

Ministère de l'enseignement Supérieur et de la recherche Scientifique  
وزارة التعليم العالي والبحث العلمي

Badji Mokhtar Annaba University  
Université Badji Mokhtar – Annaba  
Faculté de Technologie



جامعة باجي مختار – عنابة

كلية التكنولوجيا

قسم الكهروميكانيك

Département d'Electromécanique

## Thèse

Présentée pour obtenir le diplôme de

## Doctorat LMD

Spécialité : Sécurité Industrielle

Filière : Sécurité Industrielle

Par :

**BENBRAHIM Meriem**

Thème :

**Contribution à la minimisation des défauts par le diagnostic des systèmes industriels**

Thèse soutenue le 21 Avril 2026 devant le jury composé de :

N°	Nom et prénom	Grade	Etablissement	Qualité
01	BELHAMRA Ali	Prof.	Université Badji Mokhtar -Annaba	Président
02	BOURAS Hichem	MCA	Université Badji Mokhtar -Annaba	Rapporteur
04	BOUZAOUIT Azzeddine	Prof.	Université 20 Aout 1955 -Skikda	Examineur
05	OUADA Mehdi	MCA	Université Badji Mokhtar -Annaba	Examineur
06	CHOUHAL Ouahiba	MCA	Université Abbes Laghrour -Khenchela	Examinatrice

## " المساهمة في تقليل العيوب من خلال تشخيص الأنظمة الصناعية "

### الملخص:

يركز العمل المقدم في هذه الأطروحة على اكتشاف وتشخيص العيوب في الأنظمة الصناعية. والهدف الرئيسي كان المساهمة في تطوير طريقة التحسين العشوائي وتطبيقها على نظام كيميائي صناعي للتحقق من أدائها. ولحل مشكلة تشخيص الأخطاء في الأنظمة الصناعية، تم اقتراح طريقة استدلالية. هذه الطريقة هي خوارزمية جديدة IDEOA (خوارزمية تحسين التوازن المنفصل المحسنة)، والتي تهدف إلى حل مشاكل التحسين المنفصلة المختلفة وإنشاء نموذج اكتشاف العيوب القائم على القواعد، والذي يمكن تفسيره بسهولة من خلال قراءة قواعد التصنيف الخاصة به. كان التحسين الذي تم إجراؤه هو اعتماد استراتيجية OBL (التعلم القائم على العكس) في IDEOA من أجل تجنب الوقوع في الأمثلة المحلية. إن المساهمات الرئيسية لهذه الدراسة هي التطبيق المبتكر لهذه المنهجية على نظام Tenesse Eastman. ونتج عن هذه الدراسة نموذج لتشخيص الأخطاء يتكون من 16 قاعدة، ستة منها تتعلق بظروف التشغيل العادية والقواعد الأخرى تكشف عن حدوث الخطأ (F4). تم بعد ذلك حساب قيمة الدقة لتقييم فعالية نهجنا من خلال مقارنته بنتائج أخرى للخوارزميات المقدمة في الأدبيات. وتشير النتائج إلى أن النهج المقترح يتفوق على الطرق الأخرى.

**كلمات مفتاحية:** كشف، تشخيص، الأخطاء، التصنيف، توليد القواعد، التحسين، التفسير.

## « Contribution à la minimisation des défauts par le diagnostic des systèmes industriels »

### Résumé :

Les travaux présentés dans cette thèse portent sur la détection et le diagnostic explicatifs des défauts des systèmes industriels. L'objectif principal de cette thèse est de contribuer au développement d'une méthode d'optimisation stochastique et de l'appliquer à un processus industriel chimique pour valider sa performance.

Pour résoudre ce problème de diagnostic des défauts des systèmes industriels, une méthode méta-heuristique a été proposée. Cette méthode est un nouvel algorithme IDEOA (Improved Discrete Equilibrium Optimizer Algorithm), qui vise à résoudre différents problèmes d'optimisation discrets et à générer un modèle de détection de défauts basé sur des règles, facilement explicable par la lecture de ses règles de classification. La contribution algorithmique de ce travail est d'adopter la stratégie d'apprentissage par opposition OBL (Opposite Based Learning) dans l'IDEOA afin d'éviter de se retrouver bloqué dans des optima locaux et afin d'améliorer l'exploration du domaine de recherche de l'algorithme proposé. Une autre contribution clé de cette étude est l'application novatrice de cette méthodologie au processus Tennessee Eastman. Le résultat de cette étude est un modèle de diagnostic de défauts composé de 16 règles, dont six concernent les conditions de fonctionnement normales et les autres révèlent l'occurrence de défauts (F4). Une valeur d'Accuracy a été calculée pour évaluer l'efficacité de notre approche en comparant ce résultat avec ceux des autres algorithmes fournis dans la littérature. Les résultats indiquent que l'approche proposée surpasse les autres méthodes en termes d'Accuracy et d'interprétabilité.

**Mots clés :** Détection, Diagnostic, Défauts, Classification, Génération des règles, Optimisation, Interprétation.

## « Contribution to the fault minimization through the diagnosis of industrial systems »

### **Abstract:**

The work presented in this thesis focuses on an explainable fault detection and diagnosis of industrial systems. The main objective was to contribute to the development of a stochastic optimization method and to apply it to a chemical industrial process for validating its performance.

To address this problem of fault diagnosis in industrial systems, a meta-heuristic method was proposed. This method is a novel algorithm, IDEOA (Improved Discrete Equilibrium Optimizer Algorithm), which aims to solve different discrete optimization problems and to generate a rule-based fault detection model, easily explained by reading its classification rules. The improvement made was to adopt the Opposite Based Learning (OBL) strategy in the IDEOA to avoid being trapped in local optima. A key contribution of this study is the innovative application of this methodology to the Tennessee Eastman process. The result of this study is a fault diagnosis model composed of 16 rules, six of them concern normal operating conditions and the others reveal the fault occurrence (F4). An accuracy value was then calculated to evaluate the effectiveness of our approach by comparing it with other results of algorithms provided in the literature. The results indicate that the proposed approach outperforms other methods.

**Key words:** Detection, Diagnosis, Faults, Classification, Rule generation, Optimization, Interpretation.

## **Condoléances**

*C'est avec une immense tristesse que nous avons appris le décès de mon directeur de thèse : monsieur le Professeur : Brahim Hamaidi. Je garde en mémoire son engagement, sa bienveillance, son soutien et ses précieux conseils.*

*En cette douloureuse circonstance je tiens à exprimer à mon nom propre mes sincères condoléances.*

*Avec la disparition du défunt, notre université a perdu un de ses illustres enseignants qui ont marqué une empreinte durable au sein de leurs parcours d'enseignement.*

*Nous nous associons à la peine de sa famille, ses proches et toute la communauté universitaire, et affirmons toute notre solidarité et notre compassion dans cette pénible circonstance.*

*Que le Tout Puissant Allah l'accueille dans son paradis éternel. Amin.*

*"انا لله وانا اليه راجعون"*

## **Remerciements**

*Je tiens par la présente à exprimer mes plus sincères remerciements à mon directeur de thèse le défunt « **Monsieur le Pr. HAMAIDI Brahim** ». Je tiens aussi, à exprimer ma profonde gratitude envers lui pour son encadrement, et son engagement tout au long de mon parcours de recherche de doctorat.*

*Son expertise, sa patience, sa disponibilité et son sens critique ont grandement contribué à la réalisation de ce travail. Je tiens à lui remercier pour ses orientations, et son assistance qui font avancer mon mémoire.*

*Toutes mes sincères condoléances à sa famille.*

*Je tiens également à exprimer ma profonde gratitude envers mes enseignants : monsieur le Prof. « **MEHDAOUI Rafik** » et monsieur le Prof. « **HAOUASSI Hichem** », qui ont suivi de près mon travail. J'ai beaucoup apprécié votre rigueur scientifique et vos précieuses suggestions qui m'ont beaucoup apporté. Merci pour votre temps et votre disponibilité qui a été touché par votre honnêteté intellectuelle. Je vous remercie particulièrement pour votre soutien, vos aides et vos orientations qui ponctuent ce mémoire. Aucun mot ne saurait exprimer la reconnaissance que je vous dois.*

*Je remercie également et chaleureusement madame le Professeur « **MOUSS Leïa-Hayet** » pour son aimable bienveillance, ses orientations et son assistance.*

*J'exprime mes sincères respectueux remerciements au Mr. Le Professeur : « **BELHAMRA Ali** » pour avoir fait l'honneur de présider mon jury.*

*J'adresse mes plus vifs remerciements au monsieur **BOURAS Hichem** d'avoir accepté d'être le rapporteur de cette thèse, merci pour avoir assumé cette tâche avec beaucoup de sérieux.*

*Je remercie sincèrement Mr. le Prof. « **BOUZAOUIT Azzedine** », Mr. « **OUADA Mehdi** », et Mme « **CHOUHAL Ouahiba** » Pour avoir accepté d'examiner mon travail. Je leur suis très reconnaissante pour l'intérêt qu'ils y ont porté.*

*Je tiens également à exprimer ma reconnaissance envers mes collègues, tout le personnel et les enseignants de département électromécanique, sans oublier les membres de laboratoire Génie Electromécanique « **LGEM** ».*

*Enfin, je voudrai exprimer ma gratitude et mes chaleureux remerciements envers toute ma famille : mes parents, mon époux, sans oublier les parents de mon époux pour leurs sacrifices, leur constant soutien, et leur encouragement indéfectible tout au long de mes études. Votre amour, votre confiance et vos encouragements ont été une source d'inspiration pour moi.*

*Encore une fois, merci du fond du cœur à tous ceux qui, de près ou de loin, ont contribué à la réalisation de ce travail. Merci pour leur soutien inestimable.*

## *Dédicaces*

*Je remercie mon dieu, le tout puissant, notre créateur, qui m'a accordé ce succès et qui m'a aidé à surmonter les moments difficiles pour réaliser ce grand objectif de ma vie.*

*Je dédie ce travail avec beaucoup d'amour à toute ma famille pour leur soutien inconditionnel, pour toujours me pousser à être meilleure et à réussir ma carrière. Ma mère et mon père qui m'ont inculqué de bonnes valeurs et m'ont offert la meilleure éducation. Merci mes parents d'être à mes côtés et de m'avoir soutenu durant toute ma vie.*

*A mon mari, pour m'avoir donné sa force et son inspiration. Pour avoir toujours cru en moi et me dire chaque jour que je pouvais le faire. Cette réalisation est aussi la vôtre.*

*Merci aux parents de mon mari pour avoir toujours été à mes côtés pour me soutenir à tout moment.*

*Pour mes enfants Ghayth et Aoues, vous avez été ma plus grande joie et inspiration pour finir ce travail. Je vous aime profondément.*

*Et à tous mes frères et mes sœurs qui ont contribué à la force qui m'a amené jusqu'à ici.*

## **Liste des figures**

<i>Figure I. 1. Catégorisation des défauts en fonction de leur localisation.</i> .....	9
<i>Figure I. 2. Catégorisation des défauts en fonction de leur évolution dans le temps.</i> .....	10
<i>Figure I. 3. Illustration des notions : Erreur, Défaillance et Défaut.</i> .....	11
<i>Figure I. 4. Les catégories de défaillance dans la structure d'un arbre.</i> .....	13
<i>Figure I. 5. Les différentes étapes de la démarche du diagnostic.</i> .....	16
<i>Figure I. 6. Les méthodes de diagnostic et leur classification.</i> .....	17
<i>Figure I. 7. Principe des méthodes de diagnostic à base de modèle.</i> .....	18
<i>Figure I. 8. Diagnostic des défauts par l'estimation de paramètres.</i> .....	19
<i>Figure I. 9. Principe de diagnostic à base d'observateur.</i> .....	20
<i>Figure I. 10. Principe de l'approche de RDF.</i> .....	21
<i>Figure I. 11. Schéma d'un réseau de neurone artificiel.</i> .....	22
<i>Figure I. 12. Processus de la classification à base des règles.</i> .....	23
<i>Figure II. 1. Les étapes d'implémentation du KDD.</i> .....	30
<i>Figure II. 2. Techniques pour Data Mining.</i> .....	32
<i>Figure II. 3. L'apprentissage supervisé.</i> .....	32
<i>Figure II. 4. L'apprentissage non supervisé.</i> .....	36
<i>Figure II. 5. Cadre typique d'un entrepôt de données.</i> .....	39
<i>Figure III. 1. Photo représentant le principe de l'EOA.</i> .....	50
<i>Figure III. 2. Coopération des candidats à l'équilibre pour actualiser la position d'une particule.</i> .....	54
<i>Figure IV. 1. Organigramme du TEP [78].</i> .....	64
<i>Figure IV. 3. Variation des variables pertinentes contribuant à l'apparition de défaut F4 en fonctionnement normal et anormal.</i> .....	71
<i>Figure IV. 2. Évolution des variables de TEP.</i> .....	71
<i>Figure IV. 4. Évolution des variables Es18 et Ev10.</i> .....	72
<i>Figure IV. 5. Diagramme à barres de la précision obtenue par l'IDEOA et d'autres méthodes.</i> .....	73

## **Liste des tableaux**

<i>Tableau II- 1- Tableau représentant la forme d'une base de données relationnelle.....</i>	<i>37</i>
<i>Tableau IV- 1- Comparaison de la précision de classification du DEOA et d'IDEOA. ....</i>	<i>68</i>
<i>Tableau IV- 2- Les règles générées à l'aide de l'ensemble de données TEP. ....</i>	<i>69</i>
<i>Tableau IV- 3- Importance de la valeur de chaque caractéristique dans le classificateur à base d'IDEOA créé pour le diagnostic de TEP. ....</i>	<i>72</i>
<i>Tableau IV- 4- Valeur de précision obtenue par l'IDEOA et d'autres méthodes. ....</i>	<i>73</i>

## **Table des matières**

<i>المخلص</i> .....	<i>II</i>
<i>Résumé</i> .....	<i>III</i>
<i>Abstract</i> .....	<i>IV</i>
<i>Condoléances</i> .....	<i>V</i>
<i>Remerciements</i> .....	<i>VI</i>
<i>Dédicaces</i> .....	<i>VII</i>
<i>Liste des figures</i> .....	<i>VIII</i>
<i>Liste des tableaux</i> .....	<i>IX</i>
<i>Table des matières</i> .....	<i>X</i>
<i>Introduction Générale</i> .....	<i>1</i>
<b><i>Chapitre I. Diagnostic des défauts industriels</i></b> .....	<b><i>5</i></b>
I.1 Introduction.....	6
I.2 Historique de diagnostic industriel .....	7
I.3 Concepts et terminologies.....	8
I.4 Principe du diagnostic industriel.....	15
I.4.1 La détection .....	15
I.4.2 L'isolation des défauts.....	15
I.5 Organisation générale de la procédure du diagnostic .....	15
I.6 Les méthodes de diagnostic des défauts .....	17
I.6.1 Méthodes internes (à base de modèle) .....	18
I.6.2 Méthodes externes ou méthodes sans modèle .....	20
I.7 Critères de performance d'un système de diagnostic des défauts.....	23
I.8 Conclusion .....	24
<b><i>Chapitre II. Aperçu sur le processus du Data Mining</i></b> .....	<b><i>26</i></b>
II.1 Introduction.....	27
II.2 Historique du Data Mining .....	27
II.3 Définition du Data Mining.....	28
II.4 Processus d'implémentation du Data Mining.....	29
II.5 Techniques pour Data Mining .....	31
II.5.1 L'apprentissage supervisé .....	32
II.5.1.1 La classification.....	33
II.5.1.2 La régression .....	35
II.5.2 L'apprentissage non supervisé .....	35
II.5.2.1 L'association .....	36

II.5.2.2 La segmentation (Clustering) .....	36
II.6 Transformation et discrétisation des données.....	37
II.7 Types de données pour les applications du Data Mining .....	38
II.7.1 Les bases de données relationnelles .....	38
II.7.2 Entrepôts de données (DataWarehouses).....	39
II.7.3 Les données transactionnelles .....	39
II.7.4 Autres types de données.....	40
II.8 Conclusion .....	40
<b>Chapitre III : L'optimisation et les méta-heuristiques .....</b>	<b>42</b>
III.1 Introduction .....	43
III.2 Définition de l'optimisation .....	43
III.3 Les méthodes d'optimisation.....	44
III.3.1 Les méthodes déterministes .....	44
III.3.2 Les méthodes stochastiques.....	45
III.3.2.1 Les méthodes heuristiques .....	45
III.3.2.2 Les méta-heuristiques .....	46
III.4 Critères de choix des méthodes d'optimisation.....	48
III.5 Le nouvel algorithme amélioré « IDEOA » .....	48
III.5.1 Introduction.....	48
III.5.2 Historique .....	49
III.5.3 Description de l'algorithme EOA .....	50
III.5.4 Le DEOA.....	54
III.5.5 Principe de l'IDEOA .....	57
III.6 Conclusion.....	60
<b>Chapitre IV : Détection et diagnostic des défauts basé sur un IDEOA appliqué à un processus industriel.....</b>	<b>61</b>
IV.1 Introduction .....	62
IV.2 Description du système d'étude de cas .....	62
IV.3 Historique de diagnostic des défauts du TEP.....	65
IV.4 Présentation de l'approche de détection et diagnostic des défauts suggérée .....	66
IV.4.1 Prétraitement de l'ensemble des données.....	67
IV.4.2 Génération des règles .....	67
IV.5 Résultats du diagnostic des défauts par l'IDEOA.....	68
IV.5.1 Evaluation de l'IDEOA et réglage des paramètres.....	68
IV.5.2 Comparaison entre les résultats du DEOA et l'IDEOA .....	70
IV.5.3 Génération des règles à l'aide d'IDEOA .....	71

IV.5.4 L'importance des variables des données de TEP .....	72
IV.5.5 Comparaison des résultats de l'IDEOA avec ceux d'autres approches .....	74
IV.6 Conclusion.....	76
<b>Conclusion Générale.....</b>	<b>77</b>
<b>Références bibliographiques.....</b>	<b>79</b>
<b>Annexe A .....</b>	<b>88</b>
<b>Annexe B .....</b>	<b>91</b>
<b>Annexe C .....</b>	<b>92</b>
<b>Annexe D.....</b>	<b>93</b>

## ***Acronymes et Notations***

<b>ACP</b>	<i>Analyse en Composante Principale</i>
<b>AFNOR</b>	<i>Association Française de Normalisation</i>
<b>DL</b>	<i>Deep Learning</i>
<b>FDD</b>	<i>Fault Detection and Diagnosis</i>
<b>FN</b>	<i>Fault Negative</i>
<b>FP</b>	<i>Fault Positive</i>
<b>GA</b>	<i>Genetic Algorithm</i>
<b>IA</b>	<i>Intelligence Artificielle</i>
<b>IDEOA</b>	<i>Improved Discrete Equilibrium Optimizer Algorithm</i>
<b>IFAC</b>	<i>International Federation of Automatic Control</i>
<b>IEV</b>	<i>International Electromechanical Vocabulary</i>
<b>KDD</b>	<i>Knowledge Data Bases</i>
<b>KELM</b>	<i>Kernal Extreme Learning Machine</i>
<b>LAD</b>	<i>Logical Analysis Data</i>
<b>LRD</b>	<i>Réduction de la Diversité Linéaire</i>
<b>MEM</b>	<i>Elimination des Minimas Locaux</i>
<b>ML</b>	<i>Machine Learning</i>
<b>OBL</b>	<i>Opposite Based Learning</i>
<b>PNN</b>	<i>Probabilistic Neural Network</i>
<b>RBC</b>	<i>La Classification à Base des Règles</i>
<b>RDF</b>	<i>Reconnaissance De Forme</i>
<b>RDP</b>	<i>Réseaux De Pétri</i>
<b>RNA</b>	<i>Réseaux de Neurone Artificielle</i>
<b>RRA</b>	<i>Relations de Redondance Analytique</i>
<b>SA</b>	<i>Simulated Annealing</i>
<b>SGBD</b>	<i>Système de Gestion de Base de Données</i>
<b>SPE</b>	<i>Erreur quadratique de Précision</i>
<b>SVM</b>	<i>Support Vector Machines</i>
<b>TEP</b>	<i>Tennessee Eastman Process</i>
<b>TN</b>	<i>True Negative</i>
<b>TP</b>	<i>True Positive</i>

## ***Introduction Générale***

### **1.1 Contexte et objectifs**

Une augmentation considérable et en continu du nombre des accidents et des incidents industriels, provoquée principalement par des altérations des systèmes industriels, et due au progrès et développement croissant des techniques industrielles, a été constaté au cours de ces dernières années. Indépendamment de toutes les précautions et les améliorations de la sécurité dans ce domaine. De telles situations, ont incité les industriels à se concentrer davantage sur la fiabilité et la sûreté de fonctionnement des systèmes, pour minimiser autant que possible le risque d'un comportement intempestif et inattendu qui peut causer des conséquences plus au moins graves ; allant de la défaillance qui peut causer l'arrêt partiel d'une partie du système jusqu'à l'arrêt total de la fonction du système. Il apparait donc très nécessaire de doter des moyens de prédiction de ces comportements imprévus pour intervenir en amont dans le but d'éviter tous risques matériels, humains ou environnementaux.

Notre travail s'appuie principalement sur le diagnostic des défauts des systèmes industriels. Au cours de ces dernières décennies, ce module de diagnostic de défauts industriels qui consiste principalement à détecter et de localiser les défauts [1] ; s'est imposé comme une préoccupation incontournable pour les entreprises. Il consiste ; dans une première phase ; explicative, à identifier et localiser l'élément défectueux et à en rechercher les différentes causes qui peuvent altérer le système et conduire à l'interruption ou à l'arrêt complet de la production. Dans une phase ultérieure, plus prospective, le diagnostic vise à examiner les impacts qu'un défaut peut provoquer sur l'ensemble du système de manière rapide et précoce. Ce qui permet d'apporter une contribution plus significative au système décisionnel, en améliorant ainsi la disponibilité, la fiabilité, ainsi que la productivité [2]. De ce fait, le diagnostic est devenu un atout majeur, sa maîtrise au sein d'une entreprise contribue à l'amélioration de la compétitivité de l'outil de production.

Ce module de diagnostic repose aujourd'hui largement sur des approches basées sur les données, notamment les modèles d'apprentissage automatique. Toutefois, ces méthodes présentent une limite majeure : les modèles les plus performants, tels que les approches de type boîte noire, manquent d'interprétabilité, ce qui limite leur utilisation dans des environnements industriels critiques. Ainsi, la problématique de cette thèse consiste à répondre à la question suivante :

Comment concevoir un modèle de diagnostic des défauts industriels à la fois performant et interprétable, capable d'exploiter efficacement les données tout en fournissant des règles compréhensibles pour les experts ?

Durant ces dernières années, de nombreux travaux ont été consacré à résoudre ce problème de diagnostic des défauts des systèmes industriels, ce qui a suscité un grand intérêt auprès des industriels. Il implique l'utilisation de toutes les connaissances disponibles concernant le système en question afin d'identifier et de détecter les défauts pouvant affecter le système. Afin de résoudre ce problème de diagnostic des défauts industriels ; la littérature a fait l'état sur plusieurs méthodes qui ont été développé dans ce domaine. On peut distinguer deux catégories pour ces méthodes, suivant le type de l'information utilisé : les méthodes internes et les méthodes externes. On peut citer par exemple : les RDP (Réseaux De Pétri), l'Espace de Parité, l'approche d'estimation d'état par observateur, les réseaux de neurone, les méthodes de reconnaissances de formes et les méthodes d'optimisation... Pour choisir l'une de ces techniques, ça dépend à la fois de l'information que l'on désire obtenir sur le système, et du degré de complexité de ce dernier. En fait, plusieurs méthodes sont évoquées dans la littérature, dans ce travail on s'intéresse aux méthodes d'intelligence artificielle (IA), et plus précisément à la classification à base des règles.

Dans la littérature, plusieurs techniques d'IA sont mentionnées. Les méthodes de boîte noire ; telles que les réseaux de neurones artificiels (RNA), et les méthodes à boîte blanche [3], comme les méthodes de classification à base des règles (Rule based classification RC). La différence entre les deux stratégies consiste dans l'explicabilbilité des résultats obtenus lors de diagnostic des défauts industriels. La première catégorie des méthodes (modèles de boîte noire) possède l'avantage de construire des modèles précis, mais leur processus interne de prise de décisions n'est pas lisible par les experts du domaine ; d'où la difficulté de proposer des modèles interprétables. Cependant, lorsqu'il s'agit de diagnostic de défaut, et afin de prendre des décisions précises dans ce domaine, l'utilisateur doit interpréter les résultats acquis, d'où apparait la nécessité d'utiliser les modèles à boîte blanche, grâce à leur pouvoir de générer des modèles hautement interprétables [3].

Ce modèle de classification à base des règles est généré à l'aide d'un algorithme de classification bien adapté, dans ce travail, l'algorithme développé pour résoudre le problème de FDD (Fault Detection and Diagnsis) est une méthode d'optimisation méta-heuristique.

Les méta-heuristiques développées dans cette étude, sont des approches d'optimisation computationnelle utilisées pour trouver une solution ou des solutions aux problèmes compliqués, généralement lorsque les méthodes traditionnelles sont insuffisantes. Ces méthodes sont fréquemment inspirées de phénomènes naturels, elles cherchent à examiner l'ensemble de solutions de façon intelligente afin d'obtenir des solutions plus adéquates.

Contrairement à d'autres types de méthodes, les méthodes méta-heuristiques sont largement utilisées dans ce domaine. L'atout majeur de ces techniques réside dans leur capacité de construire un modèle de diagnostic des défauts industriels en utilisant un ensemble de données collectées sans utiliser le modèle du système.

Dans ce contexte, et pour surmonter les problèmes d'optimisation, les chercheurs ont développé plusieurs méthodes méta-heuristiques comme : les algorithmes génétiques (Genetic Algorithms GA) [4], le recuit simulé (Simulated Annealing SA) [5]-[6], l'optimisation par essaim de particules (particle swarm optimization) [7]... Une nouvelle méthode méta-heuristique appelée Equilibrium Optimizer Algorithm (EOA), inspiré de la physique [8], et basée sur le principe du phénomène d'équilibre du volume de contrôle, a été utilisée afin de trouver des solutions optimales. Cette méthode a été récemment proposée par Faramarzi et al. [8] pour résoudre des problèmes d'optimisation continue. Haouassi dans son ouvrage [9] a proposé une version discrète de l'algorithme EOA, pour traiter les problèmes discontinus DEOA.

Plusieurs contributions ont été marquées dans ce travail, soit sur le plan méthodologique ou expérimental. Ces principales contributions sont les suivantes :

- Le plan méthodologique : la principale contribution sur ce plan consiste à proposer une version améliorée de l'algorithme IDEOA intégrant la stratégie Opposition-Based Learning (OBL) afin d'améliorer l'exploration de l'espace de recherche. Une autre contribution sur ce plan est de proposer une approche hybride combinant à la fois la méthode d'optimisation améliorée (IDEOA) et la technique de classification à base des règles afin de concevoir un modèle de diagnostic industriel hautement interprétable, facilitant la compréhension et l'exploitation des résultats par les experts.
- Le plan expérimental : sur ce plan, et afin de valider l'efficacité de l'approche proposée dans un contexte réaliste, nous l'avons appliquée à l'ensemble de données de TEP. La performance de cette approche était évaluée et comparée avec celle des autres approches existantes dans la littérature.

## **1.2 Organisation de mémoire**

Pour répondre à la problématique posée dans ce travail, nous avons choisi de suivre les étapes suivantes :

Dans le premier chapitre nous présentons une brève définition des différents concepts et terminologies relatifs au diagnostic des défauts industriels, en clarifiant la différence entre différents termes utilisés dans ce domaine. Ensuite, on a présenté le principe et la procédure du diagnostic industriel. Ainsi qu'un bref historique sur les techniques et les méthodes principalement évoquées dans ce domaine est présenté. Une analyse non exhaustive de ces méthodes est faite pour mieux cerner le choix de notre méthode.

Dans le deuxième chapitre, nous avons introduit et expliqué en bref les concepts de base de Data Mining, les principales techniques, et leurs domaines d'application, en détaillant la tâche de classification qui est retenue dans cette thèse pour résoudre notre problème de FDD.

Dans le troisième chapitre, l'objectif principal consiste à donner une brève présentation des méthodes d'optimisation. L'approche de « Discrete Equilibrium Optimizer Algorithm » DEOA est abordée dans cette partie. Cette approche élucidé par (MALIK & HAOUASSI) dans [9], est améliorée en introduisant la technique OBL « Opposite Based Learning » pour mieux explorer l'espace de recherche et afin de s'échapper d'être piégé dans les optima locaux.

Le quatrième et le dernier chapitre, présente l'approche de détection et de diagnostic des défauts suggéré dans ce travail. Cette approche vise à générer un modèle de diagnostic et de classification des défauts à base des règles. Les règles obtenues facilitent l'explication et l'interprétation des résultats. Ensuite, pour valider cette approche, nous l'avons appliqué sur l'ensemble de données de TEP. Dans la dernière section de ce chapitre, les résultats obtenus sont évalués et comparés avec ceux fournis dans la littérature.

Le but de notre travail est d'étudier la performance de l'IDEOA, en calculant sa fonction d'Accuracy et de présenter l'importance des variables pertinentes dans l'occurrence de défaut F4. D'autres critères ont été utilisés dans cette étude pour évaluer l'interprétation du modèle de diagnostic de défauts généré par notre approche. Ces critères incluent le nombre de règles, qui définit la taille globale du classifieur, et la taille des règles.

En conclusion de ces travaux, nous pouvons rappeler les majeurs apports de cette étude, en identifiant les principales perspectives de recherche à travailler sur.

## ***Chapitre I. Diagnostic des défauts industriels***

I.1 Introduction

I.2 Historique de diagnostic industriel

I.3 Concepts et terminologies

I.4 Principe du diagnostic industriel

I.5 Organisation générale de la procédure du diagnostic

I.6 Les méthodes de diagnostic des défauts

I.7 Critères de performance d'un système de diagnostic des défauts

I.8 Conclusion

## **I.1 Introduction**

Dans nos jours, l'importance de la sûreté de fonctionnement et la fiabilité des systèmes industriels est primordiale dans la préservation et la protection des systèmes contre les différents comportements intempestifs et inattendus qui peuvent influencer leur bon fonctionnement. Alors, le diagnostic des défauts qui est un moyen de prédiction de ces comportements imprévus, est apparu et a été développé par plusieurs savants, non seulement pour la détection précoce des défauts mais aussi pour agir et prendre la décision en amont de l'occurrence de ces défauts.

Plusieurs approches ont été développées dans le but de diagnostiquer les différents défauts qui peuvent altérer ces systèmes industriels ; en tenant compte des caractéristiques des données présentes sur le système ou le type de défauts à diagnostiquer conduisant à la mise en place de stratégies adaptées. L'accent est particulièrement mis sur l'élaboration de systèmes intelligents pour détecter et diagnostiquer les principaux défauts relatifs aux systèmes industriels, en raison des besoins grandissants d'un fonctionnement plus sécurisé, sûr et fiable. Ce qui permet aussi d'accroître la productivité et la qualité du matériel industriel tout en réduisant leur coût de maintenance.

Alors que, le diagnostic des défauts des systèmes industriels permet de détecter, d'identifier et de localiser un mode de fonctionnement anormal [1]. Le diagnostic industriel peut alors s'expliquer sous la forme de la relation de cause à effet sous l'angle d'une relation de causalité déterministe, il est mis en place lorsqu'un problème est survenu sur n'importe quel type de composant, que ce soit pendant son utilisation ou lors de sa production.

Dans ce premier chapitre, après un rappel des divers travaux réalisés dans ce domaine, une brève définition des différents concepts et terminologies, utilisés dans la littérature, propres aux défauts, défaillances, dégradation, faute et panne, qui peuvent mettre en péril et qui peuvent causer des effets graves et néfastes pour les installations industrielles, l'environnement et la vie humaine, a été présenté. Ensuite, nous allons présenter le principe de diagnostic, ainsi que sa procédure de mise en œuvre. A la fin de ce chapitre, les différents outils et techniques de diagnostic et leurs objectifs : (Les méthodes symboliques, avec modèles et sans modèles) sont abordés, une analyse non exhaustive de ces méthodes est faite pour mieux cerner le choix de notre méthode.

## I.2 Historique de diagnostic industriel

Au cours de ces dernières années, de nombreux travaux font l'objet d'étudier la problématique de détecter et de diagnostiquer les défauts relatifs aux systèmes industriels. Ce diagnostic des défauts qui est un processus visant à détecter un défaut, à localiser sa source, et à en identifier les principales causes responsables de son occurrence est considéré comme un processus d'investigation des raisons et des conséquences d'un aléa [10]. L'idée générale est de mettre en regard les données collectées lors de l'opération fonctionnelle et réelle du système avec celles issues de notre savoir sur l'opération fonctionnelle normale ou nominale.

En vue du nombre des travaux et des publications concernant ce sujet, on peut constater le grand intérêt accordé au diagnostic des défauts dans différents domaines, et qui est devenu un sujet crucial à traiter actuellement à l'échelle mondiale. Une grande variété des méthodes est adoptée par les auteurs pour traiter cette problématique, parmi ces méthodes on peut trouver les méthodes basées sur la conception d'un modèle ; ces méthodes consistent sur le concept de la connaissance physique du système à diagnostiquer, en construisant un modèle du système soit mathématique ou bien graphique ou les deux à la fois, afin de comparer l'écart entre le comportement réel et observé de ce système avec celui estimé par son modèle. Cet écart qui représente l'occurrence d'un défaut, permet la détection de ce dernier en facilitant la tâche de prise de décision.

Les méthodes à base du modèle furent apparues au début des années 70 avec Willsky, 1976 ; Basseville 1988 ; Gentil 1997 ; Isermann et Balle 1997 ; Fussel et Isermann 1998. La difficulté de ces méthodes à base du modèle, réside dans la conception du modèle soit mathématique ou bien graphique pour le système, il est donc extrêmement difficile, voire impossible, d'obtenir le modèle précis en raison de la complexité des phénomènes impliqués. Ce qui a amené plusieurs auteurs et chercheurs d'opter aux méthodes sans modèles, lorsqu'aucun modèle n'est exploitable. Ces méthodes sans modèles reposent principalement sur la connaissance à priori du système, glanée sur l'expertise humaine, en exploitant l'éventuel historique pour extraire les divers modes de fonctionnement du processus. Cette connaissance passée sur le système constitue une base de données qui est confrontée aux mesures réalisées en temps réel pour réaliser et mettre en place la procédure du diagnostic des défauts. Un grand succès a été obtenu par l'implémentation de ces méthodes dans plusieurs domaines de l'industrie, Parmi les méthodes les plus couramment utilisées on peut citer : SVM (Support Vector Machines) [11]-

[12], l'analyse par composante principale (ACP) [13], la reconnaissance des formes (RDF) [14], les réseaux de neurones artificielle (RNA) [15].

### **I.3 Concepts et terminologies**

En premier lieu, il est indispensable de définir une terminologie des différents concepts fondamentaux liés au domaine du diagnostic des défauts, vu que cette terminologie est très diversifiée et n'obéit à aucun formalisme unifié. En se reposant, principalement sur la terminologie standardisée proposée par le comité technique IFAC (International Federation of Automatic Control), et basée sur les travaux de Zwingelstein, 1995 ; Isermann et Phalle, 2000. Dans ce qui suit, pour clarifier la suite de ce document, une liste non exhaustive de définitions est fournie, étant donné qu'il est fréquemment possible de constater dans la littérature internationale plusieurs définitions différentes pour un même concept.

**a. Défaut :** tout écart ou déviation d'une caractéristique ou d'un paramètre spécifique à une entité par rapport à son fonctionnement normal ou à sa caractéristique de référence [16]. Un défaut ne constitue pas un obstacle à la réalisation de la fonction pour laquelle le système a été conçu [17]. Cependant, à certains moments, en raison d'un dysfonctionnement du à l'apparition d'un défaut, le système peut conduire à une probable défaillance et même à un arrêt total ou partiel suite à une panne [10]. Le défaut peut être aussi défini comme : « l'état d'un élément qui se caractérise par son inaptitude à accomplir une fonction pour laquelle il a été conçu, à l'exception dans le cas de son indisponibilité durant la tâche de la maintenance préventive ou d'autres interventions programmées, ou en raison d'une insuffisance des ressources externes » (IEV 191-05-01).

#### **Différents types de défauts**

Un défaut est toute déviation qui peut altérer le bon fonctionnement du système. Dans la littérature, on procède à la classification des défauts d'une part, selon leur localisation ou en fonction du composant du système qu'ils affectent ; et d'autre part selon leur évolution dans le temps ou bien selon leur dynamique. Ainsi, ils peuvent être classés par leur degré de gravité [18].

##### **a.1 Catégorisation des défauts en fonction de leur localisation**

Les défauts sont des événements qui peuvent survenir à divers emplacements du système. Définissant alors les types de défauts selon le composant ou l'élément du système qu'ils affectent. Celui-ci peut être divisé en trois principaux types distincts voir la figure I.1 :

**a.1.1 Les défauts capteurs :** peuvent se traduire par une lecture erronée donnée par les instruments qui mesurent les grandeurs physiques, qui expriment l'état interne d'un système. Alors, ce type de défauts peut s'exprimer par toute variation anormale d'une mesure qui donne une représentation inadéquate de la condition physique du système. Ils peuvent être relatifs seulement à une partie du système, générant une valeur qui correspond plus au moins la véritable valeur mesurée, ou total qui donne une valeur qui n'est pas en rapport avec la vraie valeur de la grandeur à mesurer.

**a.1.2 Les défauts actionneurs :** interviennent au niveau opérationnel du système en affectant le signal d'entrée du système [19]. Ce genre de défaut peut être dû à un arrêt partiel ou complet d'un actionneur. Un arrêt partiel d'un actionneur peut survenir simultanément en fonctionnement normal ou nominal mais en se produisant uniquement en une seule partie ; ce qui peut provoquer une certaine dégradation ou détérioration de leur fonctionnement. Par contre, un arrêt total d'un actionneur se produit lorsque celui-ci rend le système incontrôlable et entraîne une incapacité à donner les commandes.

**a.1.3 Les défauts processus ou appelés aussi défauts composants :** fréquemment les défauts qui ne sont ni des défauts capteur ni des défauts actionneur, on peut les considérer arbitrairement comme des défauts de processus. Ce genre de défaut découle du composant intrinsèque du système, en minimisant son aptitude à réaliser une certaine tâche.

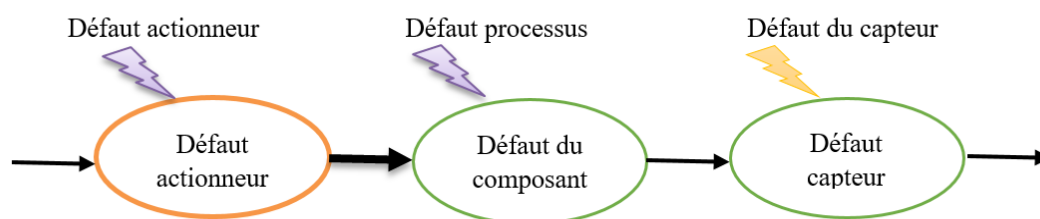


Figure I. 1. Catégorisation des défauts en fonction de leur localisation.

## a.2 Catégorisation des défauts en fonction de leur évolution dans le temps

On peut également classer les défauts en fonction de leur comportement dans le temps [20], ils peuvent être abrupt (brutal), graduel (progressif) ou intermittent, comme indique la figure I.2 [21] :

**a.2.1 Défaut abrupt :** ce type de défaut survient subitement, il se distingue par une rupture abrupte ou brutale dans la progression temporelle d'une variable de processus [19]. Lorsque

cette évolution dans le temps ne correspond pas à l'évolution normale attendue, une panne soudaine de l'élément concerné survient (interruption complète ou partielle) [2].

**a.2.2 Défaut graduel :** il est caractérisé par une évolution temporelle lente et continue des variables concernées. Il s'agit d'un défaut d'encrassement ou d'usure.

**a.2.3 Défaut intermittent :** c'est un cas particulier des défauts brusques, ce type de défaut caractérise les défauts d'alimentation, de connexion voire les faux contacts, qui apparaissent et disparaissent spontanément, puis réapparaissent et disparaissent une autre fois, ce caractère aléatoire lui rend difficile à détecter.

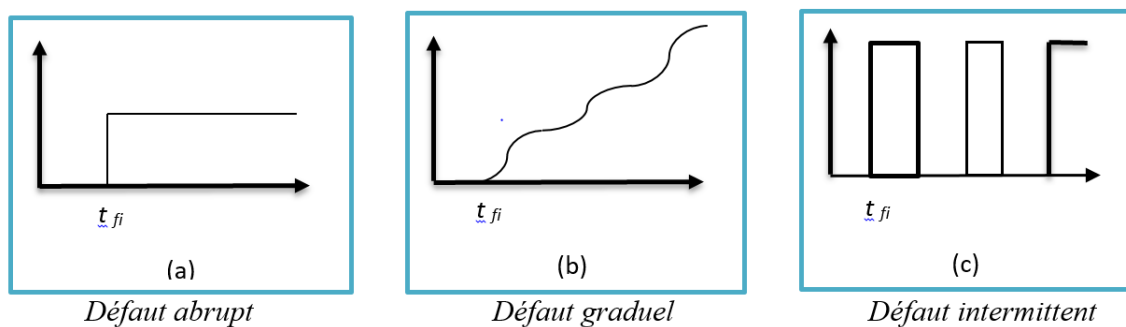


Figure I. 2. Catégorisation des défauts en fonction de leur évolution dans le temps [20].

### a.3 Catégorisation des défauts par ordre de gravité

Un défaut peut être classé selon son degré de gravité ou son effet sur l'exécution performante d'un processus, ces défauts peuvent être purement aléatoires ou progressif [22] :

- Défaut naissant : qui se manifeste progressivement et avec une ampleur grandissante.
- Défaut passager : qui apparaît pendant une durée brève et se manifeste par une déviation significative par rapport à la valeur standard.
- Défaut permanent : qui apparaît d'une manière aléatoire et se manifeste la plus part du temps par un saut brutal du signal. Ce genre de défaut a un impact durable sur le système et peut causer des dommages considérables.
- Défaut catastrophique : défaut qui peut affecter la fonction attendue du système et peut causer l'arrêt total et de manière permanente. Ce type de défauts a des conséquences plus au moins grave et peuvent être irréversibles.

**b. Défaillance :** Une défaillance survient lorsqu'un composant n'est plus capable d'accomplir une ou plusieurs de ses fonctions pour lesquelles il a été conçu. Une défaillance est définie selon

la norme AFNOR comme suit : c'est l'atteinte ou la perte de la capacité d'une entité à réaliser sa fonction demandée pour laquelle elle a été conçue selon les performances établies dans les spécifications techniques. En réalité, un défaut peut conduire à une défaillance. Toutefois, l'inverse n'est pas systématiquement faisable ; une défaillance ne mène pas forcément à un défaut. Suite à une défaillance, l'élément est considéré généralement comme défectueux, et on peut dire qu'il présente un défaut. Le passage de l'état opérationnel à l'état défectueux (présence d'une défaillance) n'est pas toujours évident et visible, et parfois on constate une dérive graduelle par rapport à un état ou à des performances jugées acceptables.

### Exemple d'une défaillance :

Considérons le cas d'une pompe destinée à pomper 100 litres par minute, avec une tolérance acceptable de  $\pm 10$  litres par minute. La pompe est considérée comme opérationnelle si le volume pompé est compris entre 90 et 110 litres par minute. La pompe est jugée comme défaillante lorsque le volume pompé est inférieur ou il dépasse cette plage. Généralement, la sévérité de la défaillance s'accroît en fonction de l'écart par rapport à la valeur souhaitée (100 litres par minute). Cependant, une erreur survient lorsque le volume pompé dévie de sa valeur souhaitée, et se rapproche par exemple, de 110 litres par minute. Si aucune intervention n'est entreprise, une erreur peut évoluer et devenir une défaillance. C'est pour cette raison qu'une erreur est parfois désignée comme « une défaillance naissante ».

Alors, une défaillance peut être due à une erreur, c'est-à-dire « écart entre une valeur ou une caractéristique mesurée, calculée ou observée et la valeur ou la caractéristique réelle, précisée ou idéalement correcte ». (IEV 191-05-24). Une erreur n'est pas considérée encore comme une défaillance, puisqu'elle se retrouve dans les marges acceptables d'écart par rapport à la performance visée (valeur souhaitée). La figure I. 3 illustre les notions d'erreur, de défaillance et de défaut.

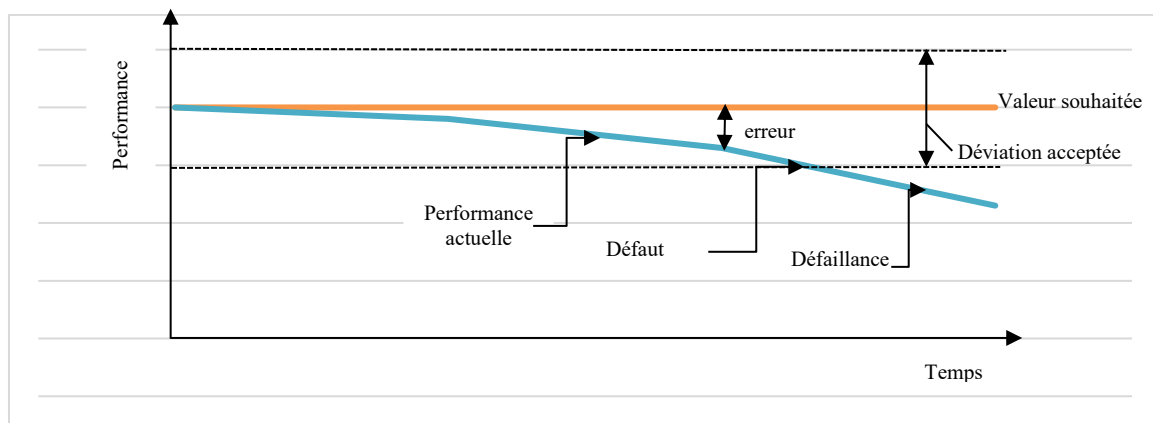


Figure I. 3. Illustration des notions : Erreur, Défaut et défaillance.

## Catégorisation des défaillances

### b.1 Catégorisation des défaillances en fonction de leurs effets

Une défaillance peut avoir des effets différents sur le système. Certaines d'entre ces défaillances n'ont aucun effet sur le système ; d'autres, en revanche peuvent affecter la disponibilité ou la sécurité. Généralement, les défaillances peuvent être classées selon la gravité de leurs effets en 4 catégories [22]:

**b.1.1 Défaillance mineure :** défaillance qui n'affecte pas l'aptitude de l'élément en question d'accomplir sa fonction principale, en causant un dommage négligeable.

**b.1.2 Défaillance majeure :** défaillance qui peut dégrader la capacité d'un élément en question à réaliser sa fonction principale ; sans lui causer de dommages notables.

**b.1.3 Défaillance critique :** défaillance qui peut altérer l'efficacité de fonctionnement d'une entité en lui causant des dommages importants, et qui peut aussi présenter des risques de graves blessures aux personnes.

**b.1.4 Défaillance désastreuse :** défaillance qui conduit à la perte de la fonction principale du système en lui causant des dommages très importants. Cette défaillance peut aussi endommager l'environnement et causer des dommages corporels et parfois mortels.

### b.2 Catégorisation des défaillances en fonction de leur vitesse d'apparition

**b.2.1 Défaillance brutale :** défaillance soudaine qui apparaît aléatoirement et non progressivement ou graduellement dans le système, et qui n'aurait pas pu être anticipée lors des inspections ou des contrôles antérieurs.

**b.2.2 Défaillance graduelle ou progressive :** défaillance causée par la variation temporelle des caractéristiques d'un système. Elle peut être anticipée lors d'une inspection ou un contrôle antérieur.

### b.3 Catégorisation de défaillances en fonction de leurs conséquences et leur détectabilité

Les défaillances matérielles peuvent être classées selon leurs conséquences comme suit :

**b.3.1 Défaillance Dangereuse (D) :** est une défaillance qui met l'équipement dans une condition où il ne peut plus garantir sa ou ses fonctions de sécurité. Quand un équipement se trouve dans cette condition, on parle d'une défaillance dangereuse (D). Lorsque l'équipement

est demandé pour assurer une fonction, alors qu'il présente un défaut, cependant il ne peut plus assurer cette fonction (c'est-à-dire qu'il ne peut pas restaurer l'état de sécurité préétabli).

**b.3.2 Défaillance Sûre (S) :** est une défaillance qui ne laisse pas l'élément dans un état où il est incapable d'assurer sa ou ses fonctions requises. Quand l'élément est dans cette condition, on dit qu'il a un défaut de sécurité.

Une défaillance de sécurité désigne généralement un dysfonctionnement imprévu de la fonction de sécurité, qui place l'élément ou seulement une partie de cet élément dans un mode sécurisé. Les défaillances peuvent être aussi catégorisés selon leur détectabilité en défauts détectés ou non-détectés :

**b.3.3 Défaillance détectée (D) :** une défaillance est définie comme étant détectée lorsqu'elle est identifiée par un test de diagnostic automatique, soit intégré à l'élément, soit relié à un résolveur logique.

**b.3.4 Défaillance non-détectée (U) :** une défaillance non-détectée ou non-diagnostiquée par un test de diagnostic automatique, intégré à l'élément ou relié à un résolveur logique. Ces défaillances non-détectées sont généralement constatées lors des contrôles de validation ou lorsque cet élément est sollicité pour assurer une fonction.

On peut alors différencier les catégories suivantes des défaillances matérielles, ces catégories de défaillance sont illustrées dans la figure I. 4 :

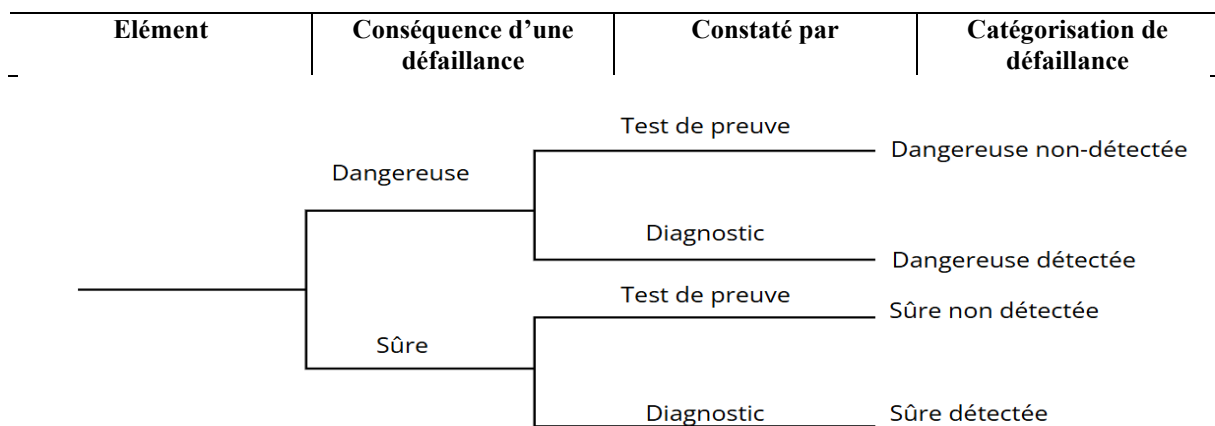


Figure I. 4. Les catégories de défaillance dans la structure d'un arbre.

**Défaillance dangereuse non-détectée (DU) :** ce type de défaillance entravent l'activation à la demande et ne sont constatées que par des tests de validation ou lors d'une sollicitation. On désigne parfois ce genre de défaillances comme des défaillances latentes ou cachées.

**Défaillance dangereuse détectée (DD) :** ce genre de défaillance est identifiée rapidement après son apparition grâce à des tests de diagnostic automatique.

**Défaillance sûre non-détectée (SU) :** ce sont les défaillances non dangereuses et non détectées par un diagnostic automatique.

**Défaillance sûre détectée (SD) :** sont identifiées comme des défaillances sécuritaires ou non-dangereuses détectées facilement par un diagnostic automatique. Dans certaines situations, la détection précoce de ce genre de défaillance peut éviter un déclenchement soudain du système.

**c. Panne :** lorsque l'entité cesse d'accomplir ses fonctions requises, suite à une défaillance. Ce dysfonctionnement peut être momentané ou accidentel [23]. On peut identifier trois types de pannes : une panne intermittente, une panne latente et une panne permanente :

- Une panne est dite intermittente lorsqu'elle est difficilement observable et le système est capable de rétablir son état de fonctionnement nominal lorsqu'il est soumis à une action de réparation.
- La panne latente est une panne cachée, qui existe mais qui ne peut pas être observable.
- Une panne permanente est une panne persistante tant qu'il n'y a pas des actions de réparation corrective.

**d. Faute :** peut être définie comme une action intentionnelle ou non, d'origine physique ou humaine. Une faute physique se présente sous forme de perturbations internes ou externes au système, elle peut être de conception ou d'exploitation du système, tandis que les fautes humaines sont dues aux imperfections humaines.

**e. Dégradation :** représente une diminution progressive ou une perte des performances d'une fonction assurée par un composant ou une entité sans avoir des conséquences fonctionnelles sur l'ensemble du système [24]. Une dégradation se distingue par une progression non-réversible des attributs caractéristiques d'une entité, elle peut être liée aux facteurs directs ; d'utilisation, du temps... ou aux facteurs indirects ; de température, d'humidité...

**f. Diagnostic des défauts :** consiste à déterminer le type, la dimension et l'emplacement d'un défaut, c'est un processus permettant de détecter la présence d'un défaut à un stade préliminaire avant la survenue d'une défaillance, à l'aide des mesures et surtout des observations effectuées sur le système à surveiller. Par conséquent, le diagnostic industriel consiste à exploiter l'ensemble des connaissances disponibles et existantes sur l'entité pour identifier les causes probables d'un défaut encouru ou survenu dans cette entité[25].

**j. Aide à la prise de décision :** permet de restaurer le fonctionnement optimal du système, en suggérant une série d'actions pour accompagner l'opérateur dans le choix adéquat de la bonne

décision face à toute circonstance. A cette étape, le système d'aide à la décision ne prend aucune mesure autonome, mais se limite à informer et à donner seulement des recommandations à l'opérateur, celui qui maîtrise toujours la situation [21].

#### **I.4 Principe du diagnostic industriel**

Le principe du diagnostic industriel consiste à l'identification du ou des défauts susceptibles d'altérer le bon fonctionnement d'un système, en se basant sur les effets occasionnés. Alors, le diagnostic est donc un processus qui implique la détection, l'identification et la localisation d'un défaut pouvant survenir dans un élément ou une partie d'un système industriel. Cette procédure repose principalement sur les deux concepts fondamentaux suivant : d'une part signaler un comportement anormal lorsqu'il y en a une déviation dans le fonctionnement nominal du système, et d'autre part, identifier ce comportement anormal et le localiser afin de pouvoir entreprendre les actions correctrices et nécessaires pour revenir au fonctionnement normal. Ces deux concepts sont spécifiés comme la détection et l'isolation des défauts.

##### **I.4.1 La détection**

La détection des défauts vise à vérifier grâce à des tests, la présence d'un défaut et leurs instants d'apparition en générant des symptômes afin de vérifier la cohérence entre les informations issues du fonctionnement réel du système rapportées par différents capteurs et le comportement attendu ou estimé. L'opération de détection de défauts permet ainsi de déterminer si le système fonctionne convenablement ou non [17].

##### **I.4.2 L'isolation des défauts**

L'étape de détection est généralement suivie par une procédure d'identification et de localisation des défauts pouvant affecter un système, ce qui permet ainsi de cerner le problème en isolant le défaut et en identifiant la partie défectueuse (le ou les composants en défaut), en permettant aussi de déterminer sa cause ainsi que sa sévérité grâce à un raisonnement logique fondé sur les observations fournies par les tests de détection [26].

#### **I.5 Organisation générale de la procédure du diagnostic**

La démarche du diagnostic des défauts passe par plusieurs étapes essentielles qui consistent à [23] :

- L'extraction des informations à base des données liées au fonctionnement normal et anormal du système à diagnostiquer, ces informations sont extraites à partir des moyens de mesure appropriés ou des observations relevées par le personnel.
- La caractérisation des défauts afin de faciliter la tâche de détection d'une anomalie en élaborant ainsi des caractéristiques associées aux symptômes révélateurs des défauts.
- La détection d'un dysfonctionnement : cette étape consiste à surveiller en continu un système ou un processus pour détecter toute déviation ou anomalie par rapport au comportement attendu. Des méthodes statistiques, des graphiques de contrôle et des systèmes d'alarme sont couramment utilisés afin de détecter les différents défauts.
- Localisation de défauts a pour objet d'indiquer quel élément physique ou composant responsable de la défaillance.
- La mise en place d'une démarche de diagnostic des défauts appropriée en utilisant l'ensemble des informations issues du système à diagnostiquer.
- La tâche d'identification a pour but d'identifier la nature spécifique du défaut, afin de le catégoriser selon une catégorie prédéfinie, en fonction de son type et sa sévérité. De plus, cette procédure vise à déterminer l'origine et les causes de déclenchement de ce défaut, en identifiant quel composant ou quel aspect spécifique du processus est affecté [27].

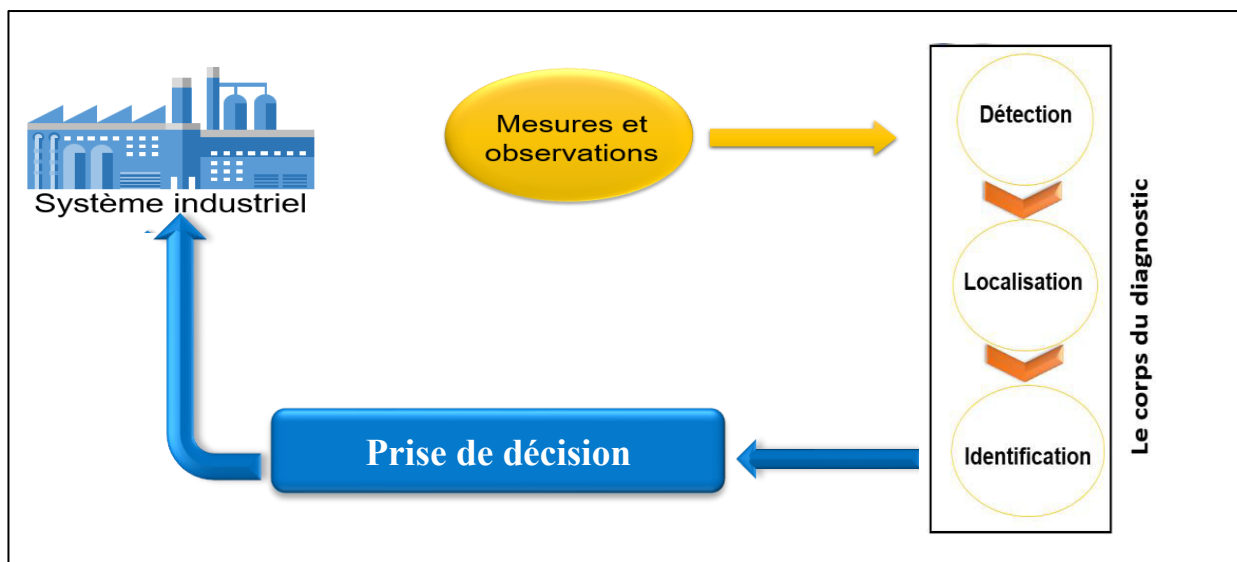


Figure I. 5. Les différentes étapes de la démarche du diagnostic.

Comme le montre la figure I. 5 les trois dernières tâches de détection, de localisation et d'identification forment le corps du diagnostic. Lorsqu'une variation dans le fonctionnement du système est constatée, alors la tâche de diagnostic doit :

- Confirmer qu'il s'agit d'un défaut ;
- Repérer la localisation de défaut pour identifier le composant défectif ;
- Déterminer les caractéristiques d'un défaut afin de l'identifier.

La dernière étape de prise de décision est basée sur les effets que le défaut peut conduire, elle peut être une simple correction qui peut aller jusqu'à l'arrêt total de l'installation.

## I.6 Les méthodes de diagnostic des défauts

Le module de diagnostic industriel s'articule principalement autour de la détection d'un élément en défaut, de le localiser, puis de l'établissement de la cause du problème, et finalement, si c'est faisable de le corriger, afin d'assurer la sécurité et la fiabilité des systèmes. Pour réaliser cette opération et dans le but de prévenir les différentes dégradations des systèmes industriels ; une grande variété des méthodes ou des techniques a été proposée dans la littérature [2], [28], [29], [30]. Dans cette section une classification non exhaustive de ces méthodes est représentée comme indique la figure I. 6. Cette classification est répartie en deux grands axes ; en fonction de la connaissance que l'on désire obtenir sur notre système et également en fonction de la complexité de ce même système. A savoir :

- Les méthodes internes (méthodes à base de modèle)
- Les méthodes externes (méthodes sans modèle)

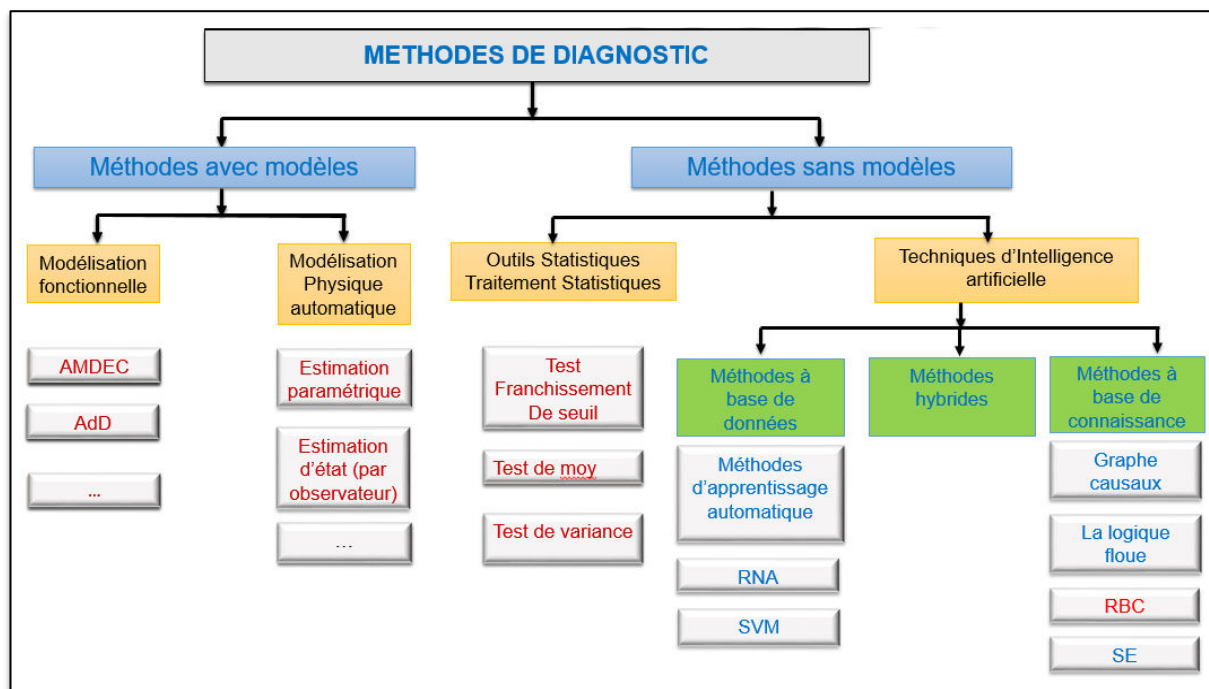


Figure I. 6. Les méthodes de diagnostic et leur classification.

Dans cette thèse, nous intéressons à la deuxième famille (diagnostic sans modèle) qui repose sur la connaissance basée sur l'histoire du processus (le retour d'expérience) et l'expertise humaine. Nous allons présenter dans ce qui suit une synthèse des principales techniques et méthodes utilisées pour le diagnostic des défauts à base de cette classification.

### I.6.1 Méthodes internes (à base de modèle)

Les méthodes basées sur la construction d'un modèle s'appuient sur le principe de comparaison du comportement réel ou observé d'un système par l'intermédiaire des capteurs avec celui-ci prédit par un modèle mathématique ou graphique décrivant son fonctionnement normal. L'écart entre ces deux comportements indique l'apparition d'une anomalie, [31] comme illustre la figure I. 7.

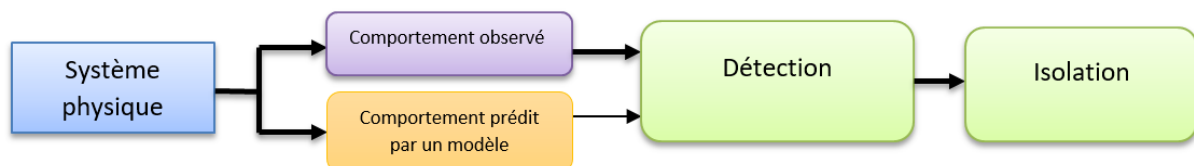


Figure I. 7. Principe des méthodes de diagnostic à base de modèle.

Cette approche de diagnostic à base de modèle implique une connaissance physique profonde sur le fonctionnement du système sous la forme d'un modèle qui doit être validé expérimentalement avant toute exploitation. Trois grandes approches peuvent être distinguées : l'estimation des paramètres, [2]-[19] ; l'estimation d'état (par observateur), et enfin l'espace de parité [32].

- **L'estimation paramétrique**

Cette méthode d'identification paramétrique repose sur le concept d'estimer de manière garantie les divers paramètres du modèle d'un procédé à l'aide des données entrée/sortie et d'en déterminer l'écart entre la valeur des paramètres calculés et ceux mesurés ou de relevés lors de fonctionnement nominal. Cet écart est utilisé comme résidu indiquant la présence ou non d'un défaut.

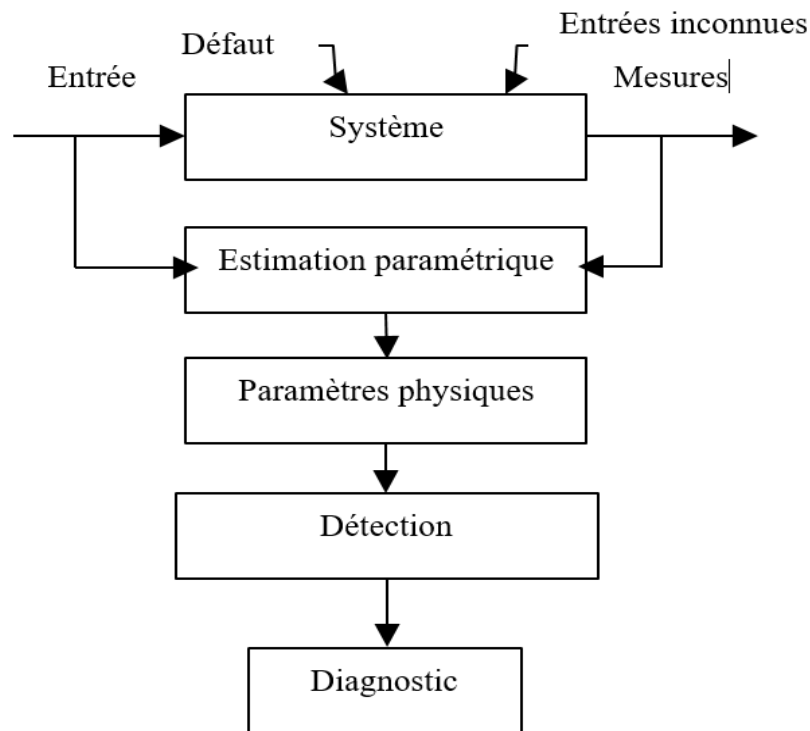


Figure I. 8. Diagnostic des défauts par l'estimation de paramètres.

L'estimation à base des paramètres a pour avantage de fournir de la connaissance sur l'amplitude d'écart. Cependant, l'un de ses principaux désavantages est qu'elle exige de posséder un système matériel constamment stimulé. Ce qui engendre des difficultés pratiques lorsqu'il s'agit de procédé qui fonctionne en mode stationnaire ou d'un procédé dangereux [33].

- **L'estimation d'état (par observateur)**

La méthode de diagnostic par observateur ou par filtre a pour principe de construire un modèle du système permettant d'estimer certaines variables. Le comportement du modèle prédit est comparé au comportement observé du système. Tout écart entre les valeurs mesurées et les valeurs estimées est appelé « résidu », ces résidus sont utilisés pour calculer l'erreur d'estimation de la sortie ce qui permet de détecter la présence ou non d'un défaut. Le principe de cette approche est illustré dans la figure I. 9. Les premiers pionniers de cette approche sont Beard [34], et Jones [35] qui ont proposé des algorithmes de détection à base d'observateurs.

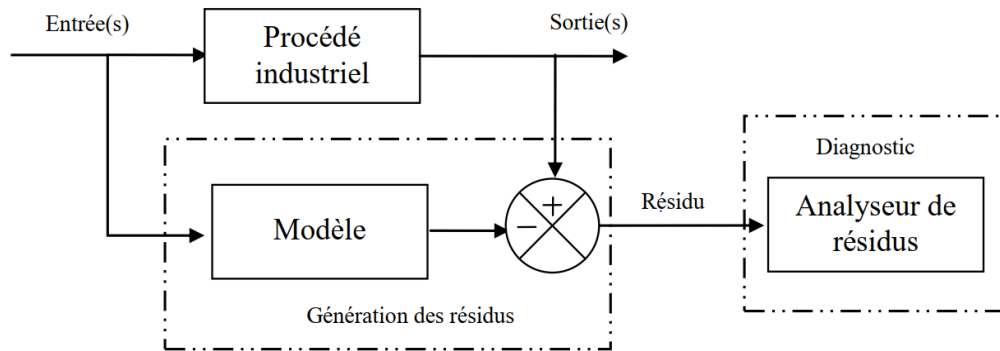


Figure I. 9. Principe de diagnostic à base d'observateur.

- **L'espace de parité**

Cette méthode a pour principe de vérifier la cohérence entre les entrées et les sorties du système à surveiller. Elle est basée principalement sur le concept de reformuler les équations du modèle afin d'établir des relations spécifiques connues sous le nom de Relations de Redondance Analytique «RRA». Ces relations ont la caractéristique de n'associer que des valeurs connues, et disponibles sur le système. Les résidus sont alors générés en remplaçant dans ces RRA les variables connues par leurs valeurs observées lors du fonctionnement du système. L'espace constitué par toutes les valeurs possibles des résidus est nommé espace de parité.

### I.6.2 Méthodes externes ou méthodes sans modèle

Dans de nombreuses applications industrielles, il peut être extrêmement difficile voire impossible d'obtenir un modèle précis du système à cause de sa complexité, de la variabilité des conditions et des interactions entre les composants. Dans tel cas, les méthodes de diagnostic qui ne nécessitent aucune modélisation sont utilisées pour identifier les différentes anomalies et les divers dysfonctionnements des systèmes, lorsqu'aucun modèle dynamique n'est exploitable. Des informations empiriques, des données historiques, des expériences préalables et des règles heuristiques sont exploitées pour tirer les données nécessaires au diagnostic. Voici quelques approches couramment utilisées dans ce contexte :

- **L'analyse en composante principale ACP**

L'Analyse en Composante Principales (ACP) est une technique qui facilite la réduction de la dimensionnalité d'un ensemble de données, en préservant autant que possible l'information contenue dans ces données [36], [37], [29]. Généralement, elle est employée pour l'exploration et la visualisation de données complexes en les projetant dans un espace de dimensions réduites, identifier des schémas et des relations cachées entre les variables. L'ACP consiste

principalement à convertir les variables interconnectés en un nouvel ensemble de variables, appelées composantes principales, non interconnectés entre elles. Lorsque une anomalie est détectée, des valeurs atypiques ou des configurations inhabituelles se manifestent dans les nouvelles dimensions créées par l'ACP.

- **La reconnaissance des formes RDF**

La méthode de diagnostic par RDF consiste en la détermination des différentes configurations ou modèles dans un ensemble de données, dans le but de les confronter avec ceux de référence. A l'aide d'un historique de mesure, cette technique sert à assigner une classe spécifique à partir des observations ou des mesures effectuées sur le système, aux différents modes de fonctionnement, soit défectueux ou non. Afin de prendre la décision d'affecter ou non une nouvelle observation à une classe, il est nécessaire de disposer d'une base de données pour déterminer la classe correspondante au défaut généré [19], [36]. L'outil de prise de décision est fondé sur le résultat de la comparaison de l'observation courante avec celle de référence liée au mode d'opération identifié préalablement. Pour le module de détection d'un dysfonctionnement par cette approche, il s'agit tout d'abord, de déterminer les paramètres et les techniques de prétraitement nécessaires. Ensuite, il s'agit, de diminuer la dimensionnalité de l'espace de caractéristiques, pour accélérer le processus de classification des nouvelles observations d'une part, et à prévenir toute dégradation des performances due à une dimension excessive dans cet espace de caractéristiques d'autre part. A la fin, l'apprentissage des classes peut être fait soit en mode supervisé ou en mode non supervisé.

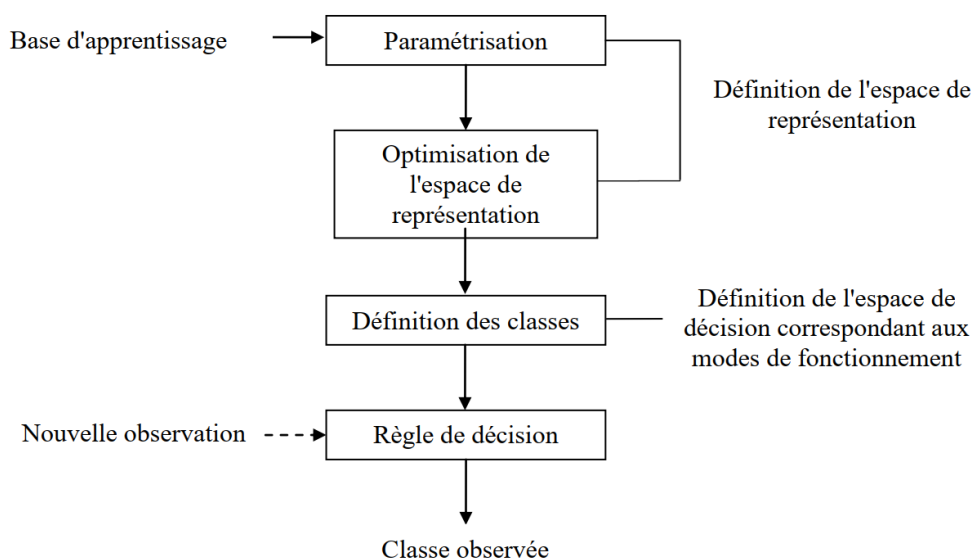


Figure I. 10.Principe de l'approche de RDF.

- **Les réseaux de neurones artificiels RNA**

Dans le cadre de diagnostic des défauts, les RNA sont utilisés soigneusement pour cerner l'intérêt de raisonnement humain par l'intelligence artificielle afin de trouver des solutions aux problèmes de défaillances. Ces réseaux de neurones artificiels RNA inspirés de neurones du cerveau humain, sont en effet construits de nœuds interconnectés, ou neurones, organisés en couches afin de communiquer à l'entrée du réseau artificiel et en phase d'apprentissage les données de simulations au lieu de mesures effectuées par les capteurs, en produisant à la sortie le résultat de diagnostic depuis la détection d'un dysfonctionnement jusqu'à la localisation de l'élément ou le composant en cause [37], [38]. En phase d'auto-apprentissage le réseau apprend à relier non seulement les informations apprises mais aussi les exemples qui leur sont similaires, ce qui reflète une certaine rigidité dans l'application de cette technique. L'architecture d'un RNA, y compris l'agencement des couches, le choix des fonctions d'activation pour les neurones et l'attribution des poids aux connexions neuronales, détermine sa capacité à saisir et à schématiser les liens complexes et non linéaires dans les données [37].

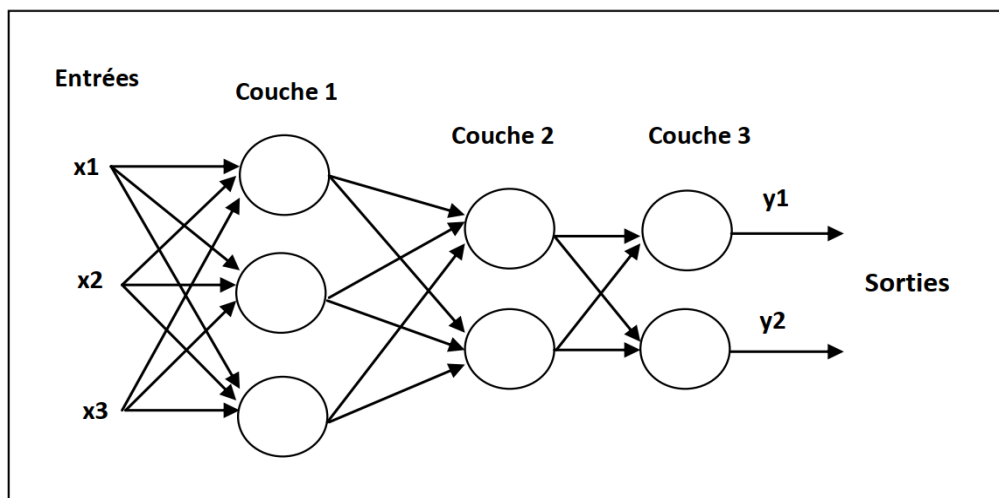


Figure I. 11. Schéma d'un réseau de neurone artificiel.

- **La classification à base des règles**

La classification est une approche d'apprentissage supervisée où un modèle d'apprentissage automatique est formé pour prédire la classe ou la catégorie d'une entrée donnée en fonction de ses caractéristiques. Cela nécessite un ensemble des données contenant à la fois les caractéristiques d'entrée (prédicteurs) et la classe ou la catégorie correspondante qui représente l'état du système (fonctionnement normal ou présence d'un défaut) [39].

Pour l'application dans le diagnostic des défauts, la classification est utilisée afin de déterminer l'état ou les conditions de fonctionnement d'un système. La réalisation de cette approche implique deux étapes principales appelées apprentissage et test. L'ensemble de données est partitionné en deux sous-ensembles : l'ensemble d'apprentissage est utilisé pour construire un modèle qui détermine la relation entre les caractéristiques d'entrée et la classe correspondante à l'aide d'un algorithme de classification approprié [39]. Pendant cette étape d'apprentissage, l'objectif est d'identifier l'ensemble de caractéristiques le plus significatif pour la détection de défauts. Cette étape produit un ensemble de règles de classification également appelé un classifieur. Une fois le classifieur est entraîné, on utilise l'ensemble de données de test pour évaluer les performances et la précision de classifieur obtenu.

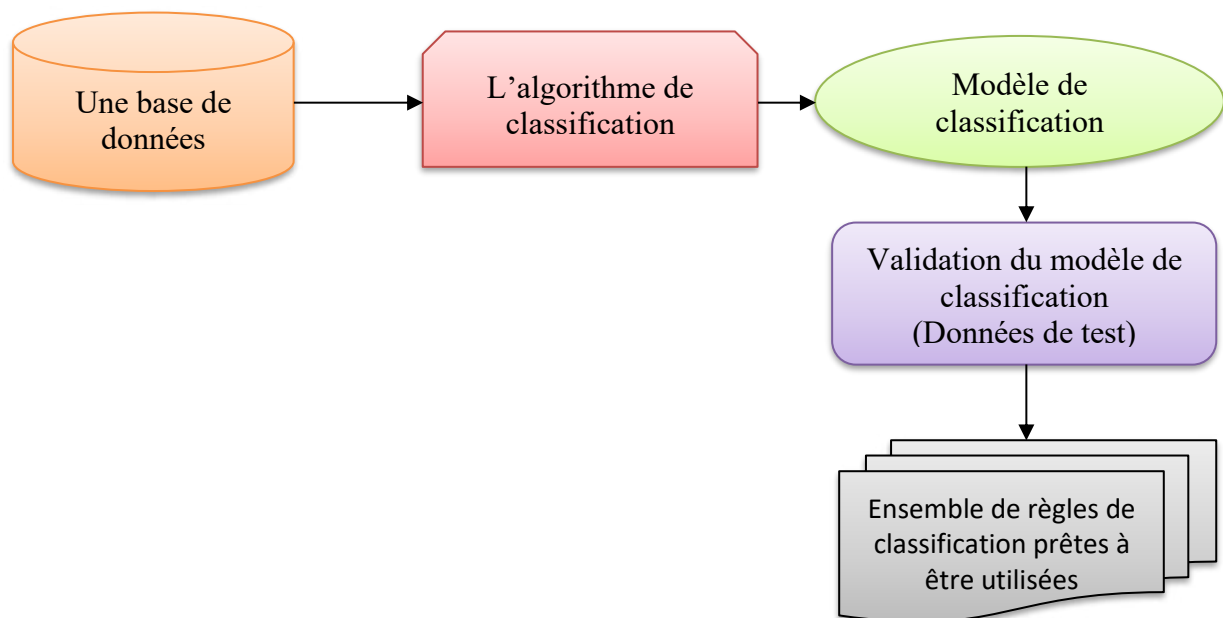


Figure I. 12. Processus de la classification à base des règles.

### I.7 Critères de performance d'un système de diagnostic des défauts

Afin d'évaluer les performances d'un système de diagnostic des défauts, quelques critères généraux sont souvent pris en compte [27]-[40] ; comme :

- La précision : qui détermine l'aptitude d'un système à déceler et à déterminer la localisation correcte des défauts. Une haute précision est essentielle pour éviter les erreurs de diagnostic.
- La sensibilité : est définie comme l'aptitude d'un système à déceler même de petits défauts ou des variations subtiles dans les signaux ou les données de mesure.

- Vitesse de diagnostic : elle désigne le délai requis afin que le système analyse les données et fournisse un résultat de diagnostic.
- La robustesse : il s'agit de l'aptitude du système à maintenir ses performances sous des conditions variables, tel qu'un environnement bruyant, des changements de charge ou des changements de conditions.
- Isolabilité : l'aptitude du système de diagnostic à identifier directement la source ou la cause principale de défaut.
- Interprétabilité : pour les procédés complexes, il s'agit de l'aptitude du système de diagnostic de défaut de donner des résultats lisibles et interprétables.

## I.8 Conclusion

Dans le présent chapitre, nous avons revisité plusieurs définitions et terminologies relatives au diagnostic de défaut, afin de pouvoir différencier les principaux termes liés à ce domaine, comme : défaut, défaillance, panne et faute. Nous aussi présenté brièvement le principe de diagnostic des défauts qui consiste principalement en deux phases consécutives : la détection et l'isolation de défauts.

Pour surmonter ce problème de diagnostic de défauts, plusieurs méthodes ont été évoquées dans la littérature, ces méthodes se composent en deux grandes familles principales : les méthodes internes à base de modèle, et les méthodes externes ou appelées aussi les méthodes sans modèles. Lorsque la modélisation du système s'avère complexe ou parfois difficile à réaliser, à cause de la complexité croissante des procédés, alors il est favorable de s'orienter aux méthodes sans modèles qui représentent une certaine robustesse et performance dans le domaine de diagnostic.

Le diagnostic des défauts, étudié dans cette thèse est vu comme un problème de classification, ce qui en fait de la classification à base des règles un choix judicieux pour cette étude. Cette technique de classification a pour utilité de classer les défauts dans des catégories précises ce qui donne lieu à la détection de défauts. Elle est considérée comme une approche d'apprentissage automatique à base d'algorithme (ML). L'algorithme retenu dans cette étude pour générer un modèle de classification à base des règles est une méthode méta-heuristiques.

Cette technique de classification utilise un ensemble de données pour le diagnostic de défaut d'un système industriel. Les informations recueillies ou les données disponibles sur ce dernier sont très immenses, d'où apparait la difficulté de les traiter pour en tirer de l'information

précieuse. La technique d'exploration de données (data-mining methods) est utilisée dans le but de traitement de cette immense quantité de données. Dans ce qui suit, nous considérons, pour le diagnostic des défauts interprétable ; une approche de classification à base des règles en utilisant une méthode méta-heuristique (Discrete Equilibrium Optimizer Algorithm) afin de pouvoir générer des règles facile à être interpréter par le lecteur.

Pour le prochain chapitre on va présenter un aperçu sur la technique d'exploration des données (Data-Mining) en détaillant la tache de la classification à base des règles.

## ***Chapitre II. Aperçu sur le processus du Data Mining***

II.1 Introduction

II.2 Historique du Data Mining

II.3 Définition du Data Mining

II.4 Processus d'implémentation du Data Mining

II.5 Techniques pour Data Mining

II.6 Transformation et discrétisation des données

II.7 Types de données pour les applications du Data Mining

II.8 Conclusion

## **II.1 Introduction**

Face à l'expansion massive des bases de données et leur utilisation répandue dans différents domaines, les entreprises doivent faire face au problème de la surcharge d'informations. Le défi d'utiliser et d'exploiter toute cette énorme masse de données stockées qui dépassent la capacité humaine, devient un problème majeur pour toutes les entreprises [41].

Dans le secteur industriel, les approches traditionnelles de traitement et de collecte de données ont montré leur insuffisance face à telle situation, ce qui a incité les chercheurs scientifiques à chercher comprendre l'immense valeur de ces données ; en conduisant à la naissance de nouvelles technologies ; telle que le Data Mining.

Le « Data Mining » également appelé la fouille de données ou encore l'exploration de données est une approche d'extraction de la connaissance intéressante [42] (règles, régularités, patterns, contraintes) de l'ensemble de données existantes sur le système. En utilisant un assemblage de techniques et algorithmes pour tirer de la connaissance ou de l'information pertinente et précieuse à partir de grande volume de données en vue d'analyses puis de prise de décisions [43], de prédiction de résultats ou d'obtention de connaissances plus approfondies des données.

Actuellement, de nombreuses techniques puissantes du Data Mining ont émergé pour faciliter l'interprétation de cette énorme quantité de données dans différents domaines tels que : le domaine d'économie, de santé, d'industrie... Où d'importantes masses de données sont recueillies et peuvent être exploitées pour prendre de meilleures décisions. D'autres applications ont été émergé aussi dans le cadre industriel comme l'aide au diagnostic des défauts des procédés de production complexes [44], [45].

Dans ce chapitre nous allons introduire et expliquer en bref les concepts de base de Data Mining, ses principales techniques, et leurs domaines d'application en détaillant la tâche de classification qui est retenue dans notre approche pour le diagnostic industriel.

## **II.2 Historique du Data Mining**

Dans la littérature, le Data Mining, également connu sous le nom de la fouille de données, est souvent aperçu comme le résultat de l'évolution naturelle des technologies de l'information : bases de données, apprentissage automatique (Machine Learning), intelligence artificielle. L'émergence du Data Mining, remonte à plusieurs décennies :

- Aux années 1970 ; à partir des systèmes de traitement de fichiers primitifs. Dans cette époque les industries utilisaient des méthodes statistiques pour l'optimisation de la

production et la détection des anomalies. Cependant, ces méthodes étaient souvent limitées par la capacité faible de traitement des ordinateurs de l'époque.

- Aux années 1980 ; l'analyse avancée des données est apparue au MIT « Massachusetts Institute of Technology » à Cambridge. Elle n'a pas suscité beaucoup d'intérêt car les autres sciences n'étaient pas suffisamment développées. A la fin de ces années, l'intelligence artificielle a été développée, des techniques comme les systèmes experts ont été introduites dans les usines, pour reproduire la prise de décision humaine et fournir des recommandations en cas de survenue d'une anomalie. Ces premiers systèmes étaient limités par leur rigidité et leur manque d'adaptabilité.
- Les fondements du Data Mining n'ont été établis qu'en 1990, résultant de la convergence de multiples facteurs, tant technologiques, économiques que sociopolitiques. Les masses considérables de données représentent des réserves d'informations stratégiques tant pour les décideurs que pour les utilisateurs. Dans le domaine industriel, l'intégration de système de contrôle et de surveillance plus sophistiqués a permis la collecte massive de données de capteurs, d'indicateurs de performance de machines, et d'autres variables critiques.

Dans le diagnostic industriel, l'évolution du data mining a une histoire de progression technologique constante, motivée par le besoin d'analyser des quantités massives de données et de fournir des diagnostics plus précis. Avec les avancées en intelligence artificielle, cette discipline continue de transformer l'industrie, rendant les usines plus intelligentes.

### **II.3 Définition du Data Mining**

Le Data Mining, également connu sous le nom de la fouille de données se présente principalement comme un processus visant à extraire et découvrir des informations utiles de l'immense quantité de données [46]. Plusieurs auteurs de la littérature ont contribué pour définir ce terme : selon Fayyad : L'exploration de données est un processus complexe qui vise à déceler des modèles de données valides, potentiellement bénéfiques et finalement compréhensibles.

Ferruzza a défini L'exploration de données comme un ensemble de méthodes utilisées dans le processus de découverte de connaissances pour identifier des relations et des modèles encore inconnus au sein des données.

Selon Parsaye : L'exploration de données est un processus d'aide à la décision qui consiste à rechercher dans de grandes bases de données des modèles d'information inconnus et inattendus.

Pour Zekulin : L'exploration de données correspond au processus visant à obtenir des informations encore inconnues, compréhensibles et exploitables, à partir de grandes bases de données, en utilisant ces informations pour prendre des décisions cruciales.

Enfin, John a défini l'exploration de données comme le processus de découverte de modèles avantageux dans les données.

Ce concept de Data Mining implique l'utilisation de plusieurs techniques développées à partir de différentes sciences et disciplines, en vue d'extraire des connaissances précieuses et pertinentes, de les analyser et de les interpréter [47].

Contrairement aux méthodes d'analyse des données traditionnelles [48], les méthodes utilisées en Data Mining aident à transformer des informations brutes en connaissances précieuses, en analysant des masses importantes de données pour découvrir les informations significatives et exploitables, qui peuvent servir à la prise de décision précises, prédire des résultats ou acquérir une connaissance plus profonde des données.

#### **II.4 Processus d'implémentation du Data Mining**

La nécessité croissante de nombreux secteurs à traiter de grande quantité de données a motivé les chercheurs scientifiques à développer et à améliorer les méthodes traditionnelles d'extraction de la connaissance. Il est primordial d'apprendre que le Data Mining ne se limite pas à la recherche de modèle au sein d'un ensemble de données. Cela ne constitue qu'une étape dans tout un processus mis par les scientifiques, les ingénieurs ou toutes autres personnes cherchant à tirer des informations précieuses à partir de données.

On parle ainsi de l'extraction de la connaissance à partir de données, également appelé en anglais : Knowledge discovery in databases (KDD), qui est un domaine pluridisciplinaire ; et qui fait la combinaison entre plusieurs méthodes de Data Mining, d'apprentissage automatique et de gestion de base des données. Le terme de Data Mining est associé au terme du KDD, et il est considéré comme une étape cruciale de celui-ci [49]. Ce processus est complexe à implémenter, il se décompose des étapes clés suivantes, comme montre la figure II.1. [50] :

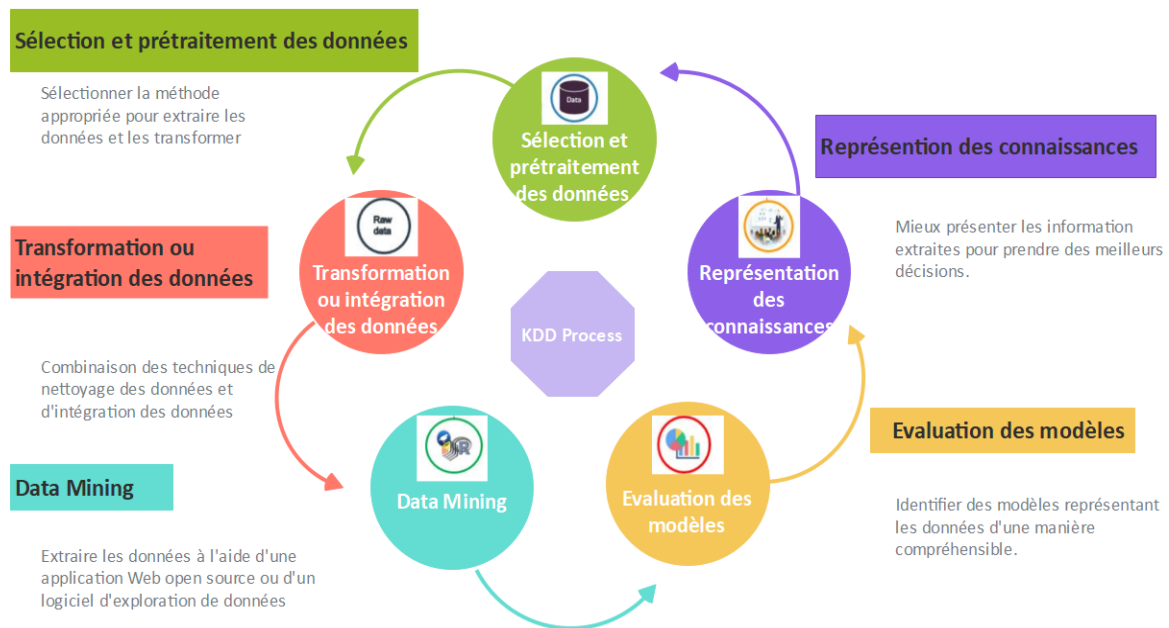


Figure II. 1. Les étapes d'implémentation du KDD.

- **Connaissance du domaine d'application** : C'est la phase initiale où les objectifs, les besoins et la compréhension du domaine d'application sont définis.
- **Sélection des données cibles** : Cette étape implique la collecte de données pertinentes qui seront utilisées pour l'analyse. Il peut s'agir de données issues de diverses ressources comme : des bases de données, des entrepôts de données, des fichiers plats, ... etc.
- **Prétraitement des données** : Concerne les données brutes, qui sont nettoyées, corrigées, intégrées et transformées pour les rendre adaptées à l'analyse. Cela peut inclure la suppression et l'élimination des données inexacts et des valeurs manquantes ou aberrantes.
- **Transformation ou intégration des données** : Pendant cette phase, les données sont transformées en un format plus approprié pour l'analyse, il s'agit de processus de regroupement de différentes sources de données dans une seule base de données. Pour y parvenir, on peut opter à des techniques et des méthodes de telles que la sélection ou l'extraction de caractéristiques, et la transformation de données.
- **La fouille de données** : C'est l'étape principale du processus KDD où les algorithmes de Data Mining sont utilisés pour extraire des schémas, des modèles, des relations ou des tendances intéressantes à partir des données. Cela peut impliquer des méthodes telles que la classification, la régression, le clustering, la découverte des règles de classification... etc.

- Évaluation des modèles : C'est l'identification de modèles représentant des connaissances basées sur des mesures données. Les modèles ou les schémas découverts sont évalués pour leur validité, leur utilité et leur adéquation aux objectifs métiers.
- Interprétation et visualisation des résultats : Les modèles et les résultats obtenus sont interprétés pour en tirer des connaissances exploitables. La visualisation peut jouer un rôle crucial pour comprendre et présenter ces informations de manière compréhensible.
- Représentation des connaissances : Les informations extraites sont présentées d'une manière significative, et sont intégrés dans un système performant pour être mis à la disposition des décideurs.
- Évaluation continue : Ce processus n'est pas statique. Il est important d'évaluer régulièrement les modèles, de réévaluer les objectifs métiers et d'améliorer les méthodes utilisées pour une meilleure extraction de connaissances.

Ces étapes ne sont pas nécessairement linéaires et peuvent impliquer des itérations pour affiner et améliorer les résultats. Le KDD est un processus itératif qui vise à extraire des informations et des connaissances significatives à partir des données pour faciliter la prise de décision informée dans divers domaines d'application [49].

## **II.5 Techniques pour Data Mining**

Dans cette section, nous présentons l'apprentissage supervisé et l'apprentissage non supervisé qui sont les deux types catégories utilisés pour faire face au problème d'exploration de données à savoir la classification, la régression, l'association et la segmentation. Généralement, l'apprentissage supervisé effectue une induction sur les données de présent pour faire des prédictions. Par contre l'apprentissage non supervisé caractérise les propriétés des données contenues dans un ensemble de données cibles.

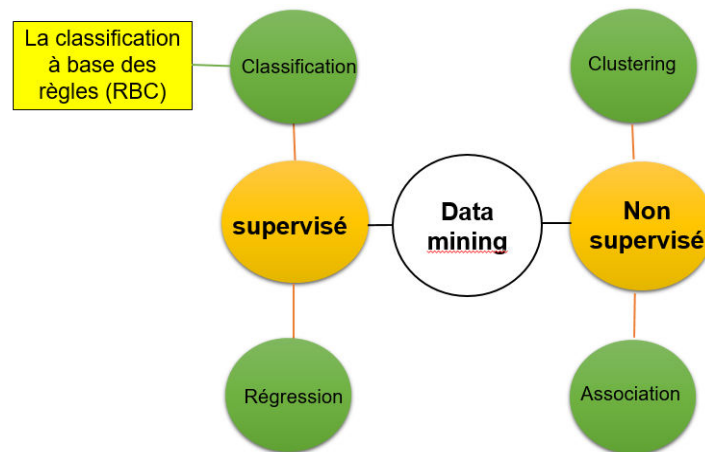


Figure II. 2. Techniques pour Data Mining.

### II.5.1 L'apprentissage supervisé

La technique supervisée dans l'exploration de données est une approche de recherche d'un modèle, dans laquelle ce modèle est entraîné sur un ensemble de données étiquetées, dont l'objectif est de prédire un événement ou d'estimer les valeurs d'un attribut numérique continues. Dans le cas d'apprentissage supervisé, la tâche analytique s'effectue par plusieurs phases. Tout d'abord, le modèle des données analysées (un classifieur), est donné ou construit en utilisant un algorithme de Data Mining. Ensuite, il faut évaluer la qualité de travail de qualitatif, et s'il ne répond pas aux exigences, un autre apprentissage du classifieur est indispensable. On répète ce processus jusqu'à ce que le niveau de qualité cible soit atteint, où il devient évident que l'algorithme choisi est inefficace en utilisant les données fournies ou que les données n'ont pas une structure identifiable. Ces types de problèmes sont liés à la classification et à la régression [47].

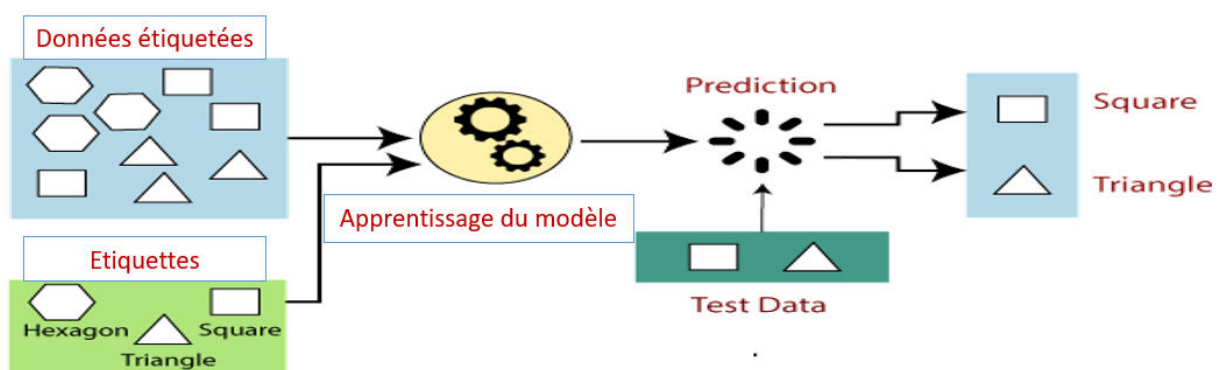


Figure II. 3. L'apprentissage supervisé.

### II.5.1.1 La classification

- **Définition**

La classification est considérée comme l'une des domaines d'apprentissage automatique. Cette technique de data mining est largement utilisée dans de nombreux domaines pour faire face aux problèmes de prédiction et de classification afin de classer les données dans des catégories qui déjà existent en fonction de leurs caractéristiques. Elle repose sur deux phases principales ;

- La phase d'apprentissage : où un classifieur est construit en utilisant l'ensemble de données d'apprentissage, cet ensemble de données comporte des objets dont on connaît leurs classes.
- La phase de classification : dans cette phase, le modèle construit dans la première phase est utilisé pour classer les nouvelles données (données de test).

La classification supervisée est définie comme le processus d'identification d'un modèle (ou d'une caractéristique) qui fait désigner et différencier des concepts ou des classes de données. L'analyse de l'ensemble de données d'apprentissage (où les éléments ou les objets de données pour lesquels les catégories de classe sont identifiées) constitue la base de développement du modèle. Alors, lorsque la catégorie de classe d'un élément ou d'un objet est inconnue, le modèle est utilisé pour deviner à quelle classe il appartient [50]. Donc, cette technique de classification consiste à prédire à quelle classe, parmi les  $k$  classes ( $C_1, C_2, \dots, C_k$ ) une observation  $x$  appartient.

- **Applications de la classification**

Cette tâche a de nombreuses applications dans différents domaines, qui a pour but d'optimiser divers processus et améliorer les décisions, parmi ces applications on peut citer : le domaine de santé et de médecine ; où la classification permet de classer les malades à partir de leurs symptômes ou résultats des analyses identiques pour faciliter la tâche de diagnostic médical. Dans le domaine de finance et de marketing : la classification des transactions financières permet de détecter les comportements frauduleux ou suspects. En marketing, elle est particulièrement adaptée à l'identification des différents profils de clients constituant la clientèle, permettant de classer ces clients en groupes homogènes pour mieux cibler les campagnes de marketing et les offres. Et enfin, dans le domaine industriel : la classification est utilisée dans le contrôle de qualité pour classer les produits en fonction de la présence des défauts ou des anomalies détectés au cours de processus de fabrication par des systèmes de détection ou d'autres méthodes de mesure. Dans la maintenance prédictive la classification

permet de prédire les pannes d'équipements en classant les équipements en fonction de leur état et de leurs données de capteurs [49].

- **Algorithmes et techniques utilisés pour la classification**

Généralement, le processus de classification suit plusieurs étapes qui consistent en premier temps à collecter les données pertinentes pour ce problème de classification. Ces données peuvent être issues de différentes ressources, comme les bases des données, les fichiers plats, les capteurs, etc. Une fois que les données sont collectées, elles sont explorées pour comprendre leur structure, leur qualité et leurs caractéristiques. Ensuite, ces données collectées peuvent nécessiter un prétraitement pour les rendre appropriées pour l'analyse. Dans un deuxième temps, et à partir des spécificités des objectifs et des données du problème donné, un algorithme de classification approprié est choisi. Les résultats obtenus sont utilisés pour construire un modèle de classification, ce modèle apprend en fonction de données en ajustant ses paramètres pour minimiser une fonction de perte ou maximiser une fonction de gain, selon l'algorithme utilisé [51]. Une fois le modèle de classification est construit, des mesures de performance sont utilisées pour l'évaluer comme : le rappel, la précision, la F-mesure. Ces mesures permettent d'évaluer l'aptitude du modèle à faire des prédictions précises des différentes classes. A la suite, de cette évaluation, les décisions prises par le modèle sont interprétées également par la classification en identifiant les caractéristiques les plus importantes pour la prédiction de chaque classe.

- **La classification à base des règles**

La classification à base des règles est une technique très connue et utilisée dans le Data Mining. Cette technique est une approche où des règles sont extraites à partir des données d'apprentissage, pour décrire des modèles et des relations entre les caractéristiques des données et les classes cibles. Chaque règle est généralement de la forme "Si [condition], Alors [classe]". La partie gauche de la règle (partie « Si ») est appelée antécédent ou condition de la règle. L'antécédent peut avoir une ou plusieurs conditions, qui sont logiquement associées par ET. La partie droite (partie « Alors ») de la règle est appelée le conséquent de la règle. La règle conséquente consiste en une prédiction de classe. La prédiction de classe est le nœud final [52].

Exemple de règle de classification :

**Débit** (X, 'augmente') et **Température** (X, 'Elevée') → Classe (X, 'A')

**Débit** (X, 'augmente') et **Température** (X, 'Basse') → Classe (X, 'B')

Débit (X, 'diminue') → Classe (X, 'C')

Ces règles sont un bon moyen pour représenter les informations qui peuvent être facilement lues et compréhensibles. Une fois que les règles sont extraites, elles doivent être évaluées pour leur précision et leur pertinence. Nous pouvons définir la précision d'une règle R comme suit :

$$\text{Précision} = \frac{n \text{ correct}}{n \text{ couvert}}$$

Après que les règles sont extraites et évaluées, elles peuvent être utilisées pour classer de nouvelles instances. Lorsqu'une nouvelle instance est présentée, les règles sont appliquées séquentiellement jusqu'à ce qu'une correspondance soit trouvée, et la classe correspondante soit prédite. L'un des avantages de la classification à base des règles est sa capacité à générer des modèles facilement interprétables.

### II.5.1.2 La régression

La régression est une technique qui consiste à modéliser et anticiper la prédiction des valeurs de données numériques (continues) au lieu des étiquettes de classe discrètes (le cas de la classification). Elle vise principalement à comprendre la liaison entre une variable cible (dépendante) et une ou plusieurs variables explicatives (indépendantes) afin de faciliter la prédiction des valeurs de la variable cible pour de nouvelles observations, permettant ainsi de tirer des connaissances significatives à partir des données [52].

Les principales difficultés rencontrées lors du traitement des problèmes de classification et de régression sont :

- La mauvaise qualité de données originales, présentant autant de données erronées et des valeurs manquantes ;
- Types d'attributs, qui ne sont pas adaptés à l'analyse des outils de data mining ;
- La présence des classes dominantes, ainsi que le problème de surajustement (recyclage) et de sous-ajustement (modèle faible).

### II.5.2 L'apprentissage non supervisé

L'apprentissage non supervisé ou non orienté, est employé lorsque les données d'entrée ne sont pas étiquetées ou ne contiennent pas de variables cibles à prédire. Contrairement aux techniques supervisées où l'algorithme est guidé par des étiquettes ou des résultats connus. L'avantage principal de ce type de problème est qu'il peut être traité sans aucune connaissance préalable des données analysées [47].

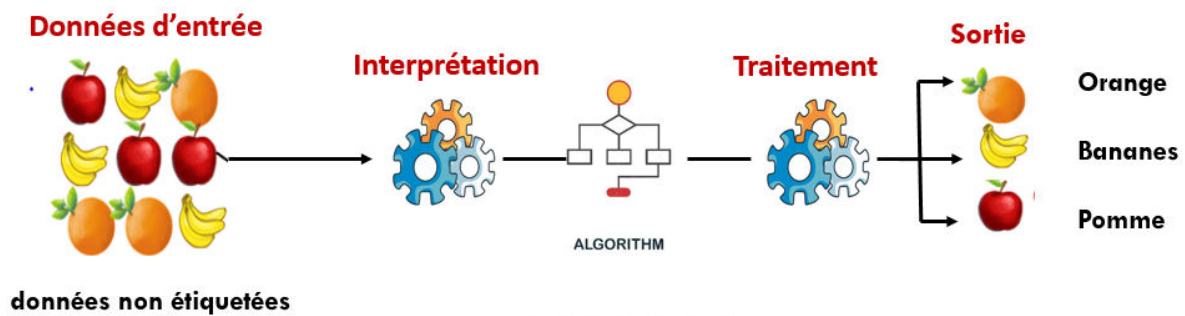


Figure II. 4. L'apprentissage non supervisé.

### II.5.2.1 L'association

Cette technique également appelée règles d'association, permet de découvrir des relations ou des connexions pertinentes entre les attributs dans des bases de données de masse importante. L'utilisation de cette technique est très répandue dans le domaine du marketing, de la recommandation de produits, de la gestion des stocks, et d'autres domaines où il est nécessaire de comprendre les associations entre éléments pour prendre des décisions informées.

L'analyse par les règles d'association est basée sur : la définition des transactions où les informations sont structurées sous la forme des transactions, et chaque transaction représente un ensemble d'articles achetés par les clients. Alors que la règle d'association est une relation de la forme « si (X), alors (Y) » ; où X et Y sont des ensembles d'articles. La règle peut s'écrire sous une autre forme :

Si 'certaines conditions sont satisfaites' alors 'prédire les valeurs pour certains autres attributs'.

### II.5.2.2 La segmentation (Clustering)

L'analyse de segmentation (clustering) est une forme d'extraction des données pour catégoriser les données associées ou l'ensemble d'objets des données dans des clusters. Cette technique permet d'examiner les différences entre les données et les similitudes, en analysant les objets de données sans recourir à la consultation des étiquettes de classes, à l'opposé de la classification et la régression, qui analysent des ensembles de données étiquetées par classes (apprentissage). Il est fort possible que, dans de nombreuses situations, les données étiquetées par classe peuvent ne pas exister au début. Le clustering peut être utilisé pour générer des étiquettes de classe pour un groupe de données. Les objets sont regroupés en se basant sur le principe de faire maximiser ou minimiser la similarité intra-classe. Autrement dit, le regroupement des objets est formé de façon que les objets inclus dans un groupe présentent une

grande similarité les uns par rapport aux autres, mais sont plutôt différents des objets des autres groupes. Chaque groupe ainsi formé peut être considéré comme une classe d'objets à partir de laquelle les règles de clustering peuvent être générées. Le clustering peut également faciliter la formation d'une taxonomie, c'est-à-dire l'organisation des observations en une hiérarchie des classes qui regroupent des événements similaires [50].

## II.6 Transformation et discrétisation de données

Afin de transformer les données, autant de méthodes sont employées. Dans cette section quelques méthodes de transformation sont présentées. La discrétisation qui est une forme de transformation de données est aussi abordée. Au cours de cette étape de prétraitement, les données sont transformées ou consolidées afin que le processus de fouille de données résultant soit plus efficace et que les modèles trouvés soient plus faciles à comprendre [50].

- **La discrétisation de données**

Lors de la fouille des données, la transformation et la discrétisation de données sont des étapes cruciales pour préparer les données en vue de l'application d'algorithmes d'exploration ou de modélisation. Alors, les données sont transformées ou consolidées sous des formes appropriées afin d'être fouillées.

La discrétisation, c'est convertir les variables continues en variables catégorielles ou discrètes, où les valeurs brutes d'un attribut numérique sont remplacées par des étiquettes d'intervalle. L'utilité de cette technique est pour simplifier le modèle ou pour traiter des données spécifiques auxquelles les algorithmes de Machine-Learning s'appliquent mieux [50].

Généralement, on peut catégoriser les techniques de discrétisation selon la façon dont elle est effectuée. On parle de la discrétisation supervisée, lorsque le processus de discrétisation utilise des informations de classe, afin de diviser des données en intervalles par la maximisation d'un critère spécifique, comme l'homogénéité des classes dans chaque intervalle. Sinon, on parle de la discrétisation non-supervisée. Si le processus commence par trouver d'abord un ou quelques points (points de partage ou points de coupure) pour diviser la totalité de l'ensemble d'attributs, puis répète tout cela d'une manière récursive sur les intervalles résultants, on parle alors de discrétisation ou de fractionnement descendant. Cela se différencie de la discrétisation ascendante, qui commence par considérer toutes les valeurs continues comme des points de partage potentiels, écartant certaines valeurs en faire fusionner les autres valeurs de voisinage pour former des intervalles, puis applique récursivement ce processus aux intervalles obtenus.

## II.7 Types de données pour les applications du Data Mining

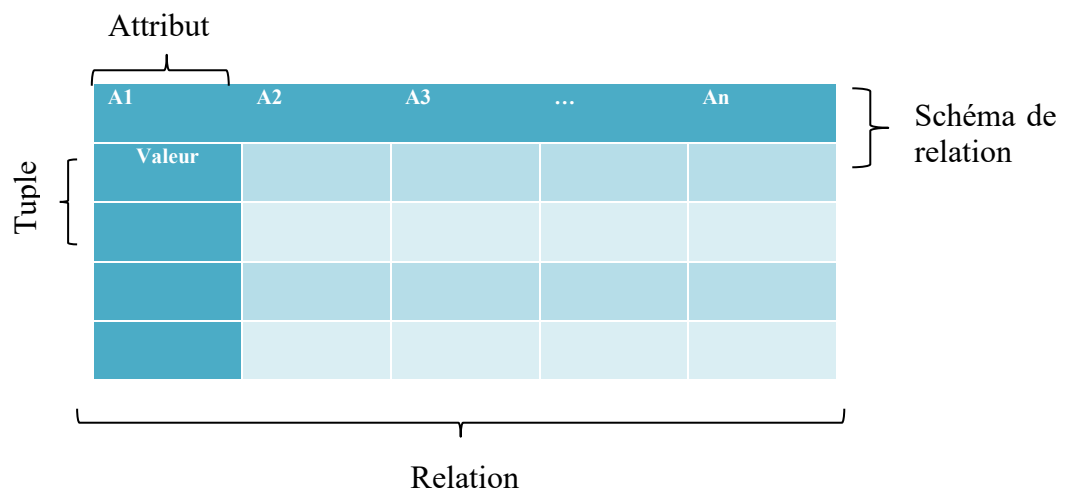
L'exploration de données est une vaste technologie qui peut être utilisée sur tous types de données, à condition qu'elle soit adaptée à l'utilisation prévue. Pour les applications du Data Mining, on peut citer : Les données issues des bases de données, les entrepôts de données, et les données transactionnelles, qui représentent les types de données les plus principaux. D'autres formes de données (tels que les données textuelles, les données multimédias, les données ordonnées/ séquentielles, les flux de données, les données géographiques, les données graphiques ou en réseau, ...etc) peuvent également être soumis à l'exploration de données. Sans doute, l'exploration de données continuera à inclure de nouveaux types de données, en progression avec leur émergence [53]. Un aperçu des types de données les plus utilisées est présenté ci-dessous :

### II.7.1 Les bases de données relationnelles

Un système de base de données, généralement appelé système de gestion de base de données (SGBD), est constitué d'un groupe de données connectées, et d'un assortiment d'applications logicielles permettant de gérer et d'accéder aux données. Ces applications logicielles offrent des méthodes et des techniques pour établir et contrôler l'accès aux données distribuées, partagées ou simultanées ; définir les structures des bases de données et leur stockage; ainsi pour garantir la cohérence et la sécurité de ces données stockées même en cas de l'apparition de défaillance dans le système, ou de tentatives d'accès indésirables [53].

Les bases de données relationnelles constituent un type de données clé dans le domaine de l'exploration de données, car elles comptent parmi les sources d'informations les plus largement disponibles, accessibles et les plus riches.

Tableau II- 1- Tableau représentant la forme d'une base de données relationnelle.



### II.7.2 Entrepôts de données (DataWarehouses)

Particulièrement, le contexte informationnel de la fouille de données concerne l'ensemble de données d'entrepôts. Un entrepôt de données, ou encore un Data Warehouse est une base de données centralisée conçue pour collecter, transformer, et stocker des données provenant de diverses sources afin de faciliter l'analyse et les requêtes. Leur processus repose sur l'ETL (Extract « extraire », Transform « transformer », et Load « charger ») facilitant la récupération de données à partir des systèmes sources, le nettoyage, la normalisation, la préparation, et l'insertion de ces données dans l'entrepôt, de façon organisée à des fins de business intelligence.

Ce Data Warehouse supervisé et géré par un administrateur de données, est généralement modélisé par un groupe de bases relationnelles ou de tableaux multidimensionnels [43]. Chaque dimension de tableau multidimensionnel correspond à un attribut ou à un ensemble d'attributs, et chaque cellule contient la valeur d'une mesure globale [53]. Ces tableaux de données multidimensionnels sont alimentés par des données non traitées et associés à un problème spécifique [43] soit : la surveillance des stocks en flux, la surveillance des données clients d'une banque ou d'une assurance, étude, surveillance en ligne les paramètres de production, Analyse de données textuelles...etc. Les informations contenues dans un entrepôt de données sont alors utilisées dans le cadre de la prise de décision dans le but de soutenir le système décisionnel. Ces informations sont conservées à des fins historiques, couvrant généralement les six à douze derniers mois, et sont généralement résumés [53].

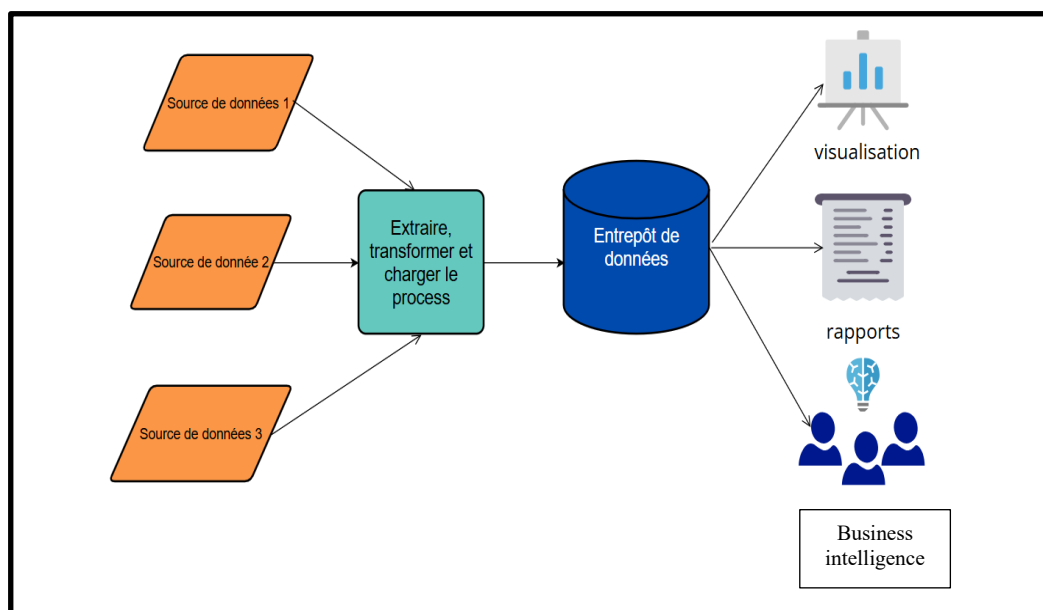


Figure II. 5. Cadre typique d'un entrepôt de données.

### II.7.3 Les données transactionnelles

En terme général, chaque enregistrement d'une base de données transactionnelle documente une transaction, comme Produits achetés par une entreprise, transaction de vente, ou l'inventaire des différents achats. Une transaction consiste généralement en une liste des éléments tels que les articles achetés, ainsi en un numéro d'identification de transaction unique (tran-ID). Une base de données transactionnelle peut contenir des tables supplémentaires contenant des informations supplémentaires sur les transactions, notamment la description de l'article, les détails du vendeur et de la branche...etc [53].

### II.7.4 Autres types de données

Plusieurs d'autres sortes de données existent en plus des différents types de données mentionnées précédemment ; et qui ont des formes et des structures diversifiées et des significations sémantiques plutôt distinctes. Une large gamme d'informations peut être extraite de ces différents types de données. Nous n'en mentionnons ci-dessous que quelques exemples. [54].

- Les données textuelles : pour évaluer l'évolution des sujets d'actualité, en extrayant de brefs commentaires textuels soumis par les utilisateurs.
- Les données multimédias, pour extraire les photos des éléments spécifiques en les classant à l'aide des étiquettes ou des tags sémantiques.
- Les données des réseaux informatiques, pour trouver des anomalies dans les flux de messages suggérant des intrusions, qui peuvent être découvertes par regroupement.
- Les données temporelles, afin d'explorer des données bancaires ou boursières.
- Les données géographiques, pour déterminer quels sous-ensembles d'objets sont géographiquement connectés.
- L'exploration du web peut révéler la dynamique du web ainsi que les liens et les connexions entre diverses pages web, utilisateurs, communautés et activités basées sur le web.

De nouveaux problèmes découlent de ces applications : comment exploiter des modèles avec des structures et de sémantiques riches, tels que la gestion des données avec des structures uniques (comme les réseaux, les arbres, les graphiques et les séquences) et une sémantique particulière (comme l'ordre, les contenus image, audio et vidéo et la connectivité).

Grace à l'amélioration et la consolidation de ces nombreuses sources de données, l'extraction de données complexes provenant de plusieurs sources donne souvent lieu à des conclusions positives.

## **II.8 Conclusion**

Afin de trouver des solutions aux différents problèmes particuliers, et liés à la surcharge d'informations, le Data Mining avec ses techniques variées, représente une excellente opportunité et un domaine avancé pour analyser et explorer d'énormes quantités de données. Cette exploration de données constitue un bon moyen pour extraire de l'information de valeur importante dans l'ensemble de données.

Cette section visait d'expliquer c'est quoi le data mining et en quoi il consiste, ses différentes techniques et leurs domaines d'application. Nous avons constaté que les techniques de data mining sont partitionnées en deux catégories, la première qui représente les techniques supervisées et la deuxième se compose des techniques non supervisé. Dans ces deux catégories, de nombreuses techniques existent avec des caractéristiques différentes qui peuvent avoir des points avantageux et même des inconvénients.

De nombreuses utilisations du data mining dans des applications industrielles ont marqué lieu, ces techniques variées de data mining sont utilisées pour extraire les informations nécessaires et importantes pour résoudre les différents problèmes rencontrés lors l'utilisation de grandes masses de données. Parmi ces techniques, la classification à base des règles qui est une très importante technique dans la prédiction des anomalies, c'est pour ça elle est retenue dans notre étude, pour donner une solution au problème mentionné dans le premier chapitre. Cette technique de Data Mining est fondée sur la construction d'un ensemble de classifieur contenant un ensemble de règles facilitant l'interprétation des résultats de détection et de diagnostic des défauts.

Dans ce qui suit, nous allons présenter la méthode d'optimisation stochastique (IDEOA) adoptée dans cette étude et utilisée par la classification des règles présentée dans cette partie pour fonder l'approche utilisée dans cette thèse.

## ***Chapitre III : L'optimisation et les méta-heuristiques***

III.1 Introduction

III.2 Définition de l'optimisation

III.3 Les méthodes d'optimisation

III.4 Critères de choix des méthodes d'optimisation

III.5 L'IDEOA

III.6 Conclusion

### III.1 Introduction

L'essor des technologies informatiques et la difficulté croissante à traiter et résoudre les problèmes d'optimisation, ont donné naissance à de nombreuses techniques d'optimisation. On peut classer ces techniques en deux catégories fondamentales : les méthodes déterministes et les méthodes stochastiques. Dans ce travail, notre attention s'est focalisée sur la deuxième catégorie et plus particulièrement aux méthodes méta-heuristique [55].

Ces méthodes sont des algorithmes qui visent à résoudre des problèmes d'optimisation difficiles et à trouver la meilleure solution qui converge vers un optimum global. En fait, les problèmes d'optimisation sont présents dans plusieurs applications qui s'intéressent par exemple à : planifier la production ; optimiser les circonstances opérationnelles des différents processus ; concevoir de nouveaux systèmes (invention ou dimensionnement) ; surveiller et contrôler les systèmes (stabilisation ou suivi de trajectoire). Ces problèmes d'optimisation peuvent être également présents dans le transport, dans la localisation, ainsi que dans le suivi et la supervision (problèmes de diagnostic et même de maintenance), comme ils peuvent être présents dans bien d'autres disciplines.

Dans un premier temps, ce chapitre a pour but de donner une brève présentation des méthodes d'optimisation. Dans un deuxième temps, nous abordons l'approche de « Discrete Equilibrium Optimizer Algorithm » DEOA. Tant que, notre problème de diagnostic des défauts industriels est vu comme un problème de classification, dans cette perspective, nous avons effectué une discrétisation des variables de données en utilisant l'outil Weka. Par la suite, la version discrète de l'algorithme d'optimisation d'équilibre DEOA élucidé par (MALIK & HAOUASSI) dans [56], est améliorée en introduisant la technique OBL « Opposite Based Learning » pour mieux explorer l'espace de recherche en fournissant des solutions prometteuses et pour s'échapper d'être piégé dans les optima locaux.

### III.2 Définition de l'optimisation

L'optimisation est un processus qui sert à atteindre la solution la plus optimale pour un problème spécifique, souvent complexe voire impossible à résoudre de manière complète dans un laps de temps raisonnable [57]. Ces méthodes d'optimisation peuvent être employées dans plusieurs applications de la vie courante comme : l'ingénierie avec ses différentes applications, la mathématique appliquée, la finance, la gestion, l'économie, l'informatique, l'industrie, l'intelligence artificielle, la médecine et bien d'autres domaines, dans le but de prendre la meilleure décision, d'augmenter l'efficacité, de réduire les coûts, et d'améliorer la performance

globale des systèmes ; tout en maximisant ou en minimisant une fonction objectif, en fonction de certaines contraintes [57].

En mathématique et surtout en ingénierie, l'optimisation joue un rôle primordial, en résolvant divers problèmes avec une ou plusieurs variables de décision ; qui représentent les valeurs à déterminer pour atteindre un objectif bien précis, sous ou sans contraintes pour trouver une solution faisable ou approximative dans l'espace de recherche. Par exemple, en industrie, on cherche à optimiser la production, les réseaux électriques, les moteurs ou les systèmes mécaniques.

Jusqu'à présent, de nombreuses méthodes d'optimisation ont été développées avec l'arrivée de l'ère de l'intelligence artificielle, et l'apprentissage automatique (Machine Learning) en s'inspirant par des phénomènes naturels ou des processus sociaux, biologiques ou physiques. Ces approches incluent les méthodes déterministes ainsi que les méthodes stochastiques (heuristique, et méta-heuristique) qui sont décrites succinctement ci-après.

### **III.3 Les méthodes d'optimisation**

Dans le contexte de la recherche en optimisation, résoudre un problème signifie améliorer le fonctionnement de n'importe quel élément par le biais de méthodes d'optimisation, qui servent à trouver la solution la plus optimale à un problème exprimé sous la forme d'une fonction donnée et qui fait intervenir un ou plusieurs critères.

Les méthodes d'optimisation, sont considérées comme des outils mathématiques qui visent à trouver la solution optimale à un problème spécifique, en minimisant les efforts requis. Ces méthodes ont connu une évolution trop rapide, principalement favorisée par la capacité puissante de calcul et de traitement des ordinateurs [58], mais aussi grâce à leur mise en œuvre dans de nombreux applications comme : la gestion et l'économie des entreprises, les chaînes commerciales, l'industrie lourde et légère, les banques, le transport, la médecine, l'agriculture, et même dans le domaine militaire. Une abondante littérature a essayé de traiter cette diversité de méthodes. On peut classer ces techniques en deux grandes catégories principales : les méthodes déterministes et les méthodes stochastiques [59].

#### **III.3.1 Les méthodes déterministes**

Les méthodes déterministes sont des méthodes particulièrement efficaces pour résoudre des problèmes d'optimisation bien définis. Leur principe consiste principalement à maximiser ou minimiser une fonction objectif tout en respectant des contraintes bien claires sans faire recours

à aucun principe stochastique. Ces méthodes sont des approches purement mathématiques qui s'appuient essentiellement sur les informations données sur une direction de recherche, garantissant une progression systématique vers la solution optimale [59]. Malheureusement, en pratique et contrairement aux méthodes stochastiques, elles risquent de converger vers un optimum local dans les problèmes non convexes, nécessitant un temps de calcul très élevé [58], elles sont aussi moins adaptées à un espace de grande dimension et aux problèmes discrets .

Il est à noter que de nombreuses méthodes d'optimisation déterministes sont basées sur le principe mathématique pour aboutir à des résultats plus certes et garantir leur succès [58], en résolvant des systèmes d'équations linéaires et non-linéaires, avec ou sans contraintes [57]. Parmi les exemples des méthodes mathématiques les plus pertinents on peut citer : Les méthodes du gradient, le gradient conjugué [60], la méthode quasi-Newton [61], et la méthode SQP [62]...etc

### **III.3.2 Les méthodes stochastiques**

Ce sont des méthodes itératives basées sur l'évènement probabiliste. Pour une même configuration initiale d'un problème d'optimisation donné, plusieurs exécutions de ces méthodes permet d'aboutir à des résultats distincts et à trouver l'optimum global d'une fonction même dans des conditions très difficiles grâce à leur propriété aléatoire [59].

Contrairement aux méthodes déterministes, les méthodes stochastiques sont moins restrictives, elles n'exigent ni la connaissance du point de départ, ni de la fonction objectif pour parvenir à un résultat optimal en conduisant à une réduction significative des temps de calcul. Cependant, le principal inconvénient de ces techniques est qu'elles ne peuvent pas garantir des solutions optimales ainsi que les méthodes déterministes [58].

Dans ce qui suit, une brève présentation des méthodes heuristiques et méta-heuristiques sera effectuée :

#### **III.3.2.1 Les méthodes heuristiques**

Le terme heuristique tire son origine du grec ancien « eurisko », signifiant « je trouve », en se référant à tout ce qui sert à l'exploration, la découverte, et l'innovation. Ces techniques heuristiques sont utilisées pour résoudre un problème posé, lorsque l'utilisation de méthodes déterministes n'est pas toujours possible à cause de leurs limitations comme leur besoin d'un temps de calcul très important [57]. Elles utilisent bien la structure du problème analysé afin d'arriver à une solution approximative, considérée comme « acceptable ». Afin de sélectionner

les directions les plus prometteuses, les méthodes heuristiques cherchent à explorer des points dans l'espace de recherche par le biais de tentatives successives [59], dans un délai de temps le plus court possible. Parmi ces méthodes les plus fréquemment utilisées on peut trouver : les algorithmes gloutons ou les approches par amélioration itérative, la méthode du Simplex de Nelder Mead [63], la stratégie de Hooke et Jeeves [64], ...etc.

### **III.3.2.2 Les méta-heuristiques**

Lors de l'élaboration de la recherche de Tabou, Fred Glover a introduit en premier le terme de « méta-heuristique ». La majorité des algorithmes méta-heuristiques s'inspirent fréquemment de la nature : de système physique comme le recuit simulé (SA) [5], [6] et la recherche tabou [65] ; de biologie tel que les algorithmes génétiques (GA) [4], ou encore d'éthologie comme l'optimisation par essaims de particules (PSO) [7]. Durant des millions d'années de progression, la nature a découvert des réponses quasi idéales à presque toutes les problématiques auxquelles elle a été confrontée. La plupart des algorithmes qui tirent leur inspiration de la nature, sont fondés sur la théorie évolutionniste de Darwin. En effet, ils ont dites bio-inspirées ou inspirées de la biologie [66].

Le terme « méta » fait référence à la capacité qu'un même algorithme possède pour résoudre divers problèmes d'optimisation complexes sans nécessiter de modifications structurelles importantes [67]. Les deux éléments essentiels de tout algorithme méta-heuristique sont : la convergence et la randomisation. La première option qui est la convergence : elle consiste à choisir les solutions les plus performantes assurant que ces dernières convergent vers le meilleur résultat possible, alors que la randomisation prévient qu'elles ne se retrouvent bloquées dans des optima locaux tout en favorisant leur diversité. L'association adéquate de ces deux éléments assure généralement l'obtention de l'optimalité globale [66].

Les méta-heuristiques sont des algorithmes itératifs qui intègrent un élément aléatoire et explorent l'espace de recherche en employant diverses méthodes pour aboutir à des solutions. Contrairement aux heuristiques, elles ne sont guidées par aucune spécificités liées au problème en question, elles sont donc indépendantes. On trouve certains théorèmes de convergence concernant la technique du recuit simulé [68] ou la stratégie tabou [69], [70] qui démontre que la probabilité de s'approcher vers la solution optimale s'accroît au fil du temps. Ainsi, si le temps se rapproche de l'infini, on a la certitude d'obtenir la solution optimale. En effet, en réalité c'est impraticable car ce délai dépasse le temps requis pour lister toutes les solutions de l'espace de recherche. Cependant, ces techniques montrent d'excellentes performances en pratique [57].

Ces algorithmes peuvent être classés de différentes façons : l'une d'entre elles consiste à les classer en: algorithmes basés sur la population, et algorithmes basés sur la trajectoire [66] :

- Les algorithmes basés sur la population, tel que l'optimisation par essaim de particules PSO, également connue sous le nom d'algorithme basé sur des agents, car il utilise plusieurs agents ou particules. Ainsi que les algorithmes génétiques sont fondés sur la population puisqu'ils utilisent un ensemble de chaînes.

- Les algorithmes basés sur la trajectoire tel que le recuit simulé qui fait appel à un agent unique, ou à une solution, en se déplaçant de façon fragmentée dans le domaine de recherche. Un déplacement ou une solution plus performante est systématiquement favorisé, tandis qu'un mouvement moins efficace peut être accepté selon un certain degré de probabilité. Les mouvements ou les déplacements tracent une trajectoire au sein de l'espace de recherche, avec une probabilité non nulle que cette trajectoire aboutit à l'optimum global.

Une autre catégorisation de ces méthodes consiste à les classer en: méthodes locales, et méthodes de recherche globales [71] :

- Les méthodes locales ont un rôle principal pour obtenir le minimum local se situant le plus près du point de départ. Elles cherchent constamment à améliorer la solution existante et par conséquent, elles ne peuvent pas se détacher d'un minimum local, car cela nécessiterait de dégrader la solution courante.

- Les méthodes globales disposent d'un procédé qui leur offre la possibilité d'échapper aux minima locaux. Ces méthodes sont généralement plus appropriées pour une exploration globale de l'espace de recherche.

Plusieurs auteurs ont pensé à combiner ces deux catégories de méthodes. L'algorithme qui en résulte possède deux caractéristiques contradictoires : L'exploration et l'exploitation de l'espace de recherche. L'exploration représente la capacité de l'algorithme à explorer de nouveaux territoires dans le domaine de recherche. Cette tâche s'avère être la plus complexe, à cause de la grande immensité de l'espace de recherche, rendant ainsi son parcours en entier trop compliqué. Cependant, ce comportement ne favorise pas la convergence en évitant d'être piégé dans un optimum local. Alors que, l'exploitation concerne la capacité de la méta-heuristique, à poursuivre la recherche dans cette zone, en utilisant les résultats obtenus lors de la phase d'exploration, pour favoriser la convergence en utilisant une solution prometteuse et aller vers l'optimum local le plus proche. Dans ce cas, le défi pour une méta-heuristique réside dans

l'établissement d'un équilibre adéquat entre ces deux comportements pour se rapprocher vers la solution optimale, tout en évitant d'être bloquer sur une valeur d'optimum local [67].

Dans le paragraphe suivant, nous aborderons les critères de choix des méthodes d'optimisation.

### **III.4 Critères de choix des méthodes d'optimisation**

Afin de choisir l'algorithme d'optimisation le plus adéquat, cet algorithme doit être robuste ou performant sur une large gamme de problèmes, pour toutes les valeurs raisonnables du point de départ. Egalement il doit être efficace : en ne demandant pas un temps de calcul important, ou une grande capacité de stockage. D'autre part, il doit avoir une certaine précision pour être en mesure de déterminer des solutions précises, sans être influencé par les erreurs inclus dans les données qui peuvent survenir lors de l'application de l'algorithme.

Dans de nombreux problèmes d'optimisation, la fonction à optimiser présente plusieurs minima. Cependant, les méthodes déterministes n'ont pas la capacité à conduire à une solution, sauf si le point de départ se situe près de la solution optimale, ce qui est très peu probable. Alors, l'estimation directe du gradient de la fonction objective lors d'une simulation par éléments finis peut poser des enjeux, limitant l'application de ces méthodes déterministes les plus performantes.

Les méthodes stochastiques sont souvent vues comme une solution alternative aux problèmes à plusieurs minima, en raison de leur capacité puissante à trouver le minimum global même sous des circonstances les plus difficiles. Cependant, le temps de calcul requis pour parvenir à cette solution peut s'avérer très élevé [59].

Pour conclure, les approches stochastiques s'avèrent être appropriées pour résoudre les problèmes d'optimisation dans le domaine de diagnostic des défauts industriels. L'IDEOA « Improved Discrete Equilibrium Optimiser Algorithm » est développé dans le paragraphe suivant.

### **III.5 Le nouvel algorithme amélioré « IDEOA »**

#### **III.5.1 Introduction**

Un nouvel algorithme d'optimisation, appelé « Equilibrium Optimizer » (EO), a récemment été introduit dans le domaine de l'optimisation. Cet algorithme a été initialement suggéré par Faramarzi et al. [8], en s'appuyant sur le principe de la physique et s'inspirant des modèles de bilan massique à volume de contrôle utilisés pour évaluer les états dynamiques et stationnaires.

La méthode proposée explore une approche de bilan massique dynamique qui prend en compte à la fois les entrées et les sorties dans un volume de contrôle dans le but de déterminer la concentration d'un constituant non réactif dans le volume de contrôle.

L'EOA a démontré une efficacité et une efficacité exceptionnelles pour obtenir des solutions optimales ou quasi optimales par rapport aux autres algorithmes actuellement disponibles dans la résolution de différents problèmes.

### **III.5.2 Historique**

Dans ce contexte, Abdel-Basset et al. [72], dans leur étude et pour améliorer la version de l'EOA, ils ont implémenté une méthode de réduction de la diversité linéaire (LRD) et d'élimination des minima locaux (MEM). Et afin d'évaluer l'efficacité de cet algorithme suggéré, une analyse comparative a été menée sur un ensemble d'algorithmes fiables, qui ont été mis en œuvre sur les cellules solaires commerciales de R.T.C France. Une autre étude, et dans le même axe de travail que celui-ci d'Abdel-Basset et al. a été réalisée ; l'objectif fondamental de cette recherche était d'évaluer les paramètres de trois modèles différents de cellules photovoltaïques. Contrairement à la version initiale de l'EO, l'IEO utilisé par Wang et al. [73] pour résoudre le problème mentionné auparavant, ils ont prévu une masse très importante de données générée par ces cellules photovoltaïques afin d'exécuter un processus d'optimisation plus efficace en utilisant une fonction objectif plus raisonnable.

Pour surmonter le problème de la sélection des caractéristiques; Too et Mirjalili, [74] ont introduit une nouvelle itération de l'EO dans l'étude de la classification des données biologiques. Ainsi, dans un essai supplémentaire, et afin de résoudre ce problème, le dernier algorithme mentionné dans la référence [75], a été utilisé avec une fonction de transfert en forme de V pour convertir les valeurs continues générées par l'« EO » en une zone de recherche binaire. Ensuite, la technique de recuit simulé (SA) a été employé pour améliorer l'exploitation du BEO. Les résultats de la méthodologie proposée démontrent une performance très compétitive pour résoudre les problèmes de sélection de caractéristiques.

Fan et al. [76], dans leur étude ont proposé une nouvelle version de l'EO connue sous le nom de (M-EO) dans le but d'améliorer sa précision d'optimisation. Ainsi, pour accroître les capacités d'exploitation de cette nouvelle version proposée et afin d'éviter se retrouver bloqué dans des optima limités, les auteurs ont intégré la technique d'apprentissage par opposition (OBL), et de nouvelles régulations de mise à jour dans ce nouvel algorithme. Contrairement à l'EOA proposé par Faramarzi et al. [8], qui traite exclusivement des problèmes d'optimisation

continue, les auteurs dans [56] ont suggéré une version discrète de l'EOA pour aborder le problème d'optimisation des enjeux discret par la classification à base des règles. L'étude visait principalement à chercher des solutions globales en empêchant être piégé dans les solutions locales, par le biais de l'association des deux approches de l'intelligence artificielle : la technique de classification associative et les méthodologies à base de population. L'introduction de nouveaux opérateurs discrets a également renforcée la puissance des capacités d'exploration et d'exploitation dans l'espace de recherche.

Dans d'autres travaux l' algorithme proposé par Farmarzi « EOA » [8], a démontré sa supériorité sur d'autres algorithmes méta-heuristiques conventionnels. À notre connaissance, l'algorithme d'optimisation (EOA) n'a jamais été adapté au problème de détection et de diagnostic des défauts industriels. Cela nous a poussés à utiliser cet algorithme pour résoudre ce problème en l'appliquant au processus chimique Tenesse Eastman (TEP). Dans cette étude, nous avons employé la version discrète (DEOA), car l'algorithme original n'est applicable qu'aux problèmes continus. Dans la section suivante, nous détaillerons la version discrète de la méthode choisie.

### III.5.3 Description de l'algorithme EOA

L'EOA est un algorithme méta-heuristique innovant et puissant qui a démontré sa supériorité pour résoudre les problèmes d'optimisation continue. Cet algorithme innovant proposé par Faramarzi et al. [8] puise son inspiration dans la physique en se reposant sur des modèles dynamiques. Il est basé sur un bilan de masse dynamique bien réparti dans un volume de contrôle.



Figure III. 1. Photo représentant le principe de l'EOA.

Une équation de bilan massique est déterminée pour démontrer les principes physiques de la conservation de la masse entrante et sortante dans un volume de contrôle. Cette équation générique de bilan massique, qui est une équation différentielle ordinaire du premier ordre,

exprime que le changement de masse au fil du temps est identique à la somme de la masse entrante dans le système et de la masse produite à l'intérieur, déduite de la masse sortante. Cette équation est représentée de la forme suivante [8] :

$$V \frac{dc}{dt} = QC_{eq} - QC + G \quad (\text{III.1})$$

Où  $C$  : signifie la concentration dans le volume de contrôle,

$V \frac{dc}{dt}$  : est le taux de variation de masse dans le volume de contrôle,

$Q$  : représente le débit volumétrique entrant et sortant du volume de contrôle,

$C_{eq}$  : représente la concentration à l'état d'équilibre, dans laquelle il n'y a pas de génération dans le volume de contrôle.

$G$  : désigne le taux de génération de masse dans le volume de contrôle.

Alors, quand :  $V \frac{dc}{dt} = 0$  l'état d'équilibre est réalisé.

Le processus d'optimisation par l'EOA commence par l'utilisation de la population initiale comme la plupart des algorithmes méta-heuristique. Ce processus débute par l'initialisation et continue à itérer jusqu'à atteindre la solution optimale ou l'état d'équilibre. Au cours de cette phase d'initialisation, les concentrations initiales sont générées arbitrairement dans un espace de recherche à partir du nombre de particules, sous la forme d'un vecteur de concentrations initiales [8], comme indiqué dans l'équation III.2 ci-dessous.

$$C_i^{initial} = C_{min} + rand_i(C_{max} - C_{min}) \quad i = 1,2,3 \dots, n \quad (\text{III.2})$$

Avec  $C_i^{initial}$  représente le vecteur de la concentration initiale de la particule  $i^{\text{ème}}$ ,

$C_{min}$  et  $C_{max}$  désignent les valeurs minimales et maximales des dimensions,

$rand_i$  est un vecteur aléatoire compris entre [0,1],

et  $n$  est le nombre de particules constituant la population.

Dans cet EOA une particule correspond à une solution et une concentration représente la position d'une particule. Et afin d'identifier les candidats à l'équilibre, les particules doivent être évaluées puis triées. Chacune de ces particules actualise sa position par l'intermédiaire de trois termes différents, comme indiqué dans l'équation III.3 [8]. Ces trois termes définissent les règles de mise à jour d'une particule.

$$C = C_{eq} + (C_0 - C_{eq})F + \frac{G}{\lambda V} (1 - F) \quad (\text{III.3})$$

$F$  est calculé comme suit :

$$F = \exp [-\lambda(t - t_0)] \quad (\text{III.4})$$

Avec  $C_0$  et  $t_0$  sont le temps et la concentration initiale respectivement.

F est le paramètre exponentiel

G représente le taux de génération

et  $\lambda$  est un vecteur aléatoire compris entre 0 et 1.

L'expression de chaque de ces termes contribuant à la règle de mise à jour des concentrations d'une solution est donné dans ce qui suit :

### III.5.3.1 la concentration à l'équilibre

Dans l'algorithme EO, les candidats les plus prometteurs connus sous le nom des candidats d'équilibre sont déterminés en premier temps, afin d'orienter les particules dans l'espace de recherche [9]. Ces candidats représentent les quatre meilleures particules et leur moyenne. Les quatre premiers candidats aident l'EOA à améliorer ses capacités d'exploration dans l'espace de recherche, et la moyenne de ces quatre particules lui permet d'atteindre une meilleure exploitation. Ces particules constituent un vecteur appelé le pool d'équilibre ou encore « the equilibrium pool » :

$$\vec{C}_{eq,pool} = \{\vec{C}_{eq(1)}, \vec{C}_{eq(2)}, \vec{C}_{eq(3)}, \vec{C}_{eq(4)}, \vec{C}_{eq(ave)}\} \quad (III.5)$$

Avec

$$C_{eq(ave)} = \frac{C_{eq(1)} + C_{eq(2)} + C_{eq(3)} + C_{eq(4)}}{4} \quad (III.6)$$

Aléatoirement et avec la même probabilité, à partir de ces candidats d'équilibre les particules mettent à jour leurs positions ou encore leurs concentrations à chaque itération.

### III.5.3.2 Le terme exponentiel

Le terme exponentiel est un paramètre très important dans la règle de mise à jour des concentrations. Il contribue principalement à l'amélioration des capacités de l'EOA de maintenir un équilibre entre l'exploration et l'exploitation. Ce terme peut être écrit sous la forme suivante [8] :

$$F = a_1 \text{sign}(r - 0.5)[e^{-\lambda t} - 1] \quad (III.7)$$

Avec  $a_1$  est une constante, qui implique une grande capacité d'exploration lorsque sa valeur est élevée,  $r$  et  $\lambda$  sont des vecteurs aléatoires compris entre 0 et 1.

Le temps  $t$  représente la fonction d'itération. (A chaque fois que le nombre d'itération augmente cette fonction décroît) :

$$t = \left(1 - \frac{iter}{\max\_iter}\right)^{\left(a_2 \frac{iter}{\max\_iter}\right)} \quad (III.8)$$

Où :  $iter$  est l'itération actuelle, et  $\max\_iter$  représente le nombre maximal d'itération,  $a_2$  est une valeur constante qui indique une meilleure capacité de recherche de l'algorithme. Plus que la valeur de  $a_2$  est élevé, plus que la capacité d'exploitation augmente et la capacité

d'exploration diminue. Les deux constantes  $a_1$  et  $a_2$  prennent les valeurs de 2 et 1 respectivement.

### III.5.3.3 Le taux de génération

Un autre terme très important de l'algorithme EO est le taux de génération, ce paramètre contribue amplement à exploiter le domaine de recherche, dont il est formé comme suit [8]:

$$G = G_0 e^{-\lambda(t-t_0)} = G_0 F \quad (\text{III.9})$$

Avec  $G_0$  est le taux de génération initial exprimé comme suit :

$$G_0 = GCP(C_{eq} - \lambda C) \quad (\text{III.10})$$

Et  $GCP$  est le paramètre qui contrôle le taux de génération, ce paramètre est défini comme :

$$GCP = \begin{cases} 0.5r_1 & \text{si } r_2 \leq GP \\ 0 & \text{si } r_2 < GP \end{cases} \quad (\text{III.11})$$

$r_1$  et  $r_2$  désignent deux nombres aléatoires dans l'intervalle  $[0,1]$ ,  $GP$  est un autre terme qui signifie la probabilité de génération. Ce terme est utilisé pour déterminer la probabilité de contribution du taux de génération au processus de mise à jour en spécifiant le nombre de particules utilisées par ce dernier pour qu'elles actualisent leurs états.

Lorsque la probabilité de génération est égale à 0.5, Les deux capacités d'exploration et d'exploitation sont équilibrées.

Finalement, le principe repose sur une mise à jour itérative des solutions permettant d'améliorer progressivement la qualité des résultats comme suit :

$$C = C_{eq} + (C - C_{eq}) \cdot F + \frac{G}{\lambda V} (1 - F) \quad (\text{III.12})$$

$C_{eq}$  qui est le premier terme dans cette équation, et comme définit précédemment est la concentration à l'équilibre. Le deuxième terme aide largement à l'exploration, en améliorant la recherche globale de l'espace afin d'atteindre un point optimal. Tandis que, le troisième terme contribue à trouver une solution précise après que le point optimal est trouvé. Ce dernier terme, aide à son tour à améliorer la capacité d'exploitation de l'EOA. Ces deux derniers termes représentent les variations de concentration, cette variation est augmentée lorsque les deux termes sont de même signe, facilitant la recherche dans l'espace global, et elle diminue dans le cas contraire facilitant la recherche locale [9].

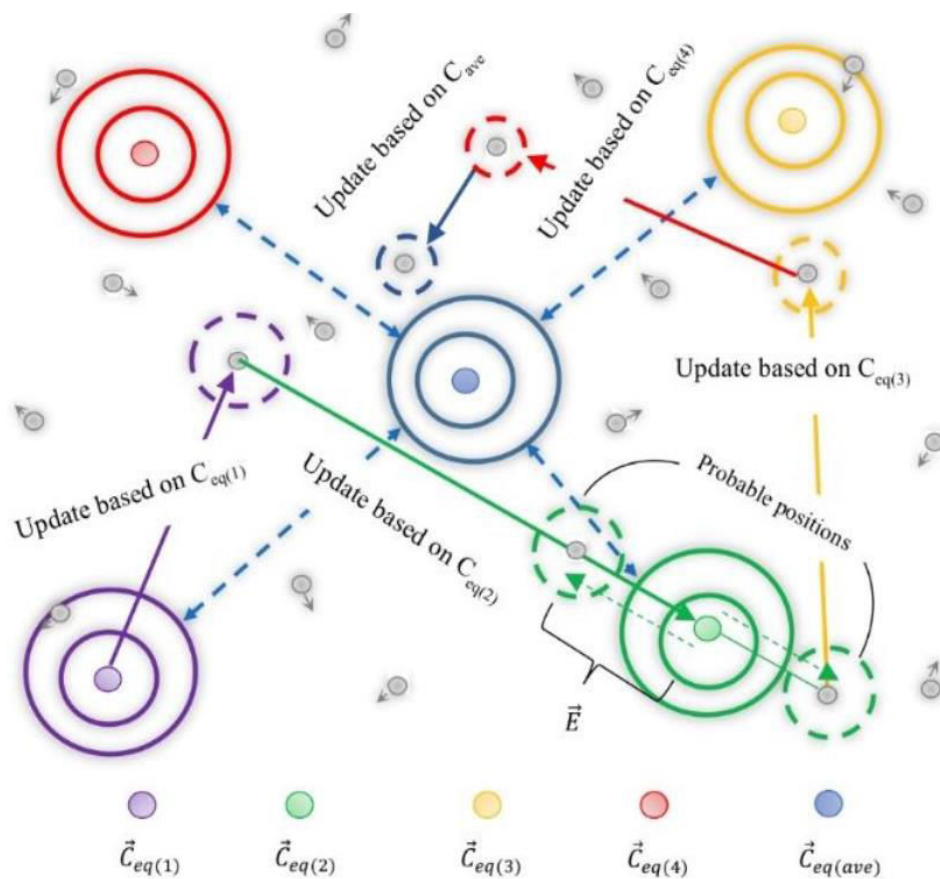


Figure III. 2. Coopération des candidats à l'équilibre pour actualiser la position d'une particule [8].

### La sauvegarde au mémoire

Ce mécanisme de sauvegarde au mémoire aide chaque particule de garder ses coordonnées dans le domaine de recherche, en donnant ainsi plus d'informations sur ses valeurs de fonction objective. Lorsque cette valeur atteinte par une particule lors de l'itération actuelle présente plus de performance que la valeur atteinte lors de l'itération précédente, alors la particule avec qui marque une meilleure valeur est celle enregistrée et sauvegardée au mémoire et l'autre sera écrasée. Ce mécanisme bénéficie des capacités d'exploration pour s'échapper des minimas locaux, et il sert aussi à améliorer l'efficacité de l'exploitation. Le pseudo code de l'algorithme EOA, ainsi que ses capacités d'exploration et d'exploitation sont présentés dans l'annexe A [8].

### III.5.4 Le DEOA

Par ailleurs, le diagnostic des défauts étudié dans ce travail est considéré comme un problème discret. Alors, nous proposons d'utiliser l'IDEOA pour le diagnostic du TEP, qui présente de bonnes performances et afin de fournir une solution plus adéquate, plus claire en termes

d'interprétation, et efficace, facilitant ainsi le processus décisionnel. Sa capacité à augmenter la précision et l'interprétation du modèle généré de diagnostic, ainsi à gérer des données de grande dimension sans être piégé dans des optima locaux, en fait un choix judicieux pour cette étude. Nous décrivons ci-dessous quelques aspects de DEOA présenté dans [3].

De nouveaux opérateurs discrets ont été proposé par les auteurs du DEOA :  $\{\odot_{x\%}, \oslash, \ominus$  et  $\otimes\}$  dans le but d'adapter l'équation de mise à jour de la position des particules aux problèmes discrets. La position de mise à jour est alors définie comme indiqué dans l'équation (III.13).

$$C_{i+1} = C_{eq} \odot_{50\%} \left( \left( \frac{G}{\lambda} \right) \odot_{F\%} \left( (C_i \ominus C_{eq}) \right) \right) \quad (\text{III.13})$$

où  $C_i$  et  $C_{i+1}$  représentent respectivement les vecteurs de position des particules actuelles et nouvelles ;  $C_{eq}$  est la position d'équilibre choisie arbitrairement dans le vecteur de pool d'équilibre. Ce vecteur est formé des premières particules représentant un quart de la population.  $\lambda$  est un vecteur discret stochastique, et le paramètre  $G^{it}$  désigne le taux de génération, qui peut être représenté mathématiquement comme indiqué dans l'équation (III.14) donnée dans [3].

$$G^{it} = \begin{cases} G_0^{it} & \text{if } r > F/100 \\ |1 - G_0^{it}| & \text{else} \end{cases} \quad (\text{III.14})$$

Ici, la notation  $|\cdot|$  désigne la valeur absolue ; F indique le paramètre exponentiel calculé à l'aide de l'équation (III.16). r est une variable qui signifie un nombre arbitraire entre [0,1] et  $G_0^{it}$  est calculée comme dans l'équation (III.15) de [3].

$$G_0^{it} = X_{eq}^{it} \ominus (\lambda \otimes X_i^{it}) \quad (\text{III.15})$$

$$F_{it} = f(it) = \frac{(N-it)^2}{N^2} \times 100 \quad (\text{III.16})$$

Ici, N désigne le nombre extrême d'itérations et indique l'itération réelle du DEOA. Les symboles  $\odot(x\%)$ ,  $\oslash$ ,  $\ominus$  et  $\otimes$  dans les équations (III.13) et (III.15) représentent des opérateurs discrets opérant sur des vecteurs discrets [3], et leurs définitions sont les suivantes :

### L'opérateur $\odot_{50\%}$

Cet opérateur permet de combiner deux vecteurs discrets afin de créer deux nouveaux vecteurs supplémentaires. Afin de déterminer le premier vecteur, on doit sélectionner le premier segment du premier vecteur et les 50% restants de l'autre. Ensuite, pour obtenir le second vecteur, il faut choisir la seconde moitié du segment du premier et la première moitié de l'autre [3].

**L'opérateur discret de soustraction ( $\ominus$ )**

$$\mathbf{X3} = \mathbf{X1} \ominus \mathbf{X2} = \begin{cases} \mathbf{X1} & \text{if } \mathbf{X1}[1] = 1 \text{ and } \mathbf{X2}[1] = 0 \\ \mathbf{X3}[1] = 0 \text{ and } \mathbf{X3}[2] = \mathbf{X1} \text{ or } \mathbf{X2} \text{ randomly} & \text{if } \mathbf{X1}[1] = \mathbf{X2}[1] \\ \mathbf{X1} \text{ or } \mathbf{X2} \text{ randomly} & \text{if } \mathbf{X1}[1] = 0 \text{ and } \mathbf{X2}[1] = 1 \end{cases} \quad (\text{III.17})$$

**L'opérateur discret de division ( $\oslash$ )**

$$\mathbf{X3} = \mathbf{X1} \oslash \mathbf{X2} = \begin{cases} \mathbf{X1} & \text{if } \mathbf{X1}[1] = 0 \\ \mathbf{X2} & \text{if } \mathbf{X1}[1] = 1 \text{ and } \mathbf{X2}[1] = 1 \\ \mathbf{X1} \text{ or } \mathbf{X2} \text{ randomly} & \text{if } \mathbf{X1}[1] = 1 \text{ and } \mathbf{X2}[1] = 0 \end{cases} \quad (\text{III.18})$$

**L'opérateur discret de multiplication ( $\otimes$ )**

$$\mathbf{X3} = \mathbf{X1} \otimes \mathbf{X2} = \begin{cases} \mathbf{X1} & \text{if } \mathbf{X1}[1] = 0 \\ \mathbf{X2} & \text{if } \mathbf{X2}[1] = 0 \\ \mathbf{X1} \text{ or } \mathbf{X2} \text{ randomly} & \text{if } \mathbf{X1}[1] \neq 0 \text{ and } \mathbf{X2}[1] \neq 0 \end{cases} \quad (\text{III.19})$$

Où X1, X2 et X3 représentent des vecteurs discrets. Le pseudocode de l'algorithme 1 démontre les spécificités de l'algorithme DEO. Dans un premier temps, toutes les particules sont initialisées arbitrairement, avant d'être déplacées dans l'espace de recherche en utilisant la fonction de mise à jour définie dans l'équation (III.13), comme détaillé dans [8].

**Algorithme 1.** L'algorithme d'optimisation d'équilibre discret (DEOA)

**Entrée :** Les données d'apprentissage (TS), Nb\_Particules, Iter\_max

**Sortie :** une règle de classification

{

Initier arbitrairement toutes les particules

Calculer la valeur de la fonction objectif des particules à l'aide de l'équation (III.21)

Pour (t = 1 à Iter\_max)

{

Choisir les meilleures particules :  $C_{eq1}$ ,  $C_{eq2}$ ,  $C_{eq3}$  and  $C_{eq4}$

$C_{eq\ avg}$  = la moyenne des  $C_{eq1}$ ,  $C_{eq2}$ ,  $C_{eq3}$  and  $C_{eq4}$

$C_{eq} = [C_{eq1}, C_{eq2}, C_{eq3}, C_{eq4}, C_{eqavg}]$

Pour ( $i = 1$  à Nb\_Particules)

{ Sélectionner un vecteur dans  $C_{eq}$ .

Calculer le vecteur aléatoire  $\lambda$ .

Générer le vecteur F à l'aide de l'équation (III.16)

Générer le vecteur  $G_0^{it}$  à l'aide de l'équation (III.15)

Générer le vecteur  $G^{it}$  à l'aide de l'équation (III.14)

Générer la nouvelle position à l'aide de l'équation (III.13)

}

Fin Pour

```

    Vérifier si la nouvelle position générée est réalisable.
    Calculer la valeur de la fonction objective de la nouvelle position générée à l'aide de
    l'équation (III.21)
    }
    Renvoyer la meilleure position considérée comme la meilleure règle.
  }.

```

---

### III.5.5 Principe de l'IDEOA

Comme toutes les techniques d'optimisation basées sur des méta-heuristiques, l'exploration et l'exploitation sont les deux principales stratégies influençant la capacité de recherche de l'algorithme [77]. Dans notre processus expérimental, nous avons constaté que le DEOA exploite bien l'espace de recherche, mais il tombe fréquemment dans les optima locaux et ne peut pas trouver une bonne exploration dans l'espace de recherche pour aboutir à des bons résultats.

Récemment, plusieurs algorithmes d'optimisation basés sur la population ont exploré l'utilisation de la technique OBL (Opposite Based Learning) pour améliorer leur taux de convergence [78]-[79]. Par conséquent, nous envisageons d'utiliser cette technique (OBL) mentionnée dans [80] pour l'algorithme IDEOA, afin d'éviter être piégé dans des optima locaux et de s'échapper à cette convergence prématurée. Dans cette section, nous nous concentrons sur l'utilisation de la stratégie OBL dans la méthode méta-heuristique DEOA pour modifier sa recherche globale (exploration de l'espace de recherche). Ainsi que, le DEOA avec l'OBL peuvent explorer davantage ce domaine de recherche, ce qui lui permet de converger rapidement vers la solution optimale. C'est pourquoi une nouvelle approche DEOA améliorée, appelée IDEOA, est proposée en adoptant la stratégie OBL.

La technique OBL est utilisée dans l'IDEOA en deux étapes : dans la population initiale et lorsque l'algorithme converge vers un optimum local. L'IDEOA et la classification à base des règles sont les deux techniques développés dans cette étude pour tirer de la connaissance à partir des données du système, dans l'objectif de proposer un modèle de diagnostic des défauts et, spécifiquement pour l'aide à la prise de décision qui constitue une étape cruciale dans ce domaine. Dans ce travail, en utilisant les données de TEP, nous avons élaboré un modèle de classification pour identifier une situation anormale ou défectueuse susceptible de menacer l'état fonctionnel de système. La méthode d'optimisation utilisée ici est un IDEOA dédié à la génération d'un modèle de diagnostic des défauts explicable. L'algorithme 2 présente l'approche globale utilisée dans ce travail.

- **Position de la particule dans l'IDEOA et l'OBL**

La stratégie OBL est appelée dans l'IDEOA pour l'amélioration de la diversité des particules, qui est définie comme suit : Soit une particule  $p$  structurée comme un vecteur à  $2N$  dimensions [56], le vecteur initial est un vecteur binaire qui indique les caractéristiques pertinentes dans la règle de classification, tandis que le second désigne la valeur de chaque vecteur choisi [3]. Dans ce travail, nous proposons de générer uniquement l'opposé du premier vecteur à l'aide de l'équation (III.20), tandis que le second vecteur est conservé sans modification car il s'agit d'un vecteur discret.

$$\textit{opposite}(p) = op = \begin{cases} op[1,j] = 1 - p[1,j] \\ op[2,j] = p[2,j] \end{cases} \quad (\text{III.20})$$

Un exemple de particule et de son opposé est donné dans ce qui suit :

Position initiale de la particule  $p$  :

1	0	0	1	1	0	1
3	4	2	1	5	3	4

Opposé de  $p$  :

0	1	1	0	0	1	0
3	4	2	1	5	3	4

- **Fonction objectif**

Cette fonction objective sert à évaluer la performance de chaque solution obtenue (règle de classification) :

$$\textit{Obj\_Func}(R) = \frac{\textit{Number of covered examples}}{\textit{Total number of examples}} \quad (\text{III.21})$$

- **Stratégie d'initialisation de l'IDEOA**

L'IDEOA génère d'abord la population initiale  $P$  arbitrairement, puis calcule la population opposée  $OP$  à partir de  $P$  en utilisant la stratégie OBL définie dans l'équation (III.20) pour chaque particule. Les meilleures particules parmi  $P$  et  $OP$  sont ensuite sélectionnées et considérées comme population initiale de l'algorithme.

- **Prévention de la stagnation dans l'optimum local par l'IDEOA**

Lorsque les particules de la population du DEOA original convergent vers un optimum local, elles ne peuvent pas modifier la direction de la recherche. Ceci est observé lorsque le résultat obtenu ne s'améliore qu'après un nombre fixe d'itérations. Cependant, dans ce cas, les particules doivent se déplacer vers une autre zone de recherche pour trouver un résultat plus performant. Ainsi, un mécanisme de recherche adaptative est proposé lorsque les particules sont piégées dans des optima locaux. Nous vérifions après chaque exécution de l'algorithme s'il existe un nombre d'itérations (seuil prédéfini) sans amélioration du résultat obtenu ; l'algorithme conclut

à une situation de piégeage dans les optima locaux et appelle l'OBL pour 50 % de la population, une réinitialisation aléatoire pour le reste des particules.

- **Aperçu de l'IDEOA**

L'algorithme 3 présente les détails de l'IDEOA :

---

**ALGORITHME 2.** L'algorithme d'optimisation d'équilibre discret et amélioré (IDEOA)

---

**Entrée :** L'ensemble de données (TS), N, Iter\_max, MT. // MT signifie le nombre maximal d'itérations sans amélioration

**Sortie :** une règle de classification

{

    Générer aléatoirement la population P

    Générer la population opposée OP en calculant l'opposé de chaque particule de P à l'aide de l'équation (III.20).

    Evaluer la fonction objective de toutes les particules P et leur OP à l'aide de l'équation (III.21)

    Parmi P et OP choisir N meilleures particules comme population initiale.

    Fit = fonction objective de la meilleure particule

    K=0

**Pour** (it = 1 à Iter\_max) faire

        //Sélection des meilleures particules

        Générer  $C_{eq1}$ ,  $C_{eq2}$ ,  $C_{eq3}$  et  $C_{eq4}$ , (les quatre meilleures particules)

**Si** la fonction objective ( $C_{eq1}$ )  $\leq$  Fit alors

            K=0

**Sinon**

            K=K+1

            Fit fonction objective ( $C_{eq1}$ )

**Fin Si**

**Si** K < MT alors

        Calculer  $C_{eq\_moy}$  (la moyenne de  $C_{eq1}$ ,  $C_{eq2}$ ,  $C_{eq3}$  et  $C_{eq4}$ )

$C_{eq} = [C_{eq1}, C_{eq2}, C_{eq3}, C_{eq4}, C_{eq\_moy}]$

**Pour** (i = 1 à N) faire

            Choisir un candidat à partir de  $C_{eq}$ .

            Générer le vecteur  $\lambda$  aléatoirement.

            A l'aide de l'équation (III.16), calculer le vecteur F

            A l'aide de l'équation (III.15), calculer le vecteur  $G_{0it}$

            A l'aide de l'équation (III.14), calculer le vecteur  $G_{it}$

            A l'aide de l'équation (III.13), générer la nouvelle position de la particule

**Fin Pour**

**Sinon**

**Pour** (i = 1 à N) faire

            A l'aide de l'équation (III.20) générer la nouvelle position // appeler la stratégie

OBL

**Fin Pour**

**Fin Si**

// Evaluation de nouvelles solutions

Vérification de la faisabilité de la nouvelle position générée.

**Pour** ( $i = 1$  à  $N$ ) faire

Evaluer la fonction objective de la nouvelle position générée à l'aide de l'équation

(III.21)

**Fin Pour**

//Mise à jour de la meilleure solution

Mettre à jour Fit si une meilleure solution est trouvée

**Fin Pour**

Renvoyer la règle la plus adéquate considérée comme la meilleure fonction objective de la position de la particule.

}.  

---

### III.6 Conclusion

Dans le présent chapitre, afin d'étudier le problème de diagnostic des défauts industriels et pour pouvoir pallier tous les risques pouvant menacer le bon fonctionnement du système, nous avons donné une brève présentation des méthodes d'optimisation ainsi que leurs critères de choix, dans un premier temps. Par la suite, nous avons donné un aperçu général sur l'algorithme EOA, son évolution dans la littérature et une brève présentation de sa version discrète.

Enfin nous avons détaillée le principe de l'approche utilisée dans cette étude qui est l'IDEOA. Dans ce qui suit, et dans le but d'évaluer l'efficacité et la performance de l'approche utilisée dans cette thèse, nous avons identifié l'approche de diagnostic des défauts suggérés en donnant les résultats de son application sur les données d'un processus industriel chimique et réaliste qui est le TEP.

## ***Chapitre IV : Détection et diagnostic des défauts basé sur un IDEOA appliqué à un processus industriel***

IV.1 Introduction

IV.2 Présentation de l'approche de détection et diagnostic des défauts suggérée

IV.3 Description du système d'étude de cas

IV.4 Historique de diagnostic des défauts du TEP

IV.5 Résultats du diagnostic des défauts

IV.6 Conclusion

## **IV.1 Introduction**

Dans cette section finale, la méthode d'optimisation présentée auparavant est appliquée à l'ensemble de données de TEP, qui est un système de référence réaliste utilisé fréquemment dans divers travaux pour évaluer l'efficacité et l'applicabilité d'une grande catégorie des méthodes de détection et de diagnostic des défauts industriels.

Dans la première section de présent chapitre, nous avons identifié l'approche de détection et de diagnostic des défauts suggéré dans cette étude. Cette méthode est destinée à générer un modèle de diagnostic et de classification des défauts à base des règles. Les règles obtenues lors de ce diagnostic favorisent l'explication et l'interprétation des résultats. Ensuite, pour évaluer l'efficacité de cette approche, nous l'avons appliqué à l'ensemble de données de TEP. Ce dernier sera présenté brièvement dans la deuxième section de ce chapitre. L'ensemble de données a été divisionné en deux parties : 80 % des instances ont été utilisé pour l'apprentissage du modèle, tandis que 20 % restantes ont été exploité pour le test de ce modèle. L'approche suggérée a été développée et implémentée à l'aide de l'outil Weka et des langages de programmation Java. Dans la dernière section de ce présent chapitre, les résultats obtenus par cette étude seront examinés et mis en perspective avec ceux mentionné dans les publications antérieures.

Notre objectif est d'évaluer l'efficacité et la performance de l'IDEOA, en déterminant sa fonction Accuracy, et mettre en évidence la pertinence des variables associées à la surveillance de défaut F4. Afin de juger l'interprétabilité du modèle de détection et de diagnostic de défauts élaboré par notre approche, d'autres critères ont été employés dans cette étude. Parmi ces critères on trouve : le nombre et la taille des règles qui définissent la taille globale du classifieur.

## **IV.2 Description du système d'étude de cas**

Notre système d'étude de cas le TEP (Tennessee Eastman Process), est un problème industriel complexe et réaliste, considéré aussi comme un problème de référence largement utilisé en génie des procédés pour tester à grande échelle les technologies de surveillance et de contrôle des procédés. L'idée initiale de ce modèle de processus chimique simulé a été suggérée par Downs et Vogel en (1993) [81]. Ce modèle se compose de cinq unités principales : un condenseur de produit, un réacteur exothermique et bi-phasique, un séparateur, une colonne de stripping et un compresseur de recyclage, comme démontré à la figure (IV.1). Le procédé implique quatre réactifs gazeux (A, C, D et E) et un réactif inerte B qui sont introduits dans le

réacteur afin de créer deux composants (G et H) et un sous-produit indésirable F. Les équations de réaction sont listées dans (IV.22). Toutes les réactions sont exothermiques, irréversibles et approximativement du premier ordre en termes de concentrations de réactifs [82].

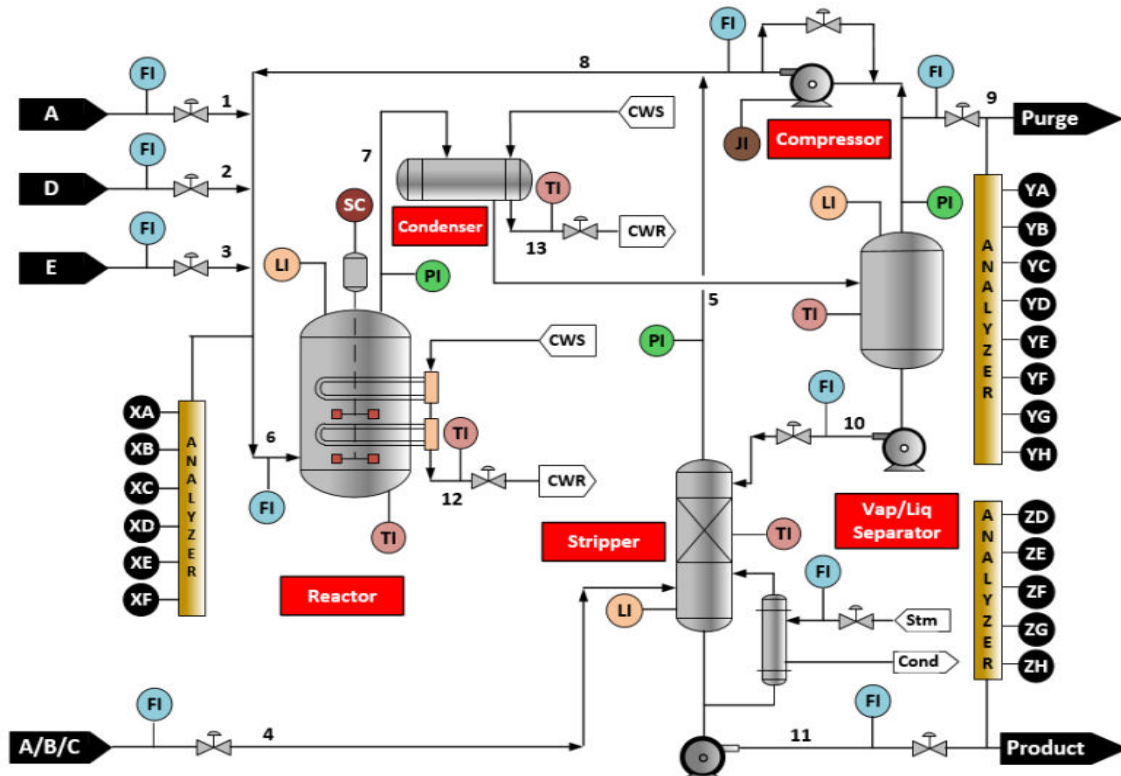
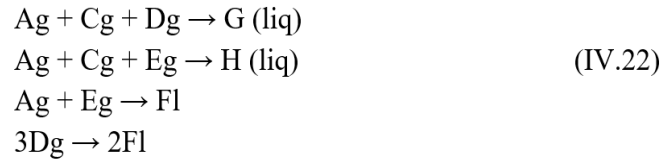


Figure IV. 1. Organigramme du TEP [78].

Dans cette étude, nous utilisons un grand échantillon de données TEP obtenu à partir de la page Web : <http://web.mit.edu/braatzgroup/links.html>. Le procédé se compose de 53 variables comme indique le tableau de l'annexe C. Où 41 des variables sont considérés comme des variables mesurées (ES1-ES41), parmi ces variables mesurées : 22 sont des variables continues, et le reste sont des variables échantillonnées. Les 12 derniers variables sont des variables manipulées (EV1-EV12), où la variable (EV12) qui est la vitesse d'agitation est négligée. Cependant, 21 défauts préprogrammés (F1-F21) sont générés par le simulateur TEP, dus à diverses perturbations, pour imiter diverses situations du fonctionnement du procédé, comme indique le tableau de l'annexe D. Parmi ces défauts ; 16 sont des défauts connus, tandis que les autres sont non-connus. Les défauts de F1 jusqu'à F7 sont des défauts qui impliquent

une variation soudaine de la variable de procédé. Alors que les défauts de F8 à F12 sont liés à un changement arbitraire de certaines variables du processus, F13 est une dérive retardée dans la cinétique de réaction et les défauts F14, F15 sont liés au blocage des vannes [83].

Sur la base de l'analyse des défauts du système, et les informations issues de la description du système, les spécialistes du domaine ont déterminé que la criticité de défaut F4 est très élevée, et que ce défaut a un impact majeur sur les performances du TEP. Par conséquent, négliger cet effet peut avoir des conséquences désastreuses sur le fonctionnement du système, ce qui rend le défaut F4 particulièrement pertinent pour notre étude.

- **Description de données de TEP**

Pour cette étude, et afin de réaliser le diagnostic des défauts industriels, nous avons utilisé l'ensemble de données TEP pour générer un modèle de détection et de diagnostic des défauts. Pour ce faire, deux types de données ont été employés : les données d'apprentissage pour construire le modèle de diagnostic et les données de test pour valider ce modèle.

Sur la base de données d'apprentissage, et dans les conditions de fonctionnement normales, le nombre total d'observations générées était de 500, avec une durée de simulation de 25 heures pour chaque exécution. Alors que, seulement 480 observations ont été utilisées après l'apparition d'un défaut. La matrice de données normale étant définie comme  $X \in \mathbb{R}^{480 \times 52}$ . En revanche, le nombre d'observations total pour les données de test était de 960, avec une durée de mesure s'étalant sur 48 heures et un intervalle d'échantillonnage de trois minutes. Par conséquent, les données issues de la matrice de test pour chaque défaut sont données comme  $X \in \mathbb{R}^{960 \times 52}$ . Ces matrices de données d'apprentissage et de test sont formées de 52 variables d'observation, en négligeant la vitesse d'agitation de l'agitateur du réacteur, qui est éliminée car elle est constante.

Dans cette étude, le fonctionnement du système est intégral et toutes les variables sont dépendantes. Pour cela, nous avons sélectionné 52 variables, car toutes les études précédentes ont montré que ces 52 variables étaient pertinentes. De plus, l'objectif principal de notre étude est de générer un ensemble de règles de classification, ces règles sont caractérisées par leur taille et leur nombre. La taille inclut les variables critiques et exclut les variables non pertinentes. Par conséquent, les résultats obtenus ne nous fournissent que les variables ayant un impact direct sur le défaut F4.

### **IV.3 Historique de diagnostic des défauts du TEP**

Dans ces dernières années, de nombreux chercheurs ont concentré leurs études sur l'amélioration des performances du diagnostic industriel des TEP en appliquant une grande variété de méthodes basées sur les données pour identifier les défauts pertinents pouvant entraîner un arrêt partiel ou total du système. Dans cette partie, nous présentons un aperçu historique des techniques et de méthodes de diagnostic utilisées dans des publications antérieures, dans le but de détecter et diagnostiquer les défauts de TEP.

On peut trouver parmi plusieurs travaux cités dans la littérature et qui traitent le problème de diagnostic des défauts dans les industries chimiques : le travail de Zou et al. [84]; ces derniers ont proposé une nouvelle technique pour récupérer les défauts de (TEP) utilisant un nouvel algorithme basé sur les réseaux de neurone pour élaborer une architecture de réseau ELM. La comparaison des résultats de cette nouvelle approche proposée dans cette étude avec ceux d'autres algorithmes fournis dans la littérature, a montré sa supériorité en termes de précision. Dans le travail de Onel et al. [85], les auteurs ont combiné la technique de sélection de caractéristiques avec l'algorithme de classification SVM (Support Vector Machine) non linéaire pour offrir un modèle de détection et de diagnostic des défauts plus précis d'un processus continu. L'objectif principal était de réduire la perte de la masse d'informations et de faciliter la modélisation et la sélection de caractéristiques simultanées. Cette étude présentait des résultats très encourageants, non seulement en termes de précision de détection, mais également en ce qui concerne la latence de détection.

Dans une autre étude, Soraya et al. ont réussi à appliquer l'analyse en composantes principales (ACP) au contrôle de procédé pour surveiller des installations de grande taille. Utilisant de nombreux types de statistiques de contrôle, comme l'erreur quadratique de prédiction (SPE), également appelée statistique Q. Cette statistique a permis d'indiquer une situation défectueuse sans fournir d'informations sur sa source. C'est pourquoi des diagrammes de contribution ont été utilisés pour isoler les défauts [86]. Il convient de noter que la méthode (ACP) ne peut pas traiter les caractéristiques non linéaires ou multimodales des données industrielles. Cependant, Zhang et al. dans leur recherche [87] ont suggéré une méthode de détection des défauts pour remédier aux limitations de l'ACP, qui révèle un taux de détection de défauts plus faible dans les processus industriels.

Pour surmonter le problème de diagnostic des défauts industriels de TEP, une autre méthode a été proposée par Han et al. Ils ont utilisé une technique itérative de réseau de neurone pour

déterminer le nombre de nœuds de couche cachés [88]. Les techniques de détection et de diagnostic des défauts en utilisant les méthodes de « Deep Learning » ont été largement appliquées à une grande variété de domaines d'ingénierie [89]-[90]. Dans ce contexte, Lomov et al. ont développé une approche plus avancée pour la prédiction précoce des défauts en utilisant l'ensemble de données de TEP étendu. Cette approche visait à étudier une variété de conceptions architecturales récurrentes et convolutives pour sélectionner l'architecture la plus optimale [91]. Ils ont également proposé d'utiliser la méthode des réseaux de neurone pour optimiser les informations utilisées dans leurs recherches.

De plus, dans le travail de Xu et al. [92] pour valider l'application technique de la méthode suggérée, les caractéristiques de l'échantillonnage sont initialement simplifiées à l'aide d'une technique de ML (Machine Learning) et de classification pilotée par les données (SVM). Les auteurs de ce travail ont utilisé les réseaux de neurone probabilistes « PNN » comme outil principal de classification du modèle de diagnostic des défauts, basé sur la sélection des caractéristiques et un optimisateur bio-heuristique afin d'optimiser le facteur de lissage caché ( $\sigma$ ) du PNN et d'améliorer ses performances de classification. Dans une autre étude, Hu et al. [93] ont proposé une autre méthode de diagnostic des défauts basée sur le Deep-Learning afin d'améliorer la précision du classificateur de défauts. Cette méthode combine la technique de sélection de caractéristiques et le classifieur KELM comme stratégie d'optimisation. Les résultats obtenus lors de cette étude confirment l'applicabilité et l'efficacité du modèle.

Il convient de noter que les stratégies proposées ci-dessus pour détecter les défauts dans le système de TEP, telles que les réseaux de neurone, SVM et l'ACP, nécessitent de grandes quantités de données pour montrer leur efficacité. Cependant, leur complexité de calcul et leur temps d'apprentissage augmentent constamment, notamment pour les architectures DL. Enfin, le plus important est qu'elles sont considérées comme des méthodes de type boîte noire, ce qui rend leur interprétation très difficile. Face à ces limitations, une approche FDD utilisant un nouvel algorithme méta-heuristique (IDEOA) a été proposée dans cette étude, cette approche sera détaillée dans ce qui suit.

#### **IV.4 Présentation de l'approche de détection et diagnostic des défauts suggérée**

Dans cette étude, une approche basée sur la génération des règles a été utilisée pour créer un modèle de FDD (détection et de diagnostic des défauts). Cette approche se compose de deux parties élémentaires, comme le montre la figure de l'annexe B. La première partie consiste à

prétraiter l'ensemble de données de TEP, tandis que, la deuxième partie implique l'utilisation de l'IDEOA pour générer des règles de classification destinées au FDD du système étudié.

#### IV.4.1 Prétraitement de l'ensemble de données

Dans cette première partie, un algorithme de discrétisation est appliqué aux variables de l'ensemble de données TEP, étant donné que la technique de classification à base des règles ne peut pas être appliquée directement aux données numériques et ne fonctionne que sur les données catégorielles. Par conséquent, dans cette étude et à l'aide de l'outil Weka, nous avons opté pour une méthode de discrétisation basée sur l'entropie afin de transformer les caractéristiques ou les variables continues en intervalles discrets [77], trouvés dans [80]. Cette méthode divise les valeurs continues de manière à minimiser l'entropie dans chaque intervalle, ce qui garantit que chaque intervalle contient des classes homogènes, pour optimiser la conservation des informations.

Les principaux paramètres de la technique de discrétisation fondée sur l'entropie ont été ajustés afin d'améliorer le processus de discrétisation. Le paramètre -B, qui contrôle le nombre de compartiments (intervalles ou valeurs catégorielles), a été défini sur 3. De plus, le paramètre -M, qui spécifie le poids minimal des instances par compartiment, a été défini sur -1. Nous activons également le paramètre de compartimentage à fréquence égale à -F, qui force les compartiments à contenir un nombre égal d'instances.

Dans l'annexe C, représente un tableau qui comporte toutes les valeurs discrètes des variables de l'ensemble de données TEP, obtenues à l'aide de la méthode de discrétisation basée sur l'entropie. Chaque variable est transformé en trois intervalles (Rangée 1, Rangée 2 et Rangée 3). Par exemple, la caractéristique ES1 a été discrétisée selon les trois intervalles suivants : intervalle 1 =  $]-\infty, 0,195023]$ , intervalle 2 =  $[0,195023-0,249177]$  et intervalle 3 =  $[0,249177, +\infty]$ . Les résultats obtenus sont présentés dans la dernière colonne du tableau de l'annexe C.

#### IV.4.2 Génération des règles

Dans cette deuxième partie, un ensemble de règles de classification est généré sous forme d'un modèle de classification améliorant en même temps : la précision et l'explication du modèle de diagnostic obtenu. Par conséquent, l'IDEO suggéré est appelé à chaque itération pour établir un classificateur à partir de l'ensemble de données prétraitées.

Cette seconde partie est réalisée par plusieurs itérations. Tant que, l'ensemble de données d'apprentissage initial ne reste pas vide. En commençant par un ensemble de règles vide

initialisant les paramètres de l'algorithme IDEOA. L'IDEOA est utilisé pour générer les règles de classification. À chaque itération, la classe C possédant le plus grand nombre d'instances est sélectionnée, et toutes les instances de cette classe sont transférées vers un nouvel ensemble de données d'apprentissage, composé uniquement d'instances appartenant à la classe choisie. Ces instances sont supprimées de l'ensemble de données d'apprentissage originales. L'algorithme IDEOA est ensuite exécuté de façon itérative pour générer une règle de classification à partir du nouvel ensemble de données d'apprentissage tant que celui-ci n'est pas vide. Une fois une nouvelle règle est générée, elle est ensuite appliquée pour classer les instances dans le nouvel ensemble de données d'apprentissage. Toutes les instances correctement classées sont ensuite supprimées du nouvel ensemble de données d'apprentissage, et la règle générée est ajoutée à l'ensemble des règles. Ce processus se poursuit jusqu'à ce que le nouvel ensemble de données d'apprentissage soit vide. Le résultat final de cette méthode est un classificateur basé sur des règles, qui est ensuite utilisé pour classer les nouvelles instances.

---

**Algorithme 3.** Méthodologie de génération du modèle de diagnostic de défaut TEP

---

**Entrée :** Les données de TEP

**Sortie :** Résultat de l'évaluation du modèle de diagnostic des défauts

**Début**

    Prétraiter l'ensemble de données de TEP

    Diviser l'ensemble de données TEP en données d'apprentissage et données de test

**Si** l'ensemble de données = les données d'apprentissage

        Générer le modèle de diagnostic des défauts à base des règles à l'aide de la méthode d'optimisation

        Evaluer le modèle de diagnostic des défauts

**Sinon**

        L'ensemble de données = les données de test

        Evaluer le modèle de diagnostic de défauts

**Fin si**

    Donner le résultat de l'évaluation du modèle de diagnostic des défauts

**Fin.**

---

## IV.5 Résultats du diagnostic des défauts par l'IDEOA

### IV.5.1 Evaluation de l'IDEOA et réglage des paramètres

Pour évaluer les performances de notre approche, nous avons calculé la fonction Accuracy à partir de la matrice de confusion. Cette dernière est utilisée principalement pour classer les prédictions d'un modèle de classification en TP, TN, FP, FN :

**TP « vrai positif »** : Si le modèle donne une classe positive pour un exemple d'une classe vraiment positive.

**TN « vrai négatif »** : Si le modèle donne une classe négative pour un exemple d'une classe négative.

**FP « faux positif »** : Lorsque l'exemple appartient à la classe négative et le modèle donne une classe positive.

**FN « faux négatif »** : Lorsque la classe de l'exemple est positive et le modèle donne une classe négative.

La structure de base d'une matrice de confusion est sous la forme d'une table qui répertorie des prédictions des résultats de classification et des résultats réels. Cette table à deux dimensions est représentée comme suit :

		Classe réelle	
		Positive	Negative
Classe prédite	Negative Positive	TP	FP
	Positive	FN	TN

La fonction Accuracy est une métrique d'évaluation communément considérée comme une mesure importante utilisée dans la littérature [9]. Elle représente le nombre d'observations correctement classées par rapport au nombre total d'observations. La fonction Accuracy est donnée par l'équation (IV.23),:

$$Accuracy = \frac{TN+TP}{TN+TP+FN+FP} \quad (IV.23)$$

D'autres critères ont été utilisés dans cette étude pour évaluer l'interprétabilité du modèle de diagnostic de défauts généré par notre approche. Ces critères incluent le nombre de règles, qui définit la taille globale du classifieur, et la taille des règles. Un classifieur plus compact et efficace est généralement caractérisé par un ensemble restreint de règles, qui réduise la complexité de calcul tout en améliorant la généralisation. En même temps, la taille de chaque règle influence le degré de précision et de spécificité de la classification des défauts, ainsi que, des règles plus brèves, courtes et plus concises contribuant amplement à une lisibilité et une interprétation améliorée.

En outre, le compromis entre la taille du classifieur et la complexité des règles joue un rôle crucial dans l'évaluation des performances du modèle. Un grand nombre de règles complexes

peut améliorer la précision de la classification, mais peut réduire l'interprétabilité. À l'inverse, un ensemble de règles plus restreint peut améliorer l'efficacité, mais risque de perdre des détails importants de la classification. Par conséquent, trouver un équilibre optimal entre ces facteurs est essentiel pour développer un système de diagnostic des défauts efficace et explicable. Afin de valider davantage notre approche, nous avons comparé ces critères à des modèles de référence afin d'évaluer les améliorations en termes de précision de classification.

#### IV.5.2 Comparaison entre les résultats du DEOA et l'IDEOA

Dans cette sous-section, la précision de classification est calculée afin d'évaluer les performances de l'approche utilisée dans cette thèse. Cette mesure de performance a également servi pour comparer les résultats de DEOA à ceux d'IDEOA, afin de déterminer les paramètres de l'algorithme permettant d'obtenir la meilleure précision. Pour obtenir les meilleures valeurs, l'expérience a été menée neuf fois ; à chaque fois, nous avons modifié le nombre de particules (50-100-200), et pour chaque particule, nous avons augmenté le nombre d'itérations (10-50-100). Les résultats de l'évaluation sont présentés dans le tableau IV-1.

Tableau IV- 1- Comparaison de la précision de classification du DEOA et d'IDEOA.

Methode			DEOA			IDEOA		
No	# Particules	# Itérations	Précision %	# Règles	Taille moy des règles	Précision %	# Règles	Taille moy des règles
1	50	10	85.71	86	6.33	82.54	43	5.76
2	50	50	80.95	87	6.75	85.65	45	5.34
3	50	100	85.71	78	6.69	87.87	35	6.23
4	100	10	85.71	77	6.12	90.47	26	6.78
5	100	50	85.71	67	6.74	95.47	16	6.62
6	100	100	85.71	68	6.97	95.23	18	6.32
7	200	10	90.47	61	6.19	95.14	22	6.56
8	200	50	85.71	56	6.56	95.19	18	5.95
9	200	100	85.71	56	6.91	95.40	18	6.78

Le tableau IV-1 montre clairement que notre algorithme proposé (IDEOA) atteint les meilleures valeurs de précision dans la plupart des cas ; il a marqué une précision plus élevée que DEOA d'une valeur de 95,47 %, pour un nombre de particules de 100 et un nombre d'itérations de 50. En revanche, chaque fois que nous exécutons le DEOA en augmentant le nombre de particules et le nombre d'itérations, nous sommes tombés au problème d'être piégé dans l'optimum local. Le DEOA ne peut pas donc améliorer les résultats obtenus. Par conséquent, il est évident que l'utilisation de l'algorithme IDEOA suggéré permet d'obtenir de meilleurs résultats en s'affranchissant des optima locaux.

### IV.5.3 Génération des règles à l'aide d'IDEOA

Comme il est indiqué dans le tableau IV-2, l'algorithme proposé dans cette étude donne comme résultat un modèle de classification, qui se compose de 16 règles, parmi ces 16 règles : six appartiennent à un fonctionnement normal du système et le reste des règles indique l'occurrence de défaut F4, avec un nombre moyen de termes de 6,62. Ces résultats révèlent que l'approche utilisée dans cette étude est très efficace pour générer des règles précises et courtes, ce qui en fait une approche explicable.

Tableau IV- 2- Les règles générées à l'aide de l'ensemble de données TEP.

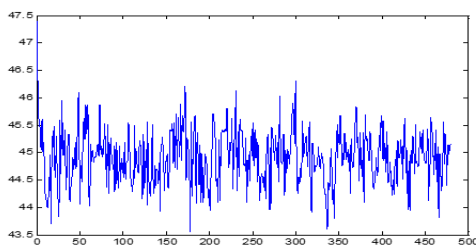
Règle N°	Règles extradites	Classe	Précision de la règle %	# Termes
R1	(ES2 ≥ 3682.6) & (ES4 ≤ 9.3008) & (26.731 ≤ ES5 ≤ 27.13) & (ES13 ≤ 2631.066667) & (339.546667 ≤ ES20 ≤ 341.323333) & (23.698333 ≤ ES31 ≤ 24.338667)	C0	10	6
R2	(74.772667 ≤ ES8 ≤ 75.761333) & (ES20 ≥ 341.323333) & (32.089 ≤ ES23 ≤ 32.526) & (23.698333 ≤ ES31 ≤ 24.338667) & (2.250567 ≤ ES34 ≤ 2.297233) & (EV10 ≤ 42.146)	C0	10	6
R3	(ES3 ≥ 4526.833333) & (65.531333 ≤ ES18 ≤ 65.937667) & (ES20 ≥ 341.323333) & (23.698333 ≤ ES31 ≤ 24.338667) & (ES36 ≥ 2.3077) & (21.428 ≤ EV3 ≤ 21.428)	C0	5	6
R4	(0.195023 ≤ ES1 ≤ 0.249177) & (79.953 ≤ ES11 ≤ 80.334) & (26.117333 ≤ ES25 ≤ 26.567667) & (ES33 ≥ 18.799667) & (0.091435 ≤ ES39 ≤ 0.104042) & (35.642667 ≤ EV7 ≤ 40.572333) & (EV8 ≥ 47.291) & (EV10 ≤ 42.146)	C0	10	8
R5	(2702.633333 ≤ ES7 ≤ 2712.866667) & (ES27 ≤ 18.755667) & (1.643167 ≤ ES28 ≤ 1.681933) & (4.7867 ≤ ES35 ≤ 4.8936) & (46.657667 ≤ EV9 ≤ 48.549333) & (EV10 ≤ 42.146) & (EV11 ≥ 19.929333)	C0	5	7
R6	(1.643167 ≤ ES28 ≤ 1.681933) & (23.698333 ≤ ES31 ≤ 24.338667) & (43.715667 ≤ ES41 ≤ 44.310333) & (21.428 ≤ EV3 ≤ 25.55) & (EV8 ≥ 47.291) & (EV10 ≤ 42.146)	C0	5	6
R7	(1.643167 ≤ ES28 ≤ 1.681933) & (ES32 ≥ 1.326107) & (53.368667 ≤ ES40 ≤ 54.195333) & (EV1 ≤ 62.563) & (53.429667 ≤ EV2 ≤ 54.230333) & (EV4 ≥ 62.39) & (EV5 ≥ 22.177667) & (EV9 ≥ 48.549333) & (EV10 ≤ 42.146) & (EV11 ≤ 17.228667)	C1	5	10
R8	(23.828333 ≤ ES14 ≤ 25.757667) & (32.089 ≤ ES23 ≤ 32.526) & (53.368667 ≤ ES40 ≤ 54.195333) & (43.715667 ≤ ES41 ≤ 44.310333) & (35.642667 ≤ EV7 ≤ 40.572333) & (EV8 ≥ 47.291)	C1	10	6
R9	(0.195023 ≤ ES1 ≤ 0.249177) & (4.7867 ≤ ES35 ≤ 4.8936) & (0.010041 ≤ ES37 ≤ 0.02523) & (0.82287 ≤ ES38 ≤ 0.83839) & (43.715667 ≤ ES41 ≤ 44.310333) & (53.429667 ≤ EV2 ≤ 54.230333)	C1	5	6
R10	(3633.5 ≤ ES2 ≤ 3682.6) & (ES16 ≥ 3109.366667) & (ES27 ≤ 18.755667) & (2.250567 ≤ ES34 ≤ 2.297233) & (0.82287 ≤ ES38 ≤ 0.83839) & (ES39 ≤ 0.091435) & (EV11 ≤ 17.228667)	C1	5	7
R11	(ES9 ≤ 120.43) & (0.33132 ≤ ES10 ≤ 0.35293) & (79.953 ≤ ES11 ≤ 80.334) & (ES22 ≤ 77.154667) & (32.089 ≤ ES23 ≤ 32.526) & (ES27 ≤ 18.755667)	C1	5	6
R12	(9.3008 ≤ ES4 ≤ 9.473) & (26.731 ≤ ES5 ≤ 27.13) & (ES13 ≥ 2641.533333) & (ES29 ≤ 32.740333) & (ES41 ≤ 43.715667) & (EV11 ≥ 19.929333)	C1	5	6
R13	(ES10 ≤ 0.35293) & (ES29 ≤ 2.740333) & (0.82287 ≤ ES38 ≤ 0.83839) & (ES39 ≥ 0.104042) & (ES41 ≤ 43.715667) & (EV10 ≥ 44.772)	C1	5	6
R14	(32.089 ≤ ES23 ≤ 32.526) & (ES27 ≤ 18.755667) & (23.698333 ≤ ES31 ≤ 24.338667) & (EV5 ≥ 22.177667) & (43.613 ≤ EV8 ≤ 47.291) & (42.146 ≤ EV10 ≤ 44.772)	C1	5	6

<b>R15</b>	$(ES18 \leq 65.531333) \& (ES19 \leq 226.816667) \& (6.8213 \leq ES26 \leq 6.9733) \& (0.82287 \leq ES38 \leq 0.83839) \& (EV10 \geq 44.772) \& (17.228667 \leq EV11 \leq 19.929333)$	C1	5	6
<b>R16</b>	$(ES28 \geq 1.681933) \& (18.235333 \leq ES33 \leq 18.799667) \& (0.010041 \leq ES37 \leq 0.02523) \& (ES41 \leq 43.715667) \& (EV1 \leq 62.563) \& (21.471333 \leq EV5 \leq 22.177667) \& (EV8 \geq 47.291) \& (46.657667 \leq EV9 \leq 48.549333)$	C1	5	8
<b>Moy</b>				6.625

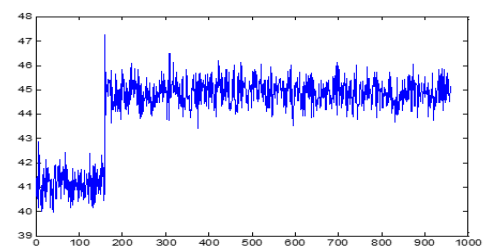
#### IV.5.4 L'importance des variables des données de TEP

Le tableau IV-3, illustre la pertinence de chaque variable de l'ensemble de données TEP, qui a été dérivée à partir des règles de classification du modèle de diagnostic obtenu dans cette étude à l'aide de l'IDEOA. Comme indique le tableau (IV.3), le débit d'eau de refroidissement du réacteur est le facteur crucial pour identifier le défaut F4, associé à une variation soudaine de la température d'entrée de l'eau de refroidissement réfrigérée. Ainsi, lorsque le défaut F4 se produit, nous observons une forte perturbation de la variable EV10, ce qui indique une augmentation significative du débit de l'eau de refroidissement réfrigérée.

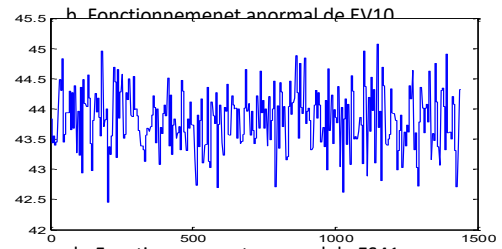
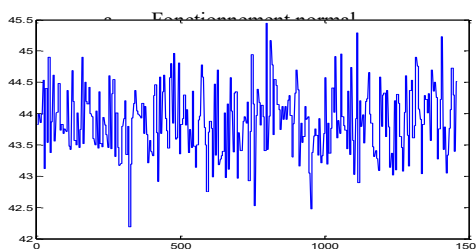
De plus, nous constatons que d'autres variables, comme ES41 et EV11, sont affectées par l'apparition du défaut F4. Les autres variables restent quasiment stables après l'apparition du défaut, ce qui permet aussi de constater un petit écart entre l'apparition de défaut et les conditions de fonctionnement normales. Les courbes de contribution de la figure 4, réalisées avec Matlab12, indiquent l'occurrence du défaut F4. Ces courbes sont obtenues lors de l'utilisation de l'ensemble de données d'apprentissage et de test de TEP, pour les variables critiques ES41, EV10 et EV11.



a. Fonctionnement normal



b. Fonctionnement anormal de EV10



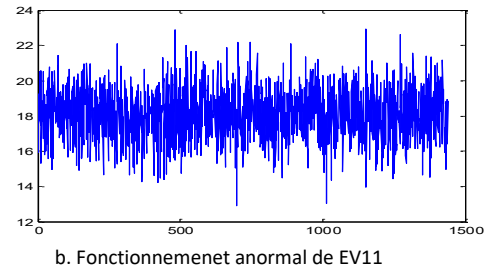
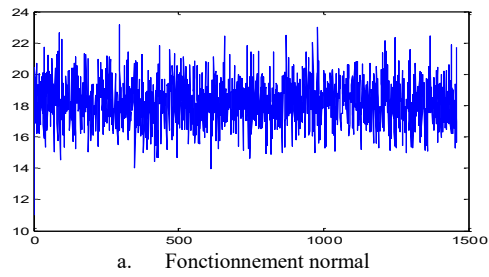


Figure IV. 3.

Variation des variables pertinentes contribuant à l'apparition de défaut F4 en fonctionnement normal et anormal.

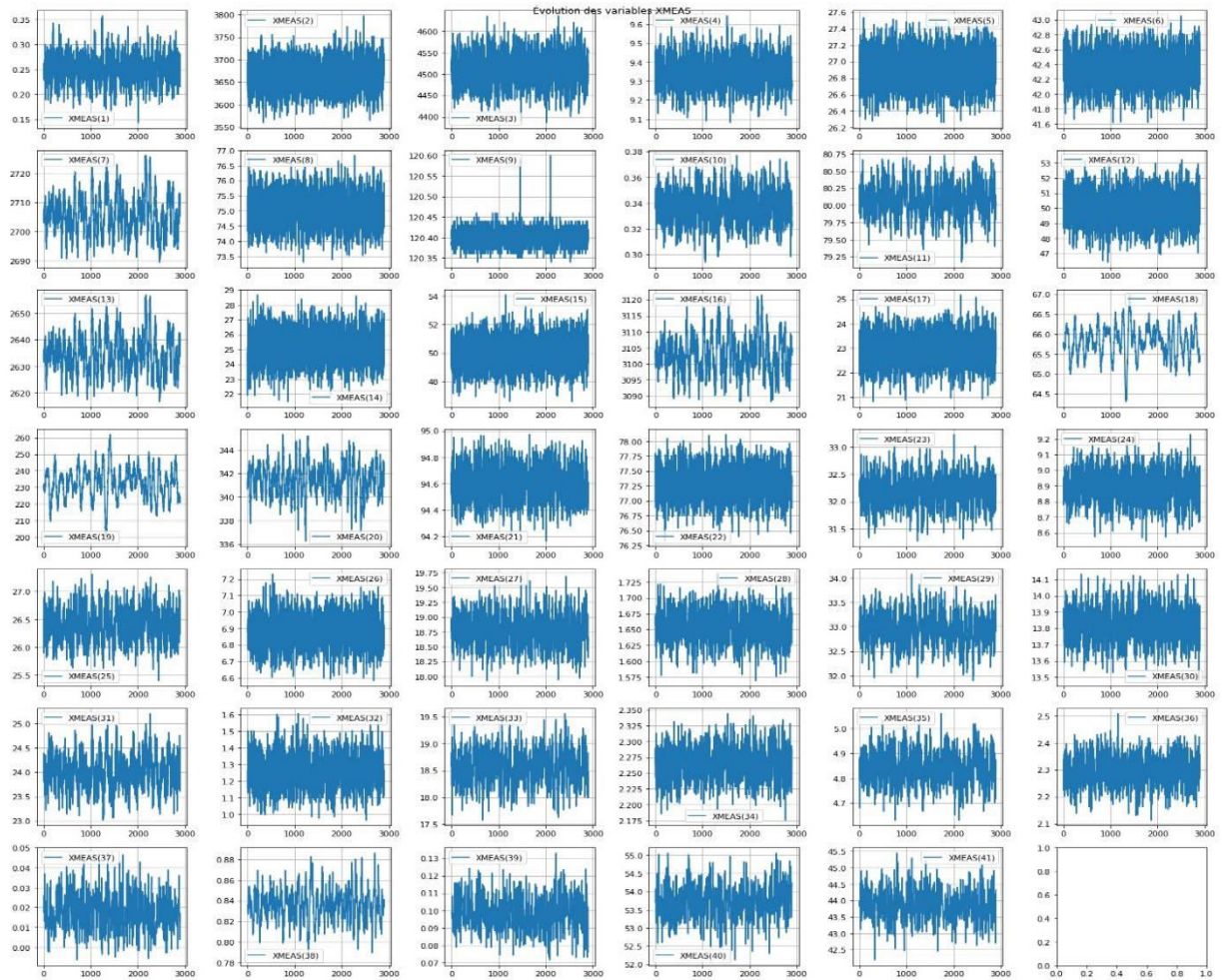


Figure IV. 2. Évolution des variables de TEP.

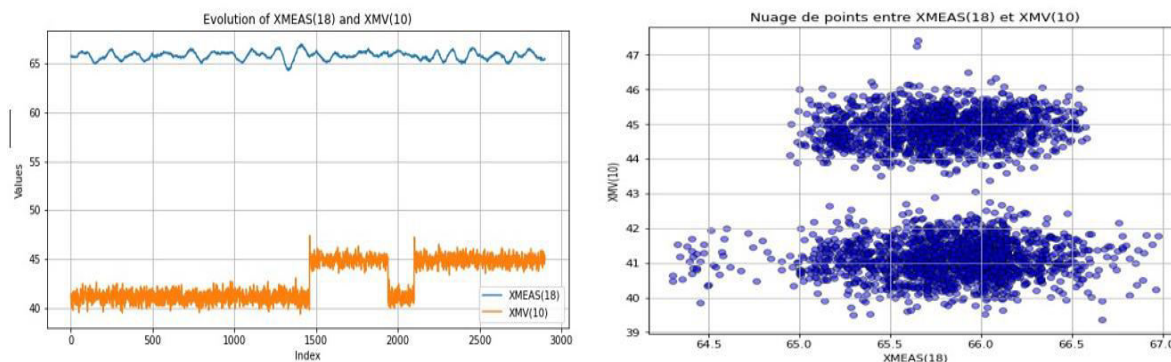


Figure IV. 4. Évolution des variables Es18 et Ev10.

Tableau IV- 3- Importance de la valeur de chaque caractéristique dans le classificateur à base d'IDEOA créé pour le diagnostic de TEP.

N°	Caractéristique	Importance des valeurs (%)	N°	Caractéristique	Importance des valeurs (%)	N°	Caractéristique	Importance des valeurs (%)
1	ES1	1.89%	19	ES19	1.89%	37	ES37	1.89%
2	ES2	1.89%	20	ES20	3.77%	38	ES38	3.77%
3	ES3	0.94%	21	ES21	0	39	ES39	2.83%
4	ES4	1.89%	22	ES22	0.94%	40	ES40	1.89%
5	ES5	1.89%	23	ES23	3.77%	41	ES41	5.66%
6	ES6	0	24	ES24	0	42	EV1	0.94%
7	ES7	0.94%	25	ES25	0.94%	43	EV2	1.89%
8	ES8	0.94%	26	ES26	0.94%	44	EV3	1.89%
9	ES9	0.94%	27	ES27	3.77%	45	EV4	0.94%
10	ES10	1.89%	28	ES28	3.77%	46	EV5	2.83%
11	ES11	1.89%	29	ES29	1.89%	47	EV6	0
12	ES12	0	30	ES30	0	48	EV7	1.89%
13	ES13	1.89%	31	ES31	4.72%	49	EV8	4.72%
14	ES14	0.94%	32	ES32	0.94%	50	EV9	2.83%
15	ES15	0	33	ES33	1.89%	51	EV10	7.55%
16	ES16	0.94%	34	ES34	1.89%	52	EV11	4.72%
17	ES17	1.89%	35	ES35	1.89%			
18	ES18	0.01%	36	ES36	0.009%			

#### IV.5.5 Comparaison des résultats de l'IDEOA avec ceux d'autres approches

Pour prouver l'efficacité de l'approche proposée, nous avons comparé les résultats obtenus par l'IDEOA lors de cette étude avec ceux des autres algorithmes de référence. L'approche proposée a été appliquée à l'ensemble de données TEP pour générer un modèle de diagnostic à base des règles. Pour cette comparaison, nous évaluons la précision de classification de chaque approche lors de l'apparition du défaut F4. Les résultats obtenus sont indiqués dans le tableau IV-4 ; d'après ces résultats, nous remarquons que l'IDEOA suggéré marque une précision plus supérieure que les autres, dans la résolution du problème de détection et de diagnostic du défaut F4 avec une meilleure valeur de 95,47 %. En revanche, LAD (Logical Analysis Data) a également montré une précision élevée de 95,04 %.

Enfin, comme indiqué dans le tableau IV-4, le résultat dans [89], utilisant l'approche Deep-Learning DDSAE ; est approximativement proche de notre résultat. La différence entre l'utilisation des approches DL et les méthodes méta-heuristiques (SWARM) ; est que notre approche est particulièrement plus efficace, idéale et plus rapide pour donner des solutions aux problèmes d'optimisation dynamique, et qu'elle peut offrir une meilleure précision globale en évitant les minima locaux. Ainsi, l'approche utilisée dans ce travail surpasse l'approche utilisée dans [89] en termes de temps d'exécution, en raison de sa simplicité, de sa robustesse et de la faible dépendance des données. De plus, ses règles générées sont faciles à comprendre par les experts du domaine, contrairement aux résultats des approches DL, qui manquent de transparence avec des explications limitées pour les applications critiques.

Par conséquent, dans cette section, les résultats obtenus par l'IDEOA ont démontré sa supériorité dans la résolution du problème de détection et de diagnostic des défauts de F4, comparativement à d'autres algorithmes de la littérature. La figure IV.5 représente le diagramme à barres de la précision obtenue par notre méthode et d'autres techniques.

Tableau IV- 4- Valeur de précision obtenue par l'IDEOA et d'autres méthodes.

Méthode	Référence	Précision %
<b>IDEOA</b>	Our approach	95.47
<b>LAD</b>	[94]	95.04
<b>SVM</b>	Weka tool	76.33
<b>ANN</b>	Weka tool	69.44
<b>ZeroR</b>	Weka tool	49.75
<b>RandomTree</b>	Weka tool	89.55
<b>DL DDSAE</b>	[89]	99.7

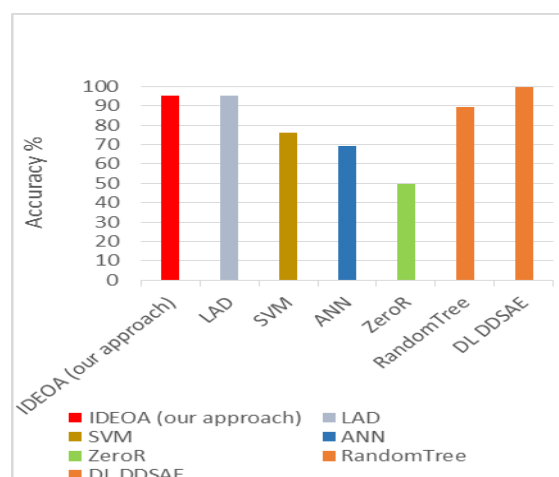


Figure IV. 5. Diagramme à barres de la précision obtenue par l'IDEOA et d'autres méthodes.

## **IV.6 Conclusion**

Ce chapitre met en lumière l'efficacité et la performance de la méthode d'optimisation développée au fil de ce travail pour s'attaquer à la problématique de détection et de diagnostic des défauts industriels. La performance de l'approche utilisée dépend de plusieurs critères ; parmi lesquels relèvent : la précision de l'algorithme, le nombre des règles et la taille des règles définissant la taille des classifieurs générés par le modèle de diagnostic des défauts.

Cette recherche est réalisée par l'application de la méthode d'optimisation IDEOA à l'ensemble de données de TEP (un processus industriel chimique et réaliste) pour l'optimisation du problème de détection et de diagnostic des défauts. Cette approche a donné des bons résultats en termes de précision et d'interprétation. Pour montrer la performance de l'approche utilisée, la fonction de précision a été calculée afin de déterminer le pourcentage des instances classées. Les résultats de cette étude ont été par la suite comparés à ceux obtenus par la version DEOA, démontrant ainsi le point fort de l'IDEOA à éviter d'être piégé dans les optima locaux.

Pour évaluer la qualité de notre approche, le nombre et la taille des règles ont été également déterminés. Ainsi que, l'importance de chaque caractéristique a été calculée afin de choisir la variable la plus pertinente affectant l'occurrence du défaut F4.

Pour conclure, nous avons comparé les résultats obtenus lors de calcul de la fonction de précision par notre approche avec ceux fournis dans la littérature, cette comparaison a montré une supériorité de la méthode choisie lors de cette étude.

## ***Conclusion Générale***

La détection et le diagnostic des défauts industriels fait un sujet crucial de nos jours. Cette thèse visait principalement à élaborer une démarche pour résoudre ce problème. Cette approche s'est concentrée sur la combinaison des deux approches : une méthode d'optimisation stochastique améliorée (IDEOA), et une technique de data mining (la classification à base des règles) pour la génération d'un modèle de diagnostic des défauts hautement interprétable.

Dans le premier chapitre, nous avons rappelé des concepts fondamentaux relatifs à la détection et le diagnostic des défauts industriels, dans le but de différencier les différents termes dans ce domaine. Plusieurs techniques de détection et de diagnostic des défauts, qui ont été mentionnées dans la littérature, sont brièvement présentées dans cette section.

Dans le deuxième chapitre, afin d'analyser et d'explorer d'énormes quantités d'informations et de données, le data mining est présenté dans cette partie. En mettant l'accent sur ses différentes techniques et leurs domaines d'application. La classification à base des règles est retenue dans cette étude car elle est considérée comme une technique puissante dans la génération des résultats interprétables et explicables par le lecteur.

Le troisième chapitre a été consacré pour présenter brièvement les différentes méthodes d'optimisation. Par la suite, nous avons donné un aperçu général sur l'algorithme EOA, son évolution dans la littérature et une brève présentation de sa version discrète. Pour conclure, certaines améliorations ont été apportées à l'algorithme EOA dans le but de traiter la problématique de détection et de diagnostic des défauts. Une autre technique appelé OBL (Opposite Based Learning) a été intégré au DEOA pour éviter d'être piégé dans les optima locaux et pour mieux explorer l'espace de recherche en améliorant la convergence de l'algorithme proposé (IDEOA).

La dernière section est dédiée à l'application de notre approche à l'ensemble de données de TEP. Cette méthode repose sur l'amélioration de la méthode d'optimisation DEOA, pour être adaptée au domaine de FDD des systèmes industriels. L'approche proposée a été améliorée par l'adoption de la technique OBL, permettant à notre algorithme d'obtenir les meilleures solutions avec une grande précision et en échappant aux optima locaux. Cette méthode d'IDEOA est considérée comme puissante, fondée sur la collaboration entre agents simples pour résoudre des problèmes complexes et optimisés. Chaque agent suit une règle simple pour résoudre le problème complexe de manière distribuée.

Les résultats obtenus lors de notre expérience comprennent : 16 règles de classification, dont six concernent les conditions de fonctionnement normales et le reste révèle l'occurrence du défaut F4. Ces résultats montrent que notre algorithme surpasse les autres algorithmes en termes de précision de classification, avec une valeur de 95,47 %. La puissance de cet algorithme réside dans la génération de règles interprétables avec un minimum de termes, ce qui rend le modèle de diagnostic généré plus explicable, ce qui en fait un outil puissant pour étudier les causes d'un défaut. En définitive, les résultats obtenus dans cette étude démontrent l'applicabilité et la capacité explicative de l'approche proposée dans le domaine du FDD des systèmes industriels.

Ces résultats ont permis aussi de démontrer qu'il est possible de concilier performance et interprétabilité dans le diagnostic des défauts industriels. Malgré les résultats prometteurs obtenus, l'approche proposée dans ce travail souffre de certaines limites qui doivent être soulignées : tout d'abord, cette approche repose sur des méta-heuristiques, qui peuvent engendrer un coût computationnel relativement élevé. De plus, bien que l'interprétabilité soit assurée via des règles, la complexité de celles-ci peut augmenter avec la dimension des données.

Cependant, l'application de cette méthode aux problèmes impliquant des scénarios de défaillance complexes, avec des ensembles de données volumineux et un espace de recherche de grande dimension, peut affecter la capacité des paramètres des agents de recherche, rendant leur coordination difficile. En outre, avec l'augmentation de la taille du problème et des ensembles de données, le nombre d'itérations s'accroît en parallèle. Cependant, l'étude du comportement de ces grands groupes d'agents peut devenir coûteuse, en termes de ressources et de temps de calcul. Malgré ces limitations, l'IDEOA montre d'excellentes performances pour de nombreux types de problèmes, notamment ceux nécessitant des solutions robustes et distribuées.

Suite aux travaux présentés dans ce document, De nombreuses perspectives ouvertes concernant plusieurs axes de recherche complémentaires peuvent être envisagées. Deux aspects sont à considérer : d'une part, notre approche nécessite une validation supplémentaire sur des problèmes concrets ; nous envisageons donc de l'appliquer aux systèmes industriels réels tels que les turbines et des échangeurs de chaleur. D'autre part, et pour développer un outil puissant en domaine du FDD, nous souhaitons lier notre approche proposée à d'autres techniques en intégrant les dernières avancées en ce domaine. Une autre perspective consiste à aborder la problématique de l'extraction de caractéristiques dans le cadre du FDD.

## ***Références bibliographiques***

1. Batteux M, Diagnosticabilit MB, Pilotés ST. Diagnosticabilité et diagnostic de systèmes technologiques pilotés To cite this version : 2011;
2. R.Isermann. Supervision FDD Methods - An Introduction. Control Eng Pract [Internet]. 1997;5(5):639–52. Available from: [https://doi.org/10.1016/S0967-0661\(97\)00046-4](https://doi.org/10.1016/S0967-0661(97)00046-4)
3. Haouassi H, Mahdaoui R, Chouhal O, Bekhouche A. An efficient classification rule generation for coronary artery disease diagnosis using a novel discrete equilibrium optimizer algorithm. J Intell Fuzzy Syst. 2022;43(3):2315–31.
4. Sampson JR. Adaptation in natural and artificial systems (John H. Holland). Society for Industrial and Applied Mathematics; 1976.
5. Kirkpatrick S, Gelatt Jr CD, Vecchi MP. Optimization by simulated annealing. Science (80- ). 1983;220(4598):671–80.
6. Kirkpatrick S. Optimization by simulated annealing: Quantitative studies. J Stat Phys. 1984;34:975–86.
7. Kennedy J, Eberhart R. Particle swarm optimization. In: Proceedings of ICNN'95-international conference on neural networks. ieee; 1995. p. 1942–8.
8. Faramarzi A, Heidarinejad M, Stephens B, Mirjalili S. Equilibrium optimizer: A novel optimization algorithm. Knowledge-Based Syst. 2020 Mar 5;191.
9. MALIK MM, HAOUASSI H. Efficient sequential covering strategy for classification rules mining using a discrete equilibrium optimization algorithm. J King Saud Univ - Comput Inf Sci [Internet]. 2021; Available from: <https://doi.org/10.1016/j.jksuci.2021.08.032>
10. Isermann R. Fault-Diagnosis Applications. Springer H. Springer Berlin Heidelberg, Berlin, Heidelberg 2011, editor. 2011.
11. Vapnik VN. An overview of statistical learning theory. IEEE Trans Neural Networks. 1999;10(5):988–99.
12. Liu B, Hsu W, Ma Y. KDD98-012.pdf. Kdd [Internet]. 1998;98:80–6. Available from: [www.aaai.org](http://www.aaai.org)

13. Monitorji G. Weihua Li. 1999;(ii):6686–91.
14. Rengaswamy R, Venkatasubramanian V. A syntactic pattern-recognition approach for process monitoring and fault diagnosis. *Eng Appl Artif Intell.* 1995;8(1):35–51.
15. Pitts W. The linear theory of neuron networks: The dynamic problem. *Bull Math Biophys.* 1943;5:23–31.
16. Touaf S. Diagnostic logique des systèmes complexes et dynamiques dans un contexte multi-agent. Université Joseph-Fourier-Grenoble I; 2005.
17. Philippot A. Contribution au diagnostic décentralisé des systèmes à événements discrets: Application aux systèmes manufacturiers. Reims; 2006.
18. Letellier C. Diagnostic robuste des systèmes incertains . Application à un système mécatronique pour l ' automobile . Clément Letellier To cite this version : HAL Id : tel-00747981 Docteur de l ' université de Rouen Application à un système mécatronique pour. 2012;
19. KHENFER R. Détection et isolation de défauts combinant des méthodes à base de données appliquées aux systèmes électro-énergétiques. 2015.
20. Isermann R, Schwarz R, Stölzl S. Fault-Tolerant Drive-by-Wire Systems – Concepts and Realizations –. *IFAC Proc Vol.* 2000;33(11):1–15.
21. Idrissi I. Contribution au Diagnostoc des Défauts de la Machine Asynchrone Doublement Alimentée de l'Eolienne à Vitesse Variable. 2019.
22. Ploix S. Diagnostic des systèmes incertains: l'approche bornée. Institut National Polytechnique de Lorraine; 1998.
23. Zwingelstein G. Failures diagnosis. Theory and practice for industrial systems; Diagnostic des defaillances. Theorie et pratique pour les systemes industriels. 1995;
24. Derbel H. Diagnostic ` a base de mod ` eles des syst ` emes temporis ´ es et d ' une sous-classe de syst ` emes dynamiques hybrides Haithem Derbel To cite this version : Université Joseph Fourier - Grenoble I Titre : Université Joseph Fourier - Grenoble I; 2009.
25. Dubuisson B. Automatique et statistiques pour le diagnostic. Hermes Sci Eur Ltd. 2001;1153.

26. Frank PM. Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy: A survey and some new results. *automatica*. 1990;26(3):459–74.
27. Patton RJ, Chen J. A review of parity space approaches to fault diagnosis. *IFAC Proc Vol*. 1991;24(6):65–81.
28. Chafouk H, Hoblos G, Langlois N, Le Gonidec S, Ragot J. Soft computing approach for data validation. *J Aerosp Comput Inf Commun*. 2007;4(1):628–35.
29. Venkatasubramanian V, Rengaswamy R, Kavuri SN, Yin K. A review of process fault detection and diagnosis: Part III: Process history based methods. *Comput Chem Eng*. 2003 Mar 15;27(3):327–46.
30. Dash S, Venkatasubramanian V. Challenges in the industrial applications of fault diagnostic systems. *Comput Chem Eng*. 2000 Jul 15;24(2–7):785–91.
31. Abdenour Soualhi. Du diagnostic au pronostic de pannes des entraînements électriques. 2013;12–9.
32. Daly KC, Gai E, Harrison J V. Generalized likelihood test for FDI in redundant sensor configurations. *J Guid Control Dyn*. 1979;2(1):9–17.
33. Fellouah R. Contribution au diagnostic de pannes pour les systèmes différentiellement plats To cite this version : HAL Id : tel-00206317 Spécialité : Automatique par Monsieur Rabah FELLOUAH Contribution au Diagnostic de Pannes pour les Systèmes Différentiellement Pla. 2008;
34. Beard RV. Failure accomodation in linear systems through self-reorganization. Massachusetts Institute of Technology; 1971.
35. Jones HL. Failure detection in linear systems. Massachusetts Institute of Technology; 1973.
36. Marzat J. Diagnostic des systèmes aéronautiques et réglage automatique pour la comparaison de méthodes [Internet]. Université Paris-Sud XI; 2011. Available from: <http://tel.archives-ouvertes.fr/tel-00647333/>
37. L.H. Chiang, E. Russell RDB. *Fault Detection and Diagnosis in Industrial Systems*. Springer-Verlag London 2001; 2001. 281 p.
38. Mohamed A. Surveillance et diagnostic d'une ligne de production par les réseaux de

- neurones artificiels. [Internet]. UNIVERSITÉ M'HAMED BOUGARA - BOUMERDES; 2012. Available from: [http://fr.wikipedia.org/w/index.php?title=R?seau\\_de\\_neurones\\_artificiels&oldid=85099731](http://fr.wikipedia.org/w/index.php?title=R?seau_de_neurones_artificiels&oldid=85099731)
39. Abdelhamid L, Razki BA. Modeling of fault detection in an industrial system using metaheuristic approaches.
40. Theilliol D. Contribution à l'étude et au développement des systèmes tolérants aux défauts: diagnostic et accommodation à base de modèles linéaires et au-delà. Habilit à Dir des Rech Univ Henri Poincaré-Nancy. 2003;1:46.
41. Pujari AK. Data Mining Techniques [Internet]. Universities Press; 2001. Available from: <https://books.google.dz/books?id=dH2KQhJboSYC>
42. Makhlouf L. Introduction à la fouille de données Réalisé par : Sommaire. 2021;
43. Besse P, Le Gall C, Raimbault N, Sarpy S. Data mining et statistique. J la société française Stat [Internet]. 2001;142(1):5–36. Available from: <http://www.math.univ-toulouse.fr/~besse/pub/sfdsmn.pdf>
44. Mieno F, Sato T, Shibuya Y, Odagiri K, Tsuda H, Take R. Yield improvement using data mining system. 1999 IEEE Int Symp Semicond Manuf ISSM 1999 - Conf Proc. 1999;391–4.
45. Gardner RM, Bieker J, Elwell S, Thalman R, Rivera E. Solving tough semiconductor manufacturing problems using data mining. ASMC (Advanced Semicond Manuf Conf Proc. 2000;2000-Janua:46–55.
46. Friedman JH. 2 What is Data Mining ? 1 Introduction. Statistics (Ber). 1997;(May).
47. Vadim K. Overview of different approaches to solving problems of data mining. Procedia Comput Sci [Internet]. 2018;123:234–9. Available from: <https://doi.org/10.1016/j.procs.2018.01.036>
48. Enzécéri JPB. Morgan, J. P. (1837-1913). Encycl Am Recessions Depress Vol 1-2. 2014;1:365–8.
49. Fayyad, Irani. Multi-Interval Discretization of Continuous Attributes for Classification Learning. 1993;

50. Han J, Kamber M, Pei J. Third Edition : Data Mining Concepts and Techniques. J Chem Inf Model [Internet]. 2012;53(9):1689–99. Available from: <http://library.books24x7.com/toc.aspx?bkid=44712>
51. Tan PN, Steinbach M, Kumar V, Pang-Ning T, Steinbach M, Kumar V. Introduction to data mining: Instructur's. Libr Congr. 2006;769.
52. Stéphane T. Data mining et statistique décisionnelle: l'intelligence des données. Editions Technip; 2012.
53. Jiawei Han MK and JP. Data Mining: Concepts and Techniques, Third Edition - Books24x7. Morgan Kaufmann Publ [Internet]. 2012;745. Available from: <http://library.books24x7.com/toc.aspx?bkid=44712>
54. FAROU B. Polycopié de cours. In 2022.
55. De Leon-Aldaco SE, Calleja H, Aguayo Alquicira J. Metaheuristic Optimization Methods Applied to Power Converters: A Review. IEEE Trans Power Electron. 2015;30(12):6791–803.
56. MALIK MM, HAOUASSI H. Efficient sequential covering strategy for classification rules mining using a discrete equilibrium optimization algorithm. J King Saud Univ - Comput Inf Sci [Internet]. 2021;34(9):7559–69. Available from: <https://doi.org/10.1016/j.jksuci.2021.08.032>
57. Gardeux V. Conception d'heuristiques d'optimisation pour les problèmes de grande dimension. Application à l'analyse de données de puces à ADN. 2011;
58. De E, Des ART, Globale O. Berthiau2001. 2002;35(2001):329–65.
59. Hajji O. Contribution au développement de méthodes d'optimisation stochastique. Application à la conception des dispositifs électrotechniques. Thèse Dr spécialité Génie électrique, Ec Cent Lille, Univ des Sci Technonlogie, Lille. 2003;1–171.
60. Shanno DF. Conjugate Gradient Methods with Inexact Searches. Math Oper Res. 1978;3(3):244–56.
61. Andrei N. Quasi-Newton Methods. Springer Optim Its Appl. 2022;195:261–314.
62. Schittkowski K. Mathematische Operationsforschung und Statistik . Series Optimization : A Journal of Mathematical Programming and Operations Research On

- inventory problems with arbitrary cost pattern , demand pattern and demand distribution. A J Math Program Oper Res. 2007;(August 2013):37–41.
63. Rao SS. Engineering optimization: Theory and practice. Eng Optim Theory Pract. 2019;1–798.
64. Al-Sultan KS, Al-Fawzan MA. A tabu search Hooke and Jeeves algorithm for unconstrained optimization. Eur J Oper Res. 1997;103(1):198–208.
65. Costa D. A tabu search algorithm for computing an operational timetable. Eur J Oper Res. 1994;76(1):98–110.
66. Yang XS. Engineering optimization: an introduction with metaheuristic applications. John Wiley & Sons; 2010.
67. Loubiere P. Amélioration des métaheuristiques d'optimisation à l'aide de l'analyse de sensibilité To cite this version: Amélioration des métaheuristiques d'optimisation à l'aide de l'analyse de sensibilité. 2017;
68. Hajek B. Mathematics of Operations Research Cooling Schedules for Optimal Annealing. 1988;(August 2015).
69. Faigle U, Kern W. Some convergence results for probabilistic Tabu Search. ORSA J Comput. 1992;4(1):32–7.
70. Glover F, Hanafi S. Tabu search and finite convergence. Discret Appl Math. 2002;119(1–2):3–36.
71. Wright SJ. Numerical optimization. 2006.
72. Abdel-Basset M, Mohamed R, Mirjalili S. A Binary Equilibrium Optimization Algorithm for 0–1 Knapsack Problems. Comput Ind Eng [Internet]. 2021;151:106946. Available from: <https://doi.org/10.1016/j.cie.2020.106946>
73. Wang J, Yang B, Li D, Zeng C, Chen Y, Guo Z, et al. Photovoltaic cell parameter estimation based on improved equilibrium optimizer algorithm. Energy Convers Manag [Internet]. 2021;236:114051. Available from: <https://doi.org/10.1016/j.enconman.2021.114051>
74. Too J, Mirjalili S. General Learning Equilibrium Optimizer: A New Feature Selection Method for Biological Data Classification. Appl Artif Intell [Internet]. 2021;35(3):247–

63. Available from: <https://doi.org/10.1080/08839514.2020.1861407>
75. Kanti K, Guha R, Kumar S, Sarkar R. BEO : Binary Equilibrium Optimizer Combined with Simulated Annealing for Feature Selection.
76. Fan Q, Huang H, Yang K, Zhang S, Yao L, Xiong Q. A modified equilibrium optimizer using opposition-based learning and novel update rules. *Expert Syst Appl* [Internet]. 2021;170(October 2020):114575. Available from: <https://doi.org/10.1016/j.eswa.2021.114575>
77. Liu SH, Mernik M, Hrnčič D, Črepinšek M. A parameter control method of evolutionary algorithms using exploration and exploitation measures with a practical application for fitting Sovova's mass transfer model. *Appl Soft Comput J*. 2013;13(9):3792–805.
78. Ul Hassan N, Bangyal WH, Ali Khan MS, Nisar K, Ibrahim AAA, Rawat DB. Improved opposition-based particle swarm optimization algorithm for global optimization. *Symmetry (Basel)*. 2021;13(12):1–23.
79. Rahab H, Haouassi H, Souidi MEH, Bakhouché A, Mahdaoui R, Bekhouche M. A Modified Binary Rat Swarm Optimization Algorithm for Feature Selection in Arabic Sentiment Analysis. *Arab J Sci Eng* [Internet]. 2022; Available from: <https://doi.org/10.1007/s13369-022-07466-1>
80. Tizhoosh HR. Opposition-based learning: A new scheme for machine intelligence. *Proc - Int Conf Comput Intell Model Control Autom CIMCA 2005 Int Conf Intell Agents, Web Technol Internet*. 2005;1:695–701.
81. Downs JJ, Vogel EF. A plant-wide industrial process control problem. *Comput Chem Eng*. 1993;17(3):245–55.
82. Evan Russell LHC, Braatz and RD. *Data-driven Methods for Fault Detection and Diagnosis in Chemical Processes*. Springer-Verlag London Berlin Heidelberg; 2000.
83. Detroja KP, Gudi RD, Patwardhan SC. Plant-wide detection and diagnosis using correspondence analysis. *Control Eng Pract*. 2007;15(12):1468–83.
84. Zou W, Xia Y, Li H. Fault diagnosis of Tennessee-eastman process using orthogonal incremental extreme learning machine based on driving amount. *IEEE Trans Cybern*. 2018;48(12):3403–10.

85. Onel M, Kieslich CA, Pistikopoulos EN. A nonlinear support vector machine-based feature selection approach for fault detection and diagnosis: Application to the Tennessee Eastman process. *AIChE J.* 2019;65(3):992–1005.
86. Soraya B, Faouzi HM, Abderrazak L. Fault Diagnosis of Tennessee Eastman Process Based on Static PCA. *Proc - 2019 1st Int Conf Sustain Renew Energy Syst Appl ICSRESA 2019.* 2019;(2).
87. Zhang C, Guo Q, Li Y. Fault detection in the Tennessee eastman benchmark process using principal component difference based on k-nearest neighbors. *IEEE Access.* 2020;8:49999–50009.
88. Han Y, Ding N, Geng Z, Wang Z, Chu C. An optimized long short-term memory network based fault diagnosis model for chemical processes. *J Process Control [Internet].* 2020;92:161–8. Available from: <https://doi.org/10.1016/j.jprocont.2020.06.005>
89. Agarwal P, Tamer M, Budman H. Explainability: Relevance based dynamic deep learning algorithm for fault detection and diagnosis in chemical processes. *Comput Chem Eng [Internet].* 2021;154:107467. Available from: <https://doi.org/10.1016/j.compchemeng.2021.107467>
90. Lv F, Wen C, Bao Z, Liu M. Fault diagnosis based on deep learning. In: *Proceedings of the American Control Conference.* 2016. p. 6851–6.
91. Lomov I, Lyubimov M, Makarov I, Zhukov LE. Fault detection in Tennessee Eastman process with temporal deep learning models. *J Ind Inf Integr [Internet].* 2021;23(September 2020):100216. Available from: <https://doi.org/10.1016/j.jii.2021.100216>
92. Xu H, Ren T, Mo Z, Yang X. A Fault Diagnosis Model for Tennessee Eastman Processes Based on Feature Selection and Probabilistic Neural Network. *Appl Sci.* 2022;12(17).
93. Hu M, Hu X, Deng Z, Tu B. Fault Diagnosis of Tennessee Eastman Process with XGB-AVSSA-KELM Algorithm. *Energies.* 2022;15(9):1–25.
94. Ragab A, El-koujok M, Poulin B, Amazouz M, Yacout S. Fault diagnosis in industrial chemical processes using interpretable patterns based on Logical Analysis of Data. *Expert Syst Appl [Internet].* 2018;95:368–83. Available from: <https://doi.org/10.1016/j.eswa.2017.11.045>

95. Maran Beena A, Pani AK. Fault Detection of Complex Processes Using nonlinear Mean Function Based Gaussian Process Regression: Application to the Tennessee Eastman Process. Arab J Sci Eng [Internet]. 2021;46(7):6369–90. Available from: <https://doi.org/10.1007/s13369-020-05052-x>

## Annexe A

### Pseudo code de l'algorithme EOA

```

Initialize the particle's populations  $i=1, \dots, n$ 
Assign equilibrium candidates' fitness a large number
Assign free parameters  $a_1=2; a_2=1; GP=0.5;$ 
While  $Iter < Max\_iter$ 
  For  $i=1$ : number of particles ( $n$ )
    Calculate fitness of  $i^{th}$  particle
    If  $fit(\vec{C}_i) < fit(\vec{C}_{eq1})$ 
      Replace  $\vec{C}_{eq1}$  with  $\vec{C}_i$  and  $fit(\vec{C}_{eq1})$  with  $fit(\vec{C}_i)$ 
    Elseif  $fit(\vec{C}_i) > fit(\vec{C}_{eq1})$  &  $fit(\vec{C}_i) < fit(\vec{C}_{eq2})$ 
      Replace  $\vec{C}_{eq2}$  with  $\vec{C}_i$  and  $fit(\vec{C}_{eq2})$  with  $fit(\vec{C}_i)$ 
    Elseif  $fit(\vec{C}_i) > fit(\vec{C}_{eq1})$  &  $fit(\vec{C}_i) > fit(\vec{C}_{eq2})$  &  $fit(\vec{C}_i) < fit(\vec{C}_{eq3})$ 
      Replace  $\vec{C}_{eq3}$  with  $\vec{C}_i$  and  $fit(\vec{C}_{eq3})$  with  $fit(\vec{C}_i)$ 
    Elseif  $fit(\vec{C}_i) > fit(\vec{C}_{eq1})$  &  $fit(\vec{C}_i) > fit(\vec{C}_{eq2})$  &  $fit(\vec{C}_i) > fit(\vec{C}_{eq3})$  &  $fit(\vec{C}_i) < fit(\vec{C}_{eq4})$ 
      Replace  $\vec{C}_{eq4}$  with  $\vec{C}_i$  and  $fit(\vec{C}_{eq4})$  with  $fit(\vec{C}_i)$ 
    End (If)
  End (For)
 $\vec{C}_{ave} = (\vec{C}_{eq1} + \vec{C}_{eq2} + \vec{C}_{eq3} + \vec{C}_{eq4})/4$ 
Construct the equilibrium pool  $\vec{C}_{eq,pool} = \{ \vec{C}_{eq(1)}, \vec{C}_{eq(2)}, \vec{C}_{eq(3)}, \vec{C}_{eq(4)}, \vec{C}_{eq(ave)} \}$ 
Accomplish memory saving (if  $Iter > 1$ )
Assign  $t = (1 - \frac{Iter}{Max\_iter})^{(a_2 \frac{Iter}{Max\_iter})}$  Eq (9)
  For  $i=1$ : number of particles ( $n$ )
    Randomly choose one candidate from the equilibrium pool (vector)
    Generate random vectors of  $\vec{\lambda}, \vec{r}$  from Eq (11)
    Construct  $\vec{F} = a_1 \text{sign}(\vec{r} - 0.5)[e^{-\vec{\lambda}t} - 1]$  Eq (11)
    Construct  $\vec{GCP} = \begin{cases} 0.5r_1 & r_2 \geq GP \\ 0 & r_2 < GP \end{cases}$  Eq (15)
    Construct  $\vec{G}_0 = \vec{GCP}(\vec{C}_{eq} - \vec{\lambda}\vec{C})$  Eq (14)
    Construct  $\vec{G} = \vec{G}_0 \cdot \vec{F}$  Eq (13)
    Update concentrations  $\vec{C} = \vec{C}_{eq} + (\vec{C} - \vec{C}_{eq}) \cdot \vec{F} + \frac{\vec{G}}{\lambda V} (1 - \vec{F})$  Eq (16)
  End (For)
   $Iter = Iter + 1$ 
End while

```

## ***La capacité d'exploration de l'EOA***

Voici une brève description des paramètres contribuant à la phase d'exploration par l'EOA :

- $a_l$  : ce paramètre gère l'amplitude de l'algorithme. Il sert à déterminer la distance entre le candidat d'équilibre et la nouvelle position. Alors, la capacité d'exploration a une relation directe avec ce paramètre : lorsque la valeur de  $a_l$  est élevée, cette capacité s'accroît à son tour. A noter que, lorsque  $a_l$  prend des valeurs supérieures à trois incitent les agents à rechercher sur les limites, ce qui dégrade les performances de l'exploration.
- $sign(r-0.5)$  : ce terme sert à contrôler la direction de l'exploration. Étant donné que  $r$  est compris entre 0 et 1 et suit une distribution uniforme, la probabilité est identique pour obtenir des signes similaires.
- $G_p$  : est la probabilité de génération qui contrôle la probabilité de participation à la mise à jour d'une concentration à l'aide du taux de génération.  $G_p=1$  indique qu'aucun terme de taux de génération ne contribuera au processus d'optimisation. Cette condition souligne une grande capacité d'exploration et conduit souvent à des solutions imprécises.  $G_p = 0$  indique que le terme de taux de génération participera toujours au processus, ce qui augmente la probabilité de stagnation dans les optima locaux. D'après des tests empiriques,  $GP = 0,5$  offre un bon équilibre entre les phases d'exploration et d'exploitation.
- $C_{eq}$  : désigne le pool d'équilibre, Ce vecteur est constitué à partir de cinq particules. Le choix de ces cinq particules se fait arbitrairement. Ces candidats choisis sont tous éloignés les uns des autres, dans la phase de l'itération initiale. La mise à jour des concentrations en fonction de ces particules candidates améliore la capacité de l'algorithme à effectuer une recherche globale dans l'espace. La particule moyenne permet également de découvrir des espaces de recherche inconnus lors des itérations initiales, lorsque les particules sont éloignées les unes des autres.

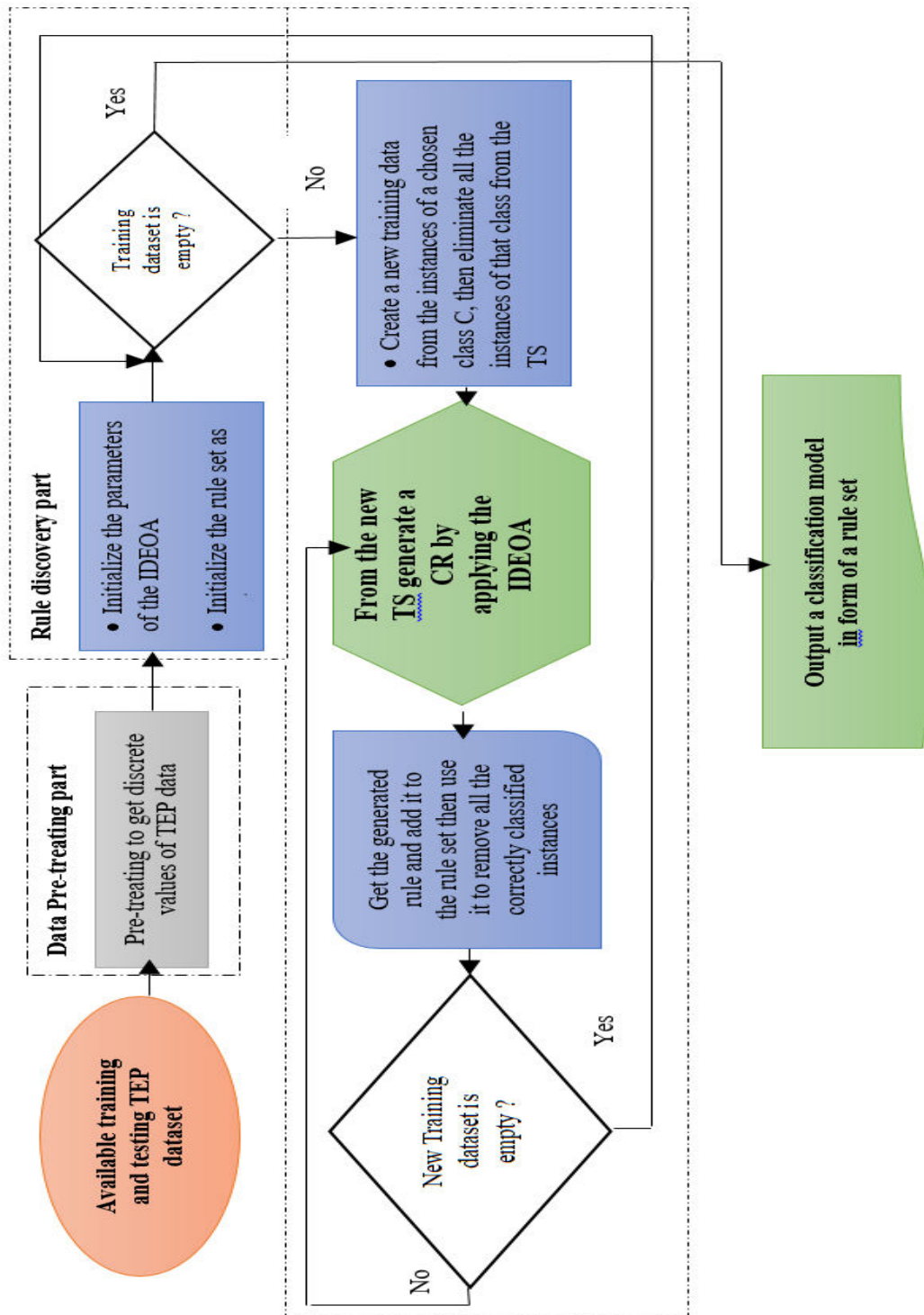
### ***La capacité d'exploitation de l'EOA***

Les principaux paramètres contribuant à la phase d'exploitation de l'EOA sont :

- $a_2$ : de même que le paramètre  $a_1$ , il a pour but de déterminer l'amplitude ou la magnitude de l'algorithme EOA, cependant, ce paramètre sert à contrôler la phase d'exploitation en recherchant la solution optimale.
- $(r - 0,5)$ : ce terme gère aussi la qualité ou la direction de l'exploitation. Il sert à déterminer la direction d'une recherche locale.
- La sauvegarde au mémoire : sert à éliminer ou à remplacer les particules les moins performantes par un certain nombre de particules les plus meilleurs. Cette fonctionnalité améliore directement la capacité d'exploitation de l'EO.
- Le pool d'équilibre : Au fil des itérations, la phase d'exploration tend à diminuer tandis que la phase de l'exploitation s'accroît. Par conséquent, lors des dernières itérations, où les candidats à l'équilibre sont proches les uns des autres, le processus de mise à jour de la concentration facilitera la recherche locale autour des candidats, menant ainsi à l'exploitation.

## Annexe B

Figure de l'annexe A. Organigramme de la méthodologie proposée à des règles utilisée pour la détection et le diagnostic des défauts.



## Annexe C

Tableau de l'annexe C. Description des variables de l'ensemble de données TEP avec leurs valeurs discrètes [95].

Variables	Type	Description	Ranks		
			Rank 1	Rank2	Rank3
ES1	Continuous Process Measurements	A Feeding	≤0.195023	[0.195023-0.249177]	≥0.249177
ES2		D Feeding	≤3633.5	[3633.5-3682.6]	≥3682.6
ES3		E Feeding	≤4458.366667	[4458.366667-4526.833333]	≥4526.833333
ES4		A and C Feeding	≤9.3008	[9.3008-9.473]	≥9.473
ES5		Flow Recycling	≤26.731	[26.731-27.13]	≥27.13
ES6		Rate of Reactor Feeding	≤42.118333	[42.118333-42.500667]	≥42.500667
ES7		Pressure of Reactor	≤2702.633333	[2702.633333-2712.866667]	≥2712.866667
ES8		Level of Reactor	≤74.772667	[74.772667-75.761333]	≥75.761333
ES9		Temp of Reactor	≤120.43	[120.43-120.51]	≥120.51
ES10		Purge Rate in	≤0.33132	[0.33132-0.35293]	≥0.35293
ES11		Temp of Product Separator	≤79.953	[79.953-80.334]	≥80.334
ES12		Level of Product Separator	≤49.165	[49.165-50.84]	≥50.84
ES13		Pressure of Product Separator	≤2631.066667	[2631.066667-2641.533333]	≥2641.533333
ES14		Underflow of Product Separator	≤23.8283	[23.828333-25.757667]	≥25.757667
ES15		Level of the Stripper	≤48.738	[48.738-50.327]	≥50.327
ES16		Pressure of the Stripper	≤3099.733333	[3099.733333-3109.366667]	≥3109.366667
ES17		Underflow of the Stripper	≤22.180333	[22.180333-23.321667]	≥23.321667
ES18		Temp of the Stripper	≤65.531333	[65.531333-65.937667]	≥65.937667
ES19		Steam Flow of the Stripper	≤226.816667	[226.816667-234.373333]	≥234.373333
ES20		Work of the Stripper	≤339.546667	[339.546667-341.323333]	≥341.323333
ES21		Outlet Temp of the Reactor Cooling Water	≤94.473333	[94.473333-94.712667]	≥94.712667
ES22	Outlet Temp of the Separator Cooling Water	≤77.154667	[77.154667-77.634333]	≥77.634333	
ES23	Sampled Process Measurements	Component A	≤32.089	[32.089-32.526]	≥32.526
ES24		Component B	≤8.842033	[8.842033-9.031467]	≥9.031467
ES25		Component C	≤26.117333	[26.117333-26.567667]	≥26.567667
ES26		Component D	≤6.8213	[6.8213-6.9733]	≥6.9733
ES27		Component E	≤18.755667	[18.755667-19.242333]	≥19.242333
ES28		Component F	≤1.643167	[1.643167-1.681933]	≥1.681933
ES29		Component A	≤32.740333	[32.740333-33.227667]	≥33.227667
ES30		Component B	≤13.767667	[13.767667-13.946333]	≥13.946333
ES31		Component C	≤23.698333	[23.698333-24.338667]	≥24.338667
ES32		Component D	≤1.150613	[1.150613-1.326107]	≥1.326107
ES33		Component E	≤18.235333	[18.235333-18.799667]	≥18.799667
ES34		Component F	≤2.250567	[2.250567-2.297233]	≥2.297233
ES35		Constituent G	≤4.7867	[4.7867-4.8936]	≥4.8936
ES36		Component H	≤2.2186	[2.2186-2.3077]	≥2.3077
ES37		Component D	≤0.010041	[0.010041-0.02523]	≥0.02523
ES38		Component E	≤0.82287	[0.82287-0.83839]	≥0.83839
ES39	Component F	≤0.091435	[0.091435-0.104042]	≥0.104042	
ES40	Component G	≤53.368667	[53.368667-54.195333]	≥54.195333	
ES41	Component H	≤43.715667	[43.715667-44.310333]	≥44.310333	
EV1	Manipulated Variables	D Feeding Flow	≤62.563	[62.563-63.682]	≥63.682
EV2		E Feeding Flow	≤53.429667	[53.429667-54.230333]	≥54.230333
EV3		A Feeding Flow	≤21.428	[21.428-25.55]	≥25.55
EV4		A and C Flow Feeding	≤60.4	[60.4-62.39]	≥62.39
EV5		Compressor Recycle Valve	≤21.471333	[21.471333-22.177667]	≥22.177667
EV6		Purge valve	≤39.244667	[39.244667-41.898333]	≥41.898333
EV7		Flow of the Separator Pot Liquid	≤35.642667	[35.642667-40.572333]	≥40.572333
EV8		Flow of the Stripper Liquid Product	≤43.613	[43.613-47.291]	≥47.291
EV9		Stripper Steam Valve	≤46.657667	[46.657667-48.549333]	≥48.549333
EV10		Flow of the Reactor Cooling Water	≤42.146	[42.146-44.772]	≥44.772
EV11		Flow of the Condenser Cooling Water	≤17.228667	[17.228667-19.929333]	≥19.929333
EV12			Agitator Speed	/	/

## Annexe D

*Tableau de l'annexe D : Ensembles des défauts de processus TEP [83].*

<b>Faults</b>	<b>Description</b>	<b>Type</b>
<b>F1</b>	A/C Feeding ratio, B composition constant	Sudden variation
<b>F2</b>	B Constituent, A/C ratio constant	Sudden variation
<b>F3</b>	D Feeding temp	Sudden variation
<b>F4</b>	Inlet temp of the Reactor cooling water (RCW)	Sudden variation
<b>F5</b>	Inlet temp of the Condenser cooling water (CCW)	Sudden variation
<b>F6</b>	A Feeding loss	Sudden variation
<b>F7</b>	C Header pressure loss–reduced validity	Sudden variation
<b>F8</b>	A, B, C Feeding consituent	Arbitrary change
<b>F9</b>	D Feeding temp	Arbitrary change
<b>F10</b>	C Feeding temp	Arbitrary change
<b>F11</b>	Inlet temp of RCW	Arbitrary change
<b>F12</b>	Inlet temp of CCW	Arbitrary change
<b>F13</b>	Reactor kinetics	Slow drift
<b>F14</b>	RCW valve	Stuck
<b>F15</b>	CCW valve	Stuck
<b>F16</b>	/	-
<b>F17</b>	/	-
<b>F18</b>	/	-
<b>F19</b>	/	-
<b>F20</b>	/	-
<b>F21</b>	The stream 4 valve was adjusted to the stable position	Constant position