ceux obtenus par Likhitha, D et al. (2019) [177] ; Zheng, S et al. (2017) [179] ; Wu, J et al. (2020) [180] et Xia, M et al. (2020) [181], comme illustré dans la section des résultats.

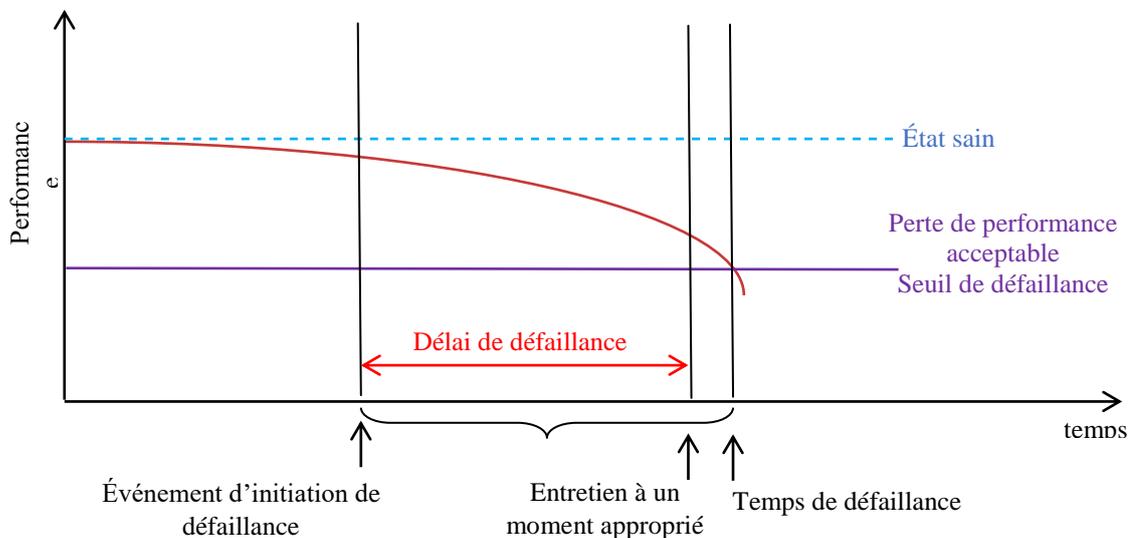


Figure III.1 : Principes de dégradation et de pronostic

III.2 Généralité :

Avant de comprendre l'apprentissage automatique (Machine Learning) et l'apprentissage profond (Deep Learning), nous devons introduire le concept d'intelligence artificielle. L'intelligence artificielle (IA) est un vaste domaine où nous essayons d'imiter le comportement humain afin de rendre les machines si puissantes pour effectuer de multiples tâches telles que la résolution de problèmes, la représentation des connaissances, la reconnaissance vocale, et d'autres. L'idée de base est de mettre des connaissances dans la machine. L'apprentissage automatique et l'apprentissage en profondeur font partie de l'intelligence artificielle. Les deux méthodes permettent aux ordinateurs de prendre des décisions intelligentes. On résume la relation entre les trois concepts d'IA, ML et DL dans la figure ci-dessous :

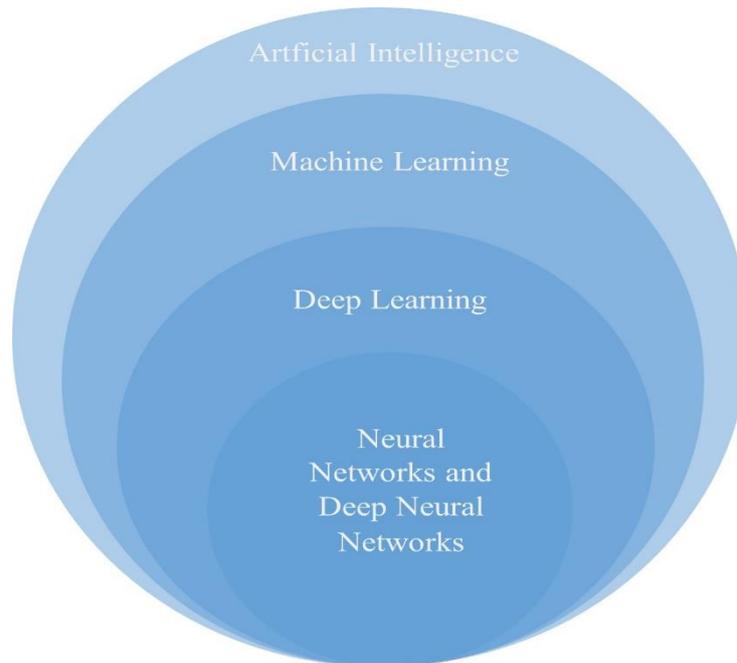


Figure III.2 : la différence entre l'intelligence artificielle, machine Learning et le Deep Learning.

III.3 Apprentissage profond :

III.3.1 Les réseaux de neurones artificiels :

Les algorithmes de l'apprentissage profond (Deep Learning) reposent essentiellement sur le mécanisme des réseaux de neurones artificiels superposés sous forme de couches successives. Ce dernier, inspiré des réseaux de neurones biologiques, peut représenter la relation entre les données dans l'espace X et l'espace de sortie Y . L'unité de calcul de base dans cette approche est le neurone, qui prend plusieurs signaux en entrée et les interprète, donne un nouveau signal aux autres neurones du réseau. Ainsi, les résultats des "neurones" de la première couche sont utilisés en entrée des calculs de la deuxième couche, et ainsi de suite. Ce principe dans lequel plusieurs couches de neurones, appelées "couches cachées", et fonctionnant à l'unisson, correspond au principe de base du Deep Learning.

Ainsi, plus formellement, nous dirions que chaque neurone a son propre état interne, interprété par la fonction d'activation. Il envoie son activation à d'autres neurones sous la forme des signaux. Les connexions entre les neurones sont établies à l'aide de liens orientés et pondérés, comme indiqué par les travaux Gurney, K. (1997) [182].

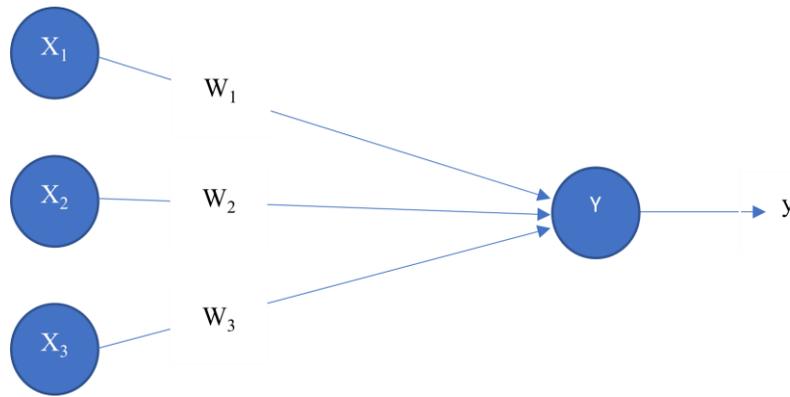


Figure III.3 : Neurone artificiel simple avec une seule sortie.

Le neurone Y reçoit X_1 , X_2 et X_3 en entrée et leur valeur de sortie est x_1 , x_2 et x_3 , les poids de liens de connexion de X_1 , X_2 et X_3 sont W_1 , W_2 et W_3 . La valeur d'entrée du neurone :

$$Y_i: y = W_1 \cdot x_1 + W_2 \cdot x_2 + W_3 \cdot x_3. \quad (\text{III.1})$$

Le signal de sortie y est déterminé par la fonction d'activation (Y).

Gurney, K. (1997) [182], souligne que les réseaux de neurones sont caractérisés par trois aspects principaux : leur architecture, leur apprentissage et leur fonction d'activation.

III.3.1.1 Architectures des réseaux de neurones

Les réseaux de neurones sont généralement divisés en deux architectures : les réseaux de neurones à un seul niveau et à multi-niveaux. Le nombre de niveaux est calculé sans tenir compte des unités.

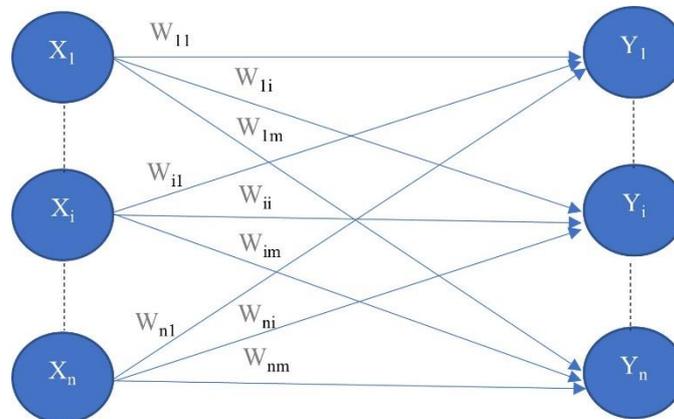


Figure III.4 : Réseau de neurones à un seul niveau. [28]

Les neurones de la couche d'entrée ne doivent transmettre et distribuer que les entrées sans effectuer de calculs. La seule véritable couche de neurones est donc celle de droite. Chaque entrée $X_1, X_2 \dots X_n$ est connectée à chaque neurone de la couche de sortie par un poids de liens. Étant donné que chaque valeur des sorties $Y_1, Y_2 \dots Y_n$ est calculée à partir du même ensemble de valeurs d'entrée, Selon Fausett, L. V. (2006) [183], chaque sortie d'un neurone dans un réseau de neurones est modifiée en fonction des poids des liens.

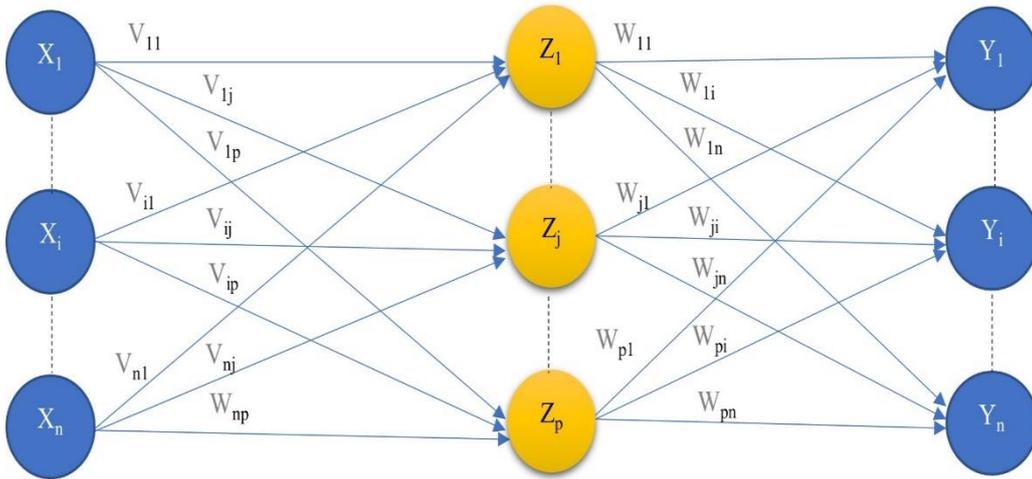


Figure III.5 : Le réseau de neurones multi-niveaux

La figure ci-dessus montre un réseau de neurones multi-niveaux, qui diffère d'un réseau à un seul niveau en ayant une ou plusieurs couches cachées. Dans cette structure, les nœuds d'entrée transmettent des informations aux unités de la première couche cachée, puis les sorties de la première couche cachée sont passées à la couche suivante, et ainsi de suite. Un réseau multi-niveau peut également être considéré comme une concaténation de groupes de réseaux à une seule couche. Le niveau de complexité se reflète dans le nombre de réseaux à une seule couche combinés dans de tels réseaux. Le concepteur d'un réseau de neurones doit tenir compte du nombre de couches cachées nécessaires, selon Goodfellow, I et al. (2016) et Maurice, B (2018), l'ajustement des poids des liens dans un réseau de neurones peut être effectué en fonction de la complexité de calcul souhaitée.

III.3.1.2 Apprentissage :

La méthode de paramétrage des poids (apprentissage) est une caractéristique importante qui distingue différents types de réseaux de neurones. Il existe deux modes d'apprentissage : Apprentissage supervisé et non supervisé.

- *Apprentissage supervisé* : dans ce type d'apprentissage, les entrées et les sorties sont fournies à l'avance. Le réseau traite ensuite l'entrée et compare son résultat avec la sortie souhaitée. Les poids sont ensuite ajustés à l'aide de l'erreur propagée dans le système. Ce processus se répète tant que les poids sont continuellement améliorés. L'ensemble de données sur lequel il est possible de s'entraîner s'appelle l'ensemble d'apprentissage.
- *Apprentissage non supervisé* : dans l'apprentissage non supervisé, le réseau reçoit l'entrée mais pas la sortie souhaitée. Le système lui-même doit alors décider quelles fonctions seront utilisées pour regrouper les données d'entrée.

III.3.1.3 Fonctionnement :

Un neurone formel gère m entrée externe (s'il s'agit du premier neurone) ou la sortie d'autres neurones (s'il s'agit des neurones intermédiaires). Le traitement donne une valeur numérique Y , le traitement consiste en quatre étapes principales :

1. La pondération : effectue un produit entre chaque entrée X_i et le poids qui lui est associé W_i .
2. La somme : une somme pondérée des coefficients synaptiques :

$$\sum_{i=1}^m (X_i * W_i) \tag{III.2}$$

3. L'ajout du biais : un biais (valeur numérique) est ajouté à la somme précédente

$$\theta + \sum_{i=1}^m (X_i * W_i) \tag{III.3}$$

4. Activation : le résultat obtenu précédemment est transformé avec une fonction d'activation φ :

$$\varphi(\theta + \sum_{i=1}^m (X_i * W_i)) \tag{III.4}$$

III.3.1.4 Fonctions d'activation :

Les fonctions d'activation sont des fonctions mathématiques utilisées au niveau de chaque neurone. Il peut être classé selon plusieurs facteurs tels que : la non-linéarité, l'étendue (plage activation), la dérivée monotone, la linéarité. Les fonctions d'activation les plus couramment utilisées sont non linéaires, selon Maurice, B (2018) [185], ce qui confère de la puissance aux réseaux de neurones est leur capacité à généraliser et à résoudre efficacement des problèmes très complexes. Le tableau suivant répertorie les fonctions d'activation les plus couramment utilisées :

Tableau III.1 : les fonctions d'activation les plus couramment utilisées

Fonction	Equation	Description
RELU	$F(X) = \begin{cases} 0 & \text{si } x < 0 \\ x & \text{si } x \geq 0 \end{cases}$	Fonction non-linéaire, très efficace, sa principale limite est la disparition du gradient.
Leaky RELU	$F(X) = \begin{cases} 0.01 & \text{si } x < 0 \\ x & \text{si } x \geq 0 \end{cases}$	Optimisation de RELU pour prévenir la perte du gradient.
ELU	$F(X) = \begin{cases} a(e^x - 1) & \text{si } x < 0 \\ x & \text{si } x \geq 0 \end{cases}$	Alternative solide a RELU, elle produit des sorties négatives, Son seul inconvénient est la Possibilité d'avoir des sorties dans $[0; \infty[$ pour $x > 0$.
Sigmoïde	$F(X) = \frac{1}{1 + e^{-x}}$	Fonction non-linéaire, très utilisée pour les problèmes de classification binaire, les sorties sont dans une plage (0 ; 1), mais elle est sensible à la disparition du gradient, et coûteuse lors de l'exécution.
Tanh	$F(X) = \frac{2}{1 + e^{-2x}} - 1$	Fonction non-linéaire, valeurs normalisées entre -1 et 1, sortie centrée sur 0, elle a les mêmes inconvénients que la fonction sigmoïde.
Softmax	$F(X)_i = \frac{e^{x_i}}{\sum_{j=1}^k e^{x_j}} / j \in \{1, \dots, k\}$	Fonction non-linéaire, généralisation de sigmoïde, valeurs normalisées entre 0 et 1, utilisée pour les problèmes de classification.

III.3.1.5 Topologie des réseaux de neurones :

Il existe deux types d'apprentissage des réseaux de neurones, selon le problème étudié :

- Avec rétro-propagation
- Sans rétro-propagation.

A. Propagation vers l'avant de l'information (Feed-forward) :

Selon Kelleher, J. D. (2019) [186], le flux de données entre les unités d'entrée et de sortie dans un réseau de neurones se déroule strictement dans une direction, de l'entrée vers la sortie. Ce principe est connu sous le nom de "strictement alimenté vers l'avant" (feedforward).

Un perceptron est un exemple typique de réseau à propagation avant. En fait, dans sa version la plus simple, un perceptron est un réseau de neurones composé d'un seul neurone qui prend n données binaires en entrée. Chacune de ses entrées i est pondérée par un poids noté w_i . Un neurone peut prendre un état "1" ou "0" (actif ou inactif, respectivement) en fonction de son entrée pondérée et d'un biais noté $\beta \in \mathbb{R}$. Ainsi le perceptron peut être représenté En tant que fonction paramétrique $f_{\theta}: \{0,1\}^n \rightarrow \{0,1\}$, où θ est son ensemble de paramètres, c'est-à-dire le biais β et les poids $w = (w_1, \dots, w_n)$. La sortie du perceptron est calculée pour le vecteur d'entrée $x \in \{0,1\}^n$ selon Nielsen, M. A. (2015) [187].

$$f(x, w) = H(z(x, w) + \beta) \tag{III.5}$$

Avec $H(t)$ la fonction de Heaviside définie pour tout $t \in \mathbb{R}$ comme $H(t) = \mathcal{J}_{\{t>0\}}$ et $z(x, w)$ la somme pondérée des entrées :

$$z(x, w) = w^T x = \sum_{j=1}^n w_j x_j \tag{III.6}$$

D'autres variantes de réseaux de neurones plus générales que les perceptrons ont émergé, comme les perceptrons multicouches. Ils sont constitués d'un grand nombre de neurones interconnectés organisés en couches successives. Ces perceptrons multicouches peuvent être représentés par un graphe acyclique, où chaque nœud représente un neurone et des arcs dirigés représentent la relation entre chaque neurone. La figure ci-dessous montre une représentation graphique de perceptron multicouche avec 3 couches ont respectivement 3, 4 et 2 neurones.

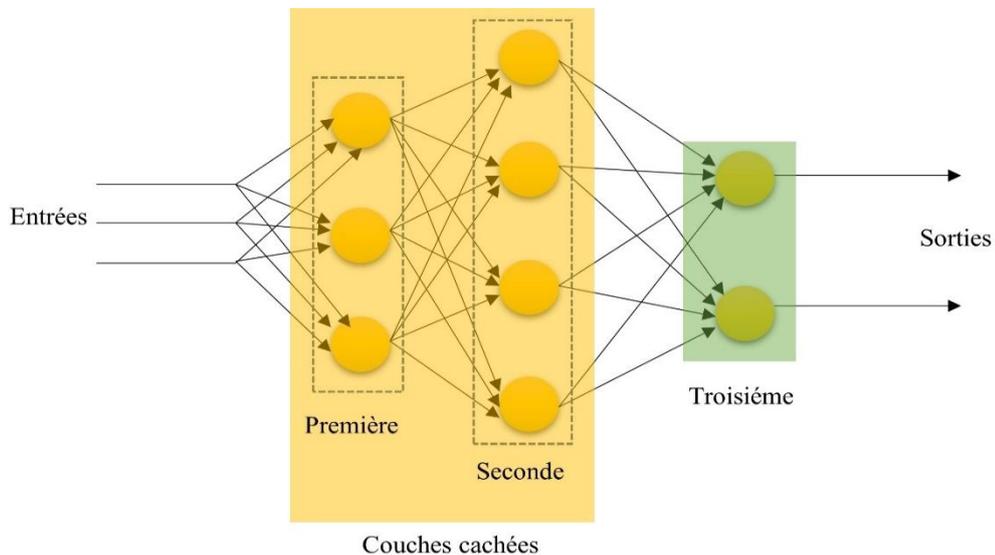


Figure III.6 : Illustration d'un perceptron multicouche selon Nielsen [187]

B. Rétro propagation de l'information (Back-propagation) :

Contrairement à la propagation vers l'avant de l'information (feed-forward), la dynamique des réseaux de neurones est importante. Dans certains cas, les valeurs d'activation des unités subissent un processus de relaxation afin que le réseau évolue vers un état stable où ces activations ne changent plus. Dans certaines applications, la variation de la valeur d'activation du neurone de sortie est importante et significative, ce qui signifie que le comportement dynamique devient la sortie du réseau, comme indiqué par Kelleher, J. D. (2019) [186].

Par conséquent, on suppose que l'objectif lors de l'apprentissage est de trouver une configuration de poids qui minimise l'erreur de prédiction. Ainsi un algorithme appelé " rétro Propagation du gradient" calcule l'erreur pour chaque neurone, de la dernière couche à la première, et tous les paramètres du réseau, l'algorithme procède en deux étapes :

1. Calculez l'erreur (sortie) en comparant la sortie au résultat attendu.
2. Propagation de l'erreur vers l'arrière

III.3.2 Modèles du Deep Learning :

Il existe plusieurs types. Ces types de réseaux sont mis en œuvre sur la base d'opérations mathématiques et d'un ensemble de paramètres nécessaires pour déterminer la sortie.

III.3.2.1 Les réseaux de neurones convolutifs (CNN) :

Les réseaux de neurones convolutifs sont conçus pour traiter des données sous la forme de tableaux de valeurs à N dimensions. Ils étaient à l'origine utilisés dans le traitement d'image, où une image couleur se compose de trois tableaux bidimensionnels contenant des intensités de pixels dans les trois canaux de couleur RVB (rouge, vert, bleu). Mais de nombreux autres types de données se présentent sous la forme de tableaux multidimensionnels, comme c'est le cas avec les séries chronologiques financières.

Il y a deux parties dans un CNN, la première partie est appelée la partie convolutive du modèle, et la deuxième partie est appelée la partie classification du modèle, qui correspond au modèle MLP (Multi layer Perceptron).

Il existe quatre types de couches dans un réseau neuronal convolutif : la couche de convolution, la couche de Pooling, la couche de correction ReLu et la couche fully connected, leur fonctionnement est comme suit :

- 1) **Couche de convolution** : La couche convolutive est l'élément central du réseau neuronal convolutif. Il forme au minimum leur première couche. Son objectif est de détecter la présence ou l'absence de caractéristiques dans l'image d'entrée. Ceci est réalisé grâce au filtrage convolutif, La méthode décrite par Aggarwal, C. C. (2018) [188], consiste à faire glisser une fenêtre représentative de la caractéristique sur l'image d'entrée et à calculer le produit de convolution entre cette caractéristique et chaque partie de l'image balayée.
- 2) **Couche de Pooling** : les couches de regroupement (Pooling) sont placées entre les couches convolutives. Il permet d'appliquer une réduction de taille à chaque carte de caractéristiques tout en préservant les caractéristiques les plus importantes (par exemple, en ne conservant que la valeur maximale). Par conséquent, le nombre de paramètres de réseau peut être réduit, réduisant ainsi le calcul nécessaire.
- 3) **Couche de correction ReLu** : La couche de correction ReLu remplace toutes les valeurs négatives reçues en entrée par des zéros. Il agit comme une fonction d'activation.

- 4) **Couche fully connected:** elle constitue toujours la dernière couche du réseau de neurones, convolutif ou non, ce n'est donc pas une caractéristique d'un CNN, elle reçoit un vecteur en entrée et produit un nouveau vecteur en sortie. Pour ce faire, il applique une combinaison linéaire aux valeurs reçues en entrée, suivie éventuellement d'une fonction d'activation.

III.3.2.2 Les réseaux de neurones récurrents (RNN)

Dans la littérature, les RNN sont principalement utilisés pour les données séquentielles, telles que les données de séries chronologiques, les données audio et vocales et le texte. Ils sont constitués de cellules RNN structurées consécutivement. Contrairement aux réseaux à propagation avant (feed-forward), selon LeCun, Y et al.(2015) [189], les réseaux de neurones récurrents (RNN) utilisent une mémoire interne pour traiter les données entrantes.

Il existe différents types de structures RNN, une à une, une-à-plusieurs, plusieurs-à-une et plusieurs-à-plusieurs (Figure III.5). En règle générale, un RNN traite la série de séquences d'entrée une par une à la fois pendant son fonctionnement. Les unités des couches suivantes (couches cachées) contiennent des informations sur l'historique des entrées dans le "vecteur d'état". Par conséquent, les informations contextuelles sont données par des connexions en boucle. Cette connexion permet de prendre en compte dans l'étape courante une ou plusieurs informations prédites à l'étape précédente.

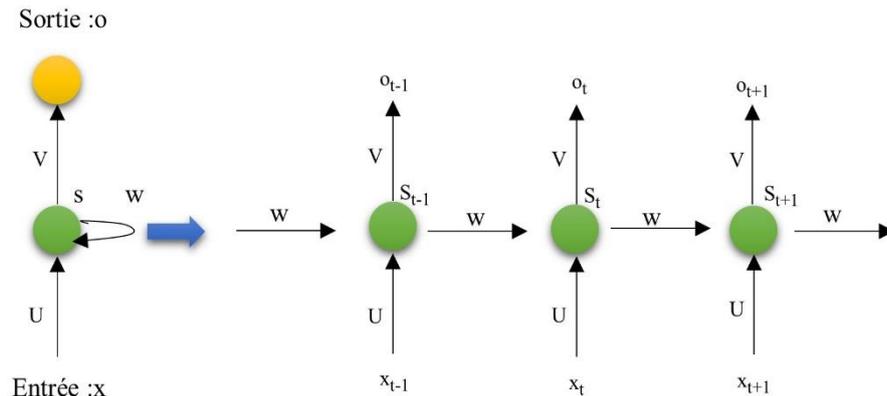


Figure III.7 : Cellule RNN à travers le temps [189].

Dans la figure ci-dessus, le flux d'informations dans la couche cachée du RNN est divisé en plusieurs laps de temps. L'état du nœud S à différents instants de t est noté s_t , la valeur d'entrée x à différents instants est x_t et la valeur de sortie o à différents instants est notée Par o_t , les valeurs des paramètres (U , W , V) sont toujours utilisées dans la même étape.

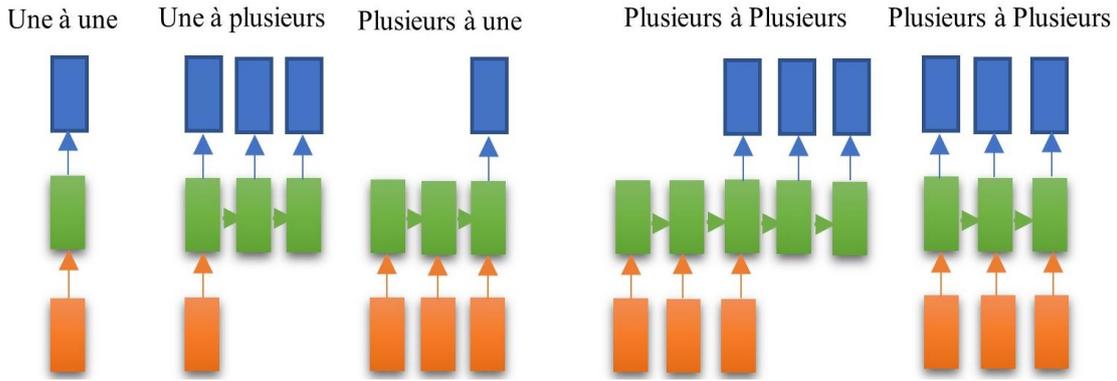


Figure III.8 : Types de séquences d'entrée pour un réseau de neurones récurrent

Cependant, les réseaux RNN introduisent des problèmes d'apprentissage lorsque les séquences à traiter deviennent trop longues. Dans les des dernières itérations du processus d'apprentissage, le gradient (utile pour tout processus d'apprentissage) peut devenir trop faible. Par conséquent, il est facile pour les RNN d'oublier les anciennes données. C'est ce qu'on le problème de la disparition de gradient (Vanishing Gradient Problem). Les réseaux LSTM tentent de résoudre ce problème.

III.3.2.3 Long Short-Term Memory (LSTM) :

Un réseau LSTM est un type spécifique de RNN couramment utilisé pour l'analyse séquentielle de données. L'avantage d'un réseau LSTM est qu'il peut mémoriser les valeurs à court et à long terme du réseau. Ils se composent d'unité LSTM, chacune constituée d'une cellule avec une porte d'entrée, de sortie et d'oubli. Ces trois portes régulent le flux d'informations. Avec ces propriétés, chaque cellule peut mémoriser la valeur souhaitée sur des intervalles de temps arbitraires. Les figures ci-dessous illustrent une cellule LSTM et une succession typique de cellules LSTM (réseau).

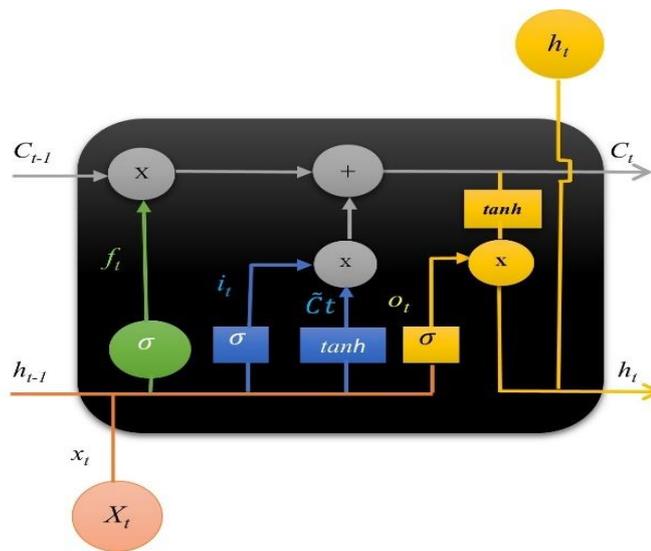


Figure III.9 : Cellules LSTM

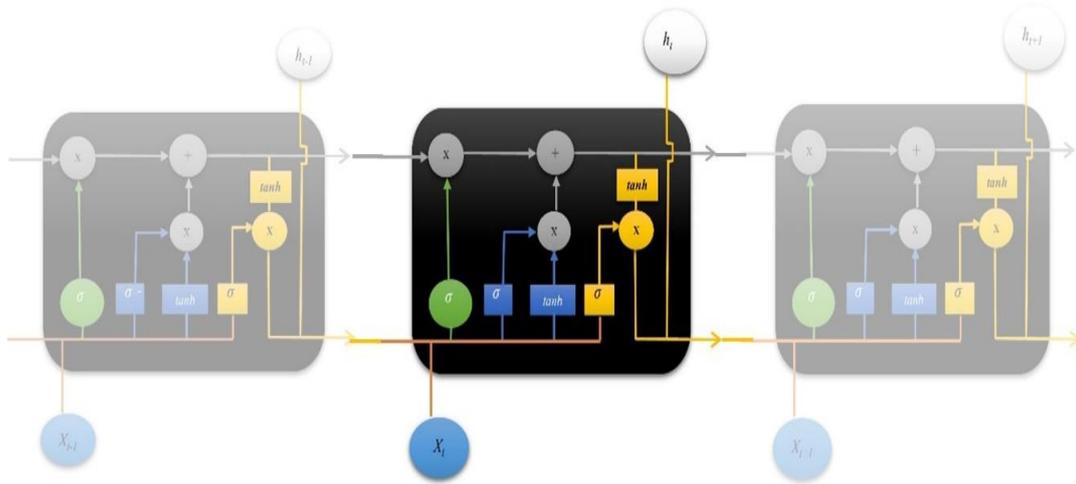


Figure III.10 : Chaîne de cellules LSTM

III.3.2.3.1 Fonctionnement du réseau LSTM

Nous allons essayer de découvrir une cellule LSTM, L'un des éléments clés de LSTM correspond à l'état de la cellule, la ligne horizontale passant par le haut du diagramme (figure III.11). Les portes est une possibilité de gérer les informations c'est-à-dire a la capacité de supprimer ou d'ajouter des informations à l'état de la cellule. Ils se composent d'une couche de réseau de neurones sigmoïde et d'une opération de multiplication ponctuelle (Figure III.12).

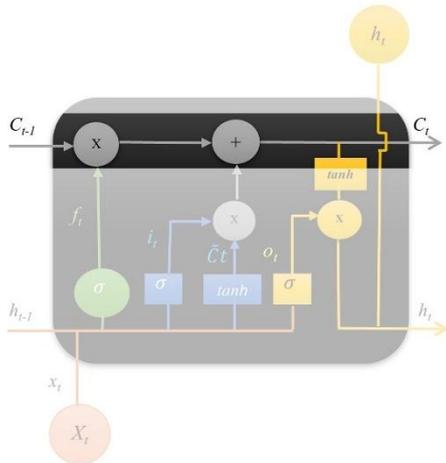


Figure III.11: état de la cellule

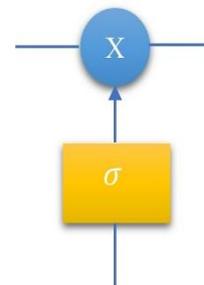


Figure III.12: Une porte LSTM

La première étape de LSTM consiste à déterminer si les informations sont conservées ou rejetées, nous concaténons les valeurs de sorties prédites dans la couche LSTM précédente avec les données entrantes, puis appliquez la fonction sigmoïde (D'après Nasekin, S et al. (2020) [190], si le résultat de la fonction d'activation d'un neurone est proche de zéro, cela indique que l'information est mémorisée, tandis que si le résultat est significativement différent de zéro, l'information risque d'être oubliée). Elle est basée sur le calcul mathématique de la formule :

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (III.7)$$

x_t : données entrantes dans le réseau de neurones.

h_{t-1} : valeur de sortie qui a été prédite dans la couche LSTM précédente.

$[h_{t-1}, x_t]$: concaténation des tableaux x_t et h_{t-1} .

W_f : poids de neurones.

b_f : biais d'un réseau de neurones.

f_t : fonction de transfert.

Pour savoir quelles valeurs C_{t-1} de la cellule précédente sont autorisées à passer, on applique la multiplication suivante :

$$C_t = f_t * C_{t-1} \tag{III.8}$$

(Si la multiplication donne une valeur proche de 0, C_{t-1} est ignorée).

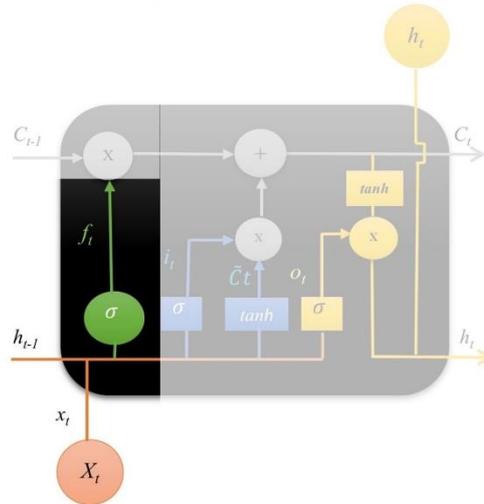


Figure III.13: porte d'oubli (Forget gate)

L'étape suivante consiste à décider quelles seront les nouvelles les informations à stocker dans l'état de la cellule, c'est-à-dire la mise à jour de l'état.

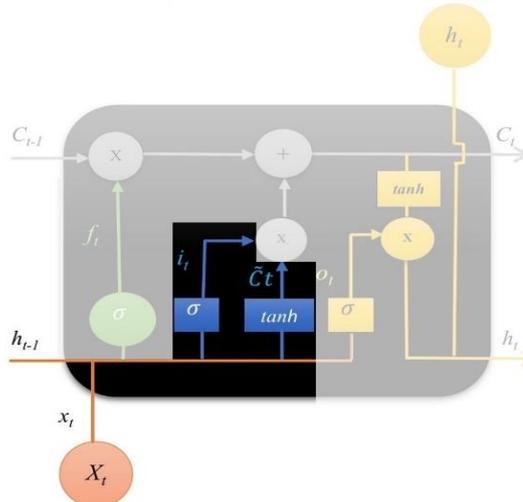


Figure III.14: Mise à jour de la cellule LSTM

Celui-ci comporte deux parties. Tout d'abord, une couche sigmoïde appelée "couche de porte d'entrée" décide des valeurs que nous mettons à jour. Ensuite, une couche tanh crée un vecteur de nouvelles valeurs candidates \tilde{C}_t , qui pourraient être ajoutées à l'état. Dans l'étape suivante, nous combinerons ces deux éléments pour créer une mise à jour de l'état.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c)$$

Avec :

\tanh : La fonction d'activation tangente hyperbolique

\tilde{C}_t : Valeur candidate

Il est maintenant temps de mettre à jour l'ancien état de cellule C_{t-1} , dans le nouvel état de cellule C_t .

Nous multiplions l'ancien état par f_t oubliant les choses que nous avons décidé d'oublier plus tôt.

Puis nous l'ajoutons $i_t * \tilde{C}_t$, il s'agit des nouvelles valeurs candidates.

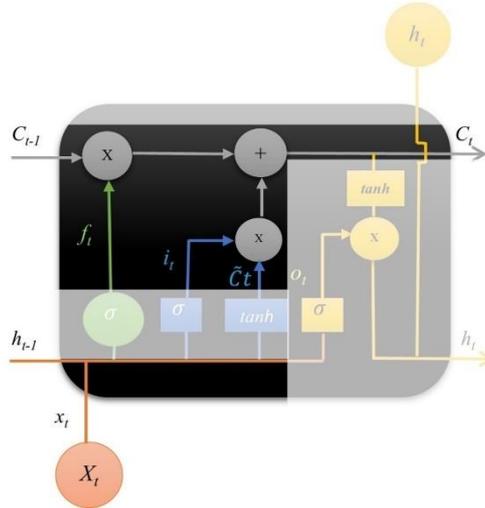


Figure III.15 : Mise à jour de l'ancien état de cellule

Sachant que :

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

Enfin, nous obtenons une sortie de l'état actuel de la cellule, la porte de sortie doit décider quel est le prochain état caché, lequel contiendra des informations sur les entrées précédentes du réseau et qui sera utilisé pour la prédiction.

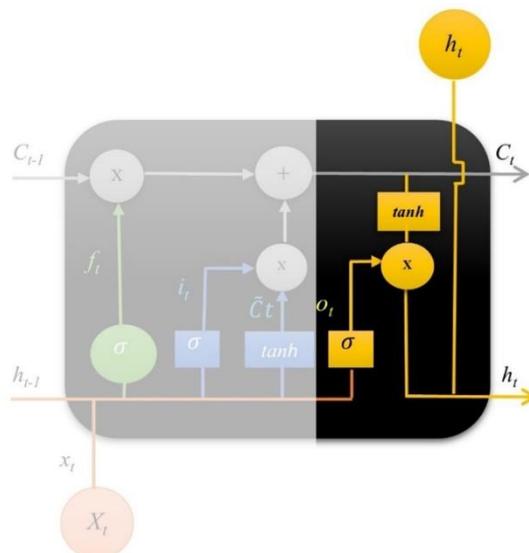


Figure III.16: la sortie de réseau LSTM

Avec

h_t : Sortie obtenue

En résumé les portes d'une cellule LSTM comme suite :

$$\text{Porte oubliée : } f_t = \sigma(W_f \cdot X_t + R_f \cdot h_{t-1} + b_f) \quad (\text{III.9})$$

$$\text{Porte d'entrée : } i_t = \sigma(W_i \cdot X_t + R_i \cdot h_{t-1} + b_i) \quad (\text{III.10})$$

$$\text{Cellule de sortie : } O_t = \sigma(W_o \cdot X_t + R_o \cdot h_{t-1} + b_o) \quad (\text{III.11})$$

$$\text{Nouvelle cellule : } \hat{C}_t = \varphi(W_c \cdot X_t + R_c \cdot h_{t-1} + b_c) \quad (\text{III.12})$$

$$\text{Mettre à jour la cellule : } C_t = f_t * C_{t-1} + i_t * \hat{C}_t \quad (\text{III.13})$$

$$\text{Sortie finale : } h_t = O_t * \phi(C_t) \quad (\text{III.14})$$

La première porte est la porte d'oubli, la seconde est la porte d'entrée et la troisième est la porte de sortie. La porte d'oubli détermine quelles informations historiques de l'état de la cellule seront rejetées, la porte d'entrée détermine quels états seront mis à jour et la porte de sortie détermine quelle partie de l'état de la cellule sera émise. Dans cette méthode, LSTM peut modifier l'état de la cellule en supprimant ou en ajoutant des informations.

(W_*, R_*, b_*) Sont les poids d'entrée, les poids récurrents et le biais pour chaque porte (* peut être i , ou o), σ est une fonction non linéaire (fonction sigmoïde dans ce cas) et "." signifie multiplication matricielle. Puis l'équation (III.12) est utilisée pour déterminer les valeurs candidates qui seront ajoutées au nouvel état (à l'instant) donné par (III.13), ainsi que les valeurs des états (à l'instant $(t - 1)$ régulés par la porte d'oubli.

Enfin, l'équation (III.14) représente les sorties finales de l'unité LSTM. φ et ϕ sont souvent des fonctions tangentes hyperboliques, mais elles peuvent aussi être d'autres fonctions non linéaires rationnelles. De plus, "*" indique la multiplication ponctuelle de deux vecteurs.

III.4 LSTM et l'estimation de RUL :

Dans cette section, nous présentons le réseau LSTM pour l'estimation RUL, y compris la structure du réseau, le flux d'informations dans les cellules LSTM et la manière dont les données du capteur sont introduites dans le réseau LSTM. Nous discuterons également de la fonction de coût pour optimiser le réseau LSTM.

III.4.1 Modèle LSTM

Supposons que nous ayons N composants du même type avec des données Run to Failure ou décrivant la fin de vie du composant. Chaque composant fait partie de l'équipement qui fournit un ensemble de séries chronologiques multivariées de données de capteur. Supposons qu'il y ait p capteurs de même type sur chaque équipement. Ensuite, les données collectées à partir de chaque équipement peuvent être représentées sous forme de matrice

$$X_n = [x_1^1, x_1^2, \dots, x_1^t, \dots, x_1^{T_n}] \in \mathbb{R}^{p \times T_n} (n = 1, \dots; N) \quad (\text{III.15})$$

Où T_n est le temps de la panne et

$$x^t = [x_1^t, \dots, x_p^t] \in \mathbb{R}^{p \times 1}, t = 1, 2, \dots, T_n \quad (\text{III.16})$$

Est le vecteur de dimension p des mesures du capteur au temps t ($t = 1, 2, \dots, T_n$) On note aussi

$$x_i = [x_i^1, x_i^2, \dots, x_i^{T_n}] \in \mathbb{R}^{p \times T} (i = 1, 2, \dots, p) \quad (\text{III.17})$$

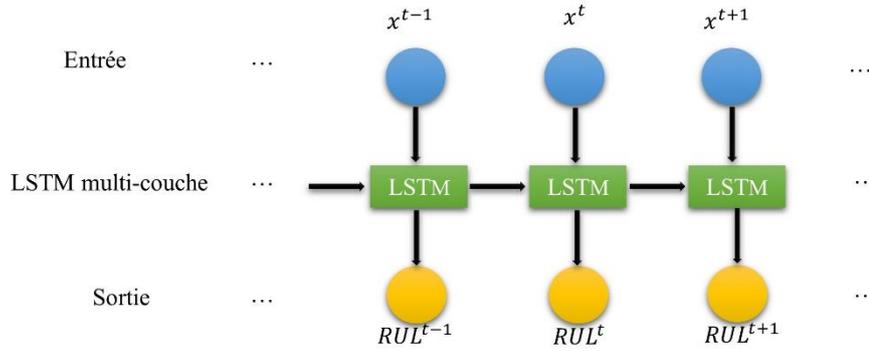


Figure III.17 : Modèle LSTM pour l'estimation RUL

Être une séquence temporelle de $i^{\text{ème}}$ mesure de capteur. Le modèle LSTM prend chaque séquence de mesures de capteur X_n et apprend comment modéliser l'ensemble de la séquence par rapport à la cible RUL.

En pratique, la séquence observée de mesures de capteur est fournie au réseau et RUL est obtenu. Notez que les séquences peuvent être de longueur différente. La structure LSTM dépliée proposée pour l'estimation RUL est illustrée à la figure III.17.

A l'instant t , le modèle LSTM prend les données du capteur p -dimensionnel x^t comme données d'entrée, et sorties estimées RUL_{est}^t .

Le modèle est composé de plusieurs couches de LSTM. Chaque couche de LSTM, qui contient une chaîne de répétition des cellules comme indiqué sur la figure III.17, est défini par la structure cellulaire et nombre de nœuds dans la cellule, les cellules répétées dans la couche LSTM ont la même structure et les mêmes valeurs de paramètre.

La structure proposée est adéquate pour l'estimation RUL en raison du fait qu'elle utilise la capacité de modélisation de complémentarité de LSTM. Plus précisément, LSTM est bon pour la modélisation temporelle et qui donne un bon résultat pour la régression.

III.4.2 Apprentissage du LSTM

L'objectif de l'apprentissage est de trouver des paramètres optimaux (poids et biais) tels que la fonction objective soit minimisée.

Soit RUL_{est}^t le RUL estimé à l'instant t . Soit RUL_{calc}^t le RUL cible à l'instant t , qui est généralement calculé comme la différence entre l'heure actuelle et la fin de vie. Le modèle LSTM proposé minimise la fonction objective suivante :

$$J = \sum_t \|(RUL_{est}^t - RUL_{calc}^t)\|^2 \quad (III.18)$$

En ce qui concerne $W_i, W_o, W_f, W_c, R_i, R_o, R_f, R_c, b_i, b_o, b_f, b_c$. Des algorithmes pour optimiser l'équation (III.18) peuvent être trouvés dans les travaux de Bengio, Y et al. (2017) [191]. Nous utilisons l'optimiseur *Adam* pour entraîner le modèle mais existe aussi Stochastic Gradient Descent (SGD), RMSprop, Adagrad, Adadelta.

Nous utilisons également le dropout et la régularisation L2 pour contrôler le sur-ajustement du modèle (overfitting). Nous nous arrêtons tôt dans le processus d'apprentissage lorsqu'il n'y a pas d'amélioration sur l'ensemble de données de validation.

De plus, pour éviter l'explosion des gradients, qui provoque un mauvais ajustement, et pour intégrer les différentes charges de travail et conditions de fonctionnement, les données fournies

au réseau LSTM doivent être préparées de manière appropriée. La quantité de données de capteur peut être très importante.

Il existe des travaux de recherche qui exploitent la réduction de dimensionnalité et le cloud computing pour traiter les données volumineuses.

Les références Zhu, Y et al. (2015) [192] ; Zheng, S et al. (2016) [193] ; Zheng, S et al. (2014) [194] ; Zheng, S et al. (2015) [195] ; Zheng, S et al. (2011) [196] ; Williams, D et al. (2014) [197] et Zhang, X et al. (2012) [198], se concentrent sur ces deux aspects, l'apprentissage en profondeur Deep Learning peut tirer parti des données volumineuses, et fonctionne mieux à l'ère du Big Data en utilisant le GPU pour le calcul parallèle.

III.5 La base de données et le prétraitement :

III.5.1 La base de données

Les jeux de données sont créés en simulant le système de propulsion aérodynamique modulaire commercial (CMAPSS) pour aborder le problème de dégradation des moteurs à turboréacteur à double flux de la NASA, étudié par Saxena, A et al. (2008) [199] ; [200].

L'ensemble de données CMAPSS contient quatre sous-ensembles de données désignés par FD001, FD002, FD003 et FD004 représentant différentes conditions de fonctionnement et modes de défaut. Chaque sous ensemble de données contient un ensemble d'apprentissage, un ensemble de test et l'ensemble RUL réel correspondant à l'ensemble de test pour la vérification. Les ensembles de données d'apprentissage comprennent des enregistrements de capteurs d'exécution jusqu'à l'échec (run to Failure) de plusieurs moteurs d'avion, ce qui signifie que le dernier enregistrement de données correspond au cycle de temps pendant lequel l'unité moteur est déclarée en panne. Les enregistrements de capteurs dans les ensembles de données de test se terminent à un certain moment avant la défaillance du système, et notre tâche vise à estimer la durée de vie utile restante de chaque moteur dans les ensembles de données de test. Il y a 26 colonnes représentant l'identification du moteur, les cycles de fonctionnement, 3 conditions de fonctionnement et 21 lectures de capteur. Chaque ligne est un instantané des données prises au cours d'un cycle d'exécution.

Le moteur à turboréacteur est constitué de divers composants (ventilateur, compresseur basse pression, compresseur haute pression, turbine basse pression, turbine haute pression, chambre de combustion etc.), et sa structure est présentée dans la Figure III.18

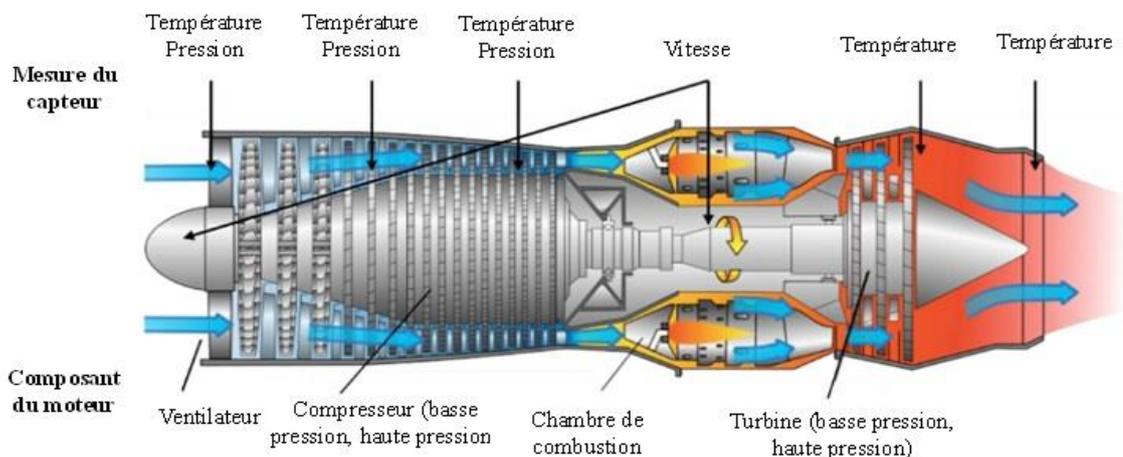


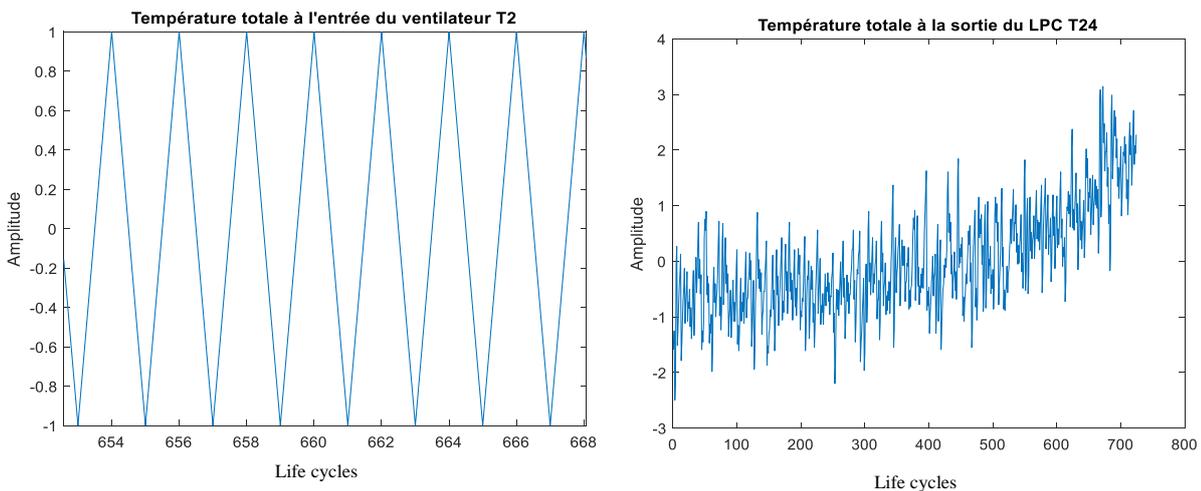
Figure III.18 : Schéma des modules du turboréacteur.

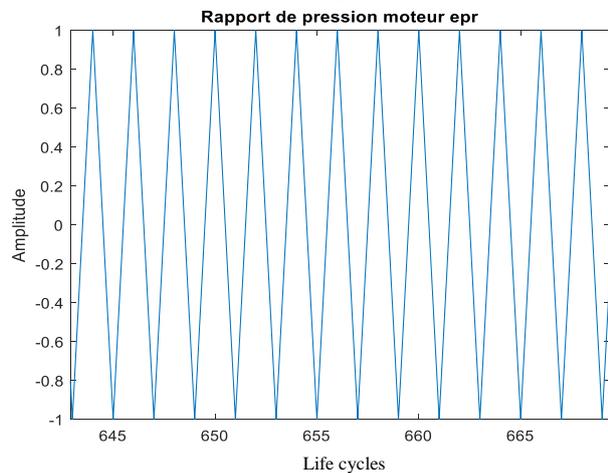
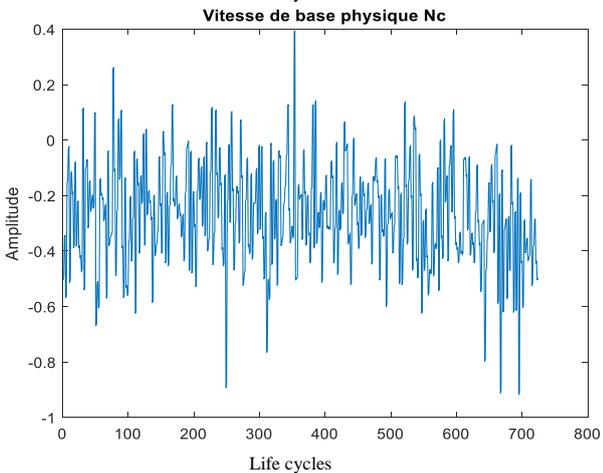
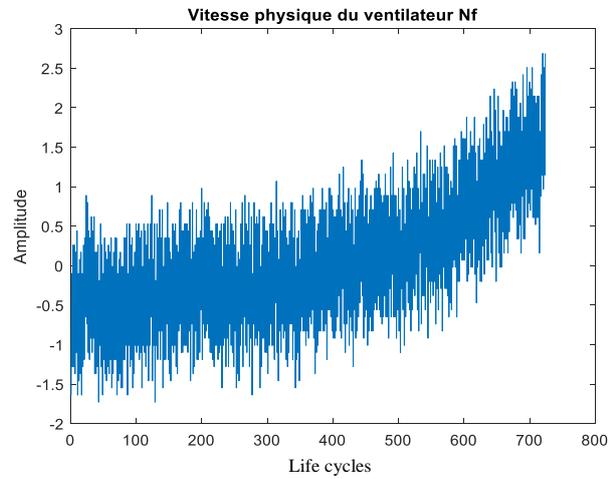
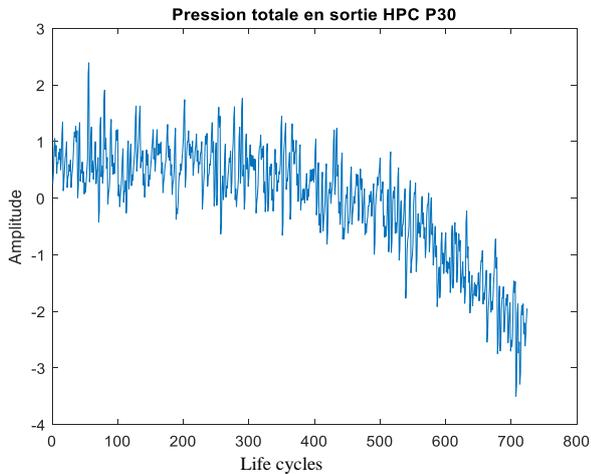
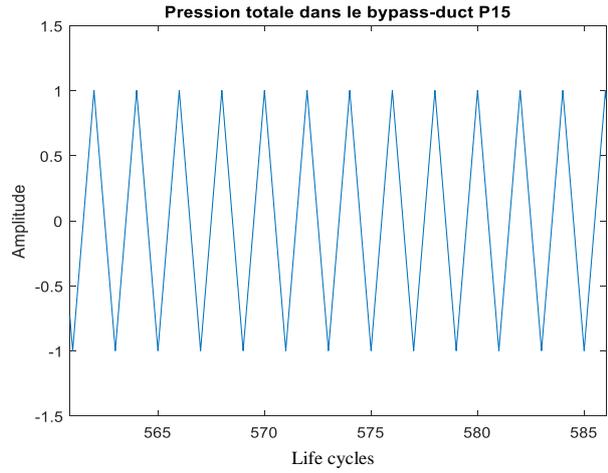
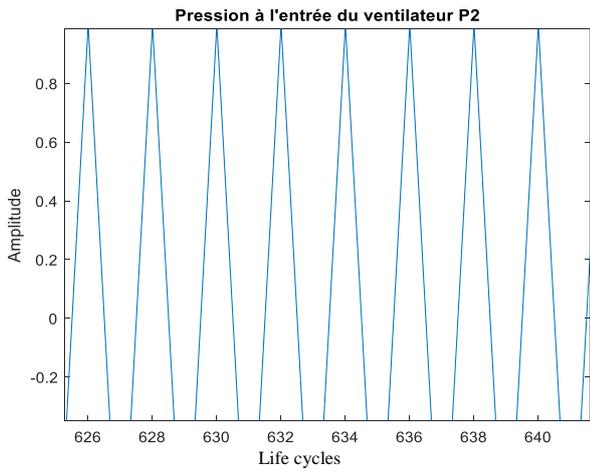
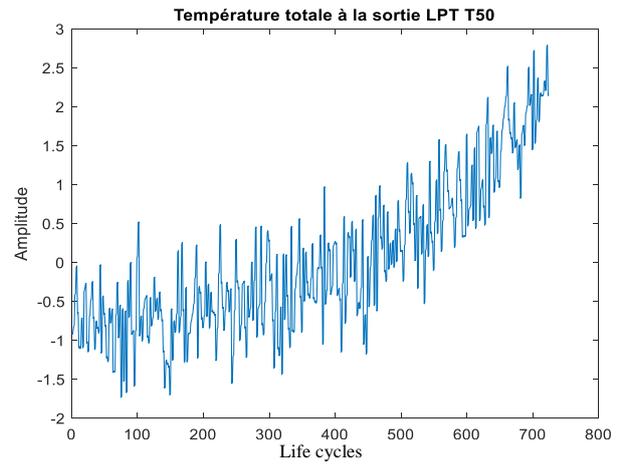
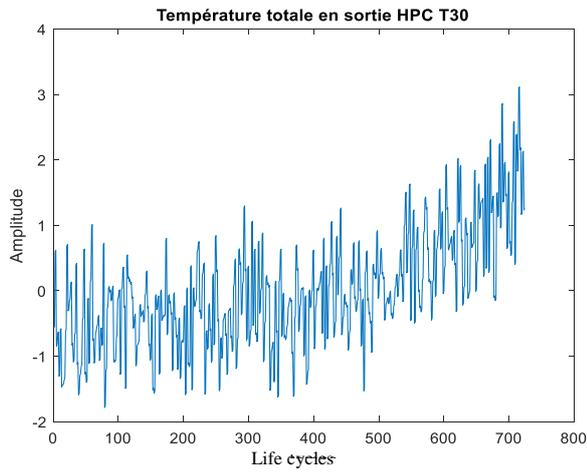
Les capteurs permettent d'acquérir les informations de chaque composant, telles que la température, la pression, la vitesse et le rapport d'air, comme présenté dans le Tableau III.2.

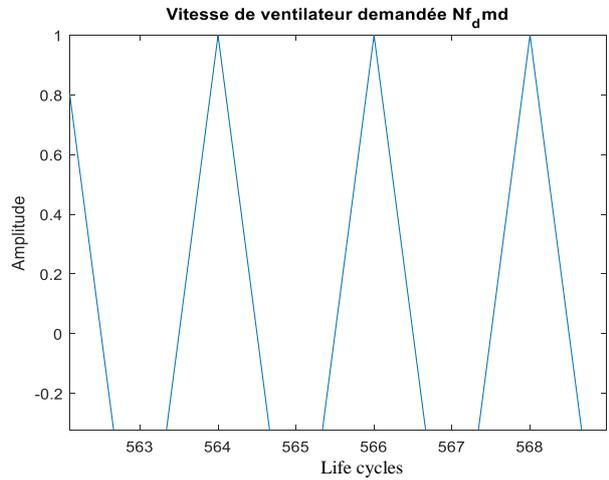
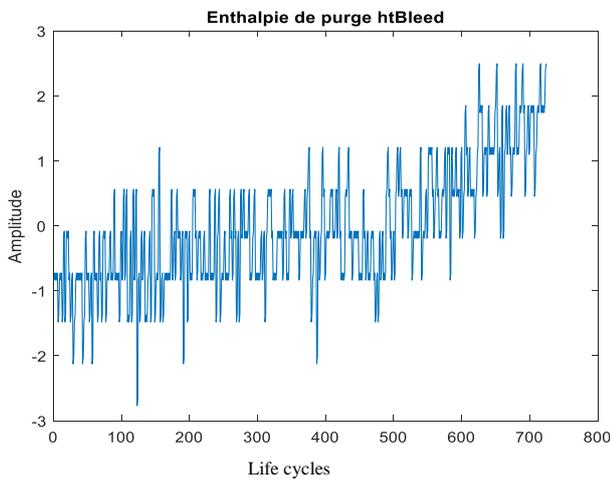
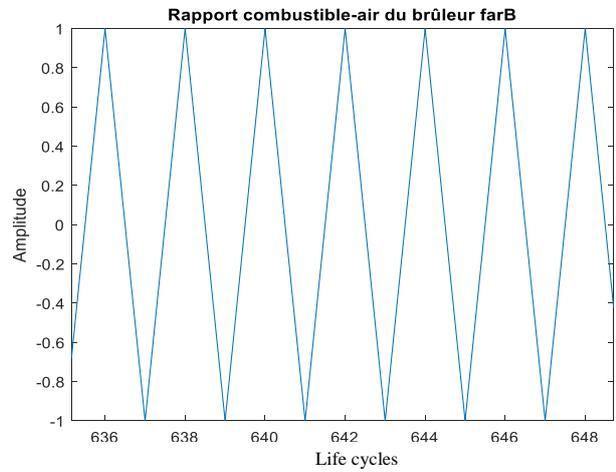
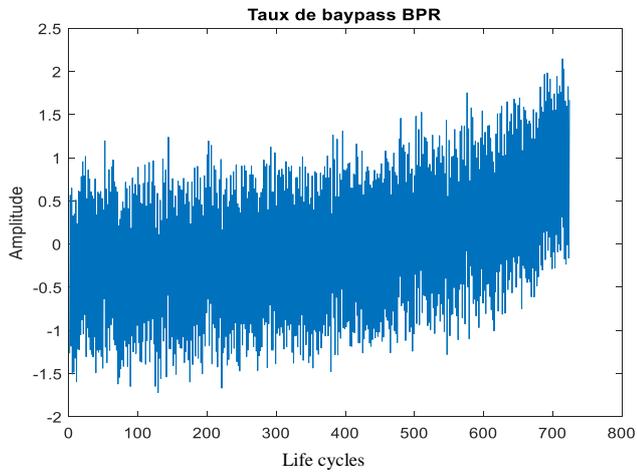
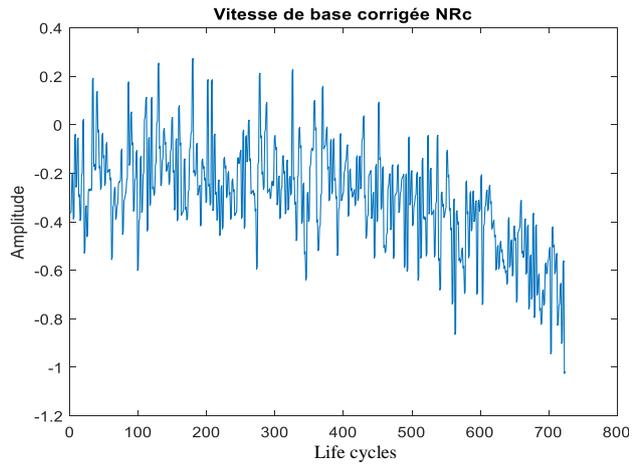
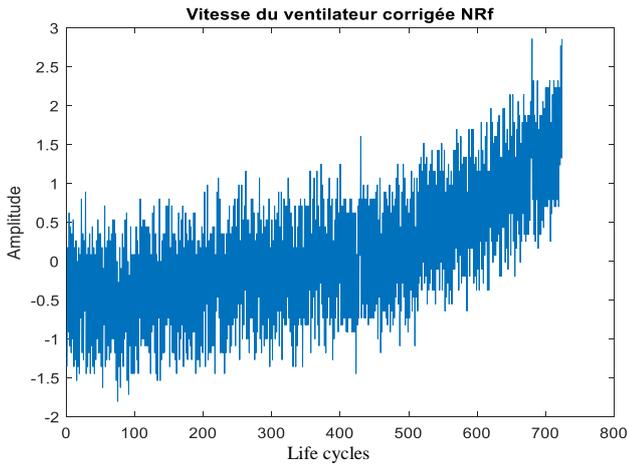
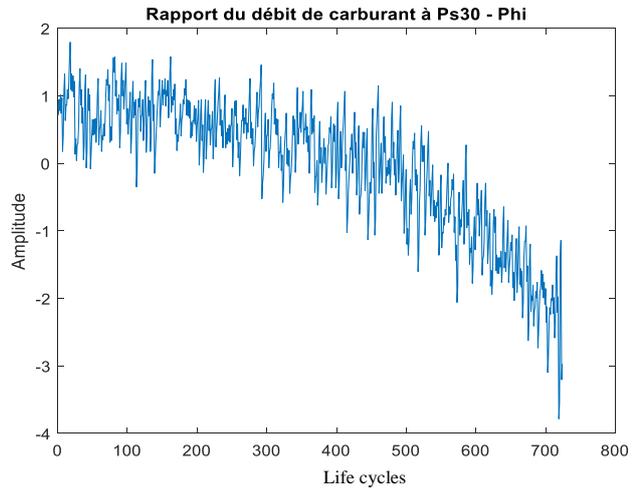
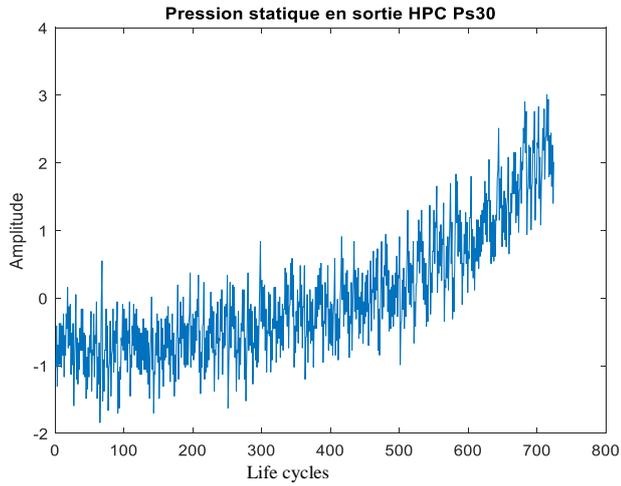
Tableau III.2. Liste et significations physiques du jeu de données C-MAPSS.

N° du capteur	Symbol	Description	Unité
1	T2	Température totale à l'entrée du ventilateur	°R
2	T24	Température totale à la sortie du LPC	°R
3	T30	Température totale en sortie HPC	°R
4	T50	Température totale à la sortie LPT	°R
5	P2	Pression à l'entrée du ventilateur	psia
6	P15	Pression totale dans le bypass-duct	psia
7	P30	Pression totale en sortie HPC	psia
8	Nf	Vitesse physique du ventilateur	rpm
9	Nc	Vitesse de base physique	rpm
10	epr	Rapport de pression moteur	-
11	Ps30	Pression statique en sortie HPC	psia
12	Phi	Rapport du débit de carburant à Ps30	pps/psi
13	NRf	Vitesse du ventilateur corrigée	rpm
14	NRc	Vitesse de base corrigée	rpm
15	BPR	Taux de baypass	-
16	farB	Rapport combustible-air du brûleur	-
17	htBleed	Enthalpie de purge	-
18	Nf_dmd	Vitesse de ventilateur demandée	rpm
19	PCNfR_dmd	Vitesse du ventilateur corrigée exigée	rpm
20	W31	Purge du liquide de refroidissement HPT	Ibm/s
21	W32	Purge du liquide de refroidissement LPT	Ibm/s

La base de données CMAPSS est illustrée dans la figure ci-dessous, montrant les lectures des 21 capteurs.







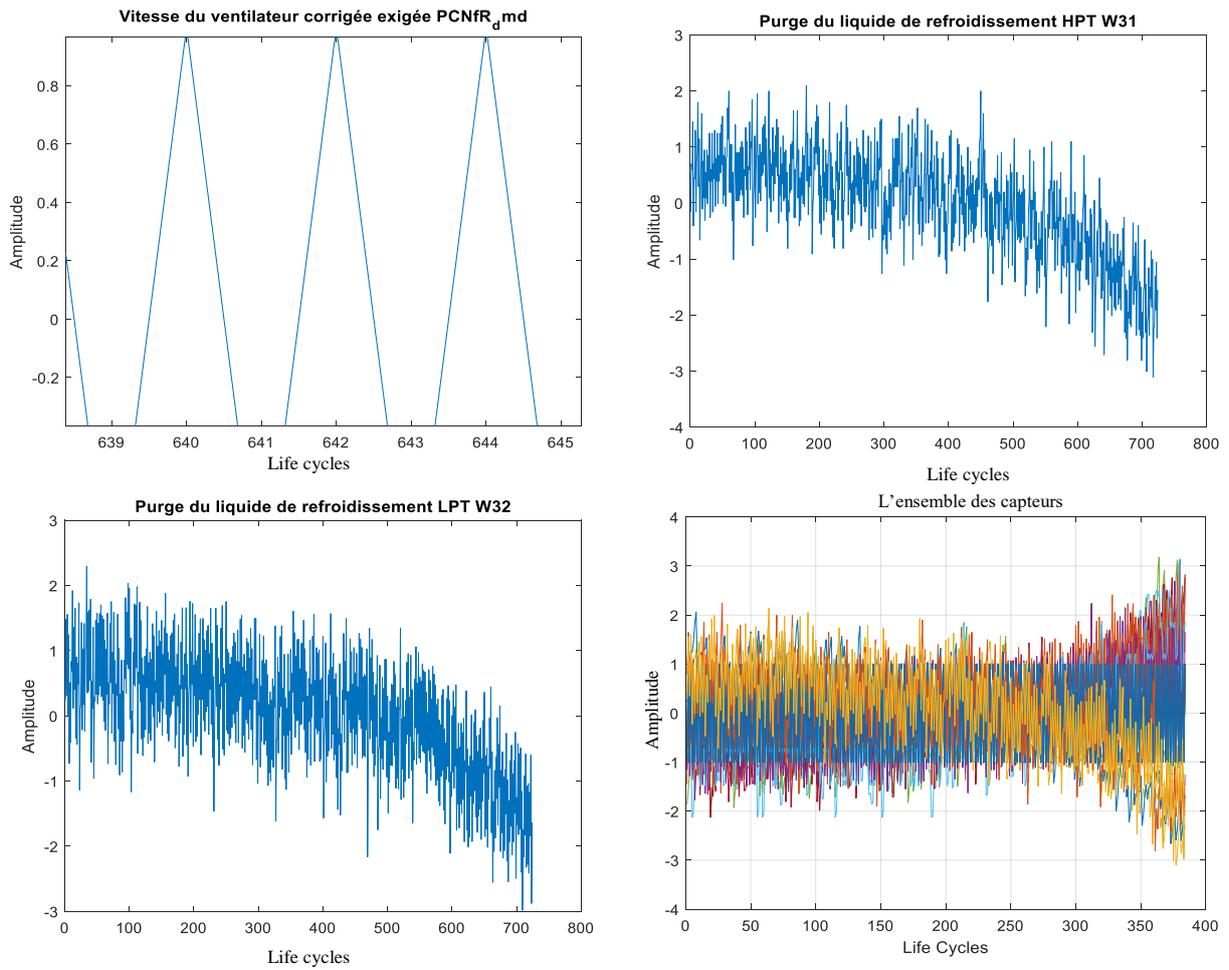


Figure III.19 lectures des 21 capteurs de base de données CMAPSS

La description détaillée de l'ensemble de données est répertoriée dans le tableau III.3.

Tableau III.3 : Description détaillée du jeu de données C-MAPSS.

Base de données		CMAPSS			
		FD001	FD002	FD003	FD004
Données d'apprentissage	Unités de moteur	100	260	100	249
	Échantillons totaux	20631	53759	24720	61249
	Cycles de vie maximum	362	378	525	543
	Cycles de vie minimaux	128	128	145	128
Données de test	Unités de moteur	100	259	100	248
	Échantillons totaux	13,096	33,991	16,596	41,214
	Cycles de vie maximum	303	367	475	486
	Cycles de vie minimaux	31	21	38	19
Conditions de fonctionnement		1	6	1	6
Modes de défaut		1	1	2	2

III.5.2 Augmentation des données comme contribution :

Le processus de génération de nouveaux échantillons en transformant les données d'apprentissage pour améliorer la précision et la robustesse des classificateurs et de la régression est connu sous le nom d'augmentation de données tel que décrit par Vachtsevanos, G et al. (2006) [46].

Nous envisageons d'utiliser des méthodes d'augmentation de données communes pour augmenter le nombre d'échantillons de la base C-MAPSS dans notre travail.

En traitement du signal, il existe deux approches de base d'augmentation des données : l'ajout de bruit et la transformation géométrique.

- *Les transformations géométriques* : L'augmentation des données par les transformations géométriques est une technique couramment utilisée pour augmenter la quantité de données d'apprentissage. Elle implique l'application de transformations géométriques aux exemples existants afin de créer de nouvelles instances de données. Ces transformations peuvent inclure la rotation, le redimensionnement, le recadrage, le décalage, le miroir horizontal/vertical, la translation, la perspective, etc. En appliquant ces transformations, on obtient des variations des exemples originaux, ce qui permet d'accroître la diversité des données d'entraînement. Cela aide les modèles d'apprentissage automatique à généraliser et à être plus robustes aux variations géométriques dans les données réelles.
- *Ajout de bruit* : L'augmentation des données par ajout de bruit est une technique utilisée pour introduire des variations aléatoires dans les exemples d'apprentissage. Ces variations aident à renforcer la capacité du modèle à généraliser et à s'adapter à des données réelles potentiellement bruitées ou imprévisibles. Voici quelques méthodes couramment utilisées pour l'ajout de bruit :
 1. *Bruit gaussien* : Le bruit gaussien est généré en ajoutant une petite quantité de valeurs aléatoires distribuées selon une distribution gaussienne (normale) aux pixels de l'image ou série temporelle. Cela crée une légère perturbation dans les valeurs données, simulant ainsi le bruit présent dans les nos données.
 2. *Floutage* : Le floutage consiste à appliquer un flou à l'image en utilisant des filtres tels que le flou gaussien ou le flou moyen. Cela peut simuler des conditions floues ou une faible résolution d'image.
 3. *Décalage de pixels* : Cette technique consiste à déplacer aléatoirement les pixels de l'image dans différentes directions. Par exemple, on peut déplacer chaque pixel d'un certain nombre de pixels horizontalement et verticalement pour créer une légère distorsion spatiale.
 4. *Altération de la luminosité/contraste* : En modifiant aléatoirement la luminosité et le contraste de l'image, on peut simuler des variations d'éclairage ou d'intensité dans les données réelles.
 5. *Compression/décompression* : L'application de compressions telles que la compression JPEG suivie de la décompression peut introduire des artefacts et du bruit dans les images, ce qui aide à augmenter la variabilité des données.

Ces techniques d'ajout de bruit sont souvent utilisées conjointement avec d'autres transformations géométriques pour augmenter davantage la diversité des exemples d'apprentissage, et à traiter des conditions réelles plus variées.

Nous considérons également la méthode d'ajout de bruit pour augmenter les échantillons de données. En théorie, nous pouvons introduire du bruit dans nos données de différentes manières (Gaussien, Poisson, Sel, Poivre, etc.).

Notre signal, d'autre part, a un grand nombre d'aléatoires et de non-stationnarité. Si nous ajoutons des bruits locaux au hasard, tels que le bruit de Poisson, le bruit de sel ou le bruit de poivre, les caractéristiques des données changeront localement.

Le bruit gaussien caractérisé par Les échantillons sont généralement considérés comme étant indépendants les uns des autres. Les valeurs du bruit blanc gaussien sont générées de manière aléatoire, ce qui signifie qu'elles ne suivent aucun modèle prévisible. Cela permet d'introduire une variabilité aléatoire dans les données d'apprentissage avec une distribution gaussienne, sur la base de ces considérations, nous concentrons nos efforts sur la génération de nouveaux échantillons d'apprentissage en ajoutant du bruit gaussien à chaque échantillon des données d'apprentissage d'origine. La fonction de densité de probabilité d'une variable aléatoire gaussienne est définie comme suit :

$$P_G(z) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(z-\mu)^2}{2\sigma^2}} \quad (\text{III.19})$$

Le niveau de gris est représenté par z , la valeur moyenne est μ , et l'écart type est σ . Dans notre travail, nous utilisons le bruit gaussien avec $\mu=0$ pour nous assurer que la valeur d'amplitude de l'échantillon n'est pas affectée par le bruit aditif. La valeur de σ est fixée à 0,5. Les ensembles de données augmentées du moteur turboréacteur simulé de la NASA sont détaillés dans le tableau III.4.

Tableau III.4 : Description détaillée de données CMAPSS après l'augmentation

Base de données		CMAPSS			
		FD001	FD002	FD003	FD004
Données d'apprentissage	Unités de moteur	100	260	100	249
	Échantillons totaux	41262	107518	49440	122498
	Cycles de vie maximum	724	756	1 050	1 086
	Cycles de vie minimaux	256	256	290	256
Données de test	Unités de moteur	100	259	100	248
	Échantillons totaux	13,096	33,991	16,596	41,214
	Cycles de vie maximum	303	367	475	486
	Cycles de vie minimaux	31	21	38	19
Conditions de fonctionnement		1	6	1	6
Modes de défaut		1	1	2	2

III.5.3 Prétraitement des données :

Pour les données temporelles multivariées, l'étape la plus importante est la normalisation des données. Les variables ont plusieurs unités et ordres de grandeur. De grandes différences entre les multi-variables mettront en évidence la plus grande tout en affaiblissant la plus petite, rendant la convergence difficile. Par conséquent, la première étape est la normalisation des données, qui supprime l'effet de la dimension sur le résultat final et rend différentes variables comparables.

Pour normaliser les données du capteur, nous utilisons la moyenne et la variance.

$$X' = \frac{x - \mu}{\sigma} \quad (III.20)$$

Où X' est la donnée normalisée des données brutes X , μ et σ dénote la moyenne et la variance de X .

Dans la figure III.20, plusieurs lectures de capteurs, telles que le capteur 8,6 et le capteur 10 du moteur 177 dans sous ensemble FD002, On observe que la lecture du capteur 8 reste constante tout au long de sa durée de vie, tandis que d'autres capteurs présentent des tendances, avec une tendance à la hausse pour le capteur 6 et une tendance à la baisse pour le capteur 10. La constance du capteur 8 indique que ses lectures ne varient pas, indépendamment de son état de santé, ce qui le rend inutile pour l'estimation et le calcul. Par conséquent, nous classons les capteurs en trois types en fonction de leurs tendances, comme illustré dans la figure III.20, et nous utilisons uniquement les capteurs à tendance ascendante et descendante pour nos estimations.

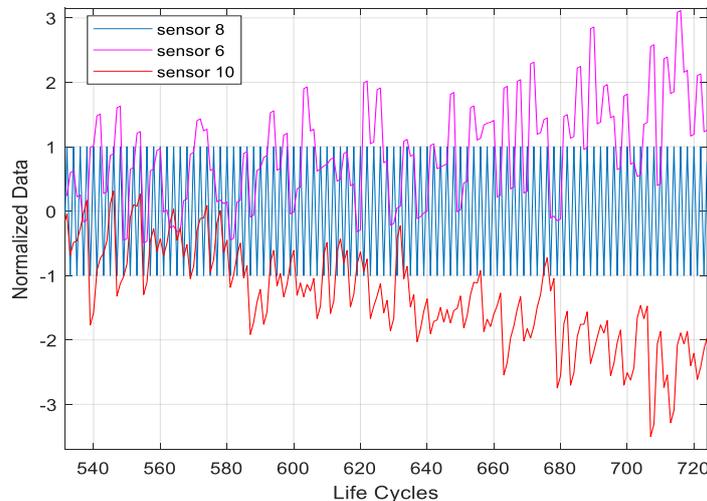


Figure III.20 : Données de capteur normalisées du moteur 177 dans FD002

III.6 Évaluation du modèle

Pour évaluer les performances de notre modèle d'estimation, nous utilisons l'erreur quadratique moyenne (RMSE) calculée comme suit :

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_{Predect} - Y_{Test})^2} \quad (III.21)$$

Et l'erreur absolue moyenne symétrique en pourcentage d'estimation (SMAPE) défini par :

$$SMAPE = \frac{1}{n} \sum \frac{|Y_{Test} - Y_{pred}|}{|Y_{Test}| + |Y_{pred}|} * 100 \quad (III.22)$$

Le nombre de séquences dans le jeu de données est n . L'erreur quadratique moyenne (RMSE) est une métrique d'évaluation largement utilisée pour les modèles d'estimation RUL qui attribue des pénalités égales aux erreurs d'estimation, quelle que soit la proximité des estimations par rapport au problème. Cependant, lorsque l'estimation est proche de la fin de vie, l'erreur absolue moyenne en pourcentage (MAPE) impose une pénalité plus importante. Dans les applications du monde réel, une estimation RUL précise est particulièrement importante lorsque le système est sur le point de tomber en panne. Par conséquent, MAPE est une meilleure métrique pour évaluer les performances d'estimation RUL que RMSE.

En conséquence, nous utilisons SMAPE pour résoudre ce problème.

III.7 Paramètres du modèle LSTM construit

Les paramètres du modèle de régression LSTM est résumé comme suite :

- L'ensemble de données est divisé en trois parties : données d'apprentissage, données cross-validation et données de test, selon un ratio de 7 :2 :1.
- Dans le processus d'apprentissage pour l'estimation RUL, nous utilisons l'erreur quadratique moyenne RMSE et l'erreur absolue moyenne symétrique en pourcentage SMAPE comme fonctions de coût.
- Régularisation : Pour éviter le surapprentissage (overfitting), nous utilisons dropout comme mécanisme d'arrêt précoce du processus d'apprentissage. La probabilité d'abandon des neurones au niveau de la couche de sortie du LSTM est définie sur 0,5, et l'objet de contrôle d'arrêt précoce est la performance de l'ensemble de données cross-validation.
- Tous les tests ont été effectués avec l'aide d'Adam selon Kingma, D. P et al. (2015) [202].
- Le nombre d'unités cachées est de 200, tandis que la couche entièrement connectée (fully connected layer) est de 50, le taux d'apprentissage initial est de 0,001 et le nombre maximal d'époques (epochs) est de 1000, avec une batch size égale à 10.

III.8 Résultats et discussion des résultats

III.8.1 Performance du pronostic de défauts

Dans le contexte de la série de données C-MAPSS (Commercial Modular Aero-Propulsion System Simulation), le "RUL" (Remaining Useful Life) est généralement exprimé en termes de cycles de fonctionnement, c'est-à-dire le nombre de fois où le moteur a été mis en marche et arrêté est fixé à 150 par le constructeur. Le RUL indique combien de cycles de fonctionnement restent avant que le moteur ne soit considéré comme défaillant. Dans les ensembles de données C-MAPSS, chaque échantillon de données est généralement associé à une valeur de RUL qui indique le nombre de cycles restants jusqu'à la défaillance du moteur. Les ensembles de données C-MAPSS peuvent inclure différentes versions, comme le C-MAPSS FD001, FD002, FD003 et FD004, qui représentent différents types de moteurs. Il est important de noter que les valeurs spécifiques de RUL et la manière dont elles sont calculées peuvent varier selon l'ensemble de données C-MAPSS utilisé et les scénarios spécifiques.

Les figures III.21(a) à III.21(d) représentent quatre moteurs sélectionnés aléatoirement dans l'ensemble de données CMAPSS. Plus précisément, les figures III.21(a) et III.21(b) illustrent les estimations de l'état des moteurs pour FD001 et FD003 respectivement. Il est remarquable que l'architecture proposée a réussi à apprendre efficacement à évaluer automatiquement le temps restant avant la défaillance (RUL), que ce soit dans un état de fonctionnement normal (visible dans la première moitié de la courbe avant la dégradation) ou dans un état de dégradation.

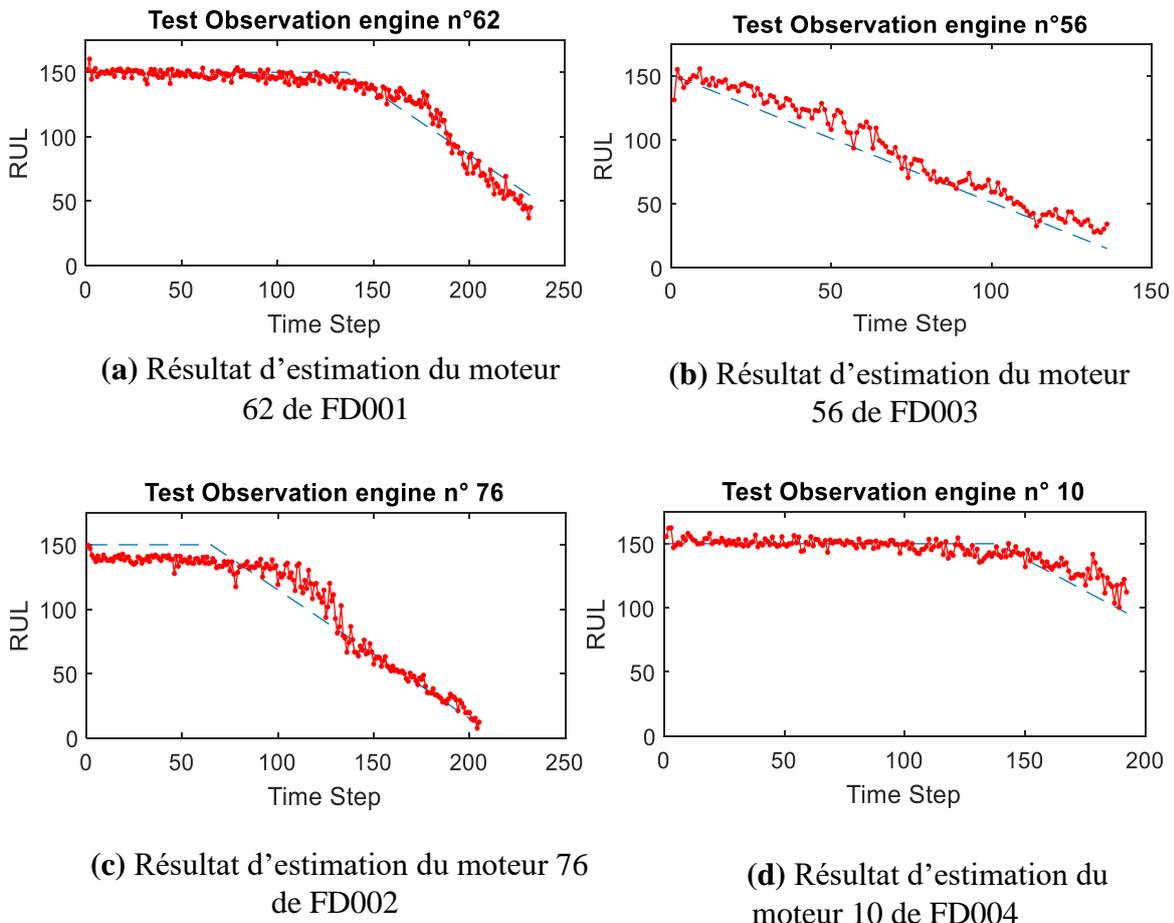


Figure III.21 Estimation RUL pour quatre moteurs aléatoires

Les résultats finaux de l'estimation pour FD002 et FD004 sont présentés dans les figures III.21(c) et III.21(d), respectivement. Ces résultats démontrent une amélioration significative de l'estimation par rapport aux modèles de prédiction précédents. Enfin, nous procédons à une comparaison de cette performance avec celle des autres modèles de prédiction.

Tableau III.5 Comparaison de RMSE et SMAPE sur le jeu de données C-MAPSS pour les modèles de prédiction RUL

Modèle de prédiction RUL	FD001		FD002		FD003		FD004	
	RMSE	SMAPE	RMSE	SMAPE	RMSE	SMAPE	RMSE	SMAPE
MLP	39.32	55.25	77.36	97.22	39.02	54.85	80.12	102.33
SVR	22.45	36.28	45.32	68.65	21.41	33.80	48.91	70.58
CNN	21.12	33.95	35.25	55.05	21.62	33.44	37.76	60.01
LSTMSAD	17.12	30.59	24.32	44.59	17.91	29.21	28.45	50.56
LSTMAAD	14.20	20.52	20.32	42.61	14.31	20.51	21.90	47.19

Le tableau III.5 met en évidence les performances des techniques suivantes : MLP, SVR, CNN, LSTM sans augmentation de données (LSTMSAD) et LSTM avec augmentation de données (LSTMAAD), en termes de RMSE et de SMAPE, sur l'ensemble de données C-MAPSS.

En effectuant une comparaison des résultats, notre modèle LSTM avec augmentation de données (LSTMAAD) a été comparé aux modèles les plus récemment publiés par Zhao, Z et al.(2017) [166] ; Zheng, S et al.(2017) [179] ; Heimes, F. O. (2008) [203] ; Yuan, M et al. (2016) [204] et Sayah, M et al. 2021) [205], confirmant ainsi la pertinence de notre approche. Les valeurs de RMSE et de SMAPE obtenues par notre méthode suggérée sont significativement inférieures à celles de ces approches.

En utilisant notre méthode avec augmentation de données, les chercheurs pourront obtenir de meilleurs résultats lorsqu'ils l'appliquent à leur modèle, tout comme nous avons pu le constater dans notre étude.

III.9 Conclusion

Dans notre étude, nous avons proposé une méthode pour estimer le temps restant avant défaillance (Remaining Useful Life - RUL) en utilisant un réseau LSTM (Long Short-Term Memory) avec l'ajout de bruit blanc gaussien pour augmenter les données d'apprentissage. Nous avons mené des expériences sur les sous-ensembles FD001, FD002, FD003 et FD004 de l'ensemble de données C-MAPSS. Les résultats ont démontré que notre approche basée sur LSTM surpassait les approches traditionnelles telles que SVR (Support Vector Regression) et CNN (Convolutional Neural Network) en termes de RMSE et SAMPE. Ces résultats positifs montrent l'efficacité de notre méthode pour améliorer les performances de prédiction du RUL en utilisant des données augmentées avec un réseau LSTM.

*Conclusion générale &
Perspectives*

Cette conclusion résume les apports de ce travail dans la première section pour répondre aux défis présentés. La section II présente des recommandations pour les travaux futurs en tant que perspectives.

Le travail présenté dans cette thèse consiste à intégrer le diagnostic et le pronostic qui sont des processus clés de la CBM dans le cadre du PHM intégré, pour optimiser la maintenance d'un système ou d'un sous-système. Les diagnostics nous permettent d'identifier et de détecter les composants défectueux qui doivent être réparés. Alors que les pronostics fournissent des informations sur l'état futur d'un composant, à partir desquelles des actions de maintenance préventive peuvent être envisagées.

Nous avons montré, au Chapitre I, les différentes formes de la maintenance ainsi que leur évolution. Ce chapitre nous a permis, entre autres, de limiter le cadre de notre travail et de l'orienter. Pour cela, nous avons restreint notre travail pour *la partie diagnostic* à l'amélioration du taux de classification correcte des défauts du moteur à induction tels que les défauts cassures de barres rotoriques.

A cette fin, nous avons proposé, au chapitre II, une nouvelle approche de classification basée sur l'utilisation de l'ondelettes par paquets comme outil d'extraction de vecteurs caractéristiques et SVM Machine à vecteurs de support (support Vector machine) multi classe comme classifieur. L'idée principale de la méthode proposée est d'utiliser la décomposition en ondelettes par paquets pour décomposer le signal de courant statorique dans l'espace-temps-fréquence et d'utiliser le résultat pour calculer l'énergie des seize premières sous-bandes, et faire la sélection des énergies des sous bandes les plus pertinentes pour les utiliser afin d'améliorer le taux de classification des défauts par la technique SVM.

Nous avons mené une comparaison entre plusieurs vecteurs des attributs en vue de déterminer ceux qui nous donnent une bonne classification et par conséquent un bon diagnostic en cas d'un défaut, avec un classifieur RNA et SVM, ce dernier qui nous donne des meilleurs résultats pour la classification.

Pour *la partie pronostic*, réserver au chapitre III. L'approche de pronostic guidée par les données est privilégiée dans notre cas pour les raisons suivantes :

- Difficulté de disposer de connaissances sur les phénomènes de dégradation complexes.
- Insuffisance d'expériences passées pour déployer des approches statistiques significatives à cause de la variabilité de conditions d'usage et de détérioration.
- Envisagement d'une stratégie de maintenance plus « dynamique » basée sur les données conditionnelles du composant au lieu d'une stratégie de maintenance préventive systématique.

L'approche que nous avons présentée dans ce chapitre est basée sur l'identification des caractéristiques de pronostic avancées susceptibles d'améliorer la prédiction de la durée de vie utile restante (RUL) en utilisant le Deep Learning et le modèle de régression LSTM pour estimer cette prédiction.

Nous avons abordé le problème du pronostic en effectuant une analyse de données de séries chronologiques. LSTM a démontré une excellente capacité à apprendre des données robustes et sensibles en introduisant une mémoire longue à court terme (LSTM) pour prédire le RUL d'un moteur d'avion.

Nous suggérons dans ce chapitre un modèle (LSTM) pour l'estimation de RUL pour les moteurs à double flux et testons ses performances sur l'ensemble de données C-MAPSS.

La principale contribution de ce chapitre est de proposer un processus d'augmentation de données pour améliorer la précision et la robustesse de la régression, qui nous donne de meilleurs résultats.

Comme perspectives : Nous proposerons,

Si le diagnostic a atteint une certaine maturité, le pronostic est, quant à lui, une discipline émergente.

Plusieurs opportunités pour continuer les recherches au-delà de l'étendue de cette thèse.

Quelques idées formulées sont présentées comme suit :

- Travailler sur les critères de performance unifiés pour évaluer les résultats de la prédiction.
- Proposition d'une nouvelle approche pour l'augmentation des données de pronostic
- Validation des modèles proposés sur les défaillances des machines (diagnostic et pronostic) dans le secteur industriel.
- En vue d'une application en monde réel de supervision, il est nécessaire de développer un outil logiciel de gestion de santé intégré qui peut être utilisé dans la détection, le diagnostic et le pronostic des défauts des composants des systèmes ou des sous-systèmes.

Références bibliographiques

- [1] Qiu, S., Cui, X., Ping, Z., Shan, N., Li, Z., Bao, X., & Xu, X. (2023). Deep Learning Techniques in Intelligent Fault Diagnosis and Prognosis for Industrial Systems: A Review. *Sensors*, 23(3), 1305.
- [2] Jarrell, D. B., Sisk, D. R., & Bond, L. J. (2004). Prognostics and condition-based maintenance: a new approach to precursive metrics. *Nuclear technology*, 145(3), 275-286.
- [3] Jardine, A. K., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical systems and signal processing*, 20(7), 1483-1510.
- [4] Venkatasubramanian, V., Rengaswamy, R., Kavuri, S. N., & Yin, K. (2003). A review of process fault detection and diagnosis: Part III: Process history based methods. *Computers & chemical engineering*, 27(3), 327-346.
- [5] Nimmo, I. (1995). Adequately address abnormal operations. *Chemical engineering progress*, 91(9).
- [6] Laser, M. (2000). Recent safety and environmental legislation. *Transactions of the Institution of Chemical Engineers*, 78(B), 419-422.
- [7] Zio, E. (2013). Prognostics and Health Management of Industrial Equipment. In S. Kadry (Ed.), *Diagnostics and Prognostics of Engineering Systems: Methods and Techniques* (pp. 333-356). IGI Global.
- [8] Gouriveau, R., Medjaher, K., Ramasso, E., & Zerhouni, N. (2013). PHM - Prognostics and health management - De la surveillance au pronostic de défaillances de systèmes complexes. *Techniques de l'Ingénieur*, MT 9570, avril 2013.
- [9] Cochetoux, P. (2010). *Contribution à la maintenance proactive par la formalisation du processus de pronostic des performances de systèmes industriels* (Thèse de doctorat). Université Henri Poincaré-Nancy 1.
- [10] Bengtsson, M., Olsson, E., Funk, P., & Jackson, M. (2004). Design of condition based maintenance system—A case study using sound analysis and case-based reasoning. *Condition Based Maintenance Systems—An Investigation of Technical Constituents and Organizational Aspects*, 57.
- [11] Garga, A. K., McClintic, K. T., Campbell, R. L., Yang, C. C., Lebold, M. S., Hay, T. A., & Byington, C. S. (2001, March). Hybrid reasoning for prognostic learning in CBM systems. In *2001 IEEE Aerospace Conference Proceedings (Cat. No. 01TH8542)* (Vol. 6, pp. 2957-2969). IEEE.
- [12] Azam, M., Tu, F., & Pattipati, K. R. (2002, July). Condition-based predictive maintenance of industrial power systems. In *Component and Systems Diagnostics, Prognostics, and Health Management II* (Vol. 4733, pp. 133-144). SPIE.
- [13] Muller, A. (2005). *Contribution à la maintenance prévisionnelle des systèmes de production par la formalisation d'un processus de pronostic* (Thèse de doctorat) Nancy 1.

- [14] Djurdjanovic, D., Lee, J., & Ni, J. (2003). Watchdog Agent—an infotronics-based prognostics approach for product performance degradation assessment and prediction. *Advanced Engineering Informatics*, 17(3-4), 109-125.
- [15] Byington, C. S., Roemer, M. J., & Galie, T. (2002, March). Prognostic enhancements to diagnostic systems for improved condition-based maintenance [military aircraft]. In *Proceedings, IEEE aerospace conference* (Vol. 6, pp. 6-6). IEEE.
- [16] Lebold, M., & Thurston, M. (2001, May). Open standards for condition-based maintenance and prognostic systems. In *Maintenance and reliability conference (marcon)* (Vol. 200).
- [17] Zhang, L., Xiong, G., Liu, H., Zou, H., & Guo, W. (2010). Bearing fault diagnosis using multi-scale entropy and adaptive neuro-fuzzy inference. *Expert Systems with Applications*, 37(8), 6077-6085.
- [18] McKee, K. K., Forbes, G. L., Mazhar, I., Entwistle, R., & Howard, I. (2014). A review of machinery diagnostics and prognostics implemented on a centrifugal pump. In *Engineering Asset Management 2011: Proceedings of the Sixth World Congress on Engineering Asset Management* (pp. 593-614). Springer London.
- [19] Heng, A., Zhang, S., Tan, A. C., & Mathew, J. (2009). Rotating machinery prognostics: State of the art, challenges and opportunities. *Mechanical systems and signal processing*, 23(3), 724-739.
- [20] McFadden, P. D. (1987). A revised model for the extraction of periodic waveforms by time domain averaging. *Mechanical systems and signal processing*, 1(1), 83-95.
- [21] Widrow, B., Glover, J. R., McCool, J. M., Kaunitz, J., Williams, C. S., Hearn, R. H., & Zeidler, J. R. (1975). Adaptive noise cancelling: Principles and applications. *Proceedings of the IEEE*, 63(12), 1692-1716. <https://doi.org/10.1109/PROC.1975.10036>
- [22] Shiroishi, J., Y Li, S. L., Danyluk, S., & Kurfess, T. (1999). Vibration analysis for bearing outer race condition diagnostics. *Journal of the Brazilian Society of Mechanical Sciences*, 21, 484-492.
- [23] Antoni, J., & Randall, R. B. (2004). Unsupervised noise cancellation for vibration signals: part II—a novel frequency-domain algorithm. *Mechanical Systems and Signal Processing*, 18(1), 103-117.
- [24] Antoni, J., & Randall, R. B. (2004). Unsupervised noise cancellation for vibration signals: part I—evaluation of adaptive algorithms. *Mechanical Systems and Signal Processing*, 18(1), 89-101.
- [25] Qiu, H., Lee, J., Lin, J., & Yu, G. (2003). Robust performance degradation assessment methods for enhanced rolling element bearing prognostics. *Advanced Engineering Informatics*, 17(3-4), 127-140.
- [26] Karimi, M., Tan, A., Mathew, J., & Senadji, B. (2006). Blind Deconvolution as a Notch Filter in Detecting Bearing Faults Corrupted by Noise. *Advances in Vibration Engineering*, 5(2), 121-132.
- [27] Han, J., Kamber, M., & Mining, D. (2006). Concepts and techniques. *Morgan kaufmann*, 340, 94104-3205.
- [28] Johnson, N. L., Kotz, S., & Balakrishnan, N. (1995). Continuous Univariate Distributions, Volume 2 (2nd ed., Vol. 289). *John Wiley & Sons*.
- [29] Bracewell, R.N. (2000). The Fourier Transform and Its Applications, 3rd edn. McGraw-Hill, Boston. ISBN 0-07-116043-4
- [30] Weaver, H. J. (1983). Applications of Discrete and Continuous Fourier Analysis. *Wiley-Interscience Publication, John Wiley & Sons*.

- [31] Tan, L., & Jiang, J. (2018). *Digital Signal Processing: Fundamentals and Applications* (3rd ed.). *Academic Press*. ISBN: 9780128150719.
- [32] Duhamel, P., & Vetterli, M. (1990). Fast Fourier transforms: a tutorial review and a state of the art. *Signal processing*, 19(4), 259-299.
- [33] Johnson, S. G., & Frigo, M. (2006). A modified split-radix FFT with fewer arithmetic operations. *IEEE Transactions on Signal Processing*, 55(1), 111-119.
- [34] Welch, P. (1967). The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on audio and electroacoustics*, 15(2), 70-73.
- [35] Covell, M., & Richardson, J. (1991, April). A new, efficient structure for the short-time Fourier transform, with an application in code-division sonar imaging. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on* (pp. 2041-2044). IEEE Computer Society.
- [36] Tan, L., & Jiang, J. (2013). *Digital Signal Processing: Fundamentals and Applications* (2nd ed.). *Oxford University Press*.
- [37] Daubechies, I. (1998). Orthonormal Bases of Compactly Supported Wavelets. *Communications on Pure and Applied Mathematics*, 41, 909-996.
- [38] Mallat, S. G. (1989). A theory for multiresolution signal decomposition: the wavelet representation. *IEEE transactions on pattern analysis and machine intelligence*, 11(7), 674-693.
- [39] Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., ... & Liu, H. H. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences*, 454(1971), 903-995.
- [40] Sharpley, R. C., & Vatchev, V. (2006). Analysis of the intrinsic mode functions. *Constructive Approximation*, 24, 17-47.
- [41] Wu, T. Y., Chen, J. C., & Wang, C. C. (2012). Characterization of gear faults in variable rotating speed using Hilbert-Huang Transform and instantaneous dimensionless frequency normalization. *Mechanical Systems and Signal Processing*, 30, 103-122.
- [42] Cohen, L. (1995). *Time-frequency analysis* (Vol. 778). New Jersey: Prentice hall.
- [43] Ville, J. (1948). Théorie et Application de la notion de Signal Analytique. *Câbles et Transmissions*, 2A(1), 61-74. Paris, France.
- [44] Kadambe, S., & Boudreaux-Bartels, G. F. (1992). A comparison of the existence of 'cross terms' in the Wigner distribution and the squared magnitude of the wavelet transform and the short-time Fourier transform. *IEEE Transactions on signal processing*, 40(10), 2498-2517.
- [45] Wang, C., Zhang, Y., & Zhong, Z. (2008). Fault diagnosis for diesel valve trains based on time-frequency images. *Mechanical Systems and Signal Processing*, 22(8), 1981-1993.
- [46] Vachtsevanos, G., Lewis, F. L., Roemer, M., Hess, A., & Wu, B. (2006). *Intelligent fault diagnosis and prognosis for engineering systems*. Hoboken, NJ: John Wiley & Sons, Inc.
- [47] Butler, S. (2012). *Prognostic algorithms for condition monitoring and remaining useful life estimation* (Doctoral dissertation, National University of Ireland Maynooth).
- [48] Ribot, P. (2009). *Vers l'intégration diagnostic/pronostic pour la maintenance des systèmes complexes* (Thèse de doctorat) Université Paul Sabatier-Toulouse III.

- [49] Mous, M. D. (2006). *Diagnostic et conduite des systèmes de production par approche à base de connaissances* (Thèse de doctorat) Batna, Université El Hadj Lakhder. Faculté des sciences de l'ingénieur).
- [50] Lyonnet, P., Thomas, M., & Toscano, R. (2012). *Fiabilité, diagnostic et maintenance des systèmes*. Lavoisier. ISBN 978-2-7430-1385-1.
- [51] Zwingelstein, G. (1995). *Diagnostic des Défaillances Théorie et Pratique pour les Systèmes Industriels*. Éditions Hermès.
- [52] Willsky, A. S. (1976). A survey of design methods for failure detection in dynamic systems. *Automatica*, 12(6), 601-611.
- [53] Chow, E. Y. E. Y., & Willsky, A. (1984). Analytical redundancy and the design of robust failure detection systems. *IEEE Transactions on automatic control*, 29(7), 603-614.
- [54] Basseville, M. (1988). Detecting changes in signals and systems—a survey. *Automatica*, 24(3), 309-326.
- [55] Patton, R. J. (1997). Fault-tolerant control: the 1997 situation. *IFAC Proceedings Volumes*, 30(18), 1029-1051.
- [56] Evsukoff, A., Montmain, J., & Gentil, S. (1997). Dynamic model and causal knowledge-based fault detection and isolation. *IFAC Proceedings Volumes*, 30(18), 687-692.
- [57] Isermann, R., & Balle, P. (1997). Trends in the application of model-based fault detection and diagnosis of technical processes. *Control Engineering Practice*, 5(5), 709-719.
- [58] Balle, P., Fischer, M., Fussel, D., Nelles, O., & Isermann, R. (1998). Integrated control, diagnosis and reconfiguration of a heat exchanger. *IEEE Control Systems Magazine*, 18(3), 52-63.
- [59] Frank, P. M. (1996). Analytical and qualitative model-based fault diagnosis—a survey and some new results. *European Journal of Control*, 2(1), 6-28.
- [60] Patan, K. (2008). *Artificial neural networks for the modelling and fault diagnosis of technical processes*.
- [61] Chiang, L. H., Russell, E. L., & Braatz, R. D. (2000). *Fault detection and diagnosis in industrial systems*. Springer Science & Business Media.
- [62] Frank, P. M. (1996). Analytical and qualitative model-based fault diagnosis—a survey and some new results. *European Journal of Control*, 2(1), 6-28.
- [63] Gertler, J. (1998). *Fault Detection and Diagnosis in Engineering Systems*. CRC Press. <https://doi.org/10.1201/9780203756126>. eBook: ISBN9780203756126
- [64] Frank, P. M. (1987). Advanced fault detection and isolation schemes using nonlinear and robust observers. *IFAC Proceedings Volumes*, 20(5), 63-68.
- [65] Staroswiecki, M., & Declerck, P. (1990). Analytical redundancy in non linear interconnected systems by means of structural analysis. In *Advanced Information Processing in Automatic Control (AIPAC'89)* (pp. 51-55). Pergamon.
- [66] Narasimhan, S., & Biswas, G. (2007). Model-based diagnosis of hybrid systems. *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans*, 37(3), 348-361.
- [67] Coquempot, V., El Meznyani, T., & Staroswiecki, M. (2005). Hybrid dynamical systems monitoring using structured analytical redundancy relations. In *proceedings of 17ème IMACS World Congress, Paris, France*.
- [68] Basseville, M., & Nikiforov, I. V. (1993). *Detection of abrupt changes: theory and application* (Vol. 104). Englewood Cliffs: Prentice Hall.
- [69] Isermann, R. (1997). Supervision, fault-detection and fault-diagnosis methods—An introduction. *Control Engineering Practice*, 5(5), 639-652.

- [70] Staroswiecki, M., & Comtet-Varga, G. (2001). Analytical redundancy relations for fault detection and isolation in algebraic dynamic systems. *Automatica*, 37(5), 687-699.
- [71] Patton, R. J., & Chen, J. (1991). A re-examination of the relationship between parity space and observer-based approaches in fault diagnosis. *European Journal of Diagnosis and Safety in Automation*, 1(2), 183-200.
- [72] Maquin, D., Cocquempot, V., Cassar, J. P., Staroswiecki, M., & Ragot, J. (1997, September). Generation of analytical redundancy relations for FDI purposes. In *IFAC Symposium on Diagnostics for Electrical Machines, Power Electronics and Drives, SDEMPED'97* (pp. 86-93).
- [73] Staroswiecki, M., Cocquempot, V., & Cassar, J. P. (1991, May). Observer based and parity space approaches for failure detection and identification. In *IMACS-IFAC international symposium, Lille, France* (Vol. 25, pp. 536-541).
- [74] Desai, M., & Ray, A. (1981, December). A fault detection and isolation methodology. In *1981 20th IEEE Conference on Decision and Control including the Symposium on Adaptive Processes* (pp. 1363-1369). IEEE.
- [75] Schwarte, A., Kimmich, F., & Isermann, R. (2003). Model-based fault detection of a diesel engine with turbo charger—a case study. *IFAC Proceedings Volumes*, 36(5), 293-306.
- [76] Magni, J. F., & Mouyon, P. (1994). On residual generation by observer and parity space approaches. *IEEE Transactions on Automatic Control*, 39(2), 441-447.
- [77] Hammouri, H., Kinnaert, M., & El Yaagoubi, E. H. (1999). Observer-based approach to fault detection and isolation for nonlinear systems. *IEEE transactions on automatic control*, 44(10), 1879-1884.
- [78] Gertler, J. (1995, June). Diagnosing parametric faults: from parameter estimation to parity relations. In *Proceedings of 1995 American Control Conference-ACC'95* (Vol. 3, pp. 1615-1620). IEEE.
- [79] Reiter, R. (1987). A theory of diagnosis from first principles. *Artificial intelligence*, 32(1), 57-95.
- [80] De Kleer, J., & Williams, B. C. (1987). Diagnosing multiple faults. *Artificial intelligence*, 32(1), 97-130.
- [81] Methnani, S. (2012). *Diagnostic, reconstruction et identification des défauts capteurs et actionneurs : application aux stations d'épuration des eaux usées* (Thèse de doctorat) Université de Toulon ; École nationale d'ingénieurs de Sfax, Tunisie).
- [82] Jolliffe, I. T. (1990). Principal component analysis: a beginner's guide—I. Introduction and application. *Weather*, 45(10), 375-382.
- [83] Pessel, N., Balmat, J. F., Lafont, F., & Bonnal, J. (2007, May). An improved PCA fault detection for the diagnosis. In *Proceedings of the 9th WSEAS International Conference on Automatic Control, Modeling & Simulation* (pp. 312-318).
- [84] Basseville, M. (1988). Detecting changes in signals and systems—a survey. *Automatica*, 24(3), 309-326.
- [85] Basseville, M. (1997). Information criteria for residual generation and fault detection and isolation. *Automatica*, 33(5), 783-803.
- [86] Assaf, E. M., Giordano, R. D. C., & Nascimento, C. A. O. D. (1996). Thermal runaway of ethylene oxidation reactors: prevision through neuronal networks. *Chemical Engineering Science*, 51(11), 3107-3112.
- [87] Barton, J. A. (1984). Runaway reactions in batch reactors. In *Inst. Chem. Eng. Symp.*

- Ser.* (Vol. 85, pp. 13-21).
- [88] Gustin, J. L. (1996). Retour d'expérience Chimie Fine. - Recommandations. *Ministère L'environnement Union Des Industries Chimiques*
- [89] Yu, D. L., Gomm, J. B., & Williams, D. (1999). Sensor fault diagnosis in a chemical process via RBF neural networks. *Control Engineering Practice*, 7(1), 49-55.
- [90] Cubillos, F. A., & Lima, E. L. (1998). Adaptive hybrid neural models for process control. *Computers & Chemical Engineering*, 22, S989-S992.
- [91] Power, Y., & Bahri, P. A. (2004). A two-step supervisory fault diagnosis framework. *Computers & Chemical Engineering*, 28(11), 2131-2140.
- [92] Venkatasubramanian, V., Vaidyanathan, R., & Yamamoto, Y. (1990). Process fault detection and diagnosis using neural networks—I. Steady-state processes. *Computers & Chemical Engineering*, 14(7), 699-712.
- [93] Hoskins, J. C., & Himmelblau, D. M. (1988). Artificial neural network models of knowledge representation in chemical engineering. *Computers & Chemical Engineering*, 12(9-10), 881-890.
- [94] Watanabe, K., Hirota, S., Hou, L., & Himmelblau, D. M. (1994). Diagnosis of multiple simultaneous faults via hierarchical artificial neural networks. *AIChE Journal*, 40, 839-848.
- [95] Villemeur, A. (1988). *Sûreté de fonctionnement des systèmes industriels : fiabilité, facteurs humains, informatisation*. Eyrolles.
- [96] Farreny, H. (1989). L'intelligence artificielle : une discipline originale avec ses problèmes et ses techniques. In *Annales des télécommunications* (Vol. 44, No. 5-6, pp. 207-217).
- [97] Wang, W. (2007). An adaptive predictor for dynamic system forecasting. *Mechanical Systems and Signal Processing*, 21(2), 809-823.
- [98] Pourahmadi, M. (2001). *Foundations of time series analysis and prediction theory* (Vol. 379). John Wiley & Sons.
- [99] Wang, W. Q., Ismail, F., & Golnaraghi, M. F. (2001). Assessment of gear damage monitoring techniques using vibration measurements. *Mechanical Systems and Signal Processing*, 15(5), 905-922.
- [100] Wang, W. (2007). A two-stage prognosis model in condition based maintenance. *European Journal of Operational Research*, 182(3), 1177-1187.
- [101] Lebold, M., & Thurston, M. (2001). Open standards for condition-based maintenance and prognostic systems. Proceedings of the 5th Annual Maintenance and Reliability Conference (MARCON), Vol. 200, Gatlinburg, TN, May.
- [102] Yan, J., Koc, M., & Lee, J. (2004). A prognostic algorithm for machine performance assessment and its application. *Production Planning & Control*, 15(8), 796-801.
- [103] Neter, J., Kutner, M. H., Nachtsheim, C. J., & Wasserman, W. (1996). *Applied Linear Statistical Models* (4th Edition). WCB McGraw-Hill.
- [104] Hush, D. R., & Horne, B. G. (1993). Progress in supervised neural networks. *IEEE Signal Proc. Mag.* 10, 8-39.
- [105] Li, C. J., & Ray, A. (1995). Neural network representation of fatigue damage dynamics. *Smart Materials and Structures*, 4(2), 126.
- [106] Vachtsevanos, G., & Wang, P. (2001, August). Fault prognosis using dynamic wavelet neural networks. In *2001 IEEE Autotestcon Proceedings. IEEE Systems Readiness Technology Conference. (cat. no. 01ch37237)* (pp. 857-870).

- [107] Yam, R. C. M., Tse, P. W., Li, L., & Tu, P. (2001). Intelligent predictive decision support system for condition-based maintenance. *The International Journal of Advanced Manufacturing Technology*, 17, 383-391.
- [108] Wang, P., & Vachtsevanos, G. (2001). Fault prognostics using dynamic wavelet neural networks. *Artificial Intelligence for Engineering Design AI EDAM*, 15(4), 349-365.
- [109] Watson, M., Byington, C. S., & Edwards, D. (2004). Dynamic signal analysis and neural network modeling for life prediction of flight control actuators. In Proceedings of the American Helicopter Society 60th Annual Forum . Alexandria, VA: AHS.
- [110] Byington, C. S., Watson, M., & Edwards, D. (2004, March). Data-driven neural network methodology to remaining life predictions for aircraft actuator components. In *2004 IEEE Aerospace Conference Proceedings (IEEE Cat. No. 04TH8720)* (Vol. 6, pp. 3581-3589). IEEE.
- [111] Wang, W. Q., Golnaraghi, M. F., & Ismail, F. (2004). Prognosis of machine health condition using neuro-fuzzy systems. *Mechanical Systems and Signal Processing*, 18(4), 813-831.
- [112] Giantomassi, A., Ferracuti, F., Benini, A., Ippoliti, G., Longhi, S., & Petrucci, A. (2011, January). Hidden Markov model for health estimation and prognosis of turbofan engines. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference* (Vol. 54808, pp. 681-689).
- [113] Bock, J. R., Brotherton, T., Grabill, P., Gass, D., & Keller, J. A. (2006, March). On false alarm mitigation. In *2006 IEEE Aerospace Conference* (pp. 15-pp). IEEE.
- [114] Clifton, D., & Tarassenko, L. (2006). Condition monitoring of gas-turbine engines. *Transfer Report, Department of Engineering Science, University of Oxford*.
- [115] Banjevic, D., & Jardine, A. K. S. (2006). Calculation of reliability function and remaining useful life for a Markov failure time process. *IMA Journal of Management Mathematics*, 17(2), 115-130. doi:10.1093/imaman/dpi029.
- [116] Bossio, G., De De Angelo, C., Solsona, J., Garcia, G. O., & Valla, M. I. (2006). Application of an additional excitation in inverter-fed induction motors for air-gap eccentricity diagnosis. *IEEE Transactions on Energy Conversion*, 21(4), 839-847.
- [117] Razik, H. (2002). Le contenu spectral du courant absorbé par la machine asynchrone en cas de défaillance, état de l'art. *La Revue 3EI*, 29, 48-52.
- [118] Deleroi, W. (1982, September). Squirrel cage motor with broken bar in the rotor—physical phenomena and their experimental assessment. In *Proc. of Int. Conf. On Electrical Machines* (pp. 767-770).
- [119] Vieira, M. (1995). Induction Motor Fault Detection and Localization Using Stator Current Advanced Signature Processing Technique. *IEEE Trans. Industrial Application*, 3(1).
- [120] Ho, S. L., Chan, W. L., & Leung, H. W. (1993). Application of statistical signal processing for condition monitoring of rotor faults in induction motors. In *Proceedings of the International Conference on Electrical Machines* (pp. 97-102). Oxford, U.K.
- [121] De Castro Parga, J. L., Manzanedo García, J. F., Novo Ramos, B., & Pérez Donsión, M. (2000). Detection of Broken Rotor Bars in an Induction Machine Based on Rotor Resistance Estimation. *International Conference on Electrical Machines (ICEM'2000), ESPOO (Finlandia)*, (Vol. II, pp. 859-862). Helsinki Univ. Of Tech. Finlandia. ISBN: 951-22-5097-7

- [122] Henao, H., Capolino, G. A., & Martis, C. (2003, October). On the stray flux analysis for the detection of the three-phase induction machine faults. In *38th IAS Annual Meeting on Conference Record of the Industry Applications Conference, 2003*. (Vol. 2, pp. 1368-1373). IEEE.
- [123] Abbaszadeh, K., Milimonfared, J., Haji, M., & Toliyat, H. A. (2001). Broken bar detection in induction motor via wavelet transformation. In *IECON'01. 27th Annual Conference of the IEEE Industrial Electronics Society* (Cat. No. 37243), Denver, CO, USA (pp. 95-99 vol.1). IEEE. doi: 10.1109/IECON.2001.976461.
- [124] Hubbard, B. B. (1995). Ondes et ondelettes - *La saga d'un outil mathématique. Pour la Science*/Belin.
- [125] Wickerhauser, M. V. (1994). *Adapted Wavelet Analysis, from Theory to Software*. AK Peters. Boston.
- [126] Daubechies, I., & Lagarias, J. C. (1991). Two-scale difference equations. I. Existence and global regularity of solutions. *SIAM Journal on Mathematical Analysis*, 22(5), 1388-1410.
- [127] Derin, H., & Elliott, H. (1987). Modeling and segmentation of noisy and textured images using Gibbs random fields. *IEEE Transactions on pattern analysis and machine intelligence*, (1), 39-55.
- [128] Muzy, J. F., Bacry, E., & Arneodo, A. (1994). The multifractal formalism revisited with wavelets. *International Journal of Bifurcation and Chaos*, 4(02), 245-302.
- [129] Mallat, S., & Zhong, S. (1992). Characterization of signals from multiscale edges. *IEEE Transactions on pattern analysis and machine intelligence*, 14(7), 710-732.
- [130] Bijaoui, A. (1991). Algorithmes de la transformation en ondelettes. Applications en astronomie. *Ondelettes et Paquet d'ondes*, 115-140.
- [131] Bouchereau, E. B. (1997). *Analyse d'images par transformées en ondelettes. Application aux images sismiques* (Thèse de doctorat), Université Joseph-Fourier-Grenoble I
- [132] Aslaksen, E. W., & Klauder, J. R. (1968). Unitary representations of the affine group. *Journal of Mathematical Physics*, 9(2), 206-211.
- [133] Farge, M., Kevlahan, N., Perrier, V., & Goirand, E. (1996). Wavelets and turbulence. *Proceedings of the IEEE*, 84(4), 639-669.
- [134] Mallat, S. G. (1989). Multiresolution approximations and wavelet orthonormal bases of $L^2(R)$. *Transactions of the American mathematical society*, 315(1), 69-87.
- [135] Mallat, S., & Hwang, W. L. (1992). Singularity detection and processing with wavelets. *IEEE transactions on information theory*, 38(2), 617-643.
- [136] Marroquin, J. L. (1985). *Probabilistic solution of inverse problems*. (Thèse de doctorat) Massachusetts Institute of Technology, Cambridge, MA.
- [137] Meyer, Y. (1992). *Ondelettes : Algorithmes et Applications*. Armand Colin. Paris.
- [138] Daubechies, I. (1992). *Ten Lectures on Wavelets*. CBMS-NSF Series in Applied Mathematics, No. 61. SIAM. Philadelphia.
- [139] Cohen, A. (1992). Biorthogonal wavelets. *Wavelets: A Tutorial in Theory and Applications*, 2, 123-152.
- [140] Cohen, A., Daubechies, I., & Feauveau, J. C. (1992). Biorthogonal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics*, 45(5), 485-560.
- [141] Chui, C. K. (1992). *An Introduction to Wavelets* (Vol. 1, Wavelet Analysis and its Applications). Academic Press. New York.
- [142] Meyer, Y. (1990). *Ondelettes et Opérateurs, I: Ondelettes*. Hermann. Paris.

- [143] Bazine, S. (2009). *Conception et implémentation d'un Méta-modèle de machines asynchrones en défaut* (Thèse de doctorat), Poitiers.
- [144] Bouzida, A. (2008). *Diagnostic de défauts de la machine asynchrone a cage par la technique des ondelettes* (Thèse de doctorat), Ecole nationale polytechnique.
- [145] Ye, Z., Wu, B., & Sadeghian, A. (2003). Current signature analysis of induction motor mechanical faults by wavelet packet decomposition. *IEEE Transactions on Industrial Electronics*, 50(6), 1217-1228.
- [146] Antonino-Daviu, J., Riera-Guasp, M., Roger-Folch, J., Martínez-Giménez, F., & Peris, A. (2006). Application and optimization of the discrete wavelet transform for the detection of broken rotor bars in induction machines. *Applied and Computational Harmonic Analysis*, 21(2), 268-279.
- [147] Cvetkovic, Z., & Vetterli, M. (1995). Discrete-time wavelet extrema representation: Design and consistent reconstruction. *IEEE Transactions on Signal processing*, 43(3), 681-693.
- [148] Chow, T. W., & Hai, S. (2004). Induction machine fault diagnostic analysis with wavelet technique. *IEEE Transactions on Industrial Electronics*, 51(3), 558-565.
- [149] Hartert, L. (2010). *Reconnaissance des formes dans un environnement dynamique appliquée au diagnostic et au suivi des systèmes évolutifs* (Thèse de doctorat), Université de Reims-Champagne Ardenne.
- [150] Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20, 273-297.
- [151] Schölkopf, B., & Smola, A. J. (2002). *Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond*. MIT Press. Cambridge.
- [152] Cornuéjols, A., Miclet, L., & Kodratoff, Y. (2002). *Apprentissage Artificiel, Concepts et Algorithmes*. ISBN 2-212-11020-0.
- [153] Chendeb, M. (2006). *Détection et classification des signaux non stationnaires par utilisation des ondelettes. Application aux signaux électromyographiques utérins* (Thèse de doctorat), Université de Technologie de Troyes.
- [154] Zhang, N. (2010). *Détection et isolation de pannes basées sur la platitude différentielle: application aux engins atmosphériques* ((Thèse de doctorat), Toulouse, INSA.
- [155] Casimir, R. (2003). *Diagnostic des défauts des machines asynchrones par reconnaissance des formes* ((Thèse de doctorat), Ecole centrale de Lyon.
- [156] Li, L. L., Zhang, X. B., Tseng, M. L., & Zhou, Y. T. (2019). Optimal scale Gaussian process regression model in Insulated Gate Bipolar Transistor remaining life prediction. *Applied Soft Computing*, 78, 261-273.
- [157] Wang, Y., Deng, C., Wu, J., Wang, Y., & Xiong, Y. (2014). A corrective maintenance scheme for engineering equipment. *Engineering Failure Analysis*, 36, 269-283.
- [158] Liao, L., & Köttig, F. (2016). A hybrid framework combining data-driven and model-based methods for system remaining useful life prediction. *Applied Soft Computing*, 44, 191-199.
- [159] Ramasso, E., Rombaut, M., & Zerhouni, N. (2012). Joint prediction of continuous and discrete states in time-series based on belief functions. *IEEE transactions on cybernetics*, 43(1), 37-50.
- [160] Lei, Y., Li, N., Guo, L., Li, N., Yan, T., & Lin, J. (2018). Machinery health prognostics: A systematic review from data acquisition to RUL prediction. *Mechanical Systems and Signal Processing*, 104, 799-834.

- [161] Liu, J., & Zio, E. (2016). Prediction of peak values in time series data for prognostics of critical components in nuclear power plants. *IFAC-PapersOnLine*, 49(28), 174-178. <https://doi.org/10.1016/j.ifacol.2016.11.030>
- [162] Hornik, K. (1991). Approximation capabilities of multilayer feedforward networks. *Neural networks*, 4(2), 251-257.
- [163] Peel, L. (2008, October). Data driven prognostics using a Kalman filter ensemble of neural network models. In *2008 International Conference on Prognostics and Health Management* (pp. 1-6). IEEE.
- [164] Tamilselvan, P., & Wang, P. (2013). Failure diagnosis using deep belief learning based health state classification. *Reliability Engineering & System Safety*, 115, 124-135.
- [165] Zhang, X., Xiao, L., & Kang, J. (2015). Degradation prediction model based on a neural network with dynamic windows. *Sensors*, 15(3), 6996-7015.
- [166] Zhao, Z., Liang, B., Wang, X., & Lu, W. (2017). Remaining useful life prediction of aircraft engine based on degradation pattern learning. *Reliability Engineering & System Safety*, 164, 74-83.
- [167] Ompusunggu, A. P., Papy, J. M., & Vandenplas, S. (2015). Kalman-filtering-based prognostics for automatic transmission clutches. *IEEE/ASME Transactions on Mechatronics*, 21(1), 419-430.
- [168] Li, X., Ding, O., & Sun, J. O. (2018). Remaining useful life estimation in prognostics using deep convolution neural networks. *Reliability Engineering & System Safety*, 172, 1-11.
- [169] Khelif, R., Chebel-Morello, B., Malinowski, S., Laajili, E., Fnaiech, F., & Zerhouni, N. (2016). Direct remaining useful life estimation based on support vector regression. *IEEE Transactions on Industrial Electronics*, 64(3), 2276-2285.
- [170] Wu, J., Xu, J., & Huang, X. (2017, July). An indirect prediction method of remaining life based on glowworm swarm optimization and extreme learning machine for lithium battery. In *2017 36th Chinese Control Conference (CCC)* (pp. 7259-7264). IEEE.
- [171] Morando, S., Jemei, S., Gouriveau, R., Zerhouni, N., & Hissel, D. (2014, October). Fuel cells remaining useful lifetime forecasting using echo state network. In *2014 IEEE Vehicle Power and Propulsion Conference (VPPC)* (pp. 1-6). IEEE.
- [172] Liu, H., Zhou, J., Zheng, Y., Jiang, W., & Zhang, Y. (2018). Fault diagnosis of rolling bearings with recurrent neural network-based autoencoders. *ISA Transactions*, 77, 167-178.
- [173] Chandra, R. (2015). Competition and collaboration in cooperative coevolution of Elman recurrent neural networks for time-series prediction. *IEEE Transactions on Neural Networks and Learning Systems*, 26(12), 3123-3136.
- [174] Malhi, A., & Gao, R. X. (2004, May). Recurrent neural networks for long-term prediction in machine condition monitoring. In *Proceedings of the 21st IEEE Instrumentation and Measurement Technology Conference (IEEE Cat. No. 04CH37510)* (Vol. 3, pp. 2048-2053). IEEE.
- [175] Lukoševičius, M., & Jaeger, H. (2009). Reservoir computing approaches to recurrent neural network training. *Computer Science Review*, 3(3), 127-149.
- [176] Wu, Y., Yuan, M., Dong, S., Lin, L., & Liu, Y. (2018). Remaining useful life estimation of engineered systems using vanilla LSTM neural networks. *Neurocomputing*, 275, 167-179.
- [177] Likhitha, D. L., & Nagaraja, R. (2019). Prediction of Remaining Useful Life of an Aircraft Engine Using LSTM Network. vol. 9, no. 6, pp. 329-334.
- [178] Li, H., Huang, J., Yang, X., Luo, J., Zhang, L., & Pang, Y. (2020). Fault diagnosis for rotating machinery using multiscale permutation entropy and convolutional neural networks. *Entropy*, 22(8), 851.

- [179] Zheng, S., Ristovski, K., Farahat, A., & Gupta, C. (2017, June). Long short-term memory network for remaining useful life estimation. In *2017 IEEE International Conference on Prognostics and Health Management (ICPHM)* (pp. 88-95). IEEE.
- [180] Wu, J., Hu, K., Cheng, Y., Zhu, H., Shao, X., & Wang, Y. (2020). Data-driven remaining useful life prediction via multiple sensor signals and deep long short-term memory neural network. *ISA Transactions*, 97, 241-250.
- [181] Xia, M., Zheng, X., Imran, M., & Shoaib, M. (2020). Data-driven prognosis method using hybrid deep recurrent neural network. *Applied Soft Computing*, 93, 106351.
- [182] Gurney, K. (1997). *An introduction to neural networks*. CRC press.
- [183] Fausett, L. V. (2006). *Fundamentals of neural networks: architectures, algorithms and applications*. Pearson Education India.
- [184] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.
- [185] Deeply Learning. (2018). Cours théoriques - Deep Learning - Fonction d'activation. <https://deeplylearning.fr/cours-theoriques-deep-learning/fonction-dactivation/>.
- [186] Kelleher, J. D. (2019). *Deep learning*. MIT press.
- [187] Nielsen, M. A. (2015). *Neural Networks and Deep Learning* (Vol. 25). Determination Press. San Francisco, CA, USA.
- [188] Aggarwal, C. C. (2018). Neural networks and deep learning. *Springer*, 10(978), 3.
- [189] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, 521(7553), 436-444.
- [190] Nasekin, S., & Chen, C. Y. H. (2020). Deep learning-based cryptocurrency sentiment construction. *Digital Finance*, 2(1-2), 39-67.
- [191] Bengio, Y., Goodfellow, I., & Courville, A. (2017). *Deep learning* (Vol. 1). Cambridge, MA, USA: MIT press.
- [192] Zhu, Y., Song, E., Zhou, J., & You, Z. (2005). Optimal dimensionality reduction of sensor data in multisensor estimation fusion. *IEEE Transactions on Signal Processing*, 53(5), 1631-1639.
- [193] Zheng, S., Nie, F., Ding, C., & Huang, H. (2016, November). A harmonic mean linear discriminant analysis for robust image classification. In *2016 IEEE 28th International Conference on Tools with Artificial Intelligence (ICTAI)* (pp. 402-409). IEEE.
- [194] Zheng, S., & Ding, C. (2014). Kernel alignment inspired linear discriminant analysis. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2014, Nancy, France, September 15-19, 2014. Proceedings, Part III 14* (pp. 401-416). Springer Berlin Heidelberg.
- [195] Zheng, S., Cai, X., Ding, C., Nie, F., & Huang, H. (2015, February). A closed form solution to multi-view low-rank regression. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 29, No. 1).
- [196] Zheng, S., Shae, Z. Y., Zhang, X., Jamjoom, H., & Fong, L. (2011). Analysis and modeling of social influence in high performance computing workloads. In *Euro-Par 2011 Parallel Processing: 17th International Conference, Euro-Par 2011, Bordeaux, France, August 29-September 2, 2011, Proceedings, Part I 17* (pp. 193-204). Springer Berlin Heidelberg.
- [197] Williams, D., Zheng, S., Zhang, X., & Jamjoom, H. (2014, April). Tidewatch: Fingerprinting the cyclicity of big data workloads. In *IEEE INFOCOM 2014-IEEE Conference on Computer Communications* (pp. 2031-2039). IEEE.
- [198] Zhang, X., Shae, Z. Y., Zheng, S., & Jamjoom, H. (2012, April). Virtual machine migration in an over-committed cloud. In *2012 IEEE Network Operations and Management Symposium* (pp. 196-203). IEEE.
- [199] Saxena, A., Goebel, K., Simon, D., & Eklund, N. (2008). *Damage propagation modeling for aircraft engine run-to-failure simulation*. In 2008 International Conference on Prognostics and Health Management (pp. 1-9). Posted: 2008.

- [200] National Aeronautics and Space Administration. Prognostics Center of Excellence Data Set Repository. NASA. Retrieved from <https://www.nasa.gov/content/prognostics-center-of-excellence-data-set-repository/>
- [201] Fawzi, A., Samulowitz, H., Turaga, D., & Frossard, P. (2016, September). Adaptive data augmentation for image classification. In *2016 IEEE international conference on image processing (ICIP)* (pp. 3688-3692). IEEE.
- [202] Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- [203] Heimes, F. O. (2008, October). Recurrent neural networks for remaining useful life estimation. In *2008 International Conference on Prognostics and Health Management* (pp. 1-6). IEEE.
- [204] Yuan, M., Wu, Y., & Lin, L. (2016, October). Fault diagnosis and remaining useful life estimation of aero engine using LSTM neural network. In *2016 IEEE international conference on aircraft utility systems (AUS)* (pp. 135-140). IEEE.
- [205] Sayah, M., Guebli, D., Nouredine, Z., & Al Masry, Z. (2021). Deep LSTM enhancement for RUL prediction using Gaussian mixture models. *Automatic Control and Computer Sciences*, *55*, 15-25.