

الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne démocratique et populaire

وزارة التعليم العالي والبحث العلمي
Ministère de l'enseignement supérieur et de la recherche scientifique

جامعة سعد دحلب البليدة
Université SAAD DAHLAB de BLIDA

كلية التكنولوجيا
Faculté de Technologie

قسم الإلكترونيك
Département d'Électronique



Mémoire de Projet de Fin d'Études

présenté par

Sabri Nassim

&

Benkercha Rabah

pour l'obtention du diplôme de Master en Électronique option Automatique

Thème

Contribution au Détection et Diagnostic des défauts dans les systèmes photovoltaïques par l'intelligence artificielle

Proposé par : Dr. Chouder Aissa & Dr. Kara Kamel

Année Universitaire 2012-2013

Remerciements

Nous remercions le bon Dieu de nous avoir donné la volonté et la patience qui nous ont permis de mener à bien ce travail.

On voudrait témoigner notre reconnaissance à **Monsieur CHOUDER**, Docteur au Centre du Développement des Energies Renouvelable (CDER) à Bouzareah et Professeur à l'Université Saad Dahleb de Blida, qui a dirigé cette thèse. On lui exprime nos vifs remerciements de nous avoir assuré l'encadrement pour l'élaboration de ce projet. Ses conseils, son aide constante, son suivi attentif, ses compétences scientifiques et humaines, l'enthousiasme qu'il a su nous communiquer ainsi que la confiance qu'il nous a témoignée nous ont permis d'accomplir notre tâche dans les meilleures conditions.

On est très sensible à l'honneur que nous a fait **Monsieur KARRA**, le responsable du Master Automatique à l'Université Saad Dahleb de Blida, en acceptant de participer à l'encadrement de ce projet. On lui exprime toute notre reconnaissance pour l'intérêt porté à ce travail pour son aide et son entière disponibilité.

On remercie le président et les membres du jury pour bien vouloir évaluer notre travail.

A toutes les personnes, qui ont contribué de près ou de loin, directement ou indirectement à l'aboutissement à ce travail, on les remercie tous.

Nos sincères gratitude s'adressent à nos famille et tous nos amis sans leurs encouragements permanents et leurs soutien moral on ne serais parvenue là où nous somme.

Enfin, on ne pourrait terminer ces remerciements sans une pensée à l'ensemble de nos enseignants qui sont à l'origine de tout notre savoir.

ملخص:

يمكن أن يتعرض نظام الخلايا الضوئية خلال ممارسة نشاطها إلى مختلف العيوب وأوجه القصور مما يؤدي إلى نقص في أداء النظام و حتى إلى حد التوقف التام للمجموعة. وجميع هذه الآثار الضارة تؤدي إلى تقليل الإنتاجية للمجموعة ، وبالتالي تقليل الأرباح للنظام، ناهيك عن تكاليف الصيانة لعودة النظام إلى حالته العادية. في هذا العمل من المذكرة نقتراح مقارنة للكشف وتشخيص الأعطال في النظم الكهربية الضوئية، على أساس تقنية الذكاء الاصطناعي وهي: شجرة القرارات

كلمات المفاتيح: نظام الخلايا الضوئية، تشخيص، شجرة القرارات

Résumé : Un système photovoltaïque peut être soumis, au cours de son fonctionnement, à différents défauts et anomalies conduisant à une baisse des performances du système voir son indisponibilité totale. Toutes ces conséquences défavorables vont évidemment réduire la productivité de l'installation, et donc réduire le profit de l'installation, sans compter le coût de maintenance pour remettre le système en état normal. Dans ce travail de mémoire nous proposons une approche de détection et de diagnostic des défauts dans les systèmes photovoltaïques, basée sur une technique de l'intelligence artificielle qui est : l'arbre de décision.

Mots clés : système photovoltaïque ; diagnostic ; Arbre de Décision.

Abstract: A photovoltaic system can be subjected during its operation, various defects and deficiencies leading to a decrease in system performance see the total unavailability. All these adverse effects will obviously reduce the productivity of the plant, and therefore reduce the profit of the system, not to mention the maintenance cost to return the system to normal state. In this memory work we propose an approach for the detection and diagnosis of faults in photovoltaic systems, based on a technique of artificial intelligence that is: the decision tree.

Keywords : Photovoltaic generator ; diagnosis ;Decision Tree

Listes des acronymes et abréviations

| | |
|-------------|--|
| PV | : Photovoltaïque |
| CDER | : Centre de Développement des Énergies Renouvelables |
| CPCR | : Centrale Photovoltaïque Connecté au Réseau |
| MPPT | : Maximum Power Point Traking |
| MPP | : Point de Puissance Maximale |
| AD | : Arbre de Décision |
| N | : Jonction N de la diode |
| P | : Jonction P de la diode |
| GPV | : Générateur Photovoltaïque |
| N_s | : Nombre de module en série du champ PV |
| V_{ocN_s} | : la tension de circuit ouvert de (N_s) cellules en série [V] |
| I_{scN_s} | : le courant de court circuit de (N_s) cellule en série [A] |
| N_p | : Nombre de module en parallèle du champ PV |
| V_{ocN_p} | : la tension de circuit ouvert de (N_p) cellules en parallèle [V] |
| I_{scN_p} | : le courant de court circuit de (N_p) cellule en parallèle [A] |
| I | : Courant fournit par la cellule [A] |
| I_{ph} | : Courant généré par l'éclairement [A] |
| I_{Rsh} | : Courant circulant dans la résistance shunt [A] |
| I_d | : Curant circulant dans la diode [A] |
| I_{0d} | : Courant de saturation de la diode [A] |
| E | : L'ensoleillement / l'éclairement / rayonnement absorbé par la cellule |
| E_0 | : L'ensoleillement / l'éclairement / rayonnement de référence (1000) [W/m ²] |
| R_s | : Résistance série [Ω] |
| R_{sh} | : Résistance shunt [Ω] |
| K | : Constante de Boltzmann ($K=1,38.10^{-23}$)[J/°K] |
| q | : Charge de l'électron ($q=1,602.10^{-19}$) [C] |
| I_{sc} | : Courant de shunt [A] |
| V_{oc} | : Tension de circuit-ouvert [V] |
| T | : Température ambiante mesure à partie de thermocouple [°K] ou [°C] |
| I_{PV} | : Courant continue en MPP délivré par système PV [A] |
| V_{PV} | : Tension continue en MPP délivré par système PV [V] |

- P_{PV} : Puissance continue en MPP délivré par système PV [VA]
- I_{mpp} : Courant en MPP [A]
- V_{mpp} : Tension en MPP [V]
- CC : Court-Circuit
- P_{mpp} : Puissance en MPP [W]
- P_T : Puissance maximale théorique [W]
- DC : Courant continue [A]
- AC : Courant Alternative [A]
- X : L'ensemble des attributs
- n : Le nombre d'attributs
- C : L'ensemble de classes
- k : Le nombre de classes
- $|C_j|$: Le nombre des exemples qui appartiennent à la classe C_j
- D : L'ensemble des exemples
- $|D|$: Le nombre total des exemples
- Info : Entropie / information [Bits]
- Gain : Gain d'information [Bits/instance]
- G_Ratio : Le rapport de gain d'information normalisé
- S_j : Le seuil de test sur l'attribut
- E_M : Le nombre d'exemples mal classés sur une feuille
- Z : Le Seuil de confiance de la distribution Normale centré réduite
- N_T : Le nombre Totale d'exemples couvres sur une feuille
- f : Le taux d'erreur réel
- α : Le niveau de confiance
- e : Le taux d'erreur prédit

Liste des Figures

Chapitre I : Généralités sur les systèmes Photovoltaïques

| | | |
|-------------|--|----|
| Figure I.1 | : Schéma de fonctionnement d'une cellule photovoltaïque | 4 |
| Figure I.2 | : Les différents types de cellules..... | 5 |
| Figure I.3 | : Schéma équivalent d'une cellule photovoltaïque..... | 6 |
| Figure I.4 | : Caractéristique I-V d'une cellule PV..... | 7 |
| Figure I.5 | : Caractéristique courant tension de Ns cellules en séries..... | 8 |
| Figure I.6 | : Caractéristique courant tension de Np cellules en parallèles..... | 9 |
| Figure I.7 | : Protection d'un GPV avec diode bypass et diode anti-retour..... | 10 |
| Figure I.8 | : Caractéristiques $I=f(V)$ et $P=f(V)$ d'un module PV..... | 11 |
| Figure I.9 | : Influence de l'ensoleillement E | 12 |
| Figure I.10 | : Influence de la température de la cellule T..... | 12 |
| Figure I.11 | : Influence de la résistance série R_s | 12 |
| Figure I.12 | : Influence de la résistance shunt R_{Sh} | 13 |
| Figure I.13 | : Système PV autonome avec ou sans conversion DC/AC..... | 14 |
| Figure I.14 | : Système hybrides PV éoliens | 14 |
| Figure I.15 | : Système PV raccordé au réseau..... | 15 |
| Figure I.16 | : Différents étages du groupe convertisseurs | 15 |
| Figure I.17 | : Schéma des topologies centrales d'un système PV connecté au réseau..... | 16 |
| Figure I.18 | : Schéma des topologies modulaires d'un système PV connecté au réseau..... | 17 |

Chapitre II : Diagnostic des défauts : Etat de l'Art

| | | |
|-------------|--|----|
| Figure II.1 | : Ombrage des Modules..... | 20 |
| Figure II.2 | : Quelques exemples de la localisation de défauts par la caméra thermique..... | 22 |
| Figure II.3 | : Principe de la réflectométrie pour localiser le défaut dans un string PV | 23 |
| Figure II.4 | : Un aperçu sur PVSAT2 | 25 |

Chapitre III : Classification par l'Arbre de Décision

| | | |
|--------------|--|----|
| Figure III.1 | : Exemple des frontières des classes..... | 28 |
| Figure III.2 | : Représentation de deux classes en deux dimensions..... | 28 |

| | | |
|--|---|----|
| Figure III.3 | : Frontière de décision parfaite sur les données d'apprentissage..... | 29 |
| Figure III.4 | : Nouveaux individus mal classés..... | 29 |
| Figure III.5 | : Nouveaux individus correctement classés..... | 30 |
| Figure III.6 | : L'induction par l'arbre de décision..... | 31 |
| Figure III.7 | : Caractéristiques I-V en fonctionnement Normal et Défaut..... | 32 |
| Figure III.8 | : Modèle d'arbre de décision pour détection de défaut..... | 33 |
| Figure III.9 | : Structure de l'arbre de décision pour l'ensemble D..... | 35 |
| Figure III.10 | : Les sous-ensembles associés aux valeurs des quatre attributs..... | 37 |
| Figure III.11 | : Les sous-ensembles correspondant à la valeur ensoleillée de l'attribut Temps..... | 38 |
| Figure III.12 | : L'arbre de décision final pour la base météo..... | 39 |
| Figure III.13 | : Evolution de taux d'erreur en apprentissage et en test..... | 41 |
| Figure III.14 | : Arbres de décision avec et sans élagage | 42 |
| Figure III.15 | : Elagage d'un sous-arbre. | 44 |
| Figure III.16 | : Nœud racine | 50 |
| Figure III.17 | : Nœud racine avec une feuille | 51 |
| Figure III.18 | : Le Modèle de l'arbre finale..... | 51 |
| Chapitre IV : Expérimentation et Validation | | |
| Figure IV.1 | : Schéma d'une Centrale Photovoltaïque Connectée au Réseau..... | 53 |
| Figure IV.2 | : Schéma général de la centrale PV du CDER..... | 54 |
| Figure IV.3 | : Les Différents types de défauts considérés dans le système PV..... | 54 |
| Figure IV.4 | : Le processus pour développer un modèle d'AD..... | 55 |
| Figure IV.5 | : Synoptique globale du système d'acquisition..... | 56 |
| Figure IV.6 | : La caractéristique $P_{PV} - V_{PV}$ pour différents valeurs de la température et l'ensoleillement | 57 |
| Figure IV.7 | : L'organigramme de filtrage des données en provenance de l'acquisition..... | 58 |
| Figure IV.8 | : Exemple d'un fichier Excel contenant les données avant filtrage... .. | 59 |
| Figure IV.9 | : Une partie de Modèle de l'AD pour la détection..... | 59 |
| Figure IV.10 | : L'affichage de l'apprentissage de l'AD pour la détection..... | 60 |
| Figure IV.11 | : L'affichage de l'apprentissage de l'AD pour la classification..... | 60 |
| Figure IV.12 | : Organigramme de détection et de classification par l'AD..... | 63 |
| Figure IV.13 | : Conversion d'un Modèle d'Arbre en programme MatLab..... | 64 |

| | | |
|--------------|---|----|
| Figure IV.14 | : Résultats de détection | 65 |
| Figure IV.15 | : Résultat de détection et de classification (String coupé)..... | 65 |
| Figure IV.16 | : Résultat de détection et de classification (CC deux modules)..... | 66 |
| Figure IV.17 | : Résultat de détection et de classification (défaillance)..... | 67 |
| Figure IV.18 | : Interface pour récupération des données..... | 68 |
| Figure IV.19 | : Interface utilisateur LabView | 68 |
| Figure IV.20 | : Résultat de test pour un système sain..... | 69 |
| Figure IV.21 | : Résultat de test dans le cas d'un string coupé | 69 |
| Figure IV.22 | : Résultat de test dans le cas court-circuit 2 modules..... | 70 |

Liste des Tableaux

| | | |
|---------------|---|----|
| Tableau II.1 | : Principaux défauts et anomalies rencontrés dans un générateur PV. | 21 |
| Tableau III.1 | : Exemple de construction d'un AD : La base Météo..... | 36 |
| Tableau III.2 | : La matrice de Confusion | 45 |
| Tableau III.3 | : Calcule de l'information pour l'attribut T | 47 |
| Tableau III.4 | : Calcule de l'information pour l'attribut E | 48 |
| Tableau III.5 | : Calcule de l'information pour l'attribut V_{PV} | 48 |
| Tableau III.6 | : Calcule de l'information pour l'attribut I_{PV} | 49 |
| Tableau III.7 | : Calcule de l'information pour l'attribut P_{PV} | 50 |
| Tableau III.8 | : Les valeurs de Gain Ratio..... | 50 |
| Tableau III.9 | : Matrice de confusion de l'ensemble de test..... | 51 |
| Tableau IV.1 | : Résultat de détection..... | 61 |
| Tableau IV.2 | : Résultat de classification..... | 61 |

Table des matières

| | |
|---|----------|
| Introduction Générale | 1 |
| Chapitre I : Généralités sur les systèmes photovoltaïques | |
| I.1 Introduction | 3 |
| I.2 Conversion photovoltaïque | 3 |
| I.2.1 Les cellules photovoltaïques..... | 3 |
| I.2.1.1 Cellule au silicium monocristallin | 4 |
| I.2.1.2 Cellule au silicium poly-cristallin | 5 |
| I.2.1.3 Cellule au silicium amorphe | 5 |
| I.3 Modèle équivalent d'une cellule PV | 6 |
| I.3.1 Résolution du modèle de la cellule | 6 |
| I.3.2 Caractéristique électrique d'une cellule PV | 7 |
| I.3.3 Regroupement des cellules | 8 |
| I.3.3.1 Regroupement en série | 8 |
| I.3.3.2 Regroupement en parallèle..... | 8 |
| I.3.3.3 Module photovoltaïque | 9 |
| I.3.4 Générateur photovoltaïque | 9 |
| I.4 Protections classiques d'un GPV | 10 |
| I.5 Paramètres de performance d'un module PV | 11 |
| I.5.1 Fonctionnement du MPPT | 11 |
| I.5.2 Influence des paramètres sur les caractéristiques I-V & P-V | 11 |
| I.6 Application des systèmes photovoltaïques..... | 13 |
| I.6.1 Systèmes autonomes | 13 |
| I.6.2 Systèmes hybrides | 14 |
| I.6.3 Systèmes raccordés au réseau | 14 |
| I.7 Convertisseur | 15 |
| I.7.1 Architecture du convertisseur | 15 |
| I.8 Conclusion | 17 |
| Chapitre II : Diagnostic des défauts : Etat de l'Art | |
| II.1 Introduction | 18 |
| II.2 Les défauts les plus courants | 18 |
| II.3 Méthodes de diagnostic d'une installation PV..... | 21 |

| | | |
|----------|---|----|
| II.3.1 | Méthodes de diagnostic courantes (industrialisé)..... | 21 |
| II.3.1.1 | Méthodes non-électriques..... | 22 |
| II.3.1.2 | Méthodes électriques | 22 |
| II.3.2 | Méthodes de diagnostic proposées dans la littérature..... | 23 |
| II.3.2.1 | Méthode de réflectométrie | 23 |
| II.3.2.2 | Analyse de la puissance et de l'énergie produite..... | 23 |
| II.3.2.3 | Autres méthodes pour le diagnostic..... | 24 |
| II.4 | Conclusion | 26 |

Chapitre III : Classification par l'Arbre de Décision

| | | |
|---------|---|----|
| III.1 | Introduction | 27 |
| III.2 | Classification supervisée pour la détection et le diagnostic..... | 27 |
| III.2.1 | Généralisation d'un classificateur | 28 |
| III.3 | Arbre de Décision | 30 |
| III.3.1 | Notion générale | 30 |
| III.3.2 | Types de données..... | 31 |
| III.3.3 | Exemple introductif à l'arbre de décision..... | 32 |
| III.3.4 | Construction d'un Arbre de Décision..... | 33 |
| III.4 | Choix de l'algorithme..... | 39 |
| III.4.1 | L'algorithme C4.5..... | 39 |
| III.4.2 | Critères d'arrêt de construction d'un AD | 40 |
| III.5 | Élagage..... | 41 |
| III.5.1 | Pré-élagage..... | 42 |
| III.5.2 | Post-élagage | 42 |
| III.6 | Évaluation du modèle | 44 |
| III.6.1 | Matrice de Confusion | 45 |
| III.7 | Exemple illustratif | 46 |
| III.8 | Avantage de l'AD..... | 52 |
| III.9 | Conclusion..... | 52 |

Chapitre IV : Expérimentation et Validation

| | | |
|------|---|----|
| IV.1 | Introduction | 53 |
| IV.2 | Description de la centrale photovoltaïque connectée au réseau (CPCR)..... | 53 |
| IV.3 | Différents Défaut considérée | 54 |
| IV.4 | Etapes de mise en oeuvre d'un modèle AD..... | 55 |

| | | |
|----------|--|----|
| IV.4.1 | Système d'acquisition des données..... | 55 |
| IV.4.2 | Choix des attributs de classification | 56 |
| IV.4.3 | Prétraitement de données..... | 57 |
| IV.4.4 | Entraînement et Test du Modèle de l'Arbre de Décision(AD)..... | 59 |
| IV.4.4.1 | Construction de l'AD avec WEKA | 59 |
| IV.4.4.2 | Implémentation sous l'environnement MatLab | 64 |
| IV.4.5 | Implémentation dans l'environnement LabVIEW | 67 |
| IV.4.5.1 | Test avec des données réelles..... | 69 |
| IV.5 | Conclusion | 71 |
| | Conclusion Générale | 72 |
| | Bibliographie | 74 |
| | Annexe | 76 |

Introduction générale

Au cours de ces dernières années, les technologies des systèmes photovoltaïques se sont développées rapidement et ont eu un impact considérable sur les systèmes de cogénération électrique et se considèrent comme l'énergie verte du nouveau siècle. Actuellement, des travaux sont réalisés sur les cellules (ou modules) et sur les systèmes PV afin d'améliorer les performances électriques des panneaux photovoltaïques et réduire les pertes d'énergies dans les installations PV. Un système photovoltaïque peut être soumis, au cours de son fonctionnement, à différents défauts et anomalies conduisant à une baisse des performances du système voir son indisponibilité totale. Toutes ces conséquences défavorables vont évidemment réduire la productivité de l'installation, et donc réduire le profit de l'installation, sans compter le coût de maintenance pour remettre le système en état normal.

Le diagnostic fin et la détection prématurée des défauts dans une installation PV réduit les considérablement les coûts de maintenance et augmente la productivité en augmentant le taux de disponibilité des installations en veillant à ce que leur rendement soit optimal.

A l'heure actuelle, il existe des systèmes de monitoring qui permettent de calculer la puissance fournie par l'installation et l'énergie associée par l'intermédiaire des capteurs qui mesure aussi bien les grandeurs électrique (tension et courant) que les grandeurs radiométriques (ensoleillement, température, ...etc.) qui sont placés dans le site et reliés a une acquisition de donnée.

Parmi les méthodes les plus utilisées dans la littérature pour le diagnostic et la détection de pannes dans les systèmes photovoltaïque sont ceux basées sur un modèle de prédiction. Dans cette approche, les deux puissances simulées et mesurées sont comparées et lorsqu'un écart important apparait, l'utilisateur est informé d'une panne probable. Ce genre de service est une première étape intéressante vers une supervision précise de l'installation. Néanmoins cette approche a montrée ces limites quand à l'identification du type de défaut avec précision.

Dans ce travail de mémoire de fin d'étude nous proposons une approche de détection et de diagnostic des défauts dans les systèmes photovoltaïques, basée sur une technique de l'intelligence artificielle qui est : l'arbre de décision.

Le premier chapitre est dédié aux généralités sur les systèmes photovoltaïques (effet photovoltaïque, modèle et caractéristique d'une cellule), ainsi que les définitions classiques des différents composants qui le constituent. Nous énumérons également les différentes applications des systèmes et leurs architectures.

Le deuxième chapitre présente dans un premier temps les différents défauts associés au système photovoltaïque connecté au réseau et les causes liées à ces défauts aussi bien dans la partie DC que la partie AC (après l'onduleur). Dans un second temps, nous établirons l'état de l'art sur les différentes méthodes de diagnostic utilisées pour détecter les défauts dans un système photovoltaïque.

Le troisième chapitre est l'objet d'une présentation plus approfondie des arbres de décision. Dans la première partie de ce chapitre, nous nous intéressons au principe de la classification en générale et les problèmes qui peuvent survenir. La deuxième partie est consacrée à l'étude des arbres de décision dont laquelle nous donnerons un exemple illustratif du concept des arbres de décisions appliquées pour la détection et la classification des défauts dans les systèmes photovoltaïques.

Dans le dernier chapitre nous détaillerons les différentes phases de la construction des modèles de détection et de classification appliqués à la centrale photovoltaïque connectée au réseau du Centre de Développement des Energies Renouvelables (CDER) d'une puissance crête de 3.2 KW. Nous donnerons également l'implémentation des concepts développés sur une plateforme graphique en utilisant la Co-simulation entre Labview et MatLab.

Chapitre I : Généralités sur les Systèmes Photovoltaïques

I.1 Introduction

Les énergies renouvelables sont des énergies à ressource illimitée. Ils regroupent un certain nombre de filières technologiques selon l'origine de l'énergie obtenue. Il existe plusieurs types de sources d'énergies renouvelables parmi eux : hydraulique, éolienne, biomasse et photovoltaïque. L'énergie photovoltaïque est la plus jeune des énergies renouvelables, elle a l'avantage d'être non polluante, souple et fiable [1]. Les systèmes photovoltaïques sont utilisés depuis 40 ans. Les applications ont commencé avec le programme spatial pour la transmission radio des satellites. Elles se sont poursuivies ensuite avec les balises en mer et l'équipement des sites isolés. La production d'électricité photovoltaïque dépend de l'ensoleillement du lieu et de la température, donc de sa localisation géographique, de la saison et de l'heure de la journée : la production est ainsi maximale à midi (heure solaire) par ciel dégagé.

I.2 Conversion photovoltaïque

L'effet photovoltaïque est un processus de transformation de l'énergie émise par le soleil, sous forme de photons, en énergie électrique à l'aide de composants semi-conducteurs appelés cellules photovoltaïques.

I.2.1 Les cellules photovoltaïques

Une cellule photovoltaïque est assimilable à une diode photosensible, Son principe de fonctionnement repose sur l'effet photovoltaïque. En effet, une cellule est constituée de deux couches minces d'un semi-conducteur. Ces deux couches sont dopées différemment :

- Pour la couche N, apport d'électrons périphériques
- Pour la couche P, déficit d'électrons.

Lorsque la première est mise en contact avec la seconde, les électrons en excès dans le matériau N diffusent dans le matériau P. La zone initialement dopée N devient chargée positivement, et la zone initialement dopée P chargée négativement. Il se crée donc entre elles un champ électrique qui tend à repousser les électrons dans la zone N et les trous vers la zone P. Une jonction (dite P-N) a été formée [2]. En ajoutant des contacts métalliques sur les zones N et P, une diode est obtenue. Lorsque la jonction est éclairée, les photons d'énergie égale ou supérieure à la largeur de la bande interdite communiquent leur énergie aux atomes, chacun fait passer un électron de la bande de valence dans la bande de conduction et laisse aussi un trou capable de se mouvoir, engendrant ainsi une paire électron-trou. Si une charge est placée aux bornes de la cellule, les électrons de la zone N rejoindront les trous de la zone P via la connexion extérieure, donnant naissance à une différence de potentiel donc un courant électrique circule [3], voir la Figure I.1.

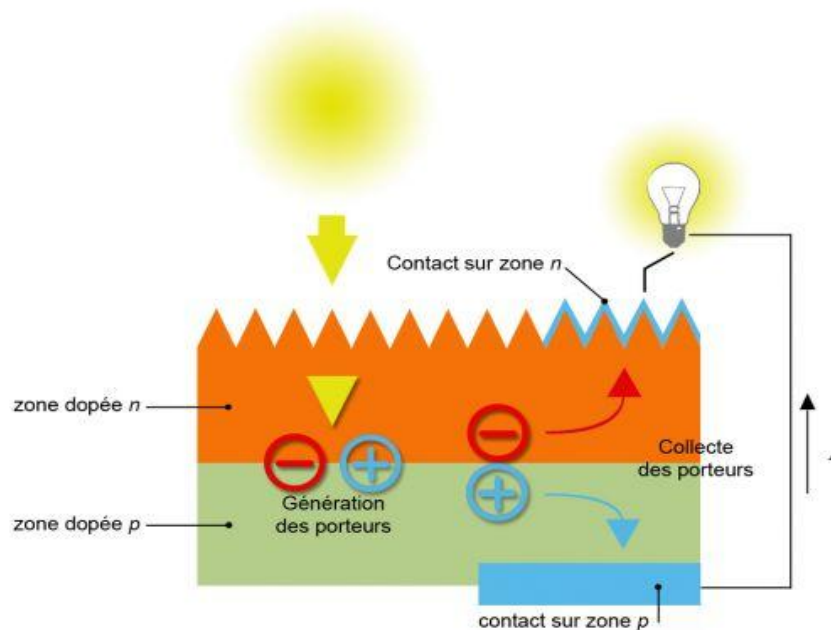


Figure I.1. Schéma de fonctionnement d'une cellule photovoltaïque.

Il existe trois grands types de cellules: mono cristallin, poly cristallin et amorphe.

I.2.1.1 Cellule au silicium monocristallin

Pour ce genre d'applications technologiques, le silicium pur est obtenu à partir de la silice de quartz ou de sable par transformation chimique métallurgique. Le silicium a un rendement électrique et une durée de vie de l'ordre de deux fois celle du silicium amorphe, mais il est nettement plus cher [2].

I.2.1.2 Cellule au silicium poly-cristallin

Le silicium poly-cristallin est un matériau composé de cristaux juxtaposés obtenus par moulage. Ce matériau est moins coûteux (que le monocristallin). Les cellules carrées ou rectangulaires sont faciles à utiliser [2].

I.2.1.3 Cellule au silicium amorphe

Le silicium absorbe le rayonnement solaire jusqu'à 100 fois mieux qu'en état cristallin ; les cellules sont constituées par des couches très minces [4], mais Leur rendement est plus bas que celui des cellules cristallines.

La Figure I.2 montre les différents types de cellules.



Silicium monocristallin

Silicium poly-cristallin

Silicium amorphe

Figure I.2. Les différents types de cellules.

Notons également que de nouvelles technologies viennent concurrencer le silicium : il s'agit des couches minces. Parmi elles, les cellules CIS, CIGS, H...

La tension générée par une cellule photovoltaïque est limitée à la valeur du gap du matériau dont elle est issue. Elle est de l'ordre de 0.6 V pour les cellules de type cristallin et amorphe. Pour le courant, il est en fonction de la surface de la cellule et pour une même surface, il dépend du rendement de la cellule.

I.3 Modèle équivalent d'une cellule PV

Le modèle mathématique associé à une cellule se trouve à partir de celui d'une jonction PN. On y ajoute le courant I_{ph} , ainsi qu'un terme modélisant les phénomènes internes.

Le schéma équivalent est montré dans la Figure I.3.

$$I = I_{ph} - I_d - I_{Rsh} \quad (I.1)$$

$$I_{ph} = I_{ph0} \frac{E}{E_0} \quad (I.2)$$

$$I_d \equiv I_{0d} \left(e^{\frac{q \cdot (V + R_s \cdot I)}{K \cdot T}} - 1 \right) \quad (I.3)$$

$$I_{Rsh} = \frac{V + R_s \cdot I}{R_{sh}} \quad (I.4)$$

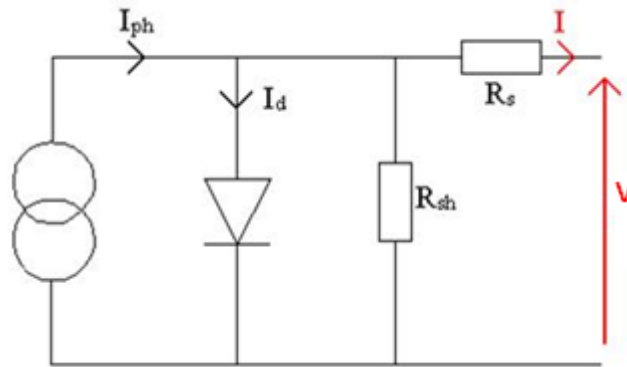


Figure I.3. Schéma équivalent d'une cellule photovoltaïque.

$$I = I_{ph} - I_{0d} \left(e^{\frac{q \cdot (V + R_s \cdot I)}{K \cdot T}} - 1 \right) - \frac{V + R_s \cdot I}{R_{sh}} \quad (I.5)$$

I.3.1 Résolution du modèle de la cellule

Le point de fonctionnement de la cellule peut être déterminé par la résolution de l'équation (I.5), ceci consiste à calculer la valeur d'une des inconnues (V ou I) en fonction de l'autre en supposant que les autres paramètres de cette équation sont connus. Il est préférable de calculer I à partir de V . La caractéristique I - V d'une cellule peut être obtenue en calculant I pour plusieurs valeurs de V sur une plage souhaitée.

L'équation (I.5) ne peut être résolue analytiquement cependant elle peut être mise sous la forme $f(I, V) = 0$. La résolution de l'équation de cette forme peut être effectuée en utilisant la méthode de Newton-Raphson classique [5].

I.3.2 Caractéristique électrique d'une cellule PV

Une cellule photovoltaïque peut être caractérisée par sa caractéristique statique courant/tension, souvent appelée caractéristique I-V. Une telle caractéristique est illustrée dans la Figure I.4

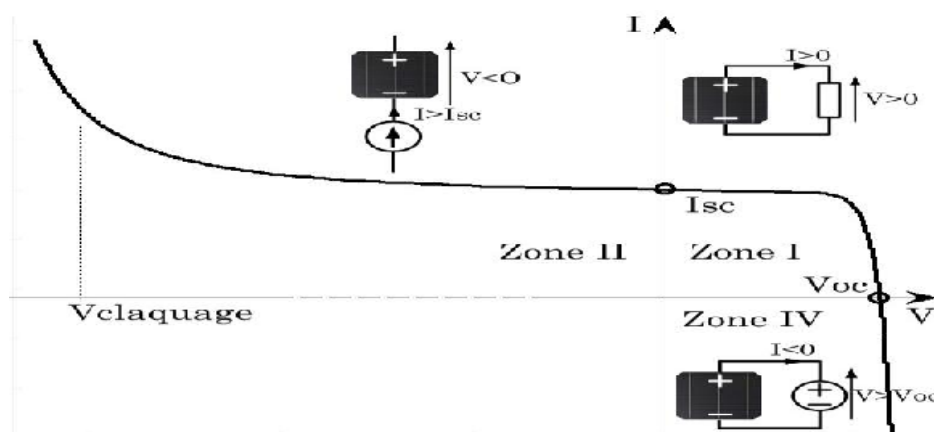


Figure I.4. Caractéristique I-V d'une cellule PV.

Dans cette Figure, on peut remarquer deux points de fonctionnement importants : le courant de court-circuit (I_{sc}) et la tension de circuit ouvert (V_{oc}). Comme son nom l'indique, le courant de court-circuit correspond au courant produit lorsque la cellule est court-circuitée. La tension de circuit ouvert est la tension aux bornes de la cellule quand il n'y a pas de courant.

La zone I de la Figure I.4 correspond au fonctionnement en générateur de la cellule avec $0 \leq V \leq V_{oc}$ et $0 \leq I \leq I_{sc}$. C'est le fonctionnement dit normal de la cellule.

La zone II correspond au fonctionnement en récepteur de la cellule avec $V < 0$ et $I > I_{sc}$. Quand le courant traversant la cellule est forcé par un circuit extérieur à dépasser sa valeur en court-circuit, la cellule produit une tension négative à ses bornes. La croissance en courant provoquera l'endommagement de la cellule si la tension à ses bornes atteint une limite : la tension de claquage [6].

La zone IV correspond de nouveau au fonctionnement en récepteur de la cellule, mais avec $V > V_{oc}$ et $I < 0$. Quand la tension aux bornes de la cellule est amenée à dépasser sa valeur en circuit ouvert, un courant inverse circule dans la cellule. Si ce courant inverse dépasse une limite, la cellule sera irrémédiablement endommagée.

I.3.3 Regroupement des cellules

I.3.3.1 Regroupement en série

Une association de N_s cellule en série Figure I.5 permet d'augmenter la tension du générateur photovoltaïque. Les cellules sont alors traversées par le même courant et la caractéristique résultant du groupement série est obtenues par addition des tensions élémentaires de chaque cellule. Les équations suivantes résument les caractéristique électrique d'une association série de N_s cellules [7].

$$V_{ocN_s} = N_s \times V_{oc} \quad (I.6)$$

$$I_{scN_s} = I_{sc} \quad (I.7)$$

V_{ocN_s} : La tension en circuit ouvert N_s cellules en série.

I_{scN_s} : Courant de court circuit de N_s cellules en série.

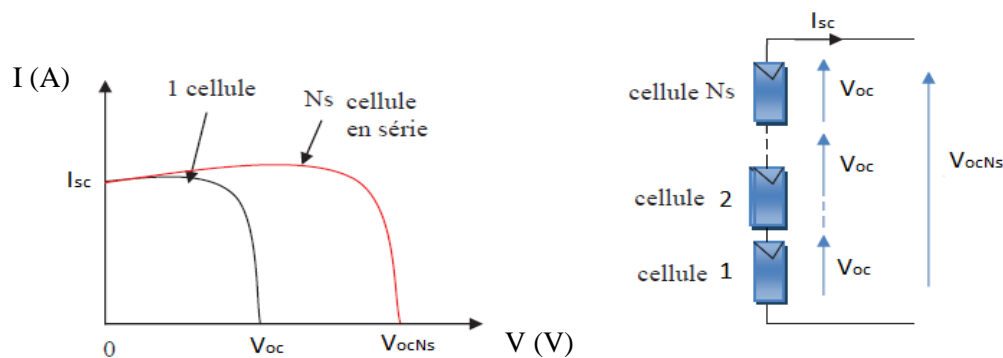


Figure I.5. Caractéristique courant tension de N_s cellule en série.

I.3.3.2 Regroupement en parallèle

Une association parallèle de N_p cellules Figure I.6 est possible et permet d'accroître le courant de sortie du générateur ainsi créé. Dans un groupement de cellules identiques connectées en parallèle, les cellules sont soumises à la même tension et la caractéristique résultante du groupement est obtenue par addition des courants [8].

$$I_{scN_p} = N_p \times I_{sc} \quad (I.8)$$

$$V_{ocN_p} = V_{oc} \quad (I.9)$$

I_{scN_p} : Le courant de court circuit de (N_p) cellule en parallèle.

V_{ocN_p} : La tension de circuit ouvert de (N_p) cellules en parallèle.

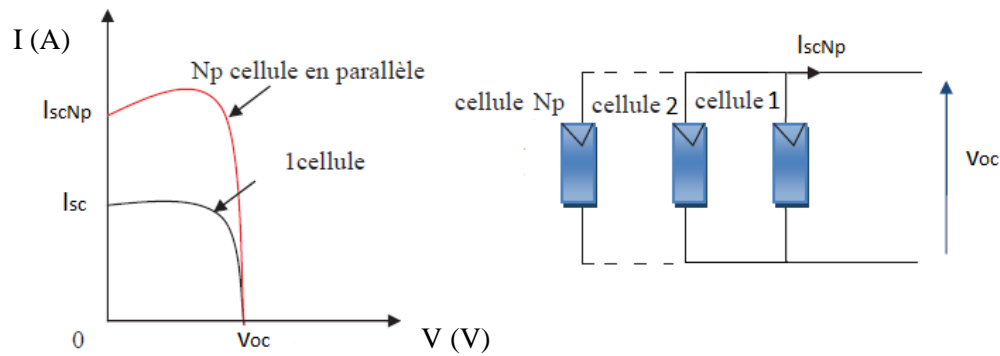


Figure I.6. Caractéristique courant tension de N_p cellule en parallèle.

I.3.3.3 Module photovoltaïque

La puissance générée par une cellule n'est pas suffisante aux applications courantes, donc il est nécessaire d'associer plusieurs cellules en série et en parallèle, ces cellules sont fragiles et sensibles à l'environnement extérieure, elles sont donc munies d'une protection mécanique (l'encapsulation). Pour toutes ces raisons, les cellules sont assemblées en modules photovoltaïques.

I.3.4 Générateur photovoltaïque

Le générateur photovoltaïque est l'unité de production d'énergie électrique sous forme de courant continu.

i) String PV

Un string PV, appelé également chaîne PV, est un ensemble de modules connectés en série afin de générer la tension de sortie spécifiée.

ii) Champ PV

Afin d'obtenir des puissances de quelques kW avec une tension convenable, il est nécessaire de regrouper les modules en série et en parallèle. Ce groupement forme un champ photovoltaïque. Il existe plusieurs configurations possibles pour interconnecter les modules dans un champ photovoltaïque : connexion *série parallèle simple*, connexion *Total Cross Tied*, connexion *Bridge Linked* [9].

I.4 Protections classiques d'un GPV

Dans cette partie nous nous intéressons qu'aux composants servant à la protection du générateur PV.

i) Diode de bypass

La diode de bypass est connectée en antiparallèle avec un groupe de cellules (voir la Figure I.7), Cette diode sert à protéger les cellules contre leur fonctionnement dans le régime inverse, plusieurs architectures de connexion des diodes de bypass dans le module ont été proposées [10], [11], [12].

ii) Diode anti-retour

La tension produite par chaque string peut être différente. Lors de la mise en parallèle de ces strings pour former un champ, le string avec la tension la plus faible peut absorber un courant inverse provenant des autres strings. Cela conduit donc à une baisse de production et les modules du string traversés par le courant inverse pourraient être également susceptibles de la défaillance. Pour éviter ces courants inverses, une diode anti-retour est placée au bout de chaque string.

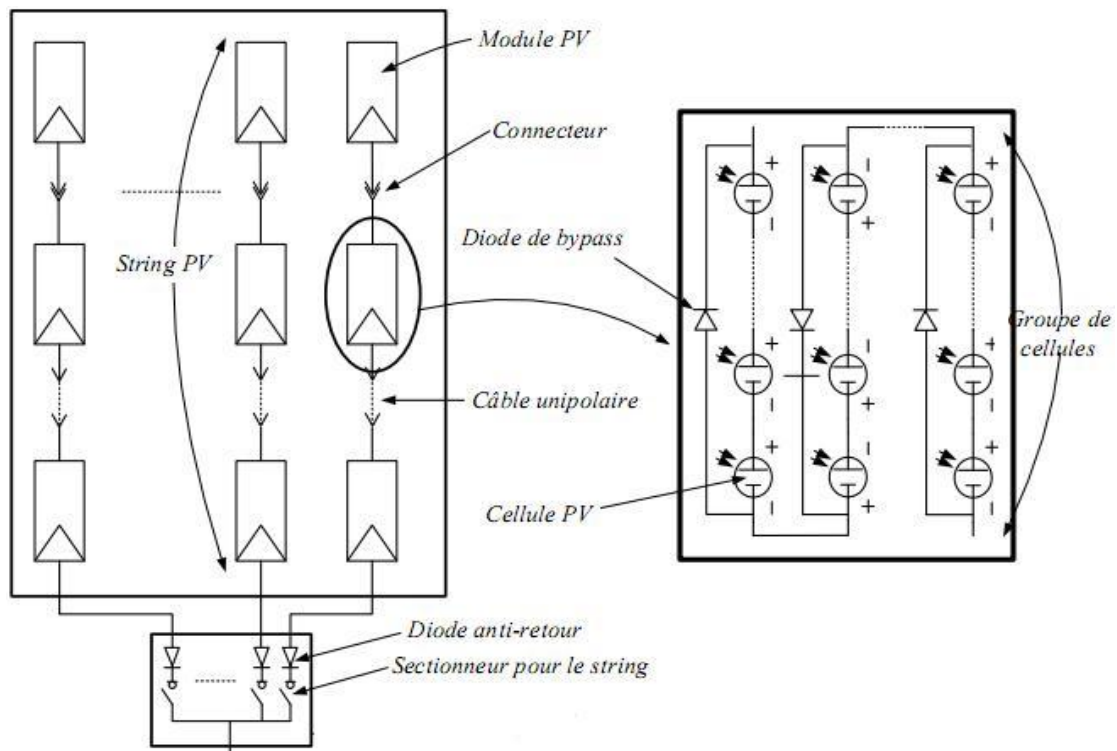


Figure I.7. Protection d'un GPV avec diode bypass et diode anti-retour.

I.5 Paramètres de performance d'un module PV

Les caractéristiques $I=f(V)$ et $P=f(V)$ d'un module PV à la forme représentée dans la Figure I.8.

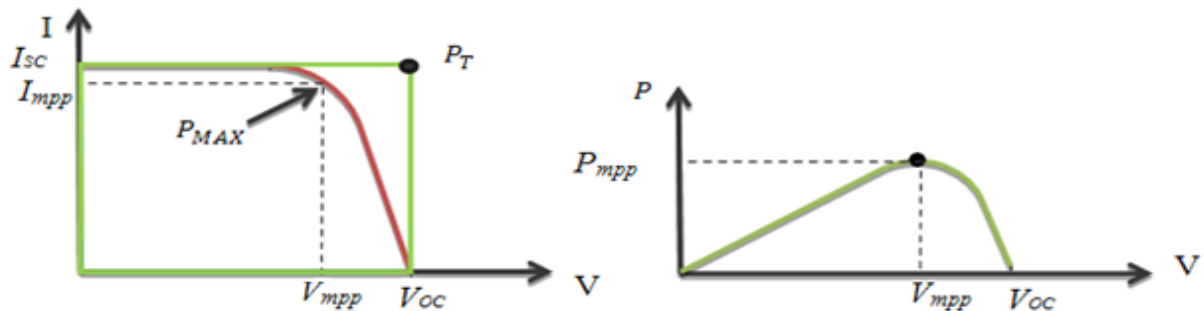


Figure I.8. Caractéristiques $I=f(V)$ et $P=f(V)$ d'un module PV.

De nombreux paramètres de performance pour un module PV peuvent être déterminés à partir de ces deux caractéristiques de la Figure I.8. Ces paramètres sont décrits ci-dessous [13-14].

I.5.1 Fonctionnement du MPPT

Entre les deux extrêmes (V_{oc} et I_{sh}), il existe un optimum donnant la plus grande puissance P_{MAX} ou MPP (Maximum Power Point). Pendant l'opération, le champ PV fonctionne normalement en point de la puissance maximale (MPP). De nombreux algorithmes de MPPT ont été proposés dans la littérature pour la recherche du point du maximum de puissance [15]

I.5.2 Influence des paramètres sur les caractéristiques I-V & P-V

Les paramètres externes considérés sont : l'ensoleillement, la température T et les paramètres interne a savoir la résistance série R_S et la résistance shunt R_{SH} . Ces paramètres vont permettre la simulation de la caractéristique I-V et P-V comme montré dans les Figure(I.9 , I.10 , I.11 , I.12), tandis que les résultats sont présentés pour un panneau PV ISOFOTON (106W,12V) de 36 cellules en série et 2 en parallèle.

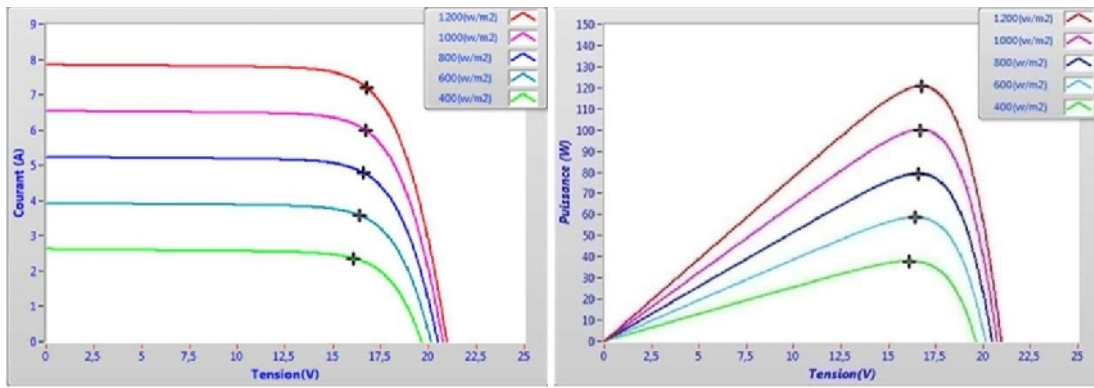


Figure I.9. Influence de l'ensoleillement E.

La Figure I.9 montre que le courant du module est proportionnel au rayonnement, tandis que la tension de circuit-ouvert change très légèrement

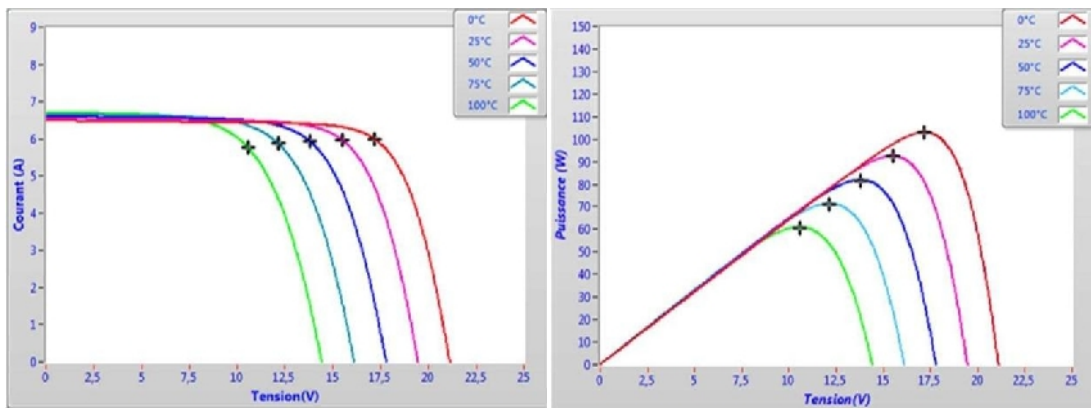


Figure I.10. Influence de la température de la cellule T.

La Figure I.10 montre que l'augmentation de la température a pour effet de diminuer la tension du module PV et donc dégradation de la puissance générée.

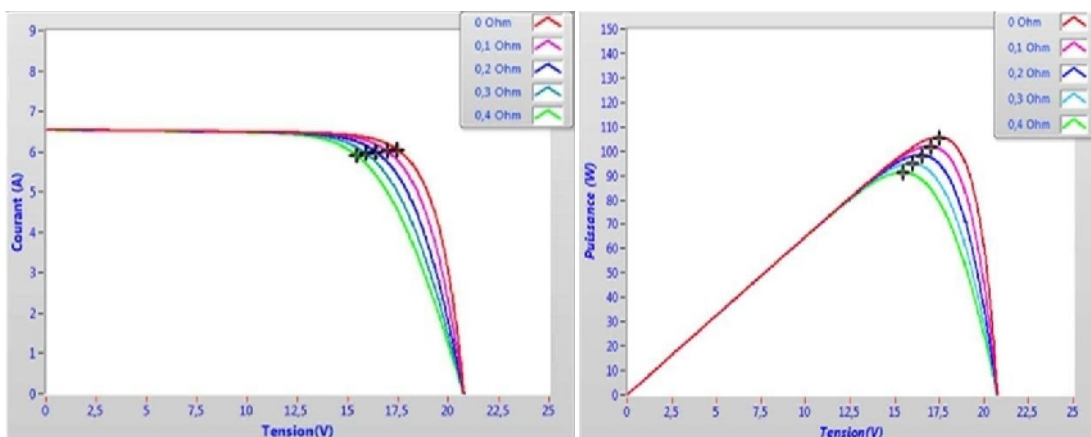


Figure I.11. Influence de la résistance série R_s .

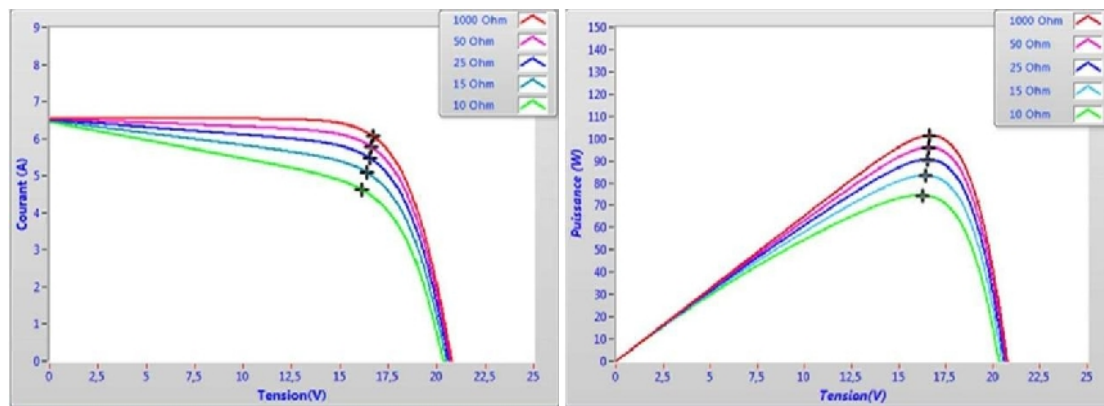


Figure I.12. Influence de la résistance shunt R_{sh} .

Pour une cellule idéale $R_S=0$ et $R_{SH}=\infty$. Ce qui traduit des pertes pratiquement nul. L'ensemble des pertes dans un module PV dépend de la résistance série. La résistance shunt représente une partie des pertes liée à une petite fuite de courant sur un chemin résistive parallèlement au diapositif intrinsèque.

I.6 Application des systèmes photovoltaïques

Les systèmes PV sont classés en fonction de trois types : autonomes, hybrides et reliés au réseau. Le type que vous choisissez dépendra de vos besoins, de votre emplacement et de votre budget.

I.6.1 Systèmes autonomes

Les systèmes autonomes sont complètement indépendants d'autres sources d'énergie. Ils servent habituellement à alimenter les maisons, les chalets ou les camps dans les régions éloignées ainsi qu'à des applications comme la surveillance à distance et le pompage de l'eau. Dans la majorité des cas, un système autonome exigera des batteries d'accumulateurs pour stocker l'énergie, comme illustré dans la figure ci-dessous.

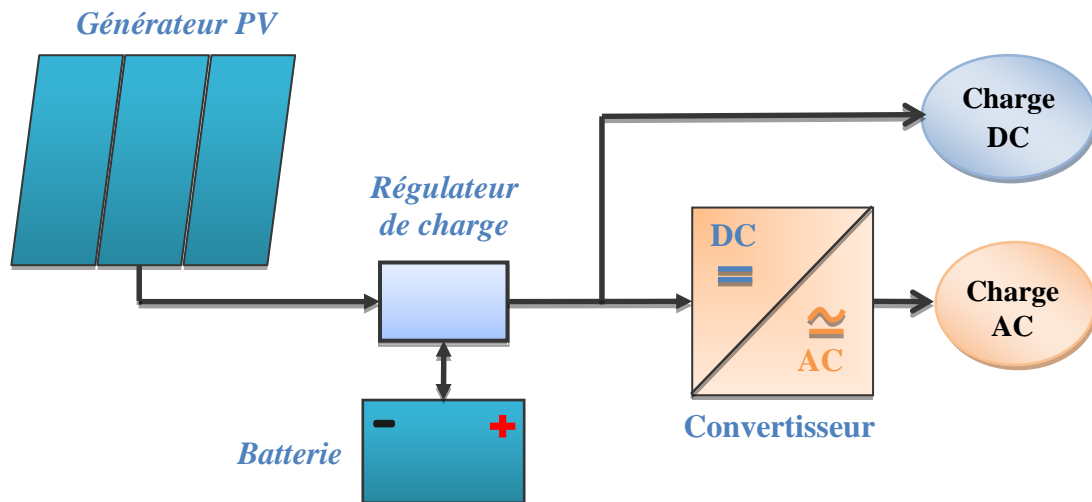


Figure I.13. Système PV autonome avec ou sans conversion DC/AC.

I.6.2 Systèmes hybrides

Les systèmes hybrides reçoivent une partie de leur énergie d'une ou de plusieurs sources supplémentaires. En pratique, les modules de systèmes PV sont souvent alliés à une éolienne ou à une génératrice à combustible. De tels systèmes ont habituellement des accumulateurs de stockage d'énergie

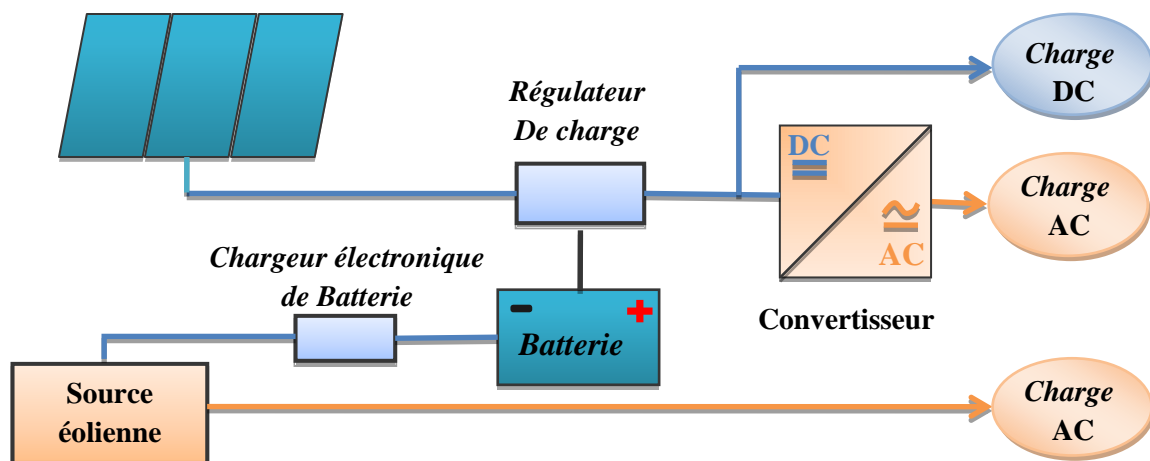


Figure I.14. Système hybrides PV éoliens.

I.6.3 Systèmes raccordés au réseau

Les installations raccordées au réseau électrique (ou reliées à une centrale de distribution) constituent généralement une solution optimale pour la production de l'électricité solaire, tant en termes d'énergie que de coûts. Ces installations se composent de modules solaires interconnectés, d'un (ou plusieurs) onduleur(s) raccordé au réseau électrique. L'onduleur

convertit le courant continu généré par les cellules solaires en un courant alternatif conforme au réseau électrique (220V, 50Hz). Il gère également des fonctions de commande et de réglage afin d'optimiser l'apport énergétique.

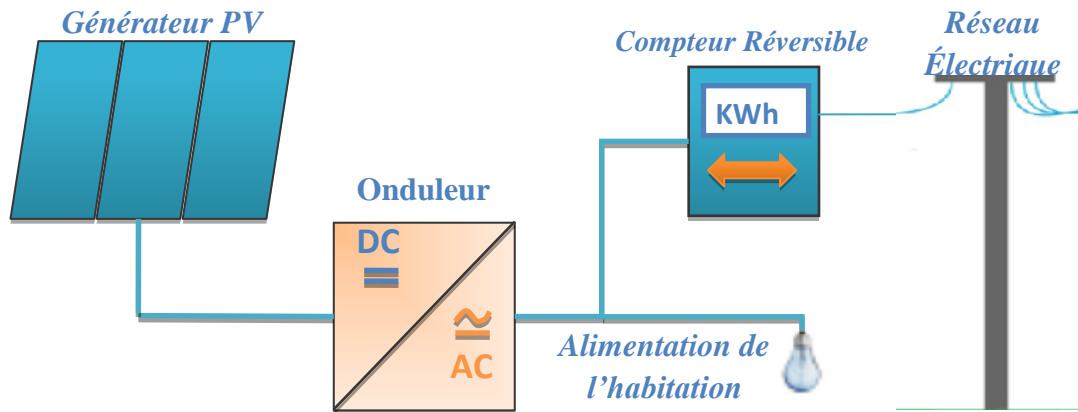


Figure I.15. Système PV raccordé au réseau.

I.7 Convertisseur

Le groupe convertisseur est composé d'un étage de hacheur suivi par un étage d'onduleur (Voir Figure I.16). Le hacheur a pour rôle d'extraire la puissance maximale du générateur PV. C'est pourquoi il est muni d'un algorithme de recherche de type MPPT (Maximum Power Point Tracker). La puissance maximale extraite est ensuite convertie en puissance alternative active par l'onduleur.

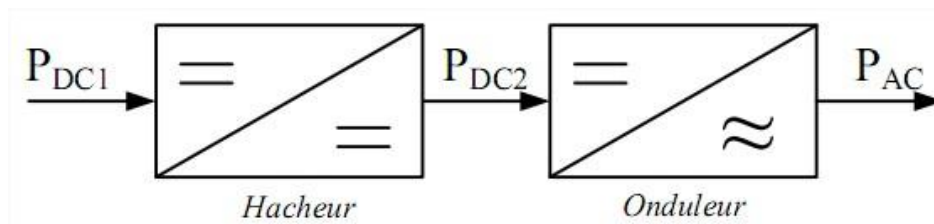


Figure I.16. Différents étages du groupe convertisseurs.

I.7.1 Architecture du convertisseur

Six architectures principales qui peuvent encore être classifiées en deux grandes topologies : topologie centrale et topologie modulaire [16].

i) Topologie centrale : Il existe trois architectures principales : onduleur central, onduleur string et onduleur multi-string (voir la Figure I.17)

Onduleur central : consiste à utiliser un seul hacheur et un seul onduleur interfaçant le

générateur PV avec le réseau électrique.

Onduleur string : cette topologie utilise un hacheur et un onduleur par string, elle permet également de mieux d'extraire la puissance maximale du générateur.

Onduleur multi-string : permet d'utiliser un seul onduleur, tout en conservant la possibilité d'utiliser un MPPT par string en utilisant un hacheur par string.

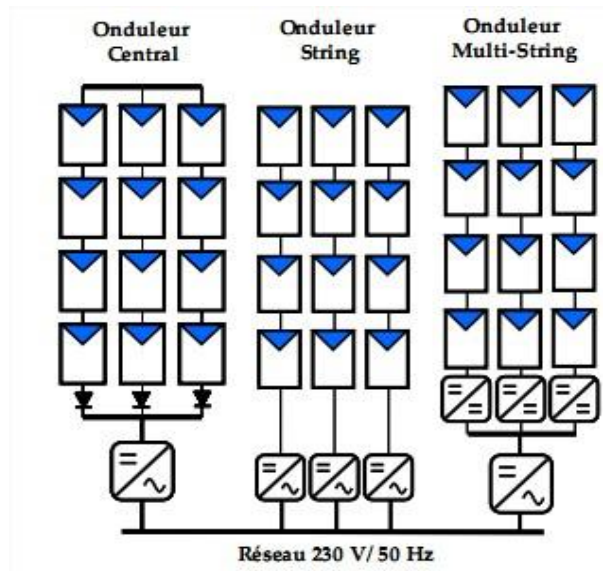


Figure I.17. Schéma des topologies centrales d'un système PV connecté au réseau.

ii) Topologie modulaire : Les topologies modulaires d'un système PV connecté au réseau sont illustrées dans la Figure I.18

Onduleur individuel : cette architecture permet directement au module de fournir de l'énergie au réseau .

Hacheur parallèle : elle utilise un hacheur par module connecté à un bus continu de tension plus élevée (400 V), auquel est relié un onduleur.

Hacheur série : elle utilise aussi un hacheur par module, mais ceux-ci sont connectés en série afin d'augmenter le rendement de conversion du hacheur.

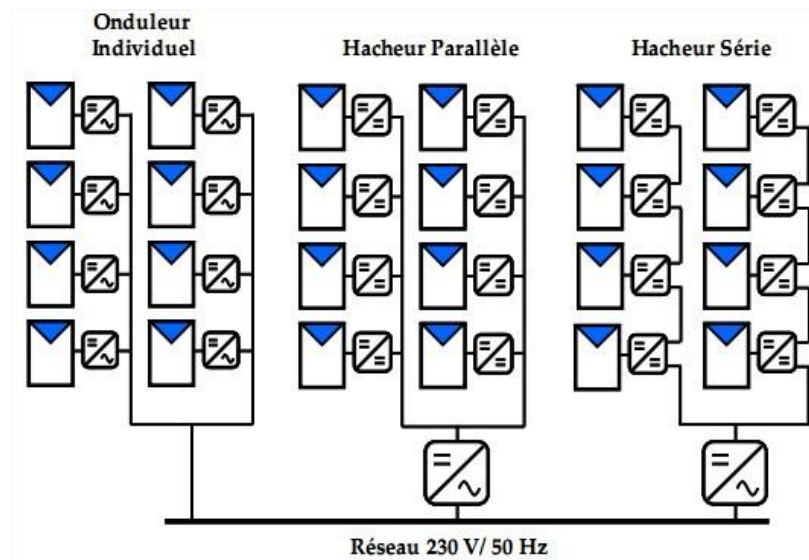


Figure I.18. Schéma des topologies modulaires d'un système PV connecté au réseau.

I.8 Conclusion

Dans ce chapitre nous avons présenté les principales caractéristiques et les technologies des éléments constitutifs d'un générateur PV. Par la suite on a expliqué brièvement comment augmenter le courant ou la tension d'un générateur photovoltaïque ainsi nous avons montré l'influence de la température et l'ensoleillement sur le rendement du module donc sur la puissance.

Chapitre II : Diagnostic des défauts : Etat de l'Art

II.1 Introduction

Les systèmes PV peuvent être généralement considérés comme fiables, mais ils ne peuvent pas être sans défauts et leur production peut être affectée de façon considérable par la présence des défauts. Un système de détection de défaut permet de réduire les pertes afin d'augmenter la production annuelle, il est défini comme un système qui est capable de distinguer entre les différentes conditions de fonctionnement de l'installation PV. Si la présence d'un défaut est détectée, le système doit être aussi en mesure d'indiquer les caractéristiques de ce défaut.

II.2 Les défauts les plus courants

Les propriétaires de systèmes PV ont tendance à penser que le système PV est un équipement sans entretien en raison de sa configuration simple [17]. En fait, l'absence d'une pièce mobile réduit considérablement la possibilité d'une défaillance. Cependant, des enquêtes et des inspections sur le terrain, comme l'étude des systèmes PV résidentiels au Japon [18] et le 1000 toits programmés en Allemagne [19], montre qu'un grand pourcentage de défauts dans les systèmes PV, nécessitent un entretien et remplacement de pièce, et chaque événement qui amène à une production de puissance plus faible que celui attendu est considéré comme un défaut. Il est possible de classer les défauts en cherchant différents aspects, en particulier : le composant qui présente le défaut, les causes, les effets et la réponse à ce défaut.

Afin de classer les défauts, il est intéressant de diviser un système PV en trois parties : le côté AC, le côté DC et entre eux l'onduleur.

Si une défaillance est enregistrée sur le côté AC, les sources possibles sont le déséquilibre du réseau, un problème dans la liaison avec l'onduleur ou la charge. La probabilité d'une défaillance due au réseau est fortement liée à l'état de ce dernier, aussi au point de raccordement.

Les défauts dus à des erreurs internes de l'onduleur sont les plus courantes et ils augmentent avec le vieillissement du composant, ces défauts peuvent être liés avec tous les composants internes de l'onduleur, comme les commutateurs, le ventilateur, le MPPT ou les varistances (en fait une cause fréquente de l'échec de l'onduleur est un condensateur électrolytique défectueux)

Bien que le côté DC est celui qui présente la plus faible quantité d'échecs, il est le plus compliqué à analyser et à détecter l'origine du défaut en raison du nombre de composants inclus et des caractéristiques de sortie non linéaires des panneaux PV, une variété de défauts peut être difficile à distinguer, pour cela une division doit être faite entre les défauts qui amènent à un arrêt complet de toute une string ou bien l'installation complète et celui qui déterminent seulement une diminution de la puissance produite.

Les facteurs qui peuvent amener à une diminution de la puissance produite sont principalement la dégradation ou la rupture d'une cellule ou un module et / ou les phénomènes d'ombrage. Les phénomènes d'ombrage (*shading*) peuvent être liés à la présence d'obstacles qui ne permettent pas aux rayonnements solaire d'atteindre le panneau. Un autre facteur qui ramène à une diminution de la puissance qui peut être une surface sale des panneaux.

Il est possible d'identifier les défauts suivants comme les plus courants :

- **Erreur interne de l'onduleur**

L'onduleur est l'élément central du système PV et représente également le plus faible élément à cause d'erreur interne qui peut avoir. Tandis que l'onduleur a toujours été l'élément qui présente le plus grand nombre de défauts, il est aussi l'élément qui a été beaucoup amélioré au cours de ces dernières années. En regardant les données statistiques, il est évident que la probabilité de défauts liés à l'onduleur subit une baisse significative.

- **Les problèmes du réseau**

La probabilité de ces défauts est fortement dépendante à l'emplacement et par conséquent la fiabilité du réseau considéré.

- **Détérioration des modules**

En raison du vieillissement du module PV un certain changement dans la puissance produite par les modules est prévu au cours des années, bien qu'il soit important de vérifier si la baisse n'est pas supérieure à celle indiquée sur la spécification du module. Le rayonnement ultraviolet du soleil provoque un vieillissement et une légère dégradation. Cependant, le test des modules sur 25 ans a montré que même les modules avec un signe

visible de vieillissement génèrent en moyenne 75% de leur puissance originale .Un groupe de recherche TÜV Rheinland a constaté que la dégradation annuelle moyenne est inférieure à 0,5%. En général, la dégradation à long terme pour les modules silicium cristallin est faible, environ 0,2% par ans [20].

- **ombrage périodique**

En raison de la présence d'obstacles générant un ombrage périodique qu'on ne peut considérer comme un propre défaut, il est important de considérer que les modules touchés par un ombrage partiel montrent une détérioration plus élevée en raison de la présence périodique de hot spot dans des conditions normale de fonctionnement [21].



Figure II.1. Ombrage des Modules.

Le reste de défauts et anomalies rencontrés dans un générateur PV sont représentés .dans le tableau ci dessous :

| Eléments du générateur PV | Origines des défauts et d'anomalies |
|---------------------------|---|
| Générateur PV | <ul style="list-style-type: none"> -Feuilles d'arbres, déjections, pollution, sable, neige etc. -Détérioration des cellules, fissure, échauffement des cellules; -Pénétration de l'humidité, dégradation des interconnexions, corrosion entre les cellules; -Modules de performances différentes; -Module arraché ou cassé; -Modules court-circuités, modules inversés. |

| | |
|--|---|
| Boîte de jonction | <ul style="list-style-type: none"> -Rupture de circuit électrique -Court-circuit du circuit électrique -Destruction de la liaison -Corrosion des connections |
| Câblage et connecteur | <ul style="list-style-type: none"> -Circuit ouvert -Court-circuit -Mauvais câblage (module inversé) -Corrosion des contacts -Rupture du circuit électrique |
| Diode de protection (diode de bypass et diode anti-retour) | <ul style="list-style-type: none"> -Destruction des diodes -Absence ou non fonctionnement de diodes -Inversion de la polarité des diodes au montage, diode mal connectée |

Tableau II.1. Principaux défauts et anomalies rencontrés dans un générateur PV.

II.3 Méthodes de diagnostic d'une installation PV

Le Contrôle de performances régulières sur le fonctionnement des systèmes PV connectés au réseau sont nécessaires pour assurer une production d'énergie fiable et optimale. L'objectif est la détection précoce de défaut dans les systèmes PV et changer les conditions de fonctionnement pour empêcher les pertes d'énergie et financières subséquentes pour l'opérateur. Ainsi, un système simple et sûr à manipuler à long terme pour vérifier les performances et l'état du système PV est nécessaire pour garantir la sécurité de ce dernier et réduire les pertes. Le but des systèmes de diagnostics est non seulement la détection des dysfonctionnements, mais aussi la mise à disposition des informations sur la source de défaillance la plus probable.

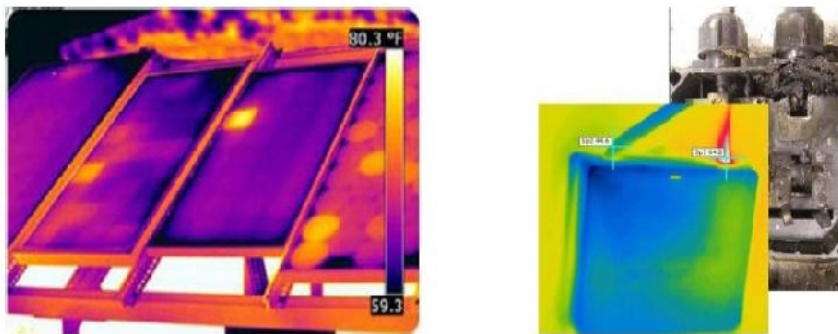
II.3.1 Méthodes de diagnostic courantes (industrialisé)

On peut distinguer deux catégories de méthodes de diagnostic courantes : des méthodes reposant sur l'analyse du courant et de la tension (que nous appellerons méthodes électriques) et des méthodes reposant sur l'analyse d'autres grandeurs que I et V (que nous appellerons méthodes non-électriques).

II.3.1.1 Méthodes non-électriques

Il existe plusieurs méthodes non-électriques, destructives ou non destructives, pour diagnostiquer le défaut au niveau de cellule PV. Le défaut principal qui peut avoir lieu à ce niveau est la fissure de la cellule. On peut citer comme méthodes: les essais mécaniques de flexion, l'imagerie par photoluminescence et électroluminescence, tests de thermographie [22], [23].

Au niveau du module PV, la méthode de l'imagerie à infrarouge (caméra thermique) est largement appliquée. Cette méthode repose sur le fait que tous les matériaux émettent un rayonnement infrarouge sur une plage de longueur d'onde qui dépend de la température du matériau. En examinant la distribution de la température au niveau du module, des anomalies (si elles ont lieu) peuvent être localisées (Voir la Figure II.2).



(a) Cellule court-circuitée. (b) Corrosion dans la boîte de jonction.

Figure II.2. Quelques exemples de la localisation de défauts par la caméra thermique.

II.3.1.2 Méthodes électriques

Les systèmes de monitoring actuels d'un GPV sont essentiellement intégrés aux onduleurs. Dans ce cadre, les données mesurées sont souvent les mêmes d'un système à l'autre.

Les grandeurs mesurées les plus courantes sont :

- Le courant et la tension coté DC du champ PV.
- Le courant et la tension coté AC du champ PV.

Il est aussi possible d'ajouter les grandeurs complémentaires que sont la température ambiante du site et l'ensoleillement aux mesures électriques. Ces données nécessitent un capteur spécifique. Les informations peuvent être traitées localement ou à distance, sur une ou plusieurs installations.

II.3.2 Méthodes de diagnostic proposées dans la littérature

De nombreuses méthodes de diagnostic ont été proposées pour détecter les défauts dans un système PV. Nous résumons dans cette partie les différentes méthodes proposées.

II.3.2.1 Méthode de réflectométrie

La méthode de réflectométrie est une méthode de diagnostic qui consiste à envoyer un signal dans le système ou le milieu à diagnostiquer. Ce signal se propage selon la loi de propagation du milieu étudié et lorsqu'il rencontre une discontinuité, une partie de son énergie est renvoyée vers le point d'injection. L'analyse du signal réfléchi permet de déduire des informations sur le système ou le milieu considéré. Cette méthode a été également appliquée pour détecter le défaut dans un string photovoltaïque [24], [25].

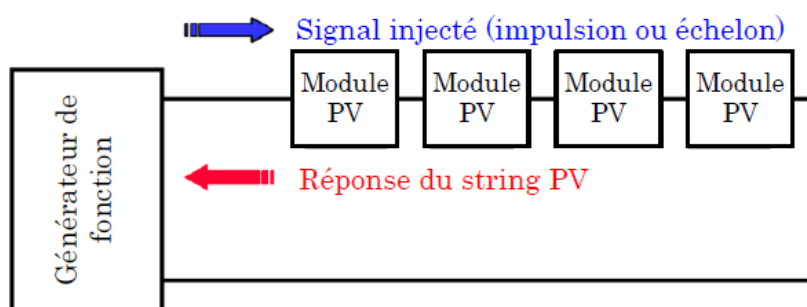


Figure II.3. Principe de la réflectométrie pour localiser le défaut dans un string PV.

L'étude expérimentale a confirmé la possibilité de localiser la position de défaut de type « circuit ouvert », « court-circuit » et « augmentation d'impédance » dans le string.

II.3.2.2 Analyse de la puissance et de l'énergie produite

Les systèmes de monitoring montrent graphiquement et compare les valeurs de la puissance réelle et simulée pour retourner des paramètres de performance et des pertes du système PV. La puissance simulée est calculée grâce à un modèle de simulation qui estime la puissance attendue comme étant la puissance que devrait produire l'installation PV dans certaines conditions météorologique. Une fois que la puissance estimée est calculée, elle doit être comparée avec la puissance réelle (mesuré). La différence entre ces deux valeurs représente les pertes dans le système.

➤ **Détection de défaut**

L'étape suivante consiste à distinguer, si la valeur de cette différence (entre la puissance réelle et simulée) représente un dysfonctionnement donc la présence d'un défaut. Lorsque la différence dépasse un seuil, qui peut être différent selon les différents modèles de simulation choisis et à leur précision, on suppose une présence de défauts.

➤ **Diagnostic de défaut**

La dernière étape est l'élément le plus avancé dans le processus est la capacité de distinguer entre les différents défauts possibles. Il existe principalement deux approches différentes qui ont été examinées jusqu'à présent : une méthode analytique et l'autre empirique. La première est basée sur les valeurs des paramètres qui décrivent le comportement de la puissance d'une installation PV telle que : la durée, l'amplitude, la fréquence et les instants de la chute. Ces mêmes attributs sont également prédéterminés pour les différents défauts considérés. Lors de leur comparaison, le défaut dont la valeur des attributs considérés est la plus proche de celle déduite des grandeurs mesurées est considéré comme le défaut responsable de la chute. La seconde est appelée empirique car elle est basée uniquement sur les résultats des tests sans regarder l'influence physique du défaut spécifique sur le système, ici l'identification du type de défaut se fait grâce à des analyses statistiques des caractéristiques électriques en présence de divers défauts.

Plusieurs études reposent sur l'analyse de la puissance et de l'énergie produite par le champ PV [26], [27]. Outre la comparaison de la puissance ou de l'énergie produite actuelle et celle attendue, la comparaison du point de la puissance maximale actuel (courant et tension correspondant à la puissance maximale) et celui attendu peut apporter plus d'information sur l'état du système PV [28].

II.3.2.3 Autres méthodes pour le diagnostic

Une étude trouvée dans [29], où une méthode basée sur la fonction de corrélation étendue et matter-element model est proposé pour identifier les défauts dans une petite installation PV. En outre, dans [30] matter-element model est combiné avec un réseau de neurone pour construire un système de diagnostic de défaut intelligent. Les deux propositions utilisent un simulateur de

système PV pour collecter les données de la puissance produite par les modules PV en fonctionnement normal et défectueux.

Dans [31] une méthodologie proposée qui permet de détecter les défauts dans un système PV, en comparant l'énergie de sortie de six sous-systèmes qui forment l'installation PV. La méthode discutée dans [28], où ils ont proposé un monitoring automatique et un système de détection de défaut basé sur l'analyse des pertes de puissance, ces dernières sont calculées en comparant les données de monitoring avec les résultats de simulation, ensuite les indicateurs de pertes sont traités pour générer un signal défectueux [32].

Finalement le projet de recherche universitaire le plus important dans le domaine de détection automatique des pannes qui a été appliqué dans un produit commercial est le projet PVSAT2. Ce projet a été mené pendant trois ans, il est le fruit d'une collaboration entre les différentes universités et entreprises. L'analyse est basée sur des données satellitaires parce que le projet vise à surveiller les conditions de fonctionnement au niveau du des surfaces ou on trouve les panneaux PV comme illustré dans la Figure II.4. Dans ce cas, les données provenant des capteurs locaux ne sont pas souvent disponibles

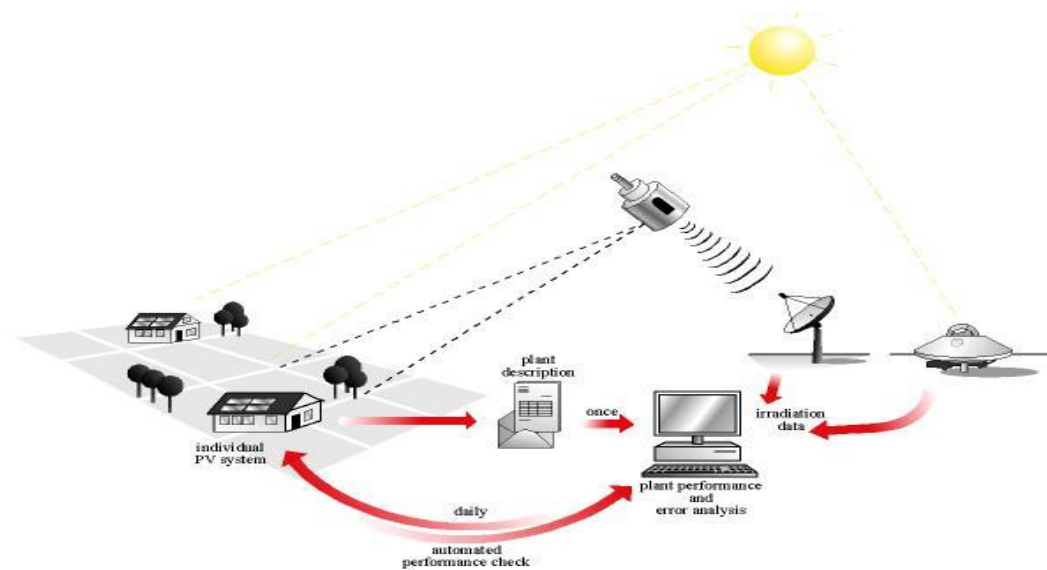


Figure II.4. Un aperçu sur PVSAT2.

Tous d'abord, les algorithmes et les procédures pour estimer les conditions météorologiques sur place (sur le site) à partir de données satellite ont été développés et améliorés. Ensuite, on associe un modèle électrique à l'installation PV réelle. Enfin, pour chaque défaut une série de paramètres est définie et la caractérisation du défaut est créée. A l'instant où un décalage entre la puissance estimée et réelle apparaît, une erreur est détectée. Ainsi, pour la détection des pannes de système PV, la méthode de « *profiling* » de défaut est l'une des approches appliquées. Elle

étudie les propriétés de la perte d'énergie comme la quantité, la durée, le changement, et la corrélation avec les systèmes PV voisins. Les résultats de cette analyse sont utilisés pour créer un profil de défaut en énumérant les attributs de la défaillance qui s'est produit [33].

La deuxième approche dans la détection de défaut est la méthode *footprint* (FDR), la méthode « *FDR* » vise également à l'identification des pertes mineures d'énergie dues à l'ombrage, les pertes de puissance permanentes et les problèmes d'onduleur. Elle analyse les modèles de la dépendance pour trois domaines différents : puissance normalisée surveillé, l'heure de fonctionnement, et la distance du soleil, par exemple, l'influence de l'ombrage sur la production d'électricité est plus élevée à basse altitude [33].

II.4 Conclusion

Dans ce chapitre on a présenté les différents défauts qui peuvent subir un système PV, ainsi que les méthodes de diagnostic de défauts proposées industrialisées ou non. Ces méthodes montre une limitation de discrimination entre les défauts, pour cela les méthodes basés sur l'exploitation des données est identifiées comme la plus prometteuse [34].

Chapitre III : Classification par l'Arbre de Décision

III.1 Introduction

Le problème d'apprentissage automatique est fondamental en intelligence artificielle. Il apparaît par exemple, dans des problèmes de classification, de prédiction,...etc. Le diagnostic peut être considéré comme une tâche de classification supervisée (c'est à dire les classes sont connues), Il existe plusieurs approches pour résoudre le problème de classification : Arbre de Décision, Réseaux de Neurones, Algorithmes Génétique, les Séparateurs à Vaste Marge, les Réseaux Bayésiens. Dans ce qui suit, nous détaillerons les fondements théoriques d'une des approches les plus efficaces et les plus utilisées en IA qui est: Arbres de Décision [35] [36] [37] [38].

III.2 Classification supervisée pour la détection et le diagnostic

Beaucoup de données sont récupérées lorsque le procédé est en fonctionnement normal, mais également lorsque le procédé subi une défaillance (ou faute). On peut catégoriser les données récoltées suivant les différentes fautes associées aux dysfonctionnements. Lorsque nous représentons graphiquement les données des différentes fautes, on peut alors chercher à dresser au mieux des frontières entre les différentes classes afin de définir les régions de chaque faute comme illustré par la Figure III.1

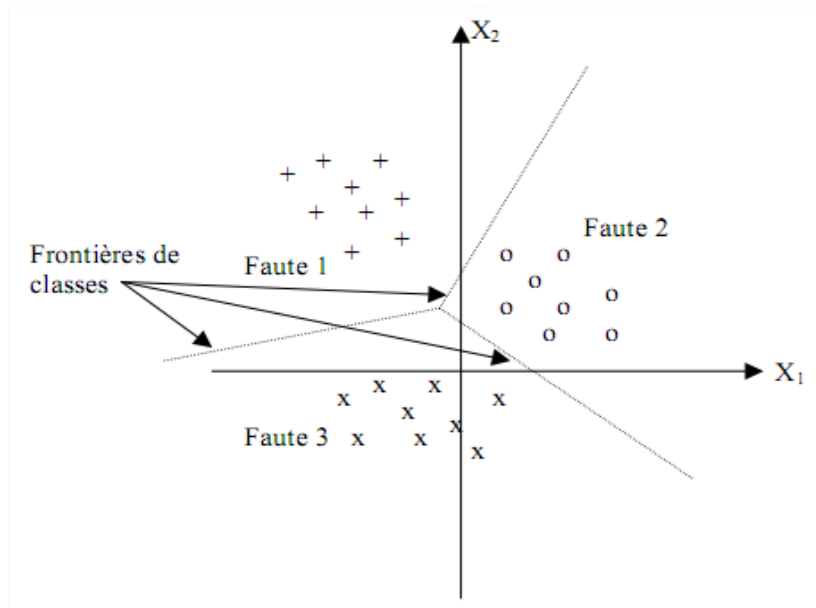


Figure III.1. Exemple des frontières des classes.

Lors de l'apparition d'une nouvelle faute (supposée détectée) dont la représentation graphique est possible, on voit tout de suite à quelle région de faute elle appartient et ainsi cette nouvelle observation est diagnostiquée. L'attribution d'une classe à une nouvelle observation est l'un des buts de la classification [35].

III.2.1 Généralisation d'un classificateur

L'intérêt d'un classificateur est qu'il donne un taux de « *Mauvaise Classification* » le plus petit possible. C'est en ce sens qu'intervient la notion de généralisation. La généralisation est la capacité à généraliser ses régions de classification afin d'obtenir un classificateur qui pourra classer correctement des individus futurs. Prenons l'exemple de la Figure III.2, on peut voir les données en deux dimensions de deux différentes classes (classe "carré" et classe "rond").

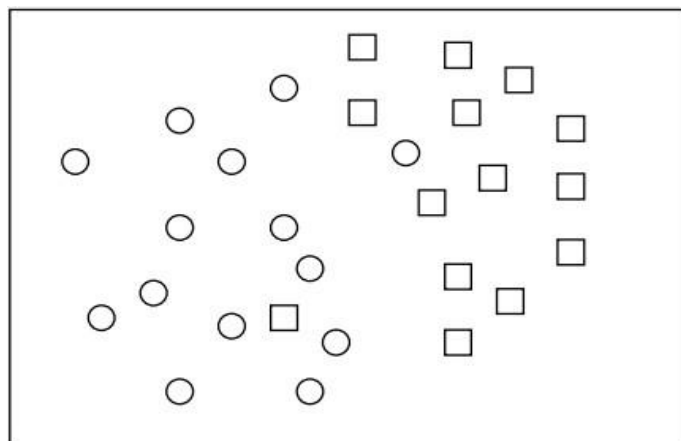


Figure III.2. Représentation de deux classes en deux dimensions.

Il est toujours possible de trouver un classificateur fournissant une frontière de décision très complexe. Grâce à cette frontière, nous pouvons classer parfaitement chaque individu de l'ensemble d'apprentissage comme montré dans la Figure III.3 ci-dessous.

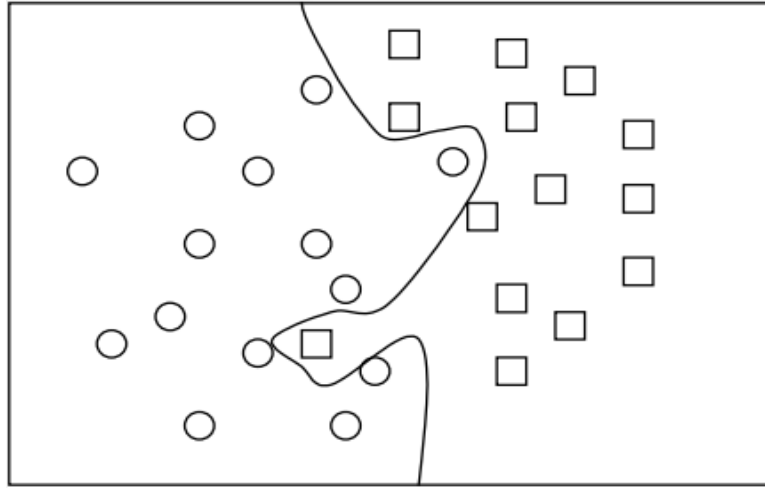


Figure III.3. Frontière de décision parfaite sur les données d'apprentissage.

Maintenant, si nous soumettons à notre classificateur des individus qu'il n'a jamais vu, comment va-t-il réagir ? Sur la figure III.4, on peut voir deux nouveaux individus représentés en pointillé. Le classificateur classe alors le rond comme étant un carré, et le carré comme étant un rond.

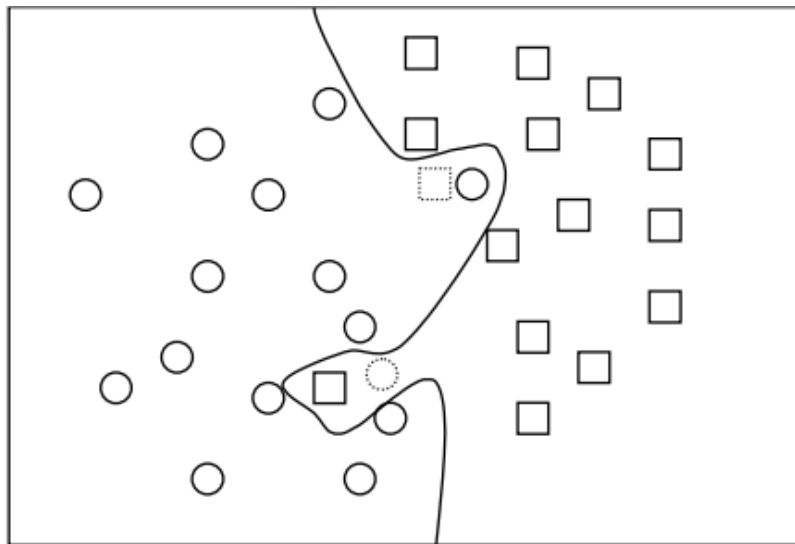


Figure III.4. Nouveaux individus mal classés.

Dans ce cas de figure, le classificateur a perdu sa capacité de généralisation. On dit, dans ce cas, qu'il ya eu un sur-apprentissage : il a trop appris de l'ensemble d'apprentissage et de ce fait ne saura reconnaître que celui-ci sans pouvoir le généraliser. Contrairement au cas donné par la

Figure III.4, celui de la Figure III.5 présente un classificateur qui a généralisé son ensemble d'apprentissage. Bien entendu, il conduit lui aussi à quelques erreurs, mais à un taux de «*Mauvaise Classifications* » plus faible que le même classificateur ayant sur-appis [39].

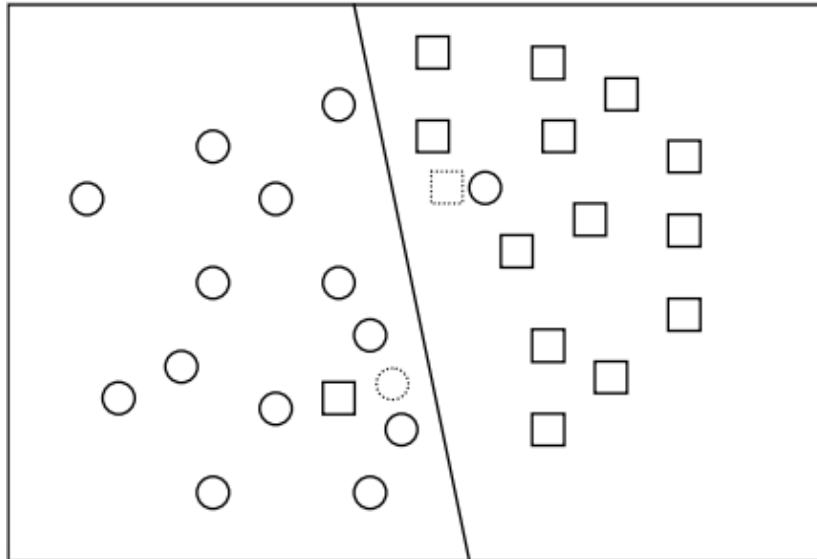


Figure III.5. Nouveaux individus correctement classés.

III.3 Arbre de Décision

III.3.1 Notion général

L'apprentissage inductif consiste à extraire un modèle à partir d'un ensemble d'exemples appelé ensemble d'apprentissage [40]. Chaque exemple de cet ensemble est représenté par un vecteur d'attributs et chaque attribut prend un ensemble de valeurs (Voir la figure III.6).

Si la classe de chaque vecteur est donnée, on se situe dans le cadre de l'apprentissage supervisé. On cherche à utiliser les exemples fournis déjà classés pour apprendre un modèle qui permet ensuite de déterminer la classe de tout nouvel exemple rencontré.

Si aucune classe n'est disponible, on se situe dans le cadre de l'apprentissage non-supervisé qui cherche à regrouper les exemples (clustering) en mesurant la similarité entre eux [37].

L'apprentissage par arbre de décision se situe dans le cadre de l'apprentissage supervisé, où la classe de chaque exemple dans la base est donnée. Le but est de construire un modèle à partir d'un ensemble d'exemples associés aux classes. Une fois ce modèle construit, on peut extraire un ensemble de règles de classement. Ce modèle ou les règles extraites sont ensuite utilisés pour classer de nouveaux exemples dont la classe est inconnue (Voir la Figure III.6).

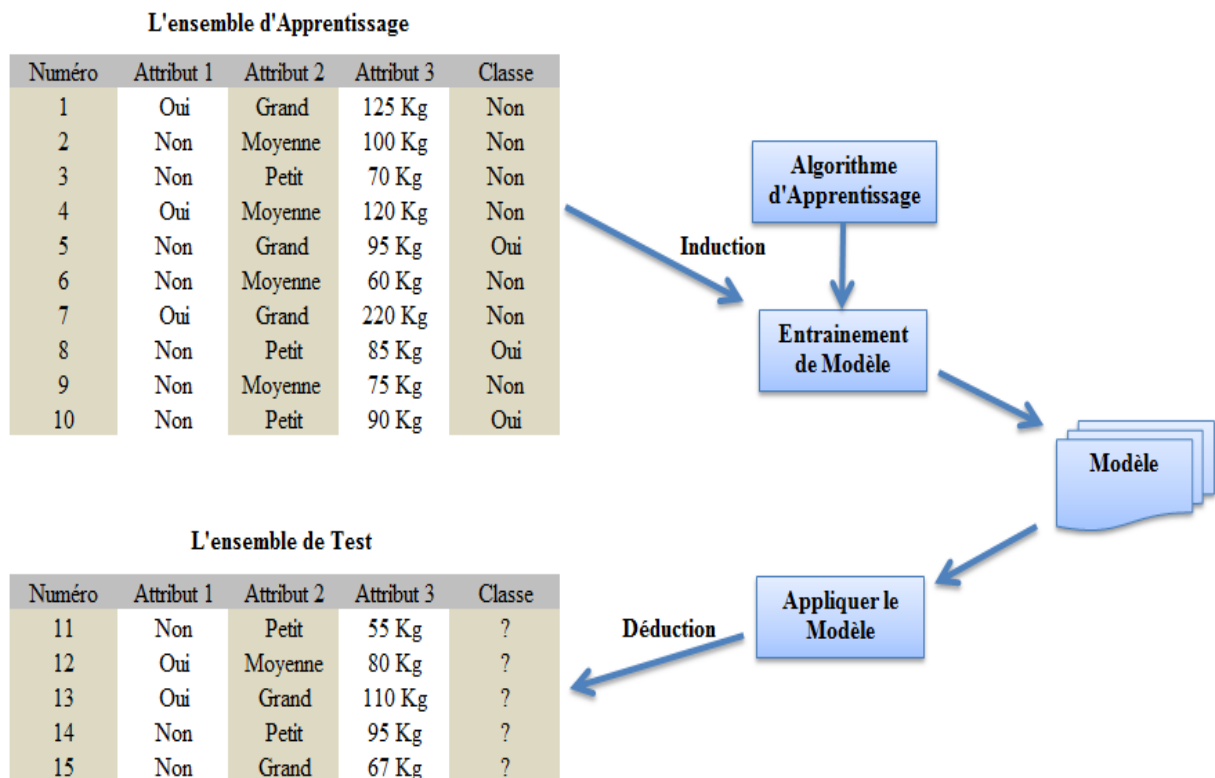


Figure III.6 :L'induction par l'arbre de décision.

III.3.2 Types de données

Nous pouvons diviser le type d'un attribut en deux grandes catégories à savoir [36]:

- **Quantitative** (Numérique) : Si l'ensemble des valeurs qu'il peut prendre est un ensemble de nombres, fini ou infini, ou un intervalle des valeurs réelles. Un attribut « X » numérique peut être discret ou continu selon sa nature :
 - **Continu** : Si l'ensemble des valeurs qu'il peut prendre est un intervalle réel. Il s'agit donc d'un ensemble infini non dénombrable: on ne peut pas énumérer systématiquement l'ensemble de tous les points d'un intervalle. Par exemple, « X » peut être l'âge d'une personne prise au hasard, sa taille, son poids, ...etc.
 - **Discret** : Si l'ensemble des valeurs qu'il peut prendre est un ensemble numérique fini (comprenant un nombre fini d'éléments) ou un ensemble infini dénombrable (comprenant une infinité de nombres que l'on peut énumérer).
- **Qualitative** (nominal) : Si l'ensemble des valeurs qu'il peut prendre est non numérique. L'attribut « X » peut être par exemple la couleur des yeux d'une personne prise au hasard, sa région de naissance, son sexe, ...etc.

D'autre part, une donnée numérique ou nominale peut être ordinale si ses valeurs sont ordonnées. Par exemple, l'attribut dont les valeurs sont {bien, très-bien, excellente} est un attribut ordinal nominal ; l'attribut dont les valeurs sont {1, 2, 3, 4, 5} est un attribut ordinal numérique (discret).

De plus, si les valeurs d'un attribut discret ou nominal sont binaires, on parle d'un attribut binaire, par exemple l'attribut nominal sexe qui prend les valeurs {masculin, féminin} ou un attribut discret qui prend les valeurs {0,1}.

Dans notre travail, nous traitons les attributs nominaux et les attributs discrets. Si les attributs dans la base d'apprentissage sont continus, on applique des méthodes de discrétisation pour les rendre discrets.

III.3.3 Exemple introductif à l'arbre de décision

Si on considère un défaut (court-circuit) dans un champ PV, généralement les caractéristiques I-V changent et affecte le point de puissance maximale (MPP), donc le MPP Normale et Défaut varient dans la courbe I-V comme illustré dans la Figure III.7. Un arbre de décision est composé de *nœud racine*, des *nœuds internes* et des *nœuds terminaux* ou des feuilles est utilisé pour détecter MPP Défaut de MPP Normale (Voir la Figure III.8). A chaque nœud il y a un test (question) qui correspond à un attribut (V_{PV} et I_{PV} correspondant à chaque MPP) et chaque feuille correspond à une classe (Normal ou Défaut).

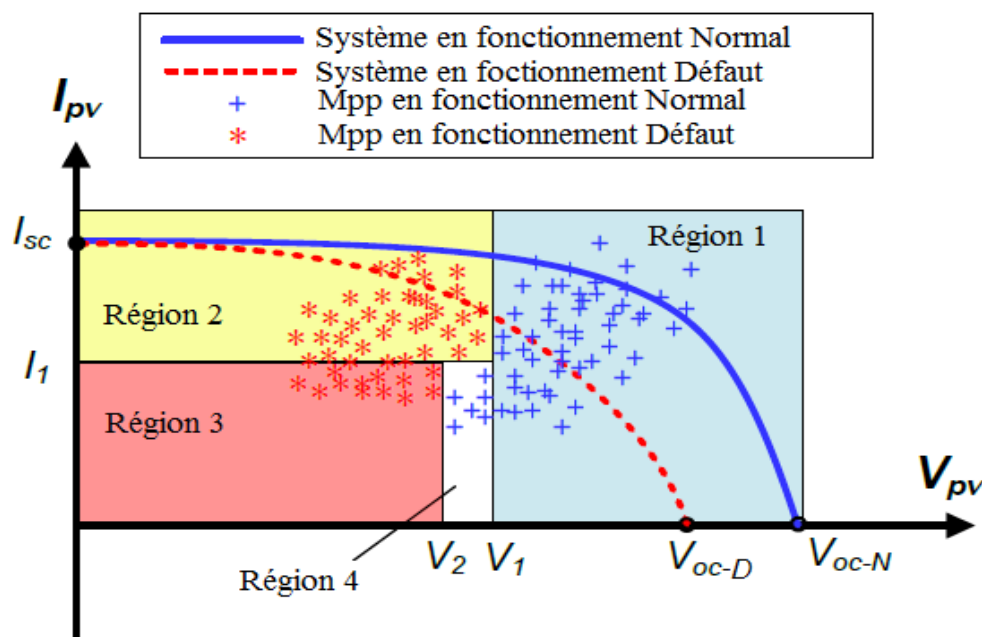


Figure III.7. Caractéristiques I-V en fonctionnement Normal et Défaut.

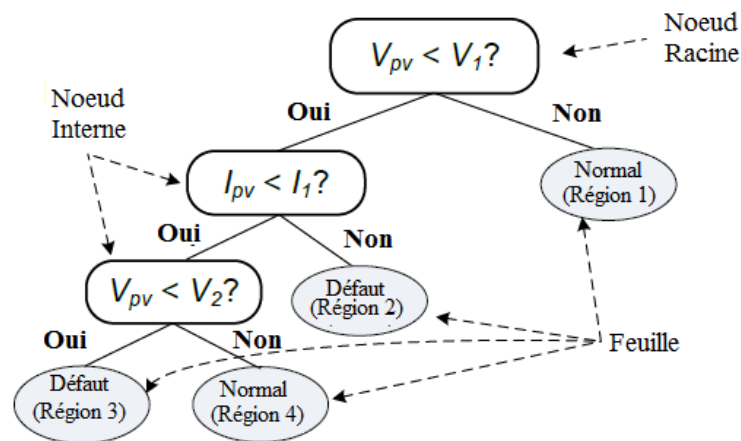


Figure III.8. Modèle d'arbre de décision pour détection de défaut.

Notons que le modèle simple dans la Figure III.8 a seulement utilisé deux attributs (V_{pv} et I_{pv}) pour détecter le défaut. Mais dans le cas réel, MPP Normale et défaut peut se chevaucher dans la courbe I-V donc il est utile que d'autre attribut soit pris en considération (puissance, température, radiation, ...etc.). Ci-après, des définitions de base relative à l'arbre de décision sont données.

- **Définition de base**

Un arbre de décision est une représentation d'une procédure de décision pour déterminer la classe d'un objet donné construis à partir d'un ensemble d'apprentissage. Un chemin de la racine à un nœud correspond a une série d'attributs (questions) avec leurs valeurs (réponses). La structure arborescente d'un arbre est équivalente à un ensemble de règles « *si – alors* ». Ceci justifie une prise décision en suivant l'un de ces chemins. Cette structure de connaissances est très proche de celle manipulée naturellement par un être humain [35].

III.3.4 Construction d'un arbre de décision

L'idée centrale de la construction d'un arbre de décision consiste à diviser récursivement les exemples de l'ensemble d'apprentissage en utilisant des tests définis à l'aide des attributs jusqu'à ce que l'on obtient des feuilles ne contenant (idéalement) que des exemples appartenant tous à la même classe. Pour diviser l'ensemble d'apprentissage, on choisit des attributs qui vont minimiser l'impureté dans les sous-arbres; autrement dit, qui maximise l'information apportée par les réponses. C'est-à-dire que pour chaque attribut qui n'a pas encore été utilisé, on calcule l'impureté qui reste après son utilisation. Celui qui laisse le moins de désordre est choisi comme étant le prochain nœud de l'arbre de décision. On répète le processus sur chaque nouveau nœud.

Le processus s'arrête quand les feuilles de l'arbre ainsi obtenu contiennent des exemples d'une seule classe ou quand aucun test n'apporte plus d'amélioration [35] [36] [37].

Dans toutes les méthodes de construction d'un arbre de décision, on trouve les trois opérateurs suivants (**Divide-and-Conquer Strategy**) expliqués ci-après [36]:

- Décider si un nœud est terminal : tous les exemples (un ou plus) appartiennent à la même classe (il y a moins d'un certain nombre d'erreurs).
- Sélectionner un test pour l'associer à un nœud.
- Affecter une classe à une feuille. On attribue la classe majoritaire à une feuille.

Formulation mathématique

Nous Adoptons les notations suivantes :

$X = \{X_1, X_2, \dots, X_i, \dots, X_n\}$; L'ensemble des attributs, avec n le nombre d'attributs ;

$C = \{C_1, C_2, \dots, C_k\}$; L'ensemble de classes, avec k le nombre de classes ;

$|C_j|$, $j = 1, 2, \dots, k$; Le nombre des exemples qui appartiennent à la classe C_j ;

D ; L'ensemble des exemples ;

$|D|$; Le nombre total des exemples.

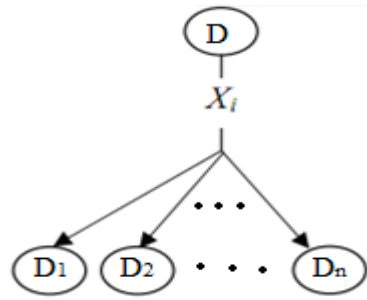
La construction d'un arbre de décision se base sur le principe d'induction de haut en bas (**Top-Down Induction**) [41]. On commence par construire le nœud racine de l'arbre en continuant jusqu'à la feuille. A chaque étape un test sur le nœud courant est choisi où le choix d'un attribut test X_i . Se fait en mesurant l'impureté (**l'incertitude**). Une des fonctions utilisée à cette fin est L'entropie [37].

- **Entropie de Shannon (Information)** : c'est la quantité d'information nécessaire pour identifier la classe d'un élément de D , définie comme suit :

$$\text{Info}(D) = - \sum_{j=1}^k \frac{|C_j|}{|D|} \log_2 \left(\frac{|C_j|}{|D|} \right) \quad (\text{III. 1})$$

S'il n'y a qu'une seule classe, l'entropie est nulle donc par convention $\log_2(0) = 0$.

Si les k classes sont équiprobables, l'entropie est maximale et vaut $\log_2(k)$. Pour partitionner l'ensemble d'apprentissage D sur la base des valeurs d'un attribut X_i en sous ensembles D_i où $i=1, \dots, n$ (Voir la Figure III.9), on va calculer l'information gagnante par cet attribut en utilisant l'entropie conditionnelle.

Figure III.9. Structure de l'arbre de décision pour l'ensemble D .

• **Entropie conditionnelle** : c'est la quantité d'informations nécessaires pour identifier la classe d'un élément de D sachant les valeurs de l'attribut X_i , il est donné par l'expression donnée ci-dessous:

$$\text{Info}(X_i, D) = - \sum_{i=1}^n \frac{|D_i|}{|D|} \cdot \text{Info}(D_i) = - \sum_{i=1}^n \left(\frac{|D_i|}{|D|} \cdot \sum_{j=1}^k \frac{|C_j|}{|D_i|} \log_2 \left(\frac{|C_j|}{|D_i|} \right) \right) \quad (\text{III. 2})$$

• **Gain d'Entropie** pour l'attribut X_i : ceci est défini par l'expression suivante :

$$\text{Gain}(X_i, D) = \text{Info}(D) - \text{Info}(X_i, D) \quad (\text{III. 3})$$

Cela représente la différence entre l'information nécessaire pour identifier un élément de D et l'information nécessaire pour identifier un élément de D sachant que la valeur de l'attribut X_i ait été obtenue.

Ce critère permet de choisir l'attribut le plus important. L'attribut choisi est celui qui a le gain le plus grand par rapport aux autres attributs. Après avoir choisi un attribut test, on doit partitionner l'ensemble en sous-ensembles selon les valeurs possibles de cet attribut. On crée les branches correspondant à chaque valeur de l'attribut, on crée des nœuds pour chaque sous-ensemble non vide et on répète le processus pour chaque nouveau nœud jusqu'à ce qu'on arrive à un critère d'arrêt. Dans ce cas, on crée une feuille et on associe à cette feuille la classe la plus probable [37].

L'algorithme de construction d'un arbre de décision est résumé ci-après [35]:

Construire-arbre (D, X): (TDIDT)

Entrées :

D: un ensemble d'apprentissage.

X: liste d'attributs.

Sortie: un arbre de décision.

Si un critère d'arrêt est vérifié

Créer une feuille avec la classe la plus probable.

Sinon:

Choisir le meilleur attribut X_i dans X (Qui maximise le gain d'information).

Partitionner D en n sous-ensembles selon les valeurs de X_i .

Créer un nœud pour X_i .

Pour toutes les valeurs de X_i : construire-arbre (D_i , $X-X_i$)

Fin si.

L'exemple illustré dans le tableau III.1 ci-dessous explicite l'application direct de l'algorithme de construction de l'arbre de décision décrit plus haut [41].

| Temps | Température | Humidité | Vent | Jouer |
|------------|-------------|----------|------|-------|
| Ensoleillé | Elevée | Haute | Faux | Non |
| Ensoleillé | Elevée | Haute | Vrai | Non |
| Couvert | Elevée | Haute | Faux | Oui |
| Pluvieux | Moyenne | Haute | Faux | Oui |
| Pluvieux | Basse | Normale | Faux | Oui |
| Pluvieux | Basse | Normale | Vrai | Non |
| Couvert | Basse | Normale | Vrai | Oui |
| Ensoleillé | Moyenne | Haute | Faux | Non |
| Ensoleillé | Basse | Normale | Faux | Oui |
| Pluvieux | Moyenne | Normale | Faux | Oui |
| Ensoleillé | Moyenne | Normale | Vrai | Oui |
| Couvert | Moyenne | Haute | Vrai | Oui |
| Couvert | Elevée | Normale | Faux | Oui |
| Pluvieux | Moyenne | Haute | Vrai | Non |

Tableau III.1. Exemple de construction d'un AD : La base Météo

Le partitionnement de l'ensemble d'apprentissage selon les valeurs des quatre attributs (Temps, Température, Humidité et Vent) est illustré sur la Figure III.10.

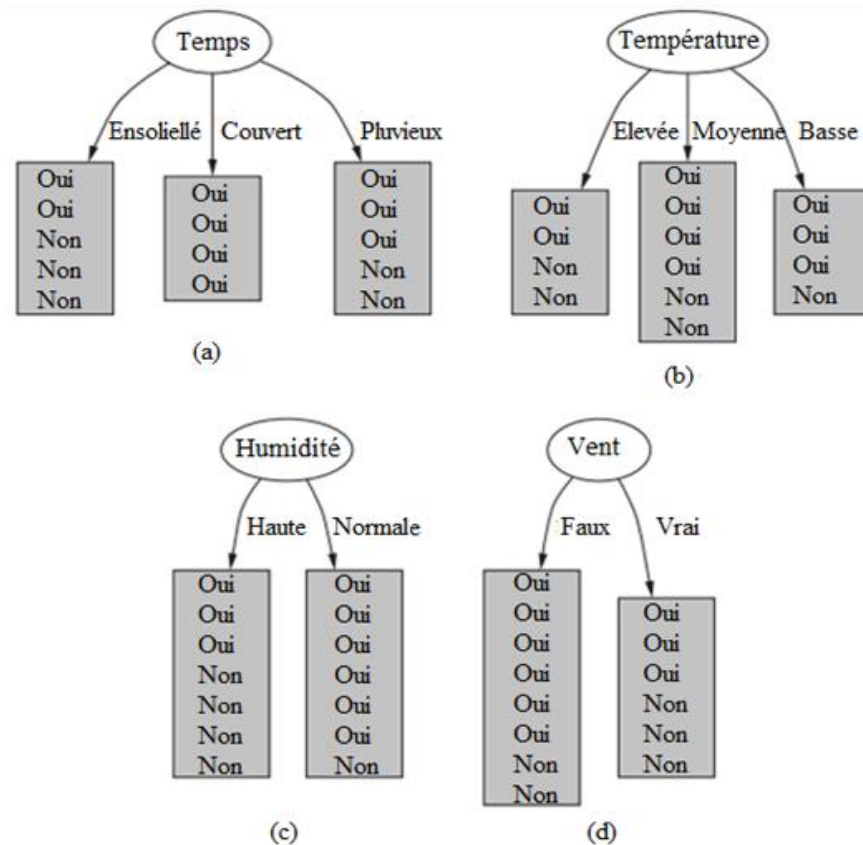


Figure III.10. Les sous-ensembles associés aux valeurs des quatre attributs.

On va chercher l'attribut le plus pertinent qui maximise le gain d'information (Equation III.3): Pour l'attribut Temps, le nombre de Oui et Non selon les valeurs : Ensoleillé, Couvert et Pluvieux sont respectivement [2, 3], [4,0], et [3, 2] (figure 3.10(a)), l'entropie correspondante est :

$$\text{Info} ([2, 3]) = 0.971 \text{ bits.}$$

$$\text{Info} ([4, 0]) = 0.0 \text{ bits.}$$

$$\text{Info} ([3, 2]) = 0.971 \text{ bits.}$$

L'entropie conditionnelle sera :

$$\begin{aligned} \text{Info} ([2, 3], [4, 0], [3, 2]) &= (5/14)*0.971+(4/14)*0+(5/14)*0.971 \\ &= 0.693 \text{ bits.} \end{aligned}$$

La base d'apprentissage contient neuf *Oui* et cinq *Non* dont l'entropie est :

$$\text{Info}(D) = \text{Info} ([9, 5]) = 0.940 \text{ bits.}$$

Info(D) correspond à l'entropie de l'ensemble d'exemples.

Finalement le gain d'information apportée par l'attribut Température est :

$$\begin{aligned}\text{Gain}(\text{Temps}) &= \text{Info}(D) - \text{Info}(\text{Temps}, D) \\ &= \text{Info}([9, 5]) - \text{Info}([2, 3], [4, 0], [3, 2]) \\ &= 0.940 - 0.693 = 0.247 \text{ bits.}\end{aligned}$$

On refait la même chose avec les autres attributs, on obtient :

$$\text{Gain}(\text{Temps}) = 0.247 \text{ bits.}$$

$$\text{Gain}(\text{Température}) = 0.029 \text{ bits.}$$

$$\text{Gain}(\text{Humidité}) = 0.152 \text{ bits.}$$

$$\text{Gain}(\text{Vent}) = 0.048 \text{ bits.}$$

L'attribut qui maximise le gain d'information est **Temps**. Donc, Temps est la racine de l'Arbre de Décision comme le montre la figure ci-dessous (Figure III.11.)

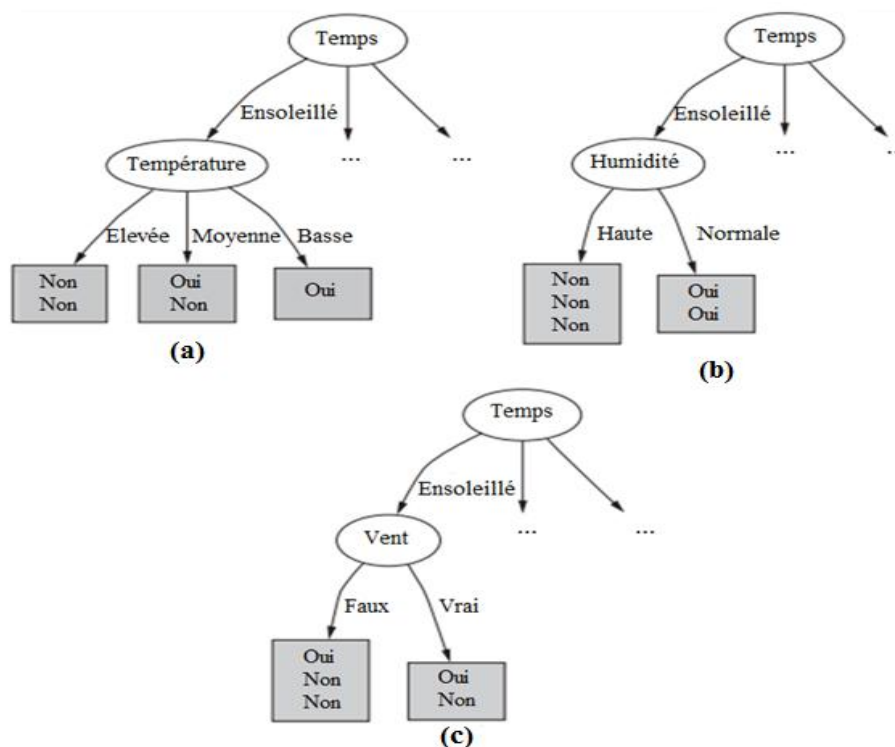


Figure III.11. Les sous-ensembles correspondant à la valeur ensoleillée de l'attribut Temps.

Pour continuer à construire l'arbre, on recommence le processus en choisissant d'autres attributs pertinents pour chaque sous-ensemble d'apprentissage (Voir la figure III.11). L'attribut le plus pertinent est *Humidité*. Donc, le sous-ensemble est partitionné en deux sous-ensembles selon les valeurs haute et normale de l'humidité. Dans le sous-ensemble correspondant à la valeur

pluvieux de l'attribut *Temps*, c'est l'attribut *Vent* qui maximise le gain et le sous-ensemble est partitionné en deux sous-ensembles selon les valeurs *vrai* et *faux* du *Vent*.

L'arbre de décision final obtenu est donné dans la figure III.12.

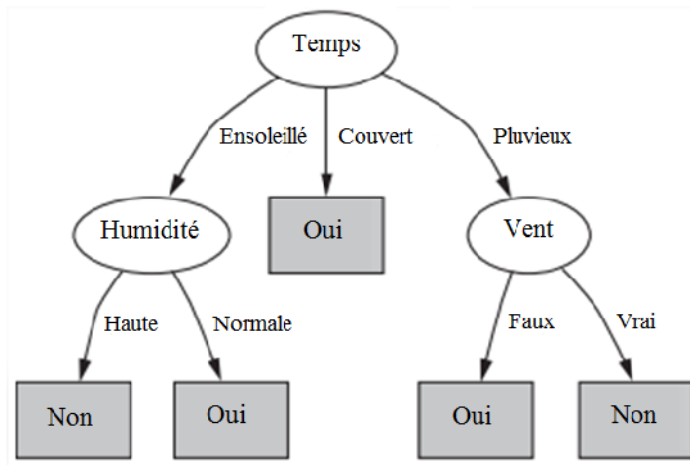


Figure III.12. L'arbre de décision final pour la base météo.

III.4 Choix de l'algorithme

L'algorithme décrit précédemment n'est pas fonctionnel pour des raisons, tels que:

- Ne prend pas en considération les attributs numériques;
- Les valeurs manquantes (inconnus) dans les attributs ne sont pas prises en compte;
- Ne traite pas l'étape d'élagage;
- Ne traite pas les données de dimensions élevées.

III.4.1 L'algorithme C4.5

C4.5 est l'algorithme proposé par Quinlan en 1993 qui est l'extension de l'algorithme de base d'ID3 [42]. Elle est l'un des algorithmes d'induction d'AD les plus connus et les plus utilisés aujourd'hui [43] [44] [45]. Elle comporte plusieurs améliorations citées ci-dessous :

- La construction d'AD selon la même méthodologie expliquée dans l'algorithme Précédente.
- L'élagage se base sur une estimation de l'erreur réelle à partir de l'ensemble d'apprentissage.
- Utilisation du gain ratio au lieu du gain d'entropie, défini comme suit :

$$G_Ratio(X_i, D) = \frac{Gain(X_i, D)}{Split\ info(X_i, D)} \quad (III. 4)$$

Avec

$$Split\ Info(X_i, D) = - \sum_{i=1}^n \frac{|D_i|}{|D|} \log_2 \left(\frac{|D_i|}{|D|} \right) \quad (III. 5)$$

- Les valeurs numériques se traitent comme suit : (soit X_i un attribut numérique)
- Toutes les valeurs de cet attribut sont classées dans l'ordre croissant.
- pour chacune des valeurs S_j on partitionne les enregistrements en deux catégories :
 - ceux qui ont une valeur inférieure ou égale à S_j .
 - ceux qui ont une valeur supérieure à S_j .

Pour chacune de ces partitions, on calcule le gain ratio, et on choisit la partition qui maximise le gain ratio.

III.4.2 Critères d'arrêt de construction d'un AD

Il y a plusieurs critères d'arrêt possibles lors de la construction d'un arbre de décision. Ils sont énumérés ci-dessous :

- Tous les exemples appartiennent à la même classe ou il n'y a plus d'attribut à utiliser.
- La profondeur de l'arbre atteint une limite fixée.
- Le nombre de feuilles atteint un maximum fixé.
- Le nombre d'instances par nœud est inférieur à un seuil fixé.
- Le gain d'information maximum obtenu est inférieur à un seuil fixé.
- La qualité de l'arbre n'augmente plus de façon sensible. Par exemple, aucun attribut n'améliore la qualité de l'arbre.

Le fait de construire l'arbre jusqu'au bout selon le premier critère ci-dessus, jusqu'à ce qu'aucun attribut ne reste à utiliser ou tous les exemples appartiennent à la même classe, favorise les feuilles ayant peu d'objets et diminue la fiabilité de l'arbre lors du classement d'un nouvel exemple. Par conséquent, le taux d'erreurs de classement augmente. Le premier critère n'aide donc pas à construire l'arbre de décision le plus simple. Les autres critères aident à stopper la construction lorsqu'un seuil est atteint, comme le nombre d'instances par feuille, la profondeur de

l'arbre, le nombre de feuilles dans l'arbre, etc. Ces critères se situent dans le cadre de l'*élagage* et plus précisément dans son premier type appelé *pré-élagage*.

III.5 Élagage

Le sur apprentissage est en général provoqué par un mauvais dimensionnement de l'arbre de décision en raison de :

- Le bruit présent dans les données d'apprentissage;
- les données ne contiennent pas certains exemples importants, ou lorsque les exemples sont trop spécifiques.

Pour limiter ce genre de problèmes on utilise l'élagage qui consiste à réduire la taille de l'arbre par suppression de quelques branches [46].

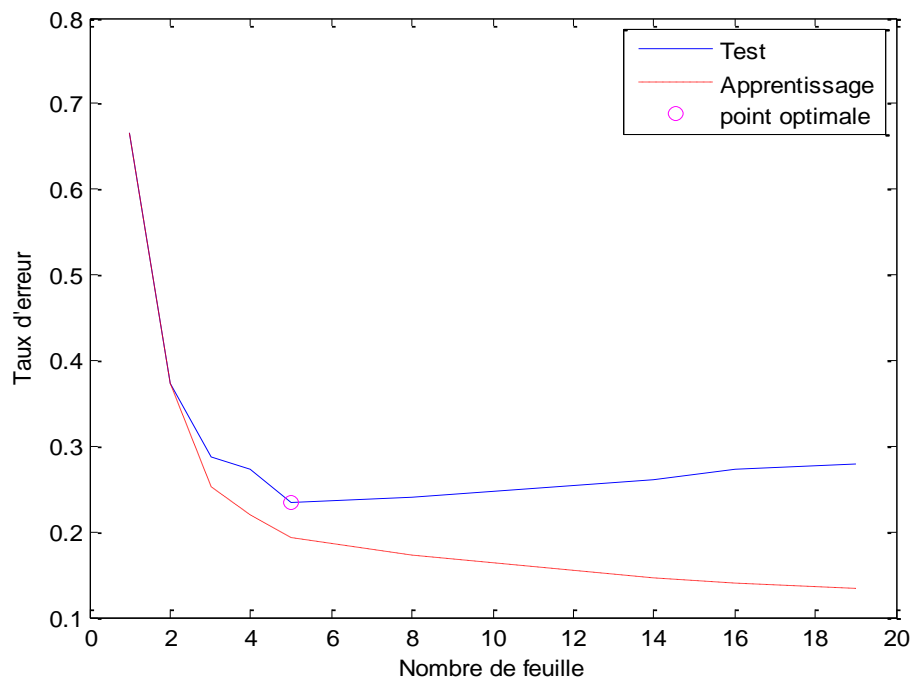
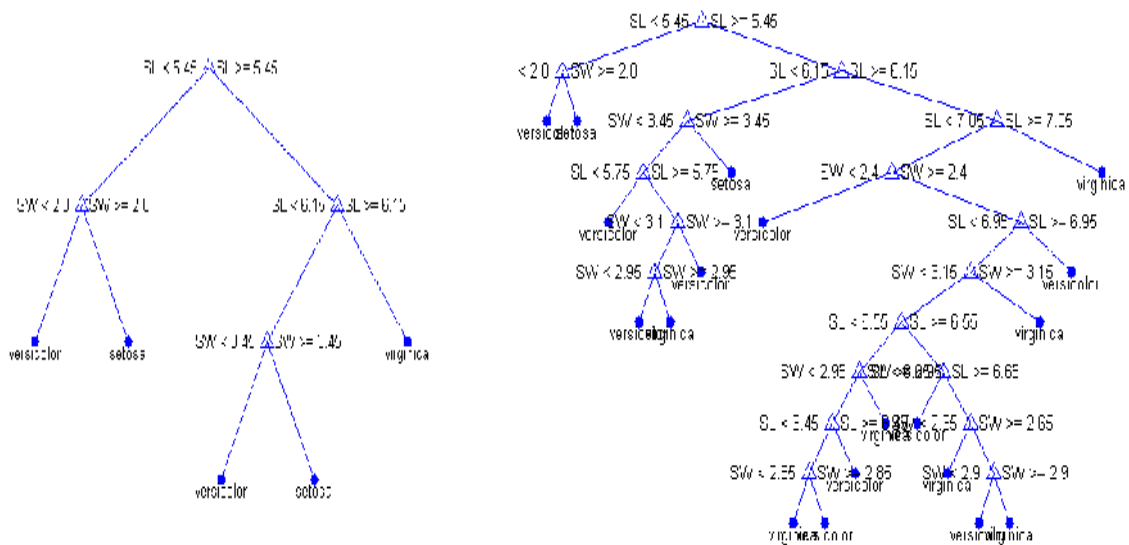


Figure III.13. Evolution de taux d'erreur en apprentissage et en test.

Dans la Figure III.13 ci-dessus le graphe en rouge, représente l'erreur sur l'ensemble d'apprentissage. En bleu, l'erreur sur l'ensemble de test.

- Si l'erreur de test augmente alors que l'erreur d'apprentissage continue à diminuer alors il y a un sur apprentissage.
- Dans le point optimal le commencement de sur apprentissage apparaît, on commence à élaguer depuis le repère indiqué sur le graphe (point optimal). Pour cela l'élagage se fait

aux niveaux de 5 feuilles, les deux structure de l'arbre avec et sans élagage sont montrée sur la Figure III.14.



(a) Arbre de décision avec 5 feuilles avec élagage. (b) Arbre de décision sans élagage avec 19 feuilles.

Figure III.14. Arbres de décision avec et sans élagage.

III.5.1 Pré-élagage

Le pré-élagage a pour but d'arrêter la construction de l'AD à l'avance même si les feuilles ne sont pas pures ; c'est-à-dire qu'on décide ou non de continuer à développer un certain nœud en utilisant par exemple les critères d'arrêts [43].

Il existe d'autre technique proposée par Quinlan et appelée « *Chi-square pruning* », citée dans [43], [47], [48].

III.5.2 Post-élagage

Dans cette approche, l'arbre de décision est simplifié en supprimant un ou plusieurs de ses sous-arbres et en les remplaçant par des feuilles. Il existe plusieurs méthodes, nous détaillerons la méthode implémentée en C4.5 :

- **Error-based pruning « EBP »**

L'idée de cette méthode est la suivante: si N_T est le nombre total d'instances parvenues à un nœud ou une feuille et E_M le nombre d'objets mal classés sur cette feuille, alors le taux d'erreur optimiste observé est $f = E_M/N_T$ (observer E événements dans N_T essais) [43]. Quinlan a utilisé l'intervalle confiance cf de la distribution Binomiale pour estimer le taux d'erreur prédit e . Il est

approxime la distribution Binomiale par la distribution Normale Centré Réduite pour trouver le seuil de confiance Z de niveau de confiance $\alpha/2$ (par défaut $\alpha = 0.25$). L'expression pour estimer la valeur de e est :

$$e(N_T, f, \alpha) = \frac{f + \frac{Z_{\alpha/2}^2}{2N_T} + Z_{\alpha/2} \sqrt{\frac{f}{N_T} - \frac{f^2}{N_T} + \frac{Z_{\alpha/2}^2}{4N_T^2}}}{1 + \frac{Z_{\alpha/2}^2}{N_T}} \quad (\text{III. 6})$$

Si $\alpha = 0,25$ alors la valeur de $Z_{\alpha/2} = 1.15$ d'après le tableau 1 voir l'annexe.

Pour décider si on élague un sous-arbre ou pas on doit d'abord :

- Calculer probabilité d'erreur prédit sur chaque feuille de ce sous-arbre $e(N_j, f_j, \alpha)$; ou j est le $j^{\text{ème}}$ feuille
- Calculer le nombre d'erreurs prédites sur chaque feuille, qui est égal au nombre de cas sur cette feuille \times le taux d'erreur prédit : $N_j \times e(N_j, f_j, \alpha)$;
- Calculer le nombre total d'erreurs prédites sur toutes les feuilles du sous-arbre (la somme de toutes les erreurs des feuilles); $\sum_j N_j \times e(N_j, f_j, \alpha)$
- Calculer le nombre d'erreurs prédites sur le nœud père de ces feuilles ; qui est le nombre de cas sur ce nœud \times le taux d'erreur prédit sur ce nœud; $N_T \times e(N_T, f_T, \alpha)$
- Comparer les valeurs obtenues lors des deux étapes précédentes: si l'erreur donnée par le nœud père est inférieure à l'erreur donnée par les feuilles alors on élague le sous-arbre de ce nœud père.

Ce processus répète d'une manière itérative de bas en haut (*bottom up*).

- La description de l'algorithme d'élagage est le suivant :

Tant qu'il existe un sous-arbre que l'on peut remplacer par une feuille sans faire croître l'erreur réelle estimée

Alors élaguer ce sous-arbre.

Dans l'exemple ci-dessous on considère un arbre de décision sans élagage, on montre dans ceux suit comment peut on élaguer

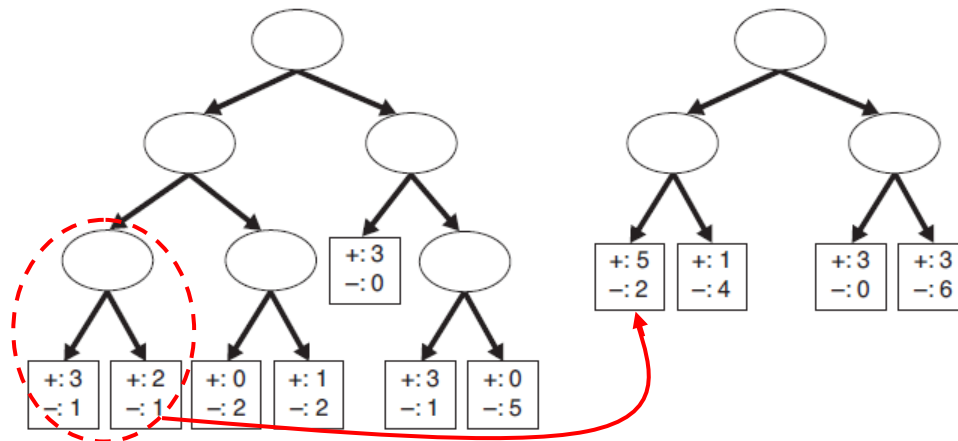


Figure III.15. Elagage d'un sous-arbre.

On fait une étude illustrative sur le sous-arbre de la Figure III.15, pour la feuille gauche $N_T = 4$ et $E = 1$ avec un intervalle de confiance $cf = 25\%$ donc le taux d'erreur estimé est

$e(4;1/4; 25\%) = 0,537$ pour la feuille droite $N_T = 3$ et $E = 1$ alors $e(3,1/3; 25\%) = 0,650$,

l'erreur globale de ces feuilles est $4 \times 0,537 + 3 \times 0,650 = 4,098$, par contre l'erreur estimée par le nœud père de cette sous-arbre est $e(7,2/7; 25\%) = 0,503$ et l'erreur globale de ce nœud sera $7 \times 0,503 = 3,521$, le taux d'erreur de nœud père est inférieur à l'erreur de ces feuilles donc on élague ce sous-arbre comme montre sur la Figure III.15

On s'intéresse aux deux feuilles situées à gauche de l'AD comme montré dans la Figure III.15

L'erreur prédite de la feuille gauche est : $e(4;1/4; 25\%) = 0,537$

L'erreur prédite de la feuille droite est : $e(3,1/3; 25\%) = 0,650$

L'erreur globale de ces feuilles est : $4 \times 0,537 + 3 \times 0,650 = \mathbf{4,098}$

L'erreur prédite par le nœud père est : $e(7,2/7; 25\%) = 0,503$

Et l'erreur globale pour ce nœud père est : $7 \times 0,503 = \mathbf{3,521}$

Il est remarquable que le taux d'erreur de nœud père est inférieur à l'erreur de ces feuilles donc on élague ce sous-arbre et on le remplace par la classe majoritaire comme montre sur la Figure III.15.

III.6 Évaluation du modèle

Pour évaluer notre approche, nous commençons par la tester sur plusieurs bases réelles. Ensuite, nous évaluons sa performance en utilisant la *Matrice de Confusion*.

III.6.1 Matrice de Confusion

La matrice de confusion [35], [36], contient des informations concernant le classement actuel dans la base de test ainsi que le classement prédit par le système de classement utilisé (C4.5). Le tableau III.2 présente la matrice de confusion pour deux valeurs de classe (positive, négative). Ainsi,

- a est le nombre de classements corrects des instances de classe **négative**.
- b est le nombre de classements incorrects des instances de classe **négative**.
- c est le nombre de classements incorrects des instances de classe **positive**.
- d est le nombre de classements corrects des instances de classe **positive**.

| | | Classe prédite | |
|---------------|----------|----------------|----------|
| | | Négative | Positive |
| Classe réelle | Négative | a | b |
| | Positive | c | d |

Tableau III.2 : La matrice de Confusion.

À partir de la matrice de confusion ci-dessus, plusieurs mesures appelées mesurent d'exactitude par classe, sont définies :

- **Taux de classification** (exactitude - taux de bon apprentissage) : la diagonale de la matrice de confusion sur toutes les éléments

$$\text{Taux de classification} = \frac{a + d}{a + b + c + d} \quad (\text{III. 7})$$

- True Positive Rate « **TP Rate** » : la proportion des instances de classe positive qui sont correctement classées.

$$\text{TP Rate} = \frac{d}{c + d} \quad (\text{III. 8})$$

C'est donc le rapport entre le nombre de bien classés et le nombre total d'instances qui devraient être bien classées. Il se calcule en utilisant la deuxième ligne de Tableau III.2.

- True Negative Rate « **TN Rate** » : la proportion des instances négatives qui sont correctement classées.

$$TNrate = \frac{a}{a + b} \quad (III. 9)$$

- False Postive Rate « **FP Rate** »: la proportion des instances négatives qui sont incorrectement classées comme positives.

$$FPrate = \frac{b}{a + b} \quad (III. 10)$$

Il se calcule en utilisant la première ligne de Tableau III.2.

- False Negative Rate « **FN Rate** »: la proportion des instances positives qui sont incorrectement classées comme négatives.

$$FNrate = \frac{c}{c + d} \quad (III. 11)$$

III.7 Exemple illustratif

Nous avons pris un exemple avec le détail pour mieux comprendre le déroulement de l'algorithme. Soit l'ensemble d'apprentissage avec cinq attributs numériques (discret) qui sont la température ambiante (T [°C]), l'ensoleillement (E [W/m²]), la tension (V_{PV} [V]), le courant (I_{PV} [A]) et la puissance (P_{PV} [W]), voir l'annexe Tableau 2.

1. Pour calculer l'entropie on utilise la table de fréquence de l'attribut classe D:

| D | |
|--------|--------|
| Normal | Défaut |
| 11 | 11 |

$$\text{Info (D)} = \text{info}[11,11]$$

$$= -\frac{11}{22} \log_2 \left(\frac{11}{22} \right) - \frac{11}{22} \log_2 \left(\frac{11}{22} \right)$$

$$\text{Info (D)} = 1 \text{ bits}$$

2. Dans l'étape suivante on range les valeurs de chaque attribut avec l'attribut classe dans un ordre croissant.

3. Calculer le gain ratio de chaque valeur d'attributs séparément pour choisir le meilleur seuil qui a la valeur de gain ratio le plus élevé (celle qui maximise le gain ratio) .

Les résultats des calculs sont montrées dans le tableau suivant :

| 22.989 | 23.34 | 23.497 | 23.55 3 | 23.691 | 23.733 | 23.791 | 23.995 | 24.14 | 24.205 | 24.537 | T (°C) | | | | | | | | | | | |
|--------|-------|--------|------------|--------|--------|--------|--------|-------|--------|--------|--------|---|---|---|---|---|---|----|---|----|---|-------------|
| ≤ | > | ≤ | > | ≤ | > | ≤ | > | ≤ | > | ≤ | > | | | | | | | | | | | |
| 1 | 10 | 2 | 9 | 3 | 8 | 4 | 7 | 5 | 6 | 6 | 5 | 7 | 4 | 8 | 3 | 9 | 2 | 10 | 1 | 11 | 0 | Normal |
| 1 | 10 | 2 | 9 | 3 | 8 | 4 | 7 | 5 | 6 | 6 | 5 | 7 | 4 | 8 | 3 | 9 | 2 | 10 | 1 | 11 | 0 | Défaut |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | Info (bits) |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Gain |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | Split Info |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | G_Ratio |

Tableau III.3. Calcul de l'information pour l'attribut T.

Pour chaque valeur de la température T, nous allons vous expliquer la méthode de calcul pour toutes les valeurs: on prend comme exemple la valeur de T = 22.989 C°

$$\begin{aligned}
 Info(T, D) &= \frac{2}{22} \times Info([T \leq 22.989]) + \frac{20}{22} \times Info([T > 22.989]) \\
 &= \frac{2}{22} \times \left[-\frac{1}{4} \times \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \times \log_2\left(\frac{1}{2}\right) \right] + \frac{20}{22} \times \left[-\frac{10}{20} \times \log_2\left(\frac{10}{20}\right) - \frac{10}{20} \times \log_2\left(\frac{10}{20}\right) \right]
 \end{aligned}$$

Donc

$$Info(T, D) = 1 \text{ bits/intance}$$

$$Gain(T, D) = Info(D) - Info(T, D) = 1 - 1 = 0$$

$$Split\ Info(T, D) = -\frac{2}{22} \times \log_2\left(\frac{2}{22}\right) - \frac{18}{22} \times \log_2\left(\frac{18}{22}\right) = 1$$

$$G_Ratio(T, D) = \frac{Gain(T, D)}{Split\ Info(T, D)} = \frac{0}{1} = 0$$

On remarque que cet attribut ne donne aucune information, puisque toutes les valeurs donnent le même gain ratio comme montrer dans le tableau III.3, on choisit une valeur au hasard qui présente le seuil (exemple T = 22.989 °C).

➤ Maintenant on passe au deuxième attribut qui est l'ensoleillement (E) :

| 368,035 | | 372,285 | | 375,014 | | 379,36 | | 382,728 | | 386,231 | | 389,954 | | 393,495 | | 396,635 | | 400,6 | | 404,063 | | E(w/m ²) |
|---------|----|---------|---|---------|---|--------|---|---------|---|---------|---|---------|---|---------|---|---------|---|-------|---|---------|---|----------------------|
| ≤ | > | ≤ | > | ≤ | > | ≤ | > | ≤ | > | ≤ | > | ≤ | > | ≤ | > | ≤ | > | ≤ | > | ≤ | > | |
| 1 | 10 | 2 | 9 | 3 | 8 | 4 | 7 | 5 | 6 | 6 | 5 | 7 | 4 | 8 | 3 | 9 | 2 | 10 | 1 | 11 | 0 | Normal |
| 1 | 10 | 2 | 9 | 3 | 8 | 4 | 7 | 5 | 6 | 6 | 5 | 7 | 4 | 8 | 3 | 9 | 2 | 10 | 1 | 11 | 0 | Défaut |
| 1 | | 1 | | 1 | | 1 | | 1 | | 1 | | 1 | | 1 | | 1 | | 1 | | 1 | | Info(bits) |
| 0 | | 0 | | 0 | | 0 | | 0 | | 0 | | 0 | | 0 | | 0 | | 0 | | 0 | | Gain |
| 1 | | 1 | | 1 | | 1 | | 1 | | 1 | | 1 | | 1 | | 1 | | 1 | | 1 | | Split Info |
| 0 | | 0 | | 0 | | 0 | | 0 | | 0 | | 0 | | 0 | | 0 | | 0 | | 0 | | G_Ratio |

Tableau III.4. Calcule de l'information pour l'attribut E.

On suit la même procédure pour l'attribut E, on remarque que les valeurs du gain ratio sont égaux, on choisit une valeur au hasard qui présente le seuil de test sur cette attribut (exemple E = 368,035 w/m²).

➤ Pour l'attribut tension V_{PV} :

| 211,299 | | 211,511 | | 227,568 | | 227,704 | | 243,217 | | 243,546 | | 243,656 | | 243,744 | | 243,807 | | V _{PV} (V) |
|----------|----|---------|----|---------|----|----------------|----|---------|---|---------|---|---------|---|---------|---|---------|---|---------------------|
| ≤ | > | ≤ | > | ≤ | > | ≤ | > | ≤ | > | ≤ | > | ≤ | > | ≤ | > | ≤ | > | |
| 0 | 11 | 0 | 11 | 0 | 11 | 0 | 11 | 2 | 9 | 3 | 8 | 4 | 7 | 4 | 7 | 5 | 6 | Normal |
| 1 | 10 | 2 | 9 | 3 | 8 | 4 | 7 | 5 | 6 | 7 | 4 | 8 | 3 | 8 | 3 | 8 | 3 | Défaut |
| 0,953 | | 0,903 | | 0,853 | | 0,783 | | 0,871 | | 0,902 | | 0,902 | | 0,902 | | 0,944 | | Info (bits) |
| 0,05 | | 0,1 | | 0,147 | | 0,217 | | 0,129 | | 0,098 | | 0,098 | | 0,098 | | 0,056 | | Gain |
| 0,267 | | 0,44 | | 0,575 | | 0,684 | | 0,902 | | 0,994 | | 0,994 | | 0,994 | | 0,976 | | Split Info |
| 0,176 | | 0,227 | | 0,256 | | 0,317 | | 0,143 | | 0,099 | | 0,099 | | 0,099 | | 0,057 | | G_Ratio |

| 243,823 | | 243,900 | | 243,969 | | 244,052 | | 244,159 | | 244,212 | | V _{PV} (V) |
|---------|---|---------|---|---------|---|---------|---|---------|---|---------|---|---------------------|
| ≤ | > | ≤ | > | ≤ | > | ≤ | > | ≤ | > | ≤ | > | |
| 6 | 5 | 7 | 4 | 8 | 3 | 9 | 2 | 10 | 1 | 11 | 0 | Normal |
| 8 | 3 | 9 | 2 | 9 | 2 | 9 | 2 | 10 | 1 | 11 | 0 | Défaut |
| 0,974 | | 0,970 | | 0,992 | | 1 | | 1 | | 1 | | Info (bits) |
| 0,026 | | 0,03 | | 0,008 | | 0 | | 0 | | 0 | | Gain |
| 0,946 | | 0,845 | | 0,773 | | 0,684 | | 0,44 | | 0 | | Split Info |
| 0,027 | | 0,036 | | 0,01 | | 0 | | 0 | | 0 | | G_Ratio |

Tableau III.5. Calcule de l'information pour l'attribut V_{PV}.

La valeur de V_{PV} qui a le gain ratio le plus élevé est 227,207(V).

➤ Pour l'attribut courant I_{PV} :

| 0 | | 2,161 | | 2,182 | | 2,207 | | 2,312 | | 4,323 | | 4,364 | | 4,415 | | $I_{PV}(A)$ |
|-------|----|-------|----|-------|----|-------|----|---------------------|----|-------|----|-------|---|-------|---|-------------|
| ≤ | > | ≤ | > | ≤ | > | ≤ | > | ≤ | > | ≤ | > | ≤ | > | ≤ | > | |
| 0 | 11 | 0 | 11 | 0 | 11 | 0 | 11 | 0 | 11 | 1 | 10 | 2 | 9 | 3 | 8 | Normal |
| 3 | 8 | 4 | 7 | 5 | 6 | 6 | 5 | 7 | 4 | 7 | 4 | 7 | 4 | 7 | 4 | Défaut |
| 0,848 | | 0,788 | | 0,724 | | 0,652 | | <u>0,570</u> | | 0,747 | | 0,839 | | 0,902 | | Info (bits) |
| 0,152 | | 0,212 | | 0,276 | | 0,348 | | <u>0,43</u> | | 0,253 | | 0,161 | | 0,098 | | Gain |
| 0,575 | | 0,684 | | 0,773 | | 0,845 | | 0,902 | | 0,946 | | 0,976 | | 0,994 | | Split Info |
| 0,264 | | 0,31 | | 0,357 | | 0,412 | | <u>0,477</u> | | 0,267 | | 0,165 | | 0,099 | | G_Ratio |

| 4,447 | | 4,499 | | 4,539 | | 4,580 | | 4,624 | | 4,666 | | 4,704 | | 4,751 | | $I_{PV}(A)$ |
|-------|---|-------|---|-------|---|-------|---|-------|---|-------|---|-------|---|-------|---|-------------|
| ≤ | > | ≤ | > | ≤ | > | ≤ | > | ≤ | > | ≤ | > | ≤ | > | ≤ | > | |
| 4 | 7 | 5 | 6 | 6 | 5 | 7 | 4 | 8 | 3 | 9 | 2 | 10 | 1 | 11 | 0 | Normal |
| 8 | 3 | 9 | 2 | 10 | 1 | 11 | 0 | 11 | 0 | 11 | 0 | 11 | 0 | 11 | 0 | Défaut |
| 0,902 | | 0,893 | | 0,871 | | 0,789 | | 0,848 | | 0,903 | | 0,953 | | 1 | | Info (bits) |
| 0,098 | | 0,107 | | 0,129 | | 0,211 | | 0,152 | | 0,097 | | 0,047 | | 0 | | Gain |
| 0,994 | | 0,946 | | 0,845 | | 0,684 | | 0,575 | | 0,44 | | 0,267 | | 1 | | Split Info |
| 0,099 | | 0,113 | | 0,153 | | 0,308 | | 0,264 | | 0,22 | | 0,176 | | 0 | | G_ratio |

Tableau III.6. Calcul de l'information pour l'attribut I_{PV} .

Le gain ratio le plus élevé de l'attribut I_{PV} est 0,477 donc le seuil de ce dernier est 2,312 (A).

➤ Pour l'attribut puissance P_{PV} :

| 0 | 527,901 | 532,865 | 538,447 | 563,437 | 960,086 | 967,9 | 1012,149 | 1024,491 | 1055,802 | $P_{PV}(w)$ | | | | | | | | | | |
|-------|---------|---------|---------|---------|---------|-------|----------|----------|----------|-------------|----|-------|-----------|-------|----|-----------|---|-------|---|-------------|
| ≤ | > | ≤ | > | ≤ | > | ≤ | > | ≤ | > | ≤ | > | ≤ | > | ≤ | > | | | | | |
| 0 | 11 | 0 | 11 | 0 | 11 | 0 | 11 | 0 | 11 | 0 | 11 | 0 | 11 | 1 | 10 | Normal | | | | |
| 3 | 8 | 4 | 7 | 5 | 6 | 6 | 5 | 7 | 4 | 8 | 3 | 9 | 2 | 10 | 1 | 11 | 0 | 11 | 0 | Défaut |
| 0,848 | | 0,789 | | 0,724 | | 0,652 | | 0,570 | | 0,477 | | 0,366 | | 0,226 | | <u>0</u> | | 0,226 | | Info (bits) |
| 0,152 | | 0,211 | | 0,276 | | 0,348 | | 0,43 | | 0,523 | | 0,634 | | 0,774 | | <u>1</u> | | 0,774 | | Gain |
| 0,575 | | 0,684 | | 0,773 | | 0,845 | | 0,902 | | 0,946 | | 0,976 | | 0,994 | | <u>1</u> | | 0,994 | | Split info |
| 0,264 | | 0,308 | | 0,357 | | 0,411 | | 0,477 | | 0,553 | | 0,65 | | 0,779 | | <u>1</u> | | 0,779 | | G_Ratio |

| | | | | | | | | | | | | | | | | | | | | |
|----------|----------|--------------|--------------|----------|--------------|----------|--------------|----------|--------------|--------------------------|---|----|---|----|---|----|---|----|---|--------|
| 1065,731 | 1076,895 | 1084,44 5 | 1097,6 69 | 1107,791 | 1116,8 08 | 1126,875 | 1137,5 21 | 1145,668 | 1155,5 54 | P_{PV}(w) | | | | | | | | | | |
| ≤ | > | ≤ | > | ≤ | > | ≤ | > | ≤ | > | ≤ | > | ≤ | > | ≤ | > | ≤ | > | ≤ | > | |
| 2 | 9 | 3 | 8 | 4 | 7 | 5 | 6 | 6 | 5 | 7 | 4 | 8 | 3 | 9 | 2 | 10 | 1 | 11 | 0 | Normal |
| 11 | 0 | 11 | 0 | 11 | 0 | 11 | 0 | 11 | 0 | 11 | 0 | 11 | 0 | 11 | 0 | 11 | 0 | 11 | 0 | Défaut |
| 0,366 | 0,477 | 0,570 | 0,652 | 0,724 | 0,789 | 0,848 | 0,903 | 0,967 | 1 | <i>Info (bits)</i> | | | | | | | | | | |
| 0,634 | 0,523 | 0,43 | 0,348 | 0,276 | 0,211 | 0,152 | 0,1 | 0,03 | 0 | <i>Gain</i> | | | | | | | | | | |
| 0,976 | 0,946 | 0,902 | 0,845 | 0,773 | 0,684 | 0,575 | 0,44 | 0,267 | 1 | <i>Split Info</i> | | | | | | | | | | |
| 0,65 | 0,553 | 0,477 | 0,411 | 0,357 | 0,308 | 0,264 | 0,227 | 0,112 | 0 | <i>G_Ratio</i> | | | | | | | | | | |

Tableau III.7. Calcul de l'information pour l'attribut P_{PV}.

D'après ce dernier tableau le seuil de P_{PV} est 1024,491 (W) .

Le résumé de toutes ces valeurs est présenté dans le tableau III.8:

| L'attribut | T | E | V _{PV} | I _{PV} | P _{PV} |
|---------------|--------|---------|-----------------|-----------------|-----------------|
| Le seuil | 22.989 | 368,035 | 227,704 | 2,312 | 1024,491 |
| Le Gain Ratio | 0 | 0 | 0,317 | 0,477 | 1 |

Tableau III.8. Les valeurs de Gain Ratio.

D'après le tableau III.8, l'attribut qui a le gain ratio le plus élevé est l'attribut P_{PV}, donc cet attribut présent le nœud racine comme illustré dans la figure ci-dessous. On a deux branches (droite et gauche) pour ce nœud (P_{PV} ≤ 1024,491 et P_{PV} > 1024,491). Pour décider si le nœud prochain est une feuille ou non, il faut voir la classe majoritaire sur ce nœud, si elle est au moins supérieur ou égale à 2 alors ce nœud est une feuille

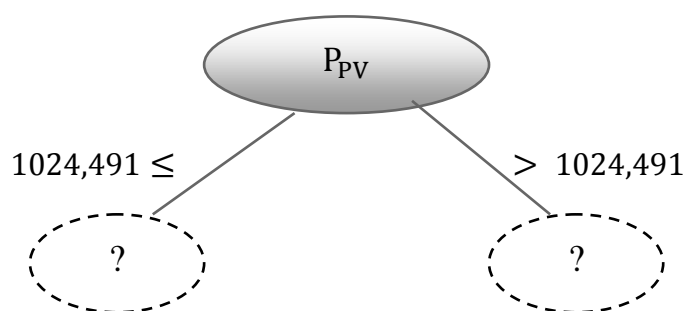


Figure III.16. Nœud racine.

Sur la branche gauche la classe majoritaire supérieure à 2 (voir le tableau III.7 pour la valeur 1024.491), donc la branche gauche se termine par une feuille d'étiquette Défaut comme illustré dans la figure ci-dessous :

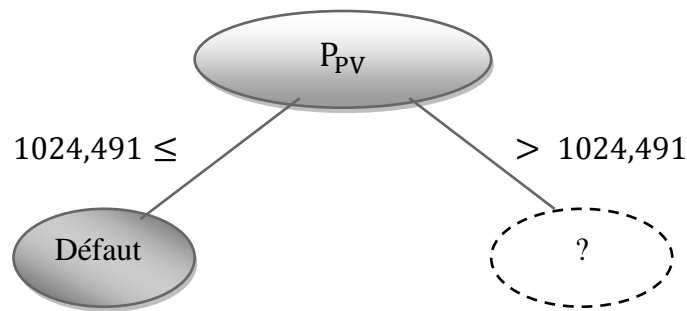


Figure III.17. Nœud racine avec une feuille.

La même chose sur la branche droite ou la classe majoritaire supérieur à 2, donc la branche droite se termine par une feuille d'étiquette Normal, voir la figure ci-dessous:

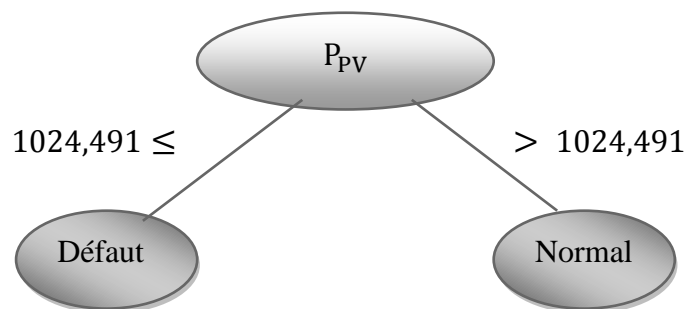


Figure III.18. Le Modèle de l'arbre finale

On peut traduire ce modèle par des règles de décisions :

Si $P_{PV} \leq 1024,491$ **alors** il y a un défaut, **Sinon** $P_{PV} > 1024,491$ **alors** il est normal.

4. D'après la construction de l'arbre on passe à la deuxième phase qui est l'élagage :

- l'erreur globale de ces feuilles est : $11 \times 0,107 + 11 \times 0,107 = 2,354$

- l'erreur globale pour ce nœud père est : $22 \times 0,619 = 13,618$

On constate que l'erreur globale par nœud père est supérieure à celle des feuilles, donc on n'élague pas l'arbre.

5. Pour valider le modèle d'AD il faut utiliser des nouvelles données qui ne sont pas utilisé dans la base d'apprentissage, l'ensemble de test est illustré dans le tableau 3 sur l'annexe, on a déduit la matrice de confusion donné dans le tableau III.9 :

| | | La classe prédite | |
|---------------|--------|-------------------|--------|
| | | Normal | Défaut |
| Classe réelle | Normal | 3 | 3 |
| | Défaut | 0 | 6 |

Tableau III.9 : Matrice de confusion de l'ensemble de test.

D'après la matrice de confusion de tableau III.9, on déduit les valeurs des paramètres qui

Nous intéressons par qui ce suit à (*Taux de classification, TN rate et TP rate*):

$$\text{Taux de classification} = \frac{3 + 6}{6 + 3 + 3} = 75\%$$

$$\text{TN Rate} = \frac{3}{3 + 3} = 50\%$$

$$\text{TP Rate} = \frac{6}{6 + 0} = 100\%$$

III.8 Avantage de l'AD

a) Les données:

C'est une des seules méthodes permettant de traiter de manière homogène pratiquement tous les types d'attributs (binaires, nominal, ordinal, numériques).

b) Intelligibilité:

La représentation des arbres de décision sont dans une forme qui sont intuitifs pour la personne moyenne, et donc nécessitent peu d'explications. Juste en regardant le diagramme on est capable de suivre le flux d'informations.

c) Efficacité:

L'apprentissage et la classification des mesures d'induction d'arbre de décision sont simples et rapides.

d) Extensibilité:

Arbres de décision peuvent traiter des données de grande dimension et sont faciles à développer sur la base de l'ensemble des données et étant que leurs classificateurs ne nécessitent aucune connaissance du domaine, ils peuvent être utilisés pour exploratoire découverte de connaissances.

e) Flexibilité:

Les arbres de décision sont utilisés dans différents domaines de la classification, y compris la fabrication et la production, l'analyse financière, la biologie moléculaire, l'astronomie et de la médecine.

III.9 Conclusion

Dans ce chapitre nous avons expliqué la méthode de construction d'un arbre de décision par l'algorithme C4.5 et la technique d'élagage utilisée par cette algorithme. Dans un tel arbre, les nœuds représentent les attributs et les valeurs de ces attributs forment les branches de chaque nœud. Les feuilles représentent la classe.

Chapitre IV : Expérimentation et Validation

IV.1 Introduction

Nous considérons dans ce chapitre l'application de l'AD, présentée dans le chapitre précédent, au diagnostic et à la détection des défauts dans une installation photovoltaïque. En effet, nous commençons par décrire la centrale photovoltaïque connectée au réseau considérée dans notre travail où nous mettons en évidence les différents défauts envisagés. Ensuite nous présentons les différentes étapes pour la construction de l'arbre de décision et les différents résultats obtenus. Pour la mise en œuvre de l'arbre de décision, nous avons utilisé trois outils logiciels différents à savoir : WEKA, MatLab et LabView.

IV.2 Description de la centrale photovoltaïque connectée au réseau (CPCR)

La CPCR installée au CDER sise à Bouzereah-Alger, fruit d'une convention entre l'Agence Internationale de Coopération Espagnole (AECI) et le CDER, est un projet pilote dédié à la recherche. Comme l'indique le schéma fonctionnel de la figure IV.1, cette centrale est composée de quatre blocs fonctionnels :

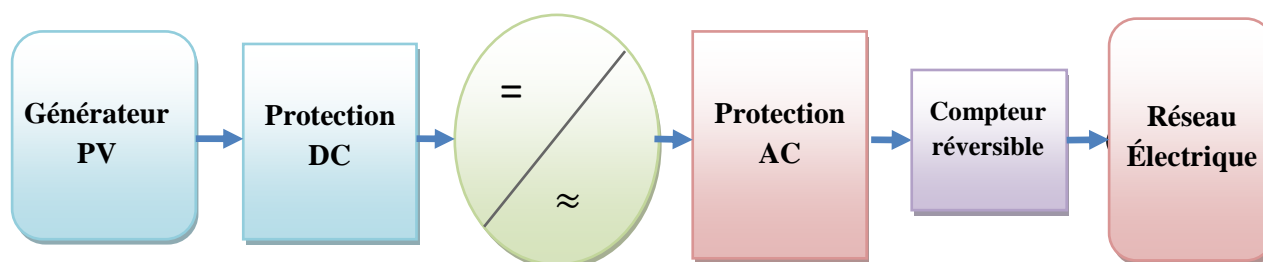


Figure IV.1. Schéma d'une Centrale Photovoltaïque Connectée au Réseau

Un champ photovoltaïque composé de 30 modules reliés équitablement sur deux branches parallèles et est connecté à un onduleur monophasé de 2,5 KW. Les caractéristiques de l'onduleur sont présentées dans le tableau 4 de l'annexe.

Chaque onduleur est branché sur une phase du réseau basse tension (220/380V~ & 50Hz) à travers un compteur réversible (consommation/production). Des circuits pour la protection de la centrale PV ainsi que le réseau sont nécessaires. La figure IV.2 présente le schéma générale de la CPR du CDER.

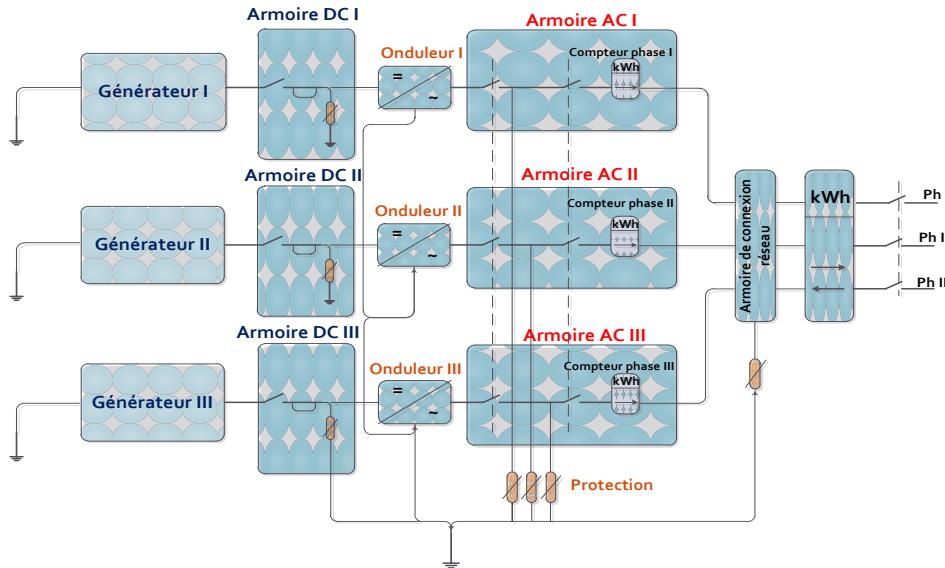


Figure IV.2. Schéma général de la centrale PV du CDER.

IV.3 Différents défauts considérés

Dans le présent travail, nous avons provoqués intentionnellement des défauts dans la centrale PV afin de vérifier notre approche. Les défauts considérés sont les suivants : string coupé (open string), court circuit des modules (1, 2 ou 3 modules sont court-circuités) et enfin le cas d'une défaillance totale (arrêt de production). Dans le schéma ci-dessous nous donnons une illustration graphique de ces défauts envisagé plus haut au sein de la centrale.

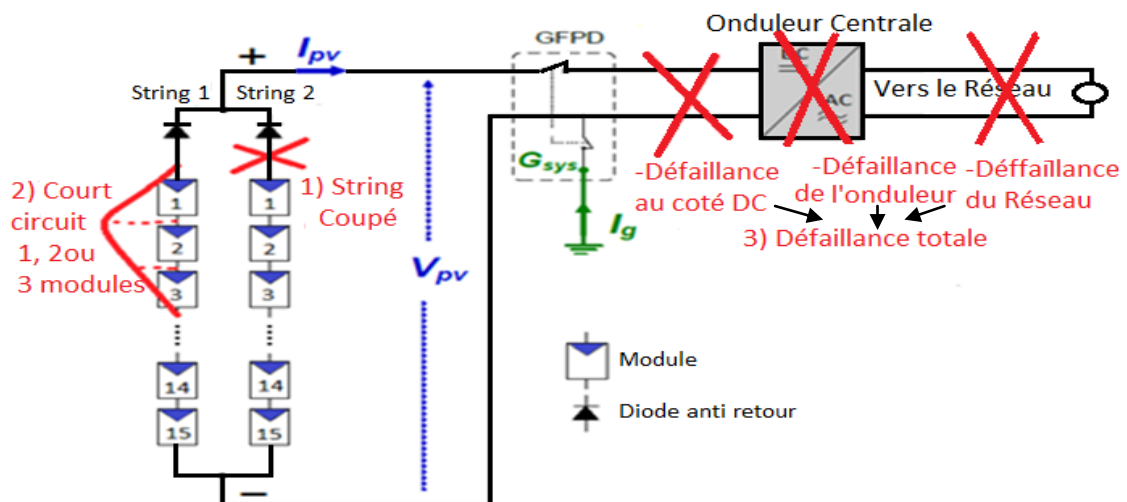


Figure IV.3. Les Différents types de défauts considérés dans le système PV.

- **String ouvert** : Ce défaut peut surgir lorsque les éléments de protection d'un string sont défectueux, un module du string est brisé ou un câble d'interconnexion entre modules est coupé (diode anti-retour, fusible, switch).
- **Modules court-circuités** : Ce défaut peut avoir lieu lorsqu'un court-circuit se produit entre les conducteurs des modules ou entre un module et le câble de terre.
- **Défaillance totale** : Ce défaut se produit lorsque l'onduleur s'arrête, soit à cause d'une erreur interne, défaillance totale de la partie DC ou à cause des problèmes liés au réseau.

IV.4 Etapes de mise en œuvre du modèle AD

La construction d'un modèle d'arbre de décision nécessite, comme toute méthode de diagnostic et de détection de défauts, la formation d'une base de données (base de connaissance) regroupant des observations des périodes de fonctionnement sain, ainsi que des observations en présence des différents défauts. Une phase de prétraitement est souvent nécessaire pour l'adaptation des données et le filtrage du bruit. La base de données est généralement divisée en deux ensembles ; un ensemble pour l'entraînement du modèle et l'autre pour réaliser le test. Ces étapes sont résumées dans l'organigramme de la figure IV.4.

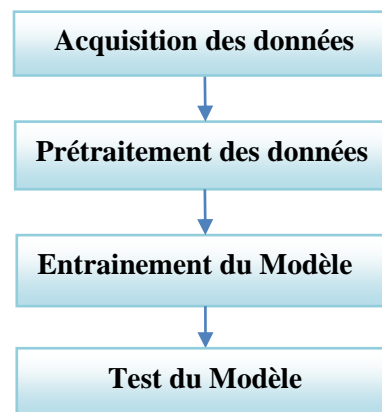


Figure IV.4. Le processus pour développer un modèle AD.

IV.4.1 Système d'acquisition des données

La surveillance d'une installation PV nécessite la réalisation d'un système permettant l'acquisition et le transfert de ses grandeurs les plus influentes tel que l'ensoleillement, la température ainsi que le courant et la tension générés. La Figure IV.5 présente les différents éléments du système d'acquisition.

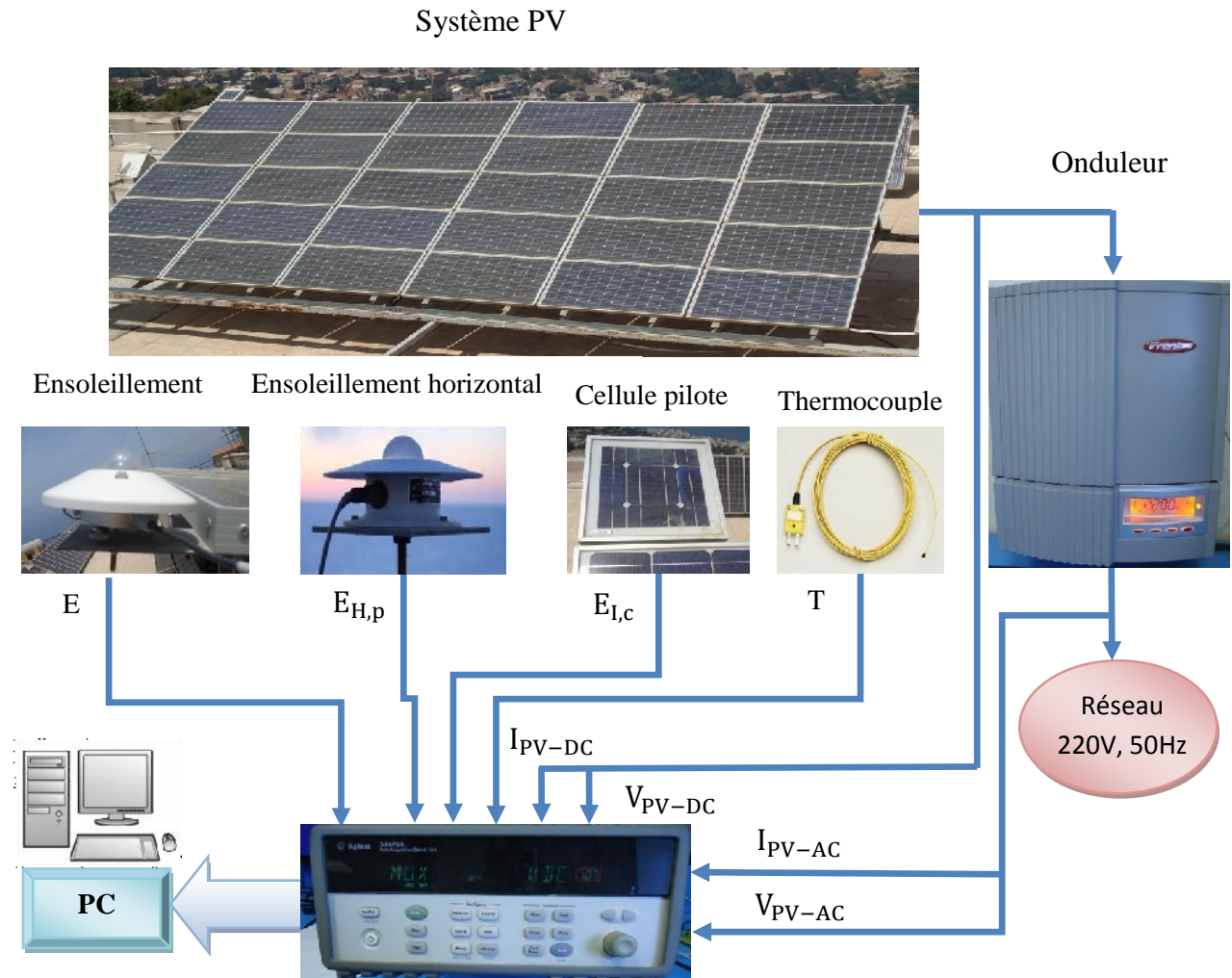


Figure IV.5. Synoptique globale du système d'acquisition.

IV.4.2 Choix des attributs de classification

Les graphes de la figure IV.6 montrent l'évolution de la puissance P_{pv} en fonction de la tension V_{pv} en fonctionnement normale et en présences de défauts pour différentes valeurs de température (T en $^{\circ}C$) et de l'ensoleillement (E en W/m^2).

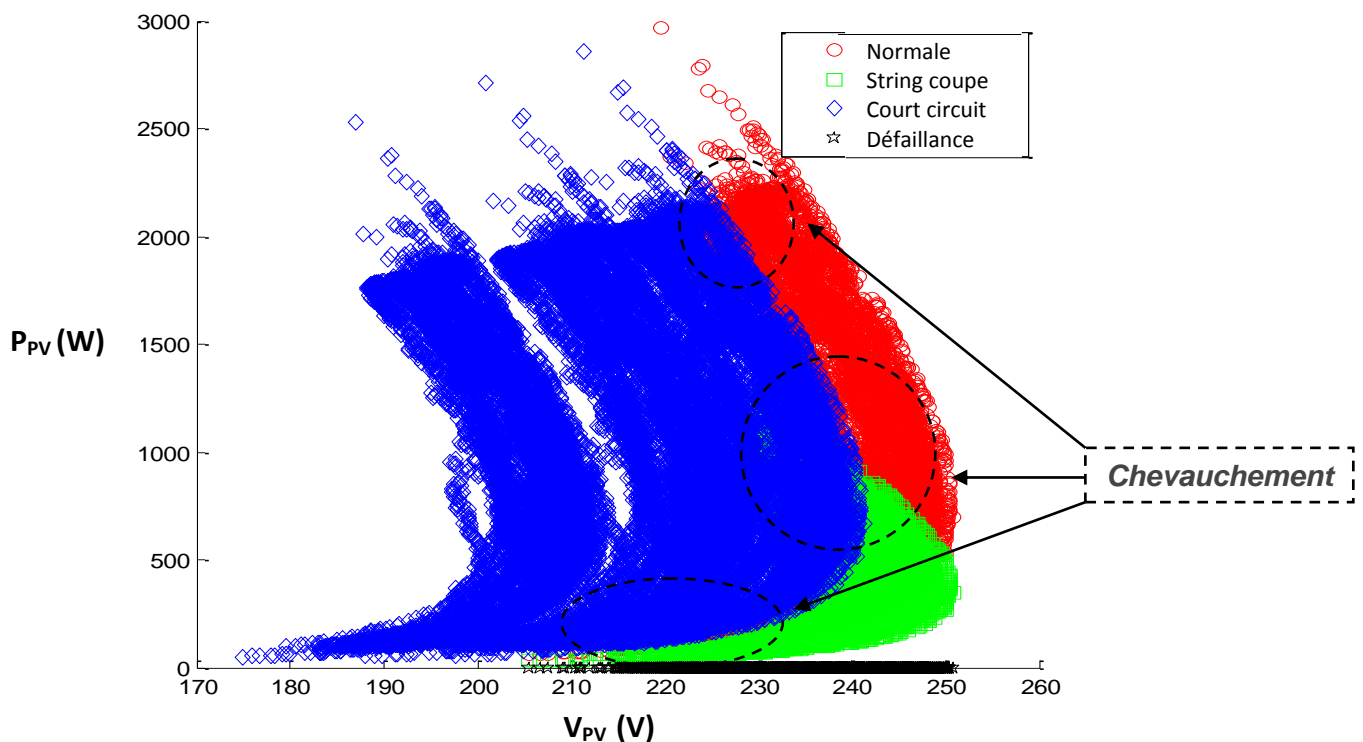


Figure IV.6. La caractéristique $P_{pv} - V_{pv}$ pour différentes valeurs de la température et de l'ensoleillement.

Nous constatons sur cette figure la présence de zones de chevauchement entre le cas d'un fonctionnement sain et celui d'un fonctionnement avec défauts. En utilisant uniquement les deux attributs P_{pv} et V_{pv} , il est donc difficile de séparer les deux classes de fonctionnement sain et avec défauts. Pour pouvoir séparer les différentes classes de fonctionnement, d'autres attributs tels que la température (T), l'ensoleillement (E) et le courant généré par le champ photovoltaïque devront être utilisés comme des attributs additionnels.

IV.4.3 Prétraitement de données

La préparation et le filtrage des données est une étape très importante. Les données sont souvent entachées de bruits et comportent diverses erreurs. En effet, dans notre cas, toutes les données correspondantes à un ensoleillement inférieur à 50 W/m^2 , qui pourraient prêter à des confusions, seront supprimées de la base de données. L'utilisation de ces données dans la phase d'apprentissage pourrait conduire à un modèle fortement biaisé et par conséquent à des conclusions erronées. Dans la figure IV.8 on a mis en évidence deux ensembles de valeurs qui doivent être supprimés de la base de données. Le premier ensemble correspond à un ensoleillement faible tandis que la seconde est une erreur de l'enregistrement. Pour filtrer les

données indésirables nous avons mis au point un programme MatLab dont l'organigramme est donné ci-dessous.

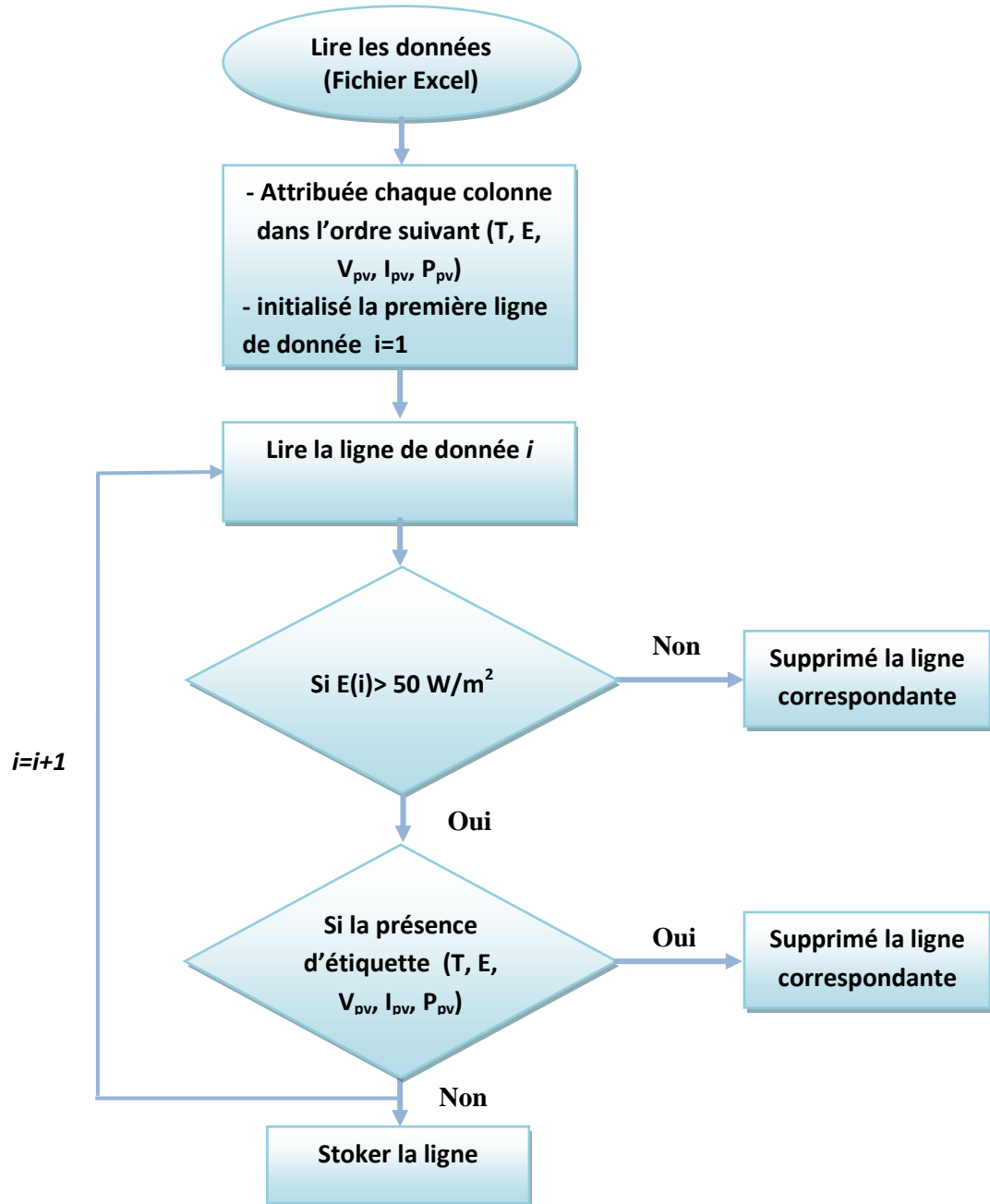


Figure IV.7. L'organigramme de Filtrage des données en provenance de l'acquisition.

A titre d'exemple, nous montrons dans la figure ci-après un fichier de données en provenance de l'acquisition comportant les données indésirables.

| | A | B | C | D | E |
|----|--------|-----------|------------|------------|------------|
| 1 | T | E | Vpv | Ipv | Ppv |
| 2 | 24,939 | 46,97702 | 223,908656 | 0,00484344 | 1,0844884 |
| 3 | 25,252 | 47,800022 | 223,84818 | 0,00733272 | 1,64141674 |
| 4 | 25,322 | 47,634094 | 224,159989 | 0,00281981 | 0,63208892 |
| 5 | 25,41 | 48,488591 | 224,444789 | 0,00212386 | 0,47668876 |
| 6 | 25,035 | 56,327832 | 226,346546 | 0,61326877 | 138,811268 |
| 7 | 24,977 | 57,369561 | 226,896398 | 0,62467432 | 141,736353 |
| 8 | T | E | Vpv | Ipv | Ppv |
| 9 | 25,22 | 0 | 227,584641 | 0 | 0 |
| 10 | 25,377 | 61,234372 | 227,898152 | 0,6667792 | 151,957748 |
| 11 | 25,622 | 62,676117 | 228,09088 | 0,68245024 | 155,660674 |
| 12 | 25,832 | 64,162126 | 228,335259 | 0,69861767 | 159,519047 |
| 13 | 26,163 | 65,534931 | 228,339502 | 0,7134981 | 162,919801 |
| 14 | 26,392 | 66,823499 | 228,453466 | 0,72749619 | 166,199027 |
| 15 | 26,413 | 68,028644 | 228,837908 | 0,74066718 | 169,492727 |
| 16 | 26,511 | 69,123154 | 229,059617 | 0,75259727 | 172,389643 |
| 17 | 26,952 | 70,101908 | 228,724454 | 0,7631147 | 174,542994 |
| 18 | 27,084 | 71,130023 | 228,859601 | 0,77430224 | 177,206503 |

Ces Valeurs vont être filtrées
($E \leq 50 \text{ W/m}^2$)

Ces lignes doivent être éliminées
(erreur d'enregistrement).

Figure IV.8. Exemple d'un fichier Excel contenant les données avant filtrage.

IV.4.4 Entraînement et Test du Modèle de l'Arbre de Décision(AD)

IV.4.4.1 Construction de l'AD avec WEKA

Le logiciel libre WEKA (Waikato Environment for Knowledge Analysis), dédié au Data Mining, est l'interface d'une librairie Java regroupant plusieurs algorithmes pour la fouille de données. WEKA est un logiciel libre. Parmi les algorithmes inclus dans ce logiciel on trouve les arbres de décision avec leurs différentes méthodes d'apprentissage automatiques.

- **Phase d'apprentissage**

Cette étape consiste à former un modèle d'AD à partir d'une base de données prétraitées de 506859 échantillons (ou exemples) englobant le fonctionnement sain et avec défauts. Une partie de ces données, choisies au hasard, est utilisée pour l'apprentissage du modèle. Le reste de données sera utilisées pour le test du modèle obtenu. La figure IV.9 montre une partie du modèle à AD obtenu en utilisant le logiciel WEKA.

```

Ppv <= 1928.088666
|
| E <= 87.406314
| | Ppv <= 179.514048: Default (32.0)
| | | Ppv > 179.514048
| | | | Vpv <= 217.835313: Default (8.0)
| | | | Vpv > 217.835313: Normal (8.0)
| | | E > 87.406314
| | | | Ppv <= 1859.687635: Default (2101.0)
| | | | Ppv > 1859.687635
| | | | | Ipv <= 8.665413
| | | | | | T <= 34.892: Default (4.0)
| | | | | | T > 34.892: Normal (4.0)
| | | | | | Ipv > 8.665413: Default (150.0)
| | Ppv > 1928.088666
| | | Vpv <= 218.23807: Default (167.0)
| | | Vpv > 218.23807: Normal (149.0)

```

Figure IV.9. Une partie du Modèle de l'AD pour la détection.

Les résultats de l'apprentissage et de test par le logiciel WEKA sont donnés dans les figures IV.10 et IV.11.

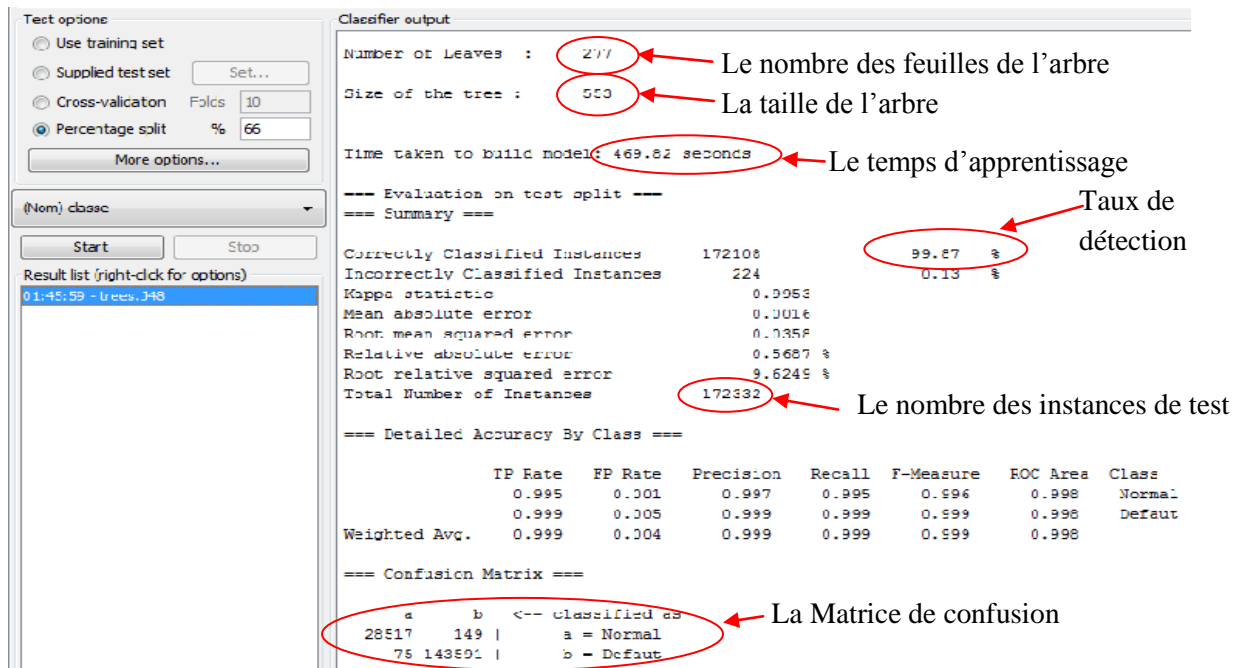


Figure IV.10 : L'affichage de l'apprentissage de l'AD pour la détection.

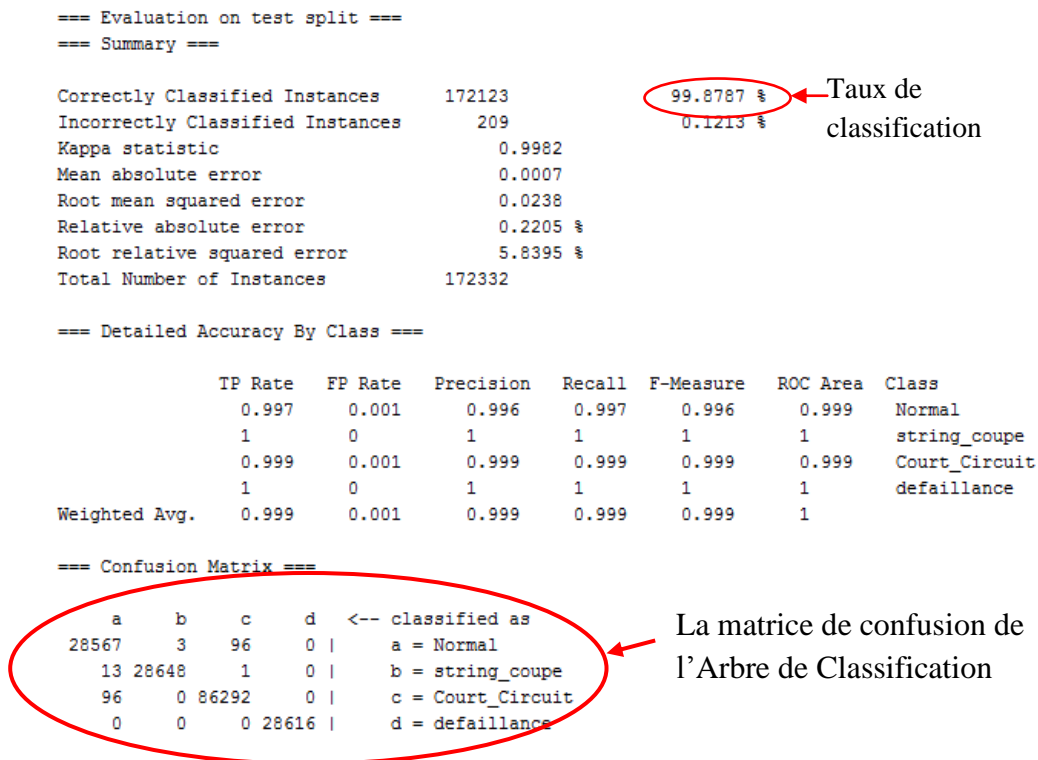


Figure IV.11. L'affichage de l'apprentissage de l'AD pour la classification.

- **Evaluation du Modèle**

Pour évaluer un arbre de décision, plusieurs bases de données de différentes tailles doivent être utilisées. Les tableaux IV.1 et IV.2 regroupent les résultats obtenus en fonction de la taille de la base de données utilisée pour les deux cas : détection et classification.

| Taille de base de données | Apprentissage | | |
|---------------------------|-------------------|----------------------|---------------------------|
| | Taille de l'arbre | Taux de détection(%) | Temps d'apprentissage (s) |
| 169230 | 255 | 79,97 | 51,04 |
| 233880 | 323 | 80,73 | 108,31 |
| 280602 | 389 | 83,01 | 140,97 |
| 433026 | 491 | 92,14 | 354,52 |
| 506859 | 553 | 99,87 | 469,82 |

Tableau IV.1.Résultats de détection

| Taille de base de données | Apprentissage | | |
|---------------------------|-------------------|----------------------------|---------------------------|
| | Taille de l'arbre | Taux de classification (%) | Temps d'apprentissage (s) |
| 169230 | 287 | 77,9521 | 48,05 |
| 233880 | 329 | 78,1905 | 101,33 |
| 280602 | 363 | 81,1681 | 151,38 |
| 433026 | 483 | 94,1299 | 384,6 |
| 506859 | 537 | 99,6299 | 522,21 |

Tableau IV.2. Résultats de classification.

Nous constatons que le meilleur taux de classification ou de détection est obtenu lorsque nous utilisons une base de données de taille importante. Cependant, une taille importante de la base de données correspond toujours à un arbre dont le nombre de nœuds est important. Nous pouvons conclure qu'un modèle performant à arbre de décision doit être construit à partir d'une base de données importante et c'est l'opération d'élagage qui permettra de réduire la taille de l'arbre de décision sans dégrader le taux de classification ou de détection.

- **Détection et classification par l'AD**

Pour maitre en œuvre l'approche de diagnostique des systèmes PV en utilisant les arbres de décisions, nous avons construit trois modèles à arbre de décision :

- Le premier modèle sert à la détection de défauts. En effet, si un défaut est détecté une alarme sera déclenchée et l'étape de classification du type de défaut détecté sera lancée.
- Le deuxième modèle sert à la classification des types de défauts détectés par le premier modèle. L'arbre de classification construite contient quatre classes (fonctionnement sain, chaine coupée, modules court-circuités, défaillance générale). Un rapport de classification indiquant la classe du défaut sera généré dans le cas d'une chaine coupée ou d'une défaillance générale.
- Si le défaut détecté appartient à la classe des modules court-circuités une autre classification pour déterminer le nombre de modules court-circuités (1, 2 ou 3 modules), doit être lancée. Cette classification est assurée par le troisième modèle. Un rapport indiquant le nombre de modules court-circuités sera généré.

L'organigramme de la figure IV.12 résume les différentes étapes de détection et de classification.

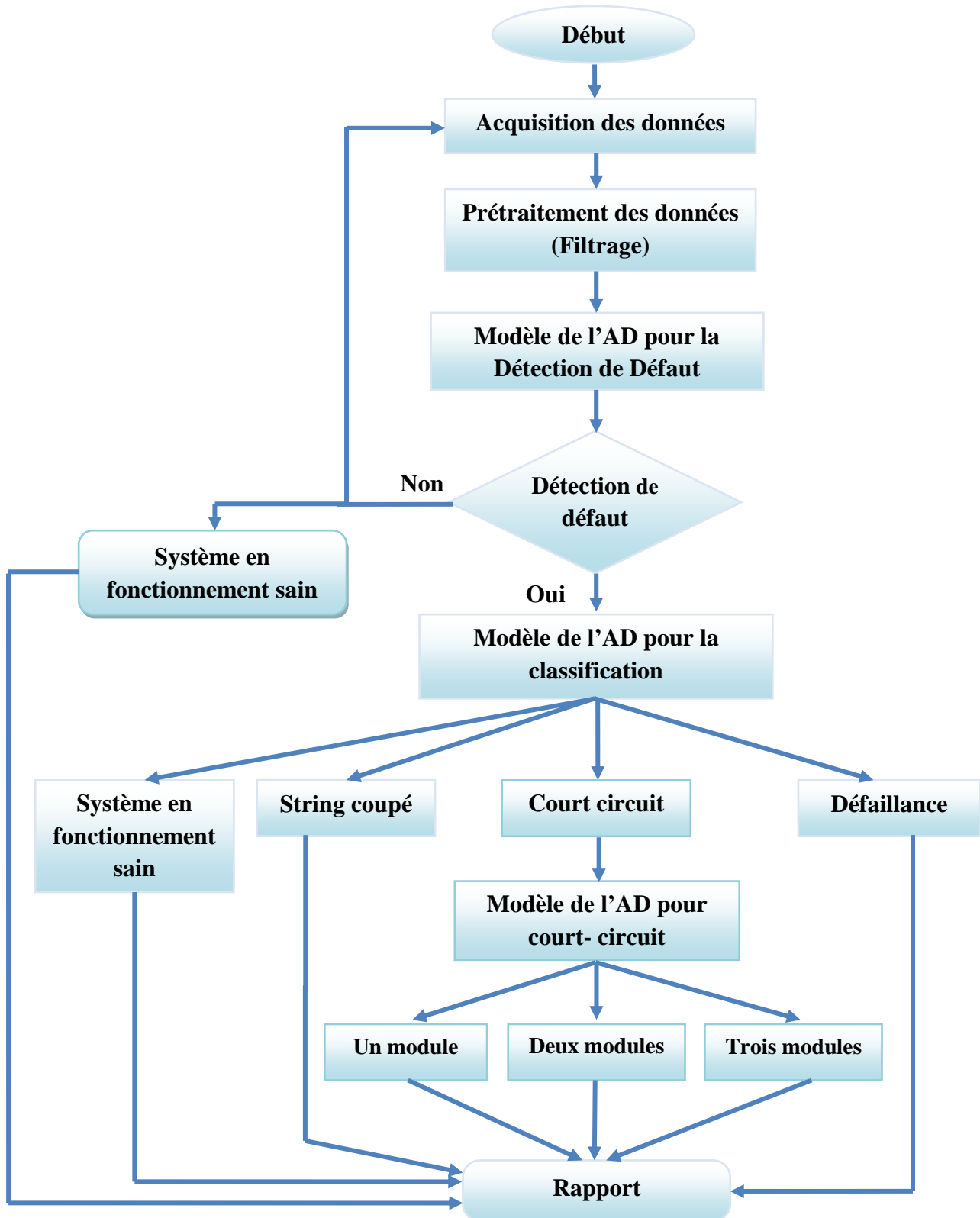


Figure IV.12. Organigramme de détection et de classification par l'AD.

IV.4.4.2 Implémentation sous l'environnement MatLab

Les trois modèles à arbre de décision créés sous l'environnement WEKA sont traduits en programmes MatLab en utilisant les instructions de type « if else ». Ce processus de conversion WEKA – Programme MatLab est illustré par la figure IV.13.

```

Ppv <= 1928.088666
|
| E <= 87.406314
| | Ppv<= 179.514048: Defaut (32.0)
| | Ppv > 179.514048
| | | Vpv <= 217.835313: Defaut (8.0)
| | | Vpv > 217.835313: Normal (8.0)
| | E > 87.406314
| | | Ppv <= 1859.687635: Defaut (2101.0)
| | | Ppv > 1859.687635
| | | | Ipv <= 8.665413
| | | | | T <= 34.892: Defaut (4.0)
| | | | | T > 34.892: Normal (4.0)
| | | | | Ipv > 8.665413: Defaut (150.0)
| | Ppv > 1928.088666
| | | Vpv <= 218.23807: Defaut (167.0)
| | | Vpv > 218.23807: Normal (149.0)

```

Transformer le code WEKA  au code MatLab

```

if Ppv(j) <= 1928.088666
    if E(j) <= 87.406314
        if Ppv(j) <= 179.514048
            fprintf('Defaut\n ');D=D+1;
        else
            if Vpv(j) <= 217.835313
                fprintf('Defaut\n ');D=D+1;
            else
                fprintf('Normal\n ');N=N+1;
            end
        end
    else
        if Ppv(j) <= 1859.687635
            fprintf('Defaut\n ');D=D+1;
        else
            if Ipv(j) <= 8.665413
                if T(j) <= 34.892
                    fprintf('Defaut\n ');D=D+1;
                else
                    fprintf('Normal\n ');N=N+1;
                end
            else
                fprintf('Defaut\n ');D=D+1;
            end
        end
    end
end
else
    if Vpv(j) <= 218.23807
        fprintf('Defaut\n ');D=D+1;
    else
        fprintf('Normal\n ');N=N+1;
    end
end
end

```

Figure IV.13. Conversion d'un Modèle d'Arbre en programme MatLab.

Les modèles que nous avons obtenus ont été testés en utilisant les données journalières mesurées de l'installation PV considérée. Les résultats de test ont été très satisfaisants. Toutes fois nous avons poussés notre investigation sur des données horaires afin de tester la validité de l'arbre de décision en cas de plusieurs états de fonctionnement de la centrale pendant une journée.

- **Etude de cas**

Dans les figures présenté ci-dessous nous montrons l'efficacité du modèle de l'arbre de décision développé à détecté et à classifié les défauts sur la base de données prise pour chaque heure (avec une période d'échantillonnage d'une minute)

Cas 1 : fonctionnement sain, la journée du 17/12/2012 de 11h32 jusqu'à 12h33

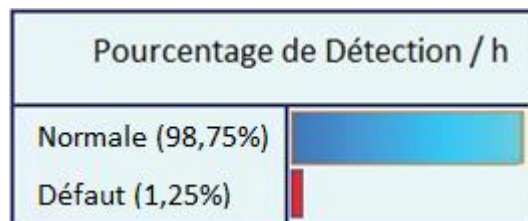


Figure IV.14. Résultats de détection.

Cas 2 : fonctionnement avec string coupé, la journée du 6/12/2012 de 14h04 jusqu'à 15h05

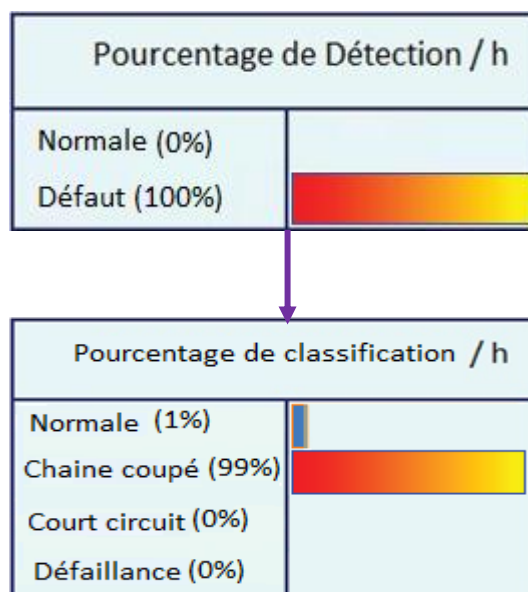


Figure IV.15. Résultats de détection et de classification (string coupé).

Cas 3 : fonctionnement avec Court circuit de deux modules, la journée du 22/12/2012 de 13h45 jusqu'à 14h46

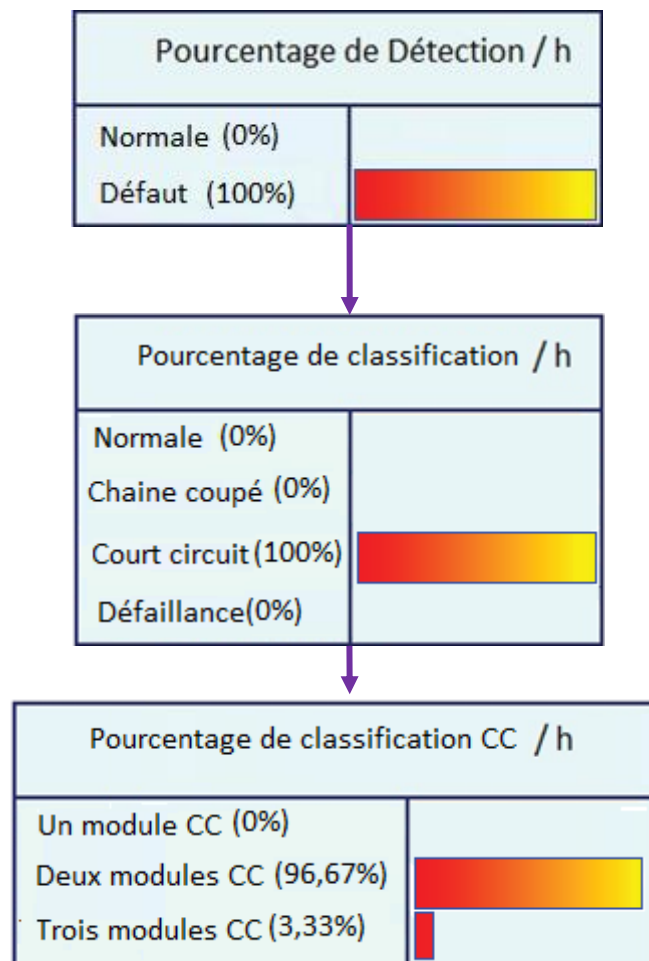


Figure IV.16. Résultats de détection et de classification (CC deux modules).

Cas 4 : défaillance totale, la journée du 1/12/2012 de 08h44 jusqu'à 09h45

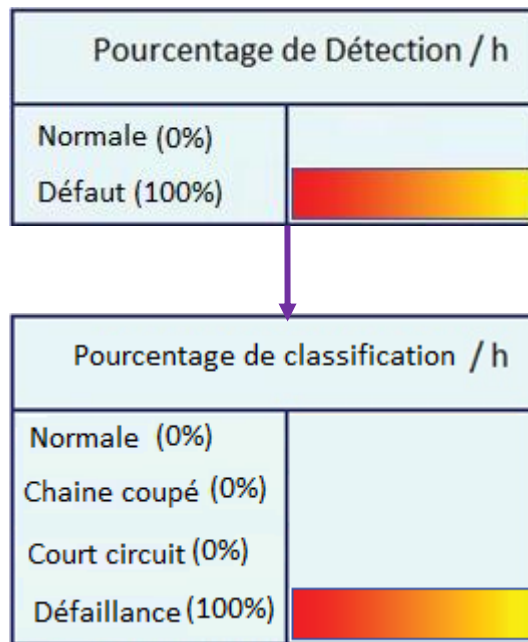


Figure IV.17. Résultats de détection et de classification (Défaillance).

IV.4.4 Implémentation des modèles de l'AD dans l'environnement LabVIEW

LabVIEW (Laboratory Virtual Instrument Engineering Workbench) est un environnement de programmation graphique qui utilise des icônes au lieu de lignes de texte pour créer des applications. Il permet la construction d'une interface utilisateur à l'aide d'un ensemble d'outils et d'objets. L'interface utilisateur d'un VI (Virtual Instrument) est appelée la face-avant. Ensuite, un code, pour commander les objets de la face-avant, est créé en utilisant les représentations graphiques de différentes fonctions. Ce code source graphique est aussi appelé code G ou code du diagramme [49].

Les possibilités de communication qui existent entre MatLab et LabVIEW nous ont permis de créer une interface pour faire appel aux sous programmes, écrits en MatLab, représentant les trois modèles à arbre de décision. L'image de la figure IV.19 illustre cette interface, mais avant ça on montre l'interface de récupération des données dans la figure IV.18 ci-dessous

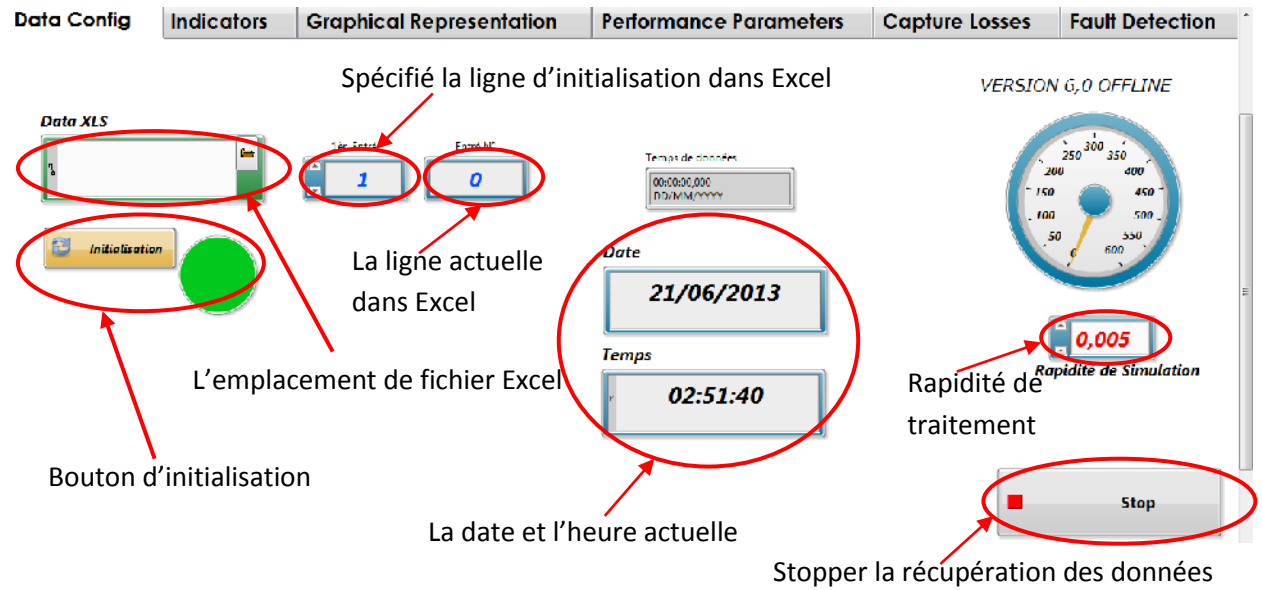


Figure IV.18. Interface pour récupération des données

Après la récupération des données on passe à l'interface utilisateur.

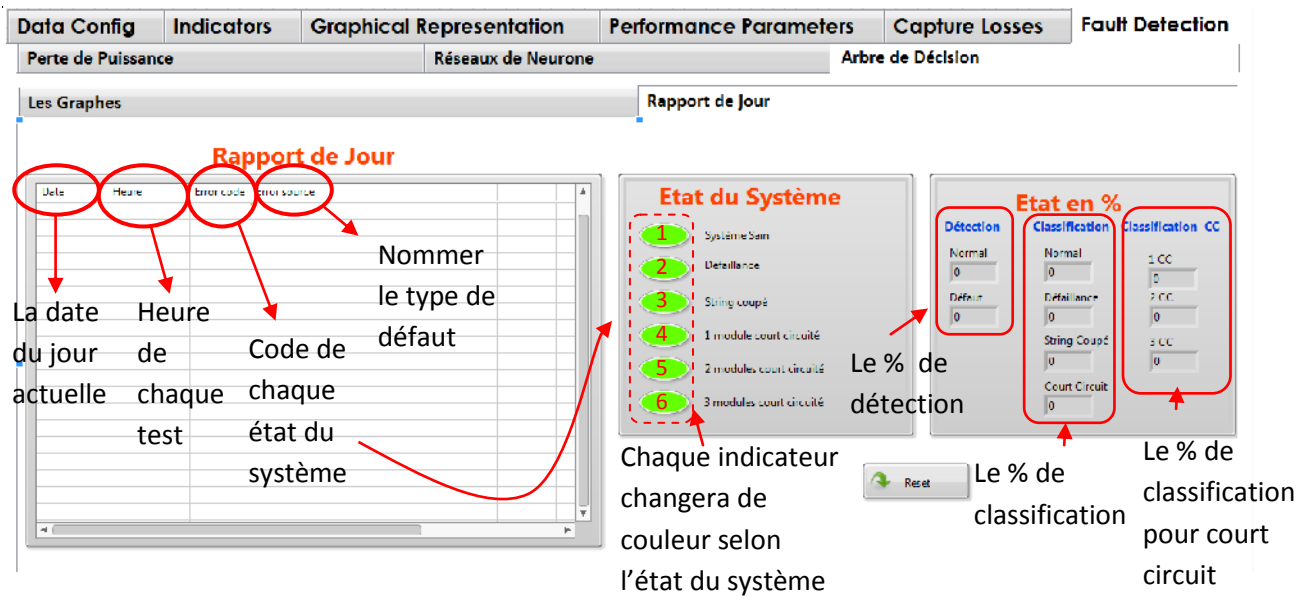


Figure IV.19. Interface utilisateur LabView

L'interface utilisateur est constituée des trois fenêtres suivantes :

- La première fenêtre affiche un rapport journalier sur l'état de l'installation indiquant les différentes classes pour chaque heure où l'ensoleillement > 50 W/m².
- La deuxième fenêtre est identique à la première. L'état de l'installation est indiqué en allumant la lampe correspondante.

- Dans la troisième fenêtre l'état de l'installation est indiqué par le taux en (%) de chaque classe.

IV.4.4.1 Test avec des données réelles

➤ Système Sain

Les résultats du teste effectué sur les données du système en fonctionnement sain récupérées le 17/12/2012 sont présentés dans la figure IV.20 ci-dessous

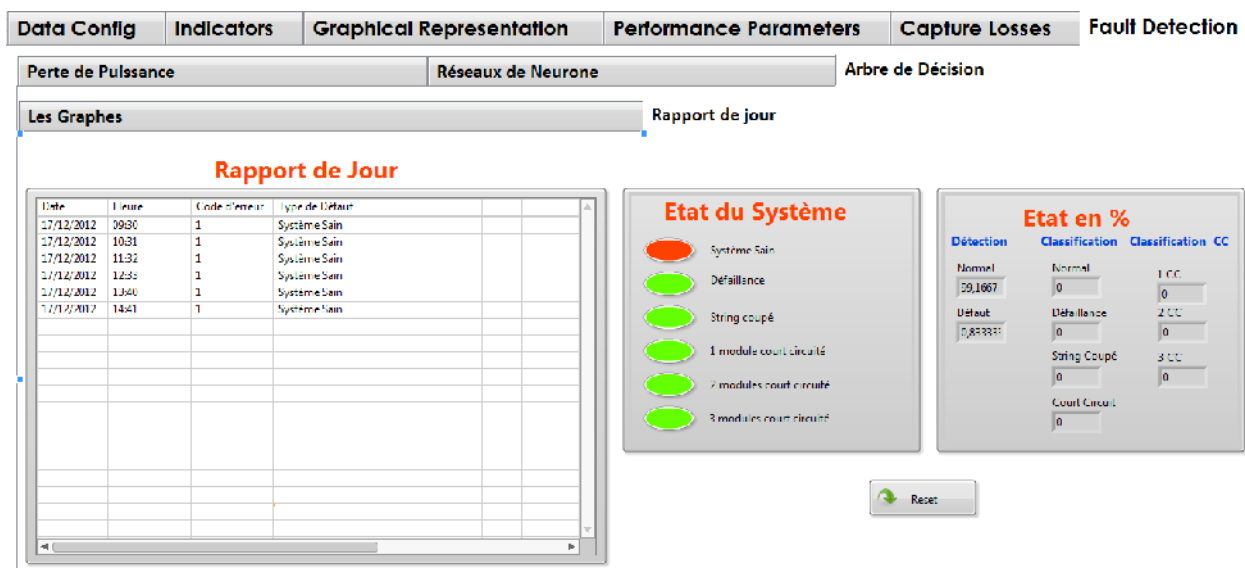


Figure IV.20. Résultats de teste pour un système sain

D'après l'affichage ci-dessus le système de diagnostic a parfaitement identifié l'état sain chaque heure.

➤ String coupé

Les résultats du teste effectué sur des données en présence du défaut «String coupé» sont indiqués par la figure IV.21 suivante :

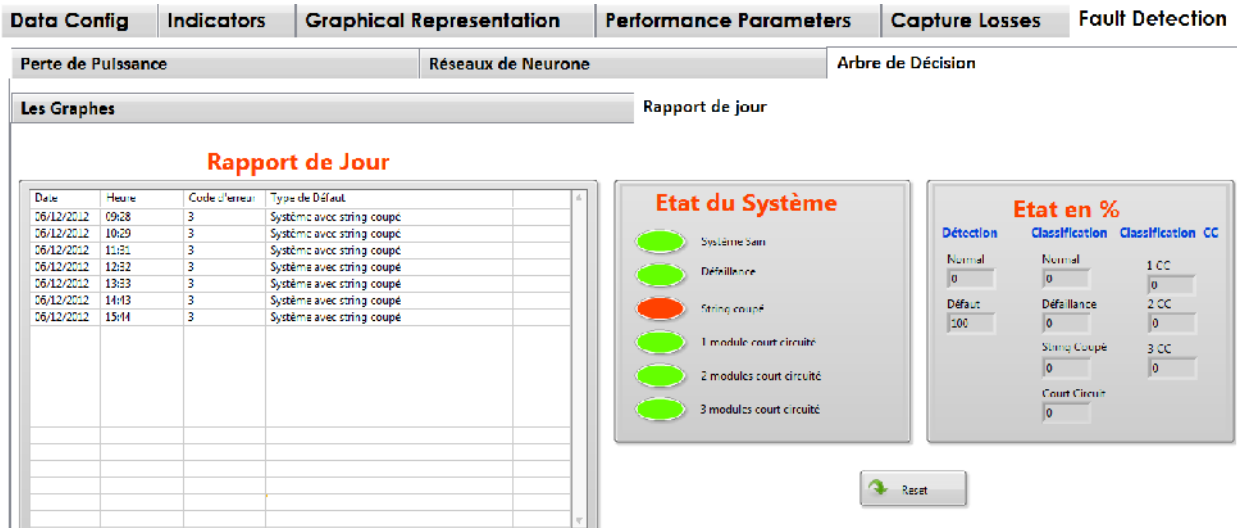


Figure IV.21. Résultats de teste dans le cas d'un string coupé.

➤ **Court-circuit de deux modules**

Les résultats obtenus dans le cas d'un court-circuit de deux modules sont données par la figure IV.22.

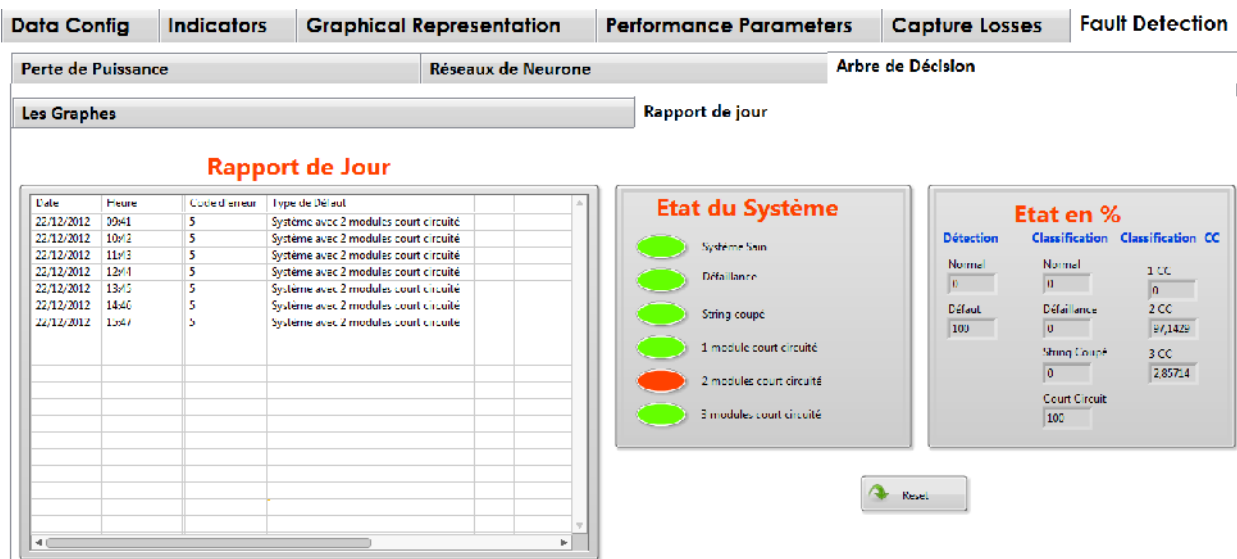


Figure IV.22. Résultats de teste dans le cas d'un court-circuit de 2 modules

IV.9 Conclusion

Dans ce chapitre, après avoir présenté une description générale de la centrale photovoltaïque sur laquelle notre travail de détection et classification de défauts c'est porté, nous avons donné toutes les étapes de la mise en œuvre du modèle à arbre de décision ainsi que les différents résultats de test obtenus. La mise au point des différents modèles développés a été réalisée en utilisant trois logiciels différents (WEKA, MatLab, Labview). L'interface graphique réalisée sous l'environnement Labview permet la surveillance et le traitement de la situation de l'installation PV d'une manière interactive et très conviviale.

Les résultats de test obtenus nous montrent que l'obtention de bons modèles qui permettraient de prendre la bonne décision nécessitent l'utilisation de bases de données de taille importante et couvrant toutes les modes de fonctionnement de l'installation photovoltaïque.

Conclusion générale

L'objectif qui a été donné au travail que nous avons réalisé, dans le cadre ce projet, est le développement d'une méthode, basée sur la classification par arbres de décision, pour le diagnostique et la détection de défauts dans les systèmes photovoltaïque. La mise en œuvre de ce travail n'était pas possible sans passer par l'étude des systèmes photovoltaïques et des différentes méthodes de construction d'un arbre de décision. Bien que la construction des arbres de décisions à partir des données est une discipline déjà ancienne, il y a peu de travaux qui ont utilisé cette approche pour le diagnostique et la détection de défauts dans les systèmes photovoltaïques.

Dans ce mémoire, après avoir donné les notions de base relatives aux systèmes photovoltaïques et le diagnostique et la détection de défauts, nous avons détaillé les points clés pour la construction d'un arbre de décision à partir de données. Ensuite, nous avons développé, en construisant plusieurs arbres de décision, un système pour le diagnostique et la détection de défauts dans une installation photovoltaïque. Pour la construction des différents arbres, nous avons utilisé la méthode de classification C4.5 implémentée dans un logiciel libre et téléchargeable sur internet, et un ensemble de données issues de l'installation photovoltaïque considérée. Cet ensemble de données, couvrant le fonctionnement sain et avec défauts de l'installation, contient des valeurs pouvant conduire un modèle biaisé. Ces valeurs, en particulier celles relatives à un ensoleillement faible, ne doivent pas être utilisées dans la phase de construction des arbres. Une opération de prétraitement est donc, nécessaire pour supprimer toutes les valeurs indésirables de la base de données ainsi que tout bruit qui pourrait être superposé à ces données.

Après avoir testé et validé les modèles à arbre de décision que nous avons construit en utilisant le logiciel WEKA, nous les avons codés par un programme écrit en MatLab. Ceci, nous a permis par la suite de développer, en utilisant le logiciel Labview, une interface graphique permettant d'appliquer les arbres de décision construits sur les données recueillies et d'afficher le résultat sous différentes formats.

Afin de conclure sur l'efficacité et la précision du système de diagnostique et de détection de défauts que nous avons développé, nous avons effectué plusieurs tests, en considérant des cas de défauts différents, sur des données réels récupérées sur une journée. Les différents tests ont été effectués chaque heure. Les résultats obtenus sont satisfaisants et encourageant et un taux de 99% de classification et de détection à été réalisé.

Le projet que nous avons réalisé a porté sur un sujet d'actualité et nous a permis d'envisager des domaines très prometteurs à savoir le diagnostic des systèmes d'une façon générale et plus particulièrement les systèmes photovoltaïques, et la classification des données en utilisant les arbres de décision. Ceci, pourrait être très bénéfique pour nous dans l'orientation de nos futurs travaux.

Le travail que nous avons réalisé pourra faire l'objet de plusieurs améliorations dans le futur. Nous pouvons citer, entre autres, la considération de d'autres attributs et l'inclusion dans la base de données de d'autres types de défauts.



Les défauts dus à des erreurs internes de l'onduleur sont les plus courantes et ils augmentent avec le vieillissement du composant, ces défauts peuvent être liés avec tous les composants internes de l'onduleur, comme les commutateurs, le ventilateur, le MPPT ou les varistances (en fait une cause fréquente de l'échec de l'onduleur est un condensateur électrolytique défectueux)

Bien que le côté DC est celui qui présente la plus faible quantité d'échecs, il est le plus compliqué à analyser et à détecter l'origine du défaut en raison du nombre de composants inclus et des caractéristiques de sortie non linéaires des panneaux PV, une variété de défauts peut être difficile à distinguer, pour cela une division doit être faite entre les défauts qui amènent à un arrêt complet de toute une string ou bien l'installation complète et celui qui déterminent seulement une diminution de la puissance produite.

Les facteurs qui peuvent amener à une diminution de la puissance produite sont principalement la dégradation ou la rupture d'une cellule ou un module et / ou les phénomènes d'ombrage. Les phénomènes d'ombrage (*shading*) peuvent être liés à la présence d'obstacles qui ne permettent pas aux rayonnements solaire d'atteindre le panneau. Un autre facteur qui ramène à une diminution de la puissance qui peut être une surface sale des panneaux.

Il est possible d'identifier les défauts suivants comme les plus courants :

- **Erreur interne de l'onduleur**

L'onduleur est l'élément central du système PV et représente également le plus faible élément à cause d'erreur interne qui peut avoir. Tandis que l'onduleur a toujours été l'élément qui présente le plus grand nombre de défauts, il est aussi l'élément qui a été beaucoup amélioré au cours de ces dernières années. En regardant les données statistiques, il est évident que la probabilité de défauts liés à l'onduleur subit une baisse significative.

- **Les problèmes du réseau**

La probabilité de ces défauts est fortement dépendante à l'emplacement et par conséquent la fiabilité du réseau considéré.

- **Détérioration des modules**

En raison du vieillissement du module PV un certain changement dans la puissance produite par les modules est prévu au cours des années, bien qu'il soit important de vérifier si la baisse n'est pas supérieure à celle indiquée sur la spécification du module. Le rayonnement ultraviolet du soleil provoque un vieillissement et une légère dégradation. Cependant, le test des modules sur 25 ans a montré que même les modules avec un signe

visible de vieillissement génèrent en moyenne 75% de leur puissance originale .Un groupe de recherche TÜV Rheinland a constaté que la dégradation annuelle moyenne est inférieure à 0,5%. En général, la dégradation à long terme pour les modules silicium cristallin est faible, environ 0,2% par ans [20].

- **ombrage périodique**

En raison de la présence d'obstacles générant un ombrage périodique qu'on ne peut considérer comme un propre défaut, il est important de considérer que les modules touchés par un ombrage partiel montrent une détérioration plus élevée en raison de la présence périodique de hot spot dans des conditions normale de fonctionnement [21].



Figure II.1. Ombrage des Modules.

Le reste de défauts et anomalies rencontrés dans un générateur PV sont représentés .dans le tableau ci dessous :

| Eléments du générateur PV | Origines des défauts et d'anomalies |
|---------------------------|---|
| Générateur PV | <ul style="list-style-type: none"> -Feuilles d'arbres, déjections, pollution, sable, neige etc. -Détérioration des cellules, fissure, échauffement des cellules; -Pénétration de l'humidité, dégradation des interconnexions, corrosion entre les cellules; -Modules de performances différentes; -Module arraché ou cassé; -Modules court-circuités, modules inversés. |

| | |
|--|---|
| Boîte de jonction | <ul style="list-style-type: none"> -Rupture de circuit électrique -Court-circuit du circuit électrique -Destruction de la liaison -Corrosion des connections |
| Câblage et connecteur | <ul style="list-style-type: none"> -Circuit ouvert -Court-circuit -Mauvais câblage (module inversé) -Corrosion des contacts -Rupture du circuit électrique |
| Diode de protection (diode de bypass et diode anti-retour) | <ul style="list-style-type: none"> -Destruction des diodes -Absence ou non fonctionnement de diodes -Inversion de la polarité des diodes au montage, diode mal connectée |

Tableau II.1. Principaux défauts et anomalies rencontrés dans un générateur PV.

II.3 Méthodes de diagnostic d'une installation PV

Le Contrôle de performances régulières sur le fonctionnement des systèmes PV connectés au réseau sont nécessaires pour assurer une production d'énergie fiable et optimale. L'objectif est la détection précoce de défaut dans les systèmes PV et changer les conditions de fonctionnement pour empêcher les pertes d'énergie et financières subséquentes pour l'opérateur. Ainsi, un système simple et sûr à manipuler à long terme pour vérifier les performances et l'état du système PV est nécessaire pour garantir la sécurité de ce dernier et réduire les pertes. Le but des systèmes de diagnostics est non seulement la détection des dysfonctionnements, mais aussi la mise à disposition des informations sur la source de défaillance la plus probable.

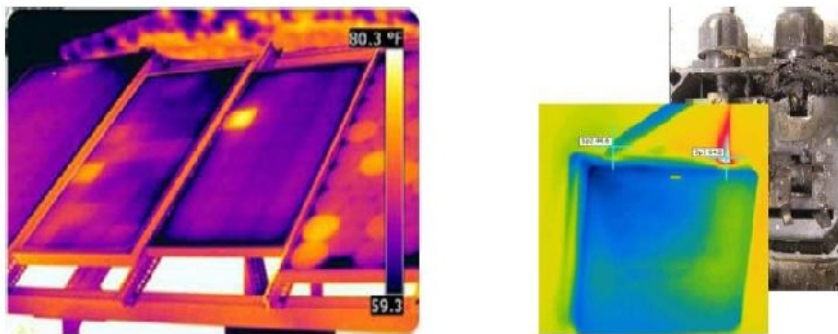
II.3.1 Méthodes de diagnostic courantes (industrialisé)

On peut distinguer deux catégories de méthodes de diagnostic courantes : des méthodes reposant sur l'analyse du courant et de la tension (que nous appellerons méthodes électriques) et des méthodes reposant sur l'analyse d'autres grandeurs que I et V (que nous appellerons méthodes non-électriques).

II.3.1.1 Méthodes non-électriques

Il existe plusieurs méthodes non-électriques, destructives ou non destructives, pour diagnostiquer le défaut au niveau de cellule PV. Le défaut principal qui peut avoir lieu à ce niveau est la fissure de la cellule. On peut citer comme méthodes: les essais mécaniques de flexion, l'imagerie par photoluminescence et électroluminescence, tests de thermographie [22], [23].

Au niveau du module PV, la méthode de l'imagerie à infrarouge (caméra thermique) est largement appliquée. Cette méthode repose sur le fait que tous les matériaux émettent un rayonnement infrarouge sur une plage de longueur d'onde qui dépend de la température du matériau. En examinant la distribution de la température au niveau du module, des anomalies (si elles ont lieu) peuvent être localisées (Voir la Figure II.2).



(a) Cellule court-circuitée. (b) Corrosion dans la boîte de jonction.

Figure II.2. Quelques exemples de la localisation de défauts par la caméra thermique.

II.3.1.2 Méthodes électriques

Les systèmes de monitoring actuels d'un GPV sont essentiellement intégrés aux onduleurs. Dans ce cadre, les données mesurées sont souvent les mêmes d'un système à l'autre.

Les grandeurs mesurées les plus courantes sont :

- Le courant et la tension coté DC du champ PV.
- Le courant et la tension coté AC du champ PV.

Il est aussi possible d'ajouter les grandeurs complémentaires que sont la température ambiante du site et l'ensoleillement aux mesures électriques. Ces données nécessitent un capteur spécifique. Les informations peuvent être traitées localement ou à distance, sur une ou plusieurs installations.

II.3.2 Méthodes de diagnostic proposées dans la littérature

De nombreuses méthodes de diagnostic ont été proposées pour détecter les défauts dans un système PV. Nous résumons dans cette partie les différentes méthodes proposées.

II.3.2.1 Méthode de réflectométrie

La méthode de réflectométrie est une méthode de diagnostic qui consiste à envoyer un signal dans le système ou le milieu à diagnostiquer. Ce signal se propage selon la loi de propagation du milieu étudié et lorsqu'il rencontre une discontinuité, une partie de son énergie est renvoyée vers le point d'injection. L'analyse du signal réfléchi permet de déduire des informations sur le système ou le milieu considéré. Cette méthode a été également appliquée pour détecter le défaut dans un string photovoltaïque [24], [25].

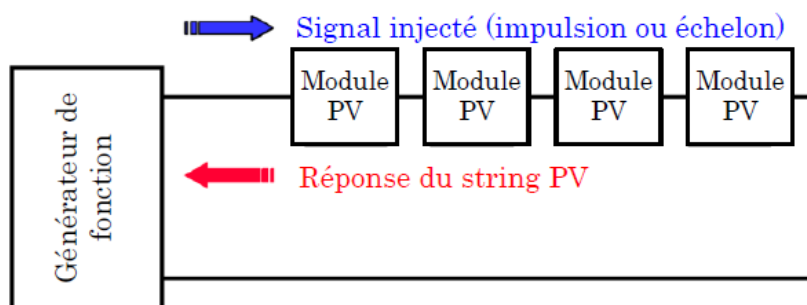


Figure II.3. Principe de la réflectométrie pour localiser le défaut dans un string PV.

L'étude expérimentale a confirmé la possibilité de localiser la position de défaut de type « circuit ouvert », « court-circuit » et « augmentation d'impédance » dans le string.

II.3.2.2 Analyse de la puissance et de l'énergie produite

Les systèmes de monitoring montrent graphiquement et compare les valeurs de la puissance réelle et simulée pour retourner des paramètres de performance et des pertes du système PV. La puissance simulée est calculée grâce à un modèle de simulation qui estime la puissance attendue comme étant la puissance que devrait produire l'installation PV dans certaines conditions météorologique. Une fois que la puissance estimée est calculée, elle doit être comparée avec la puissance réelle (mesuré). La différence entre ces deux valeurs représente les pertes dans le système.

➤ **Détection de défaut**

L'étape suivante consiste à distinguer, si la valeur de cette différence (entre la puissance réelle et simulée) représente un dysfonctionnement donc la présence d'un défaut. Lorsque la différence dépasse un seuil, qui peut être différent selon les différents modèles de simulation choisis et à leur précision, on suppose une présence de défauts.

➤ **Diagnostic de défaut**

La dernière étape est l'élément le plus avancé dans le processus est la capacité de distinguer entre les différents défauts possibles. Il existe principalement deux approches différentes qui ont été examinées jusqu'à présent : une méthode analytique et l'autre empirique. La première est basée sur les valeurs des paramètres qui décrivent le comportement de la puissance d'une installation PV telle que : la durée, l'amplitude, la fréquence et les instants de la chute. Ces mêmes attributs sont également prédéterminés pour les différents défauts considérés. Lors de leur comparaison, le défaut dont la valeur des attributs considérés est la plus proche de celle déduite des grandeurs mesurées est considéré comme le défaut responsable de la chute. La seconde est appelée empirique car elle est basée uniquement sur les résultats des tests sans regarder l'influence physique du défaut spécifique sur le système, ici l'identification du type de défaut se fait grâce à des analyses statistiques des caractéristiques électriques en présence de divers défauts.

Plusieurs études reposent sur l'analyse de la puissance et de l'énergie produite par le champ PV [26], [27]. Outre la comparaison de la puissance ou de l'énergie produite actuelle et celle attendue, la comparaison du point de la puissance maximale actuel (courant et tension correspondant à la puissance maximale) et celui attendu peut apporter plus d'information sur l'état du système PV [28].

II.3.2.3 Autres méthodes pour le diagnostic

Une étude trouvée dans [29], où une méthode basée sur la fonction de corrélation étendue et matter-element model est proposé pour identifier les défauts dans une petite installation PV. En outre, dans [30] matter-element model est combiné avec un réseau de neurone pour construire un système de diagnostic de défaut intelligent. Les deux propositions utilisent un simulateur de

système PV pour collecter les données de la puissance produite par les modules PV en fonctionnement normal et défectueux.

Dans [31] une méthodologie proposée qui permet de détecter les défauts dans un système PV, en comparant l'énergie de sortie de six sous-systèmes qui forment l'installation PV. La méthode discutée dans [28], où ils ont proposé un monitoring automatique et un système de détection de défaut basé sur l'analyse des pertes de puissance, ces dernières sont calculées en comparant les données de monitoring avec les résultats de simulation, ensuite les indicateurs de pertes sont traités pour générer un signal défectueux [32].

Finalement le projet de recherche universitaire le plus important dans le domaine de détection automatique des pannes qui a été appliqué dans un produit commercial est le projet PVSAT2. Ce projet a été mené pendant trois ans, il est le fruit d'une collaboration entre les différentes universités et entreprises. L'analyse est basée sur des données satellitaires parce que le projet vise à surveiller les conditions de fonctionnement au niveau du des surfaces ou on trouve les panneaux PV comme illustré dans la Figure II.4. Dans ce cas, les données provenant des capteurs locaux ne sont pas souvent disponibles

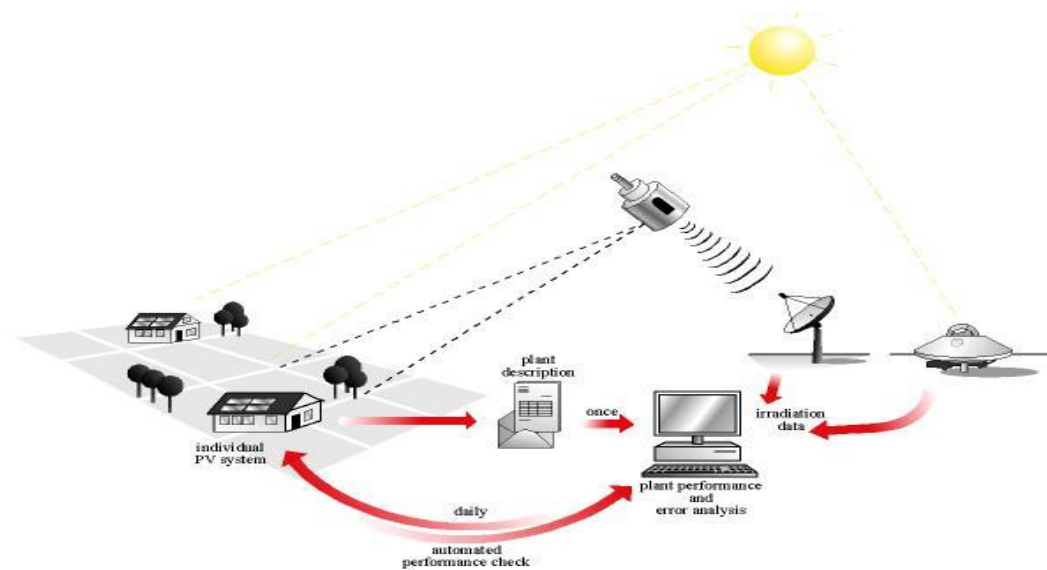


Figure II.4. Un aperçu sur PVSAT2.

Tous d'abord, les algorithmes et les procédures pour estimer les conditions météorologiques sur place (sur le site) à partir de données satellite ont été développés et améliorés. Ensuite, on associe un modèle électrique à l'installation PV réelle. Enfin, pour chaque défaut une série de paramètres est définie et la caractérisation du défaut est créée. A l'instant où un décalage entre la puissance estimée et réelle apparaît, une erreur est détectée. Ainsi, pour la détection des pannes de système PV, la méthode de « *profiling* » de défaut est l'une des approches appliquées. Elle

étudie les propriétés de la perte d'énergie comme la quantité, la durée, le changement, et la corrélation avec les systèmes PV voisins. Les résultats de cette analyse sont utilisés pour créer un profil de défaut en énumérant les attributs de la défaillance qui s'est produit [33].

La deuxième approche dans la détection de défaut est la méthode *footprint* (FDR), la méthode « *FDR* » vise également à l'identification des pertes mineures d'énergie dues à l'ombrage, les pertes de puissance permanentes et les problèmes d'onduleur. Elle analyse les modèles de la dépendance pour trois domaines différents : puissance normalisée surveillé, l'heure de fonctionnement, et la distance du soleil, par exemple, l'influence de l'ombrage sur la production d'électricité est plus élevée à basse altitude [33].

II.4 Conclusion

Dans ce chapitre on a présenté les différents défauts qui peuvent subir un système PV, ainsi que les méthodes de diagnostic de défauts proposées industrialisées ou non. Ces méthodes montre une limitation de discrimination entre les défauts, pour cela les méthodes basés sur l'exploitation des données est identifiées comme la plus prometteuse [34].



Panneaux photovoltaïques monocristallins

Isofoton I-106 12/24V

Réf. M180106

| CARACTERISTIQUES PHYSIQUES | I-106/12 | I-106/24 |
|--|-------------------|----------|
| Dimensions (mm) | 1310 x 654 x 39,5 | |
| Poids (kg) | | 11,5 |
| Nombre de cellules en série | 36 | 72 |
| Nombre de cellules en parallèle | 2 | 1 |
| TONC (800 W/m ² , 20 °C, AM 1.5, 1 m/s) | | 47 °C |



| CARACTERISTIQUES ELECTRIQUES (1 000 W/m ² , 25 °C, AM 1.5) | I-106/12 | I-106/24 |
|--|----------|----------|
| Tension nominale (Vn) | 12 V | 24 V |
| Puissance crête (± 5%) (Pmax) | | 106 W |
| Courant de court-circuit (Icc) | 6,54 A | 3,27 A |
| Tension de circuit ouvert (Vco) | 21,6 V | 43,2 V |
| Courant à puissance max. (Imax) | 6,1 A | 3,05 A |
| Tension à puissance max. (Vmax) | 17,4 V | 34,8 V |

CARACTERISTIQUES DE FABRICATION

| | |
|--------------------|---|
| Cellules | Si monocristallins, texturées et avec une couche anti-réflexion |
| Contacts | Redondants, multiples, à chaque cellule |
| Lamé | EVA (éthylène-vinyle acétate) |
| Face frontale | Verre trempé de grande transmission |
| Face postérieure | Tedlar à plusieurs couches |
| Cadre | Aluminium anodisé |
| Boîte de connexion | 2 x IP65 avec diodes anti-retour |
| Prise de terre | Oui |
| Certifications | IEC 61215 / Classe II |
| Section de câble | 4-10 mm ² |

Terminal de connexion Bornes avec possibilité de soudure, multi contact en option



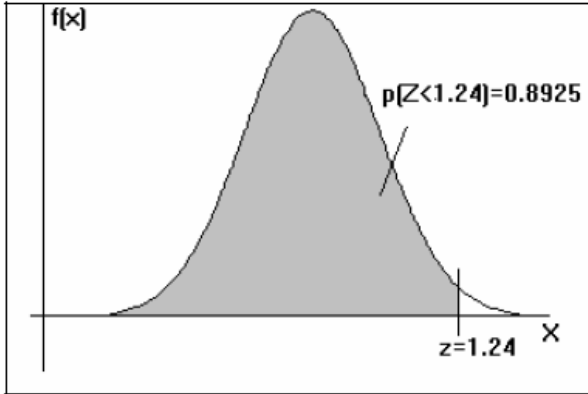
Sous réserve de modifications techniques

Systèmes solaires Photovoltaïques, assistance technique en mesure et essai, moyens de test.

ZA du Pré de l'Orme BP 14 38761 VARCES cedex SARL au capital de 7 500 Euros APE 333 Z
Tél.: 04 76 98 83 57 Fax: 04 76 98 79 04 Siret 444 688 774 00025 RM 381 ; E-Mail : commercial@surtec.fr

TABLE DE LA LOI NORMALE CENTREE REDUITE

Lecture de la table: Pour $z=1.24$ (intersection de la ligne 1.2 et de la colonne 0.04), on a la proportion $P(Z < 1,24) = 0.8925$



$P(Z > 1,96) = 0,025$
 $P(Z > 2,58) = 0,005$
 $P(Z > 3,29) = 0,0005$

Rappels:
 $1/ P(Z > z) = 1 - P(Z < z)$ et $2/ P(Z < -z) = P(Z > z)$
 Exemple: Sachant $P(Z < 1,24) = 0,8925$, on en déduit:
 $1/ (P(Z > 1,24) = 1 - P(Z < 1,24) = 1 - 0,8925 = 0,1075$
 $2/ P(Z < -1,24) = P(Z > 1,24) = 0,1075$

La valeur de $Z \alpha/2$ pour niveau de confiance de 25%

| z | 0,00 | 0,01 | 0,02 | 0,03 | 0,04 | 0,05 | 0,06 | 0,07 | 0,08 | 0,09 |
|-----|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| 0,0 | 0,5000 | 0,5040 | 0,5080 | 0,5120 | 0,5160 | 0,5199 | 0,5239 | 0,5279 | 0,5319 | 0,5359 |
| 0,1 | 0,5398 | 0,5438 | 0,5478 | 0,5517 | 0,5557 | 0,5596 | 0,5636 | 0,5675 | 0,5714 | 0,5753 |
| 0,2 | 0,5793 | 0,5832 | 0,5871 | 0,5910 | 0,5948 | 0,5987 | 0,6026 | 0,6064 | 0,6103 | 0,6141 |
| 0,3 | 0,6179 | 0,6217 | 0,6255 | 0,6293 | 0,6331 | 0,6368 | 0,6406 | 0,6443 | 0,6480 | 0,6517 |
| 0,4 | 0,6554 | 0,6591 | 0,6628 | 0,6664 | 0,6700 | 0,6736 | 0,6772 | 0,6808 | 0,6844 | 0,6879 |
| 0,5 | 0,6915 | 0,6950 | 0,6985 | 0,7019 | 0,7054 | 0,7088 | 0,7123 | 0,7157 | 0,7190 | 0,7224 |
| 0,6 | 0,7257 | 0,7291 | 0,7324 | 0,7357 | 0,7389 | 0,7422 | 0,7454 | 0,7486 | 0,7517 | 0,7549 |
| 0,7 | 0,7580 | 0,7611 | 0,7642 | 0,7673 | 0,7704 | 0,7734 | 0,7764 | 0,7794 | 0,7823 | 0,7852 |
| 0,8 | 0,7881 | 0,7910 | 0,7939 | 0,7967 | 0,7995 | 0,8023 | 0,8051 | 0,8078 | 0,8106 | 0,8133 |
| 0,9 | 0,8159 | 0,8186 | 0,8212 | 0,8238 | 0,8264 | 0,8289 | 0,8315 | 0,8340 | 0,8365 | 0,8389 |
| 1,0 | 0,8413 | 0,8438 | 0,8461 | 0,8485 | 0,8508 | 0,8531 | 0,8554 | 0,8577 | 0,8599 | 0,8621 |
| 1,1 | 0,8643 | 0,8665 | 0,8686 | 0,8708 | 0,8729 | 0,8749 | 0,8770 | 0,8790 | 0,8810 | 0,8830 |
| 1,2 | 0,8849 | 0,8869 | 0,8888 | 0,8907 | 0,8925 | 0,8944 | 0,8962 | 0,8980 | 0,8997 | 0,9015 |
| 1,3 | 0,9032 | 0,9049 | 0,9066 | 0,9082 | 0,9099 | 0,9115 | 0,9131 | 0,9147 | 0,9162 | 0,9177 |
| 1,4 | 0,9192 | 0,9207 | 0,9222 | 0,9236 | 0,9251 | 0,9265 | 0,9279 | 0,9292 | 0,9306 | 0,9319 |
| 1,5 | 0,9332 | 0,9345 | 0,9357 | 0,9370 | 0,9382 | 0,9394 | 0,9406 | 0,9418 | 0,9429 | 0,9441 |
| 1,6 | 0,9452 | 0,9463 | 0,9474 | 0,9484 | 0,9495 | 0,9505 | 0,9515 | 0,9525 | 0,9535 | 0,9545 |
| 1,7 | 0,9554 | 0,9564 | 0,9573 | 0,9582 | 0,9591 | 0,9599 | 0,9608 | 0,9616 | 0,9625 | 0,9633 |
| 1,8 | 0,9641 | 0,9649 | 0,9656 | 0,9664 | 0,9671 | 0,9678 | 0,9686 | 0,9693 | 0,9699 | 0,9706 |
| 1,9 | 0,9713 | 0,9719 | 0,9726 | 0,9732 | 0,9738 | 0,9744 | 0,9750 | 0,9756 | 0,9761 | 0,9767 |
| 2,0 | 0,9772 | 0,9778 | 0,9783 | 0,9788 | 0,9793 | 0,9798 | 0,9803 | 0,9808 | 0,9812 | 0,9817 |
| 2,1 | 0,9821 | 0,9826 | 0,9830 | 0,9834 | 0,9838 | 0,9842 | 0,9846 | 0,9850 | 0,9854 | 0,9857 |
| 2,2 | 0,9861 | 0,9864 | 0,9868 | 0,9871 | 0,9875 | 0,9878 | 0,9881 | 0,9884 | 0,9887 | 0,9890 |
| 2,3 | 0,9893 | 0,9896 | 0,9898 | 0,9901 | 0,9904 | 0,9906 | 0,9909 | 0,9911 | 0,9913 | 0,9916 |
| 2,4 | 0,9918 | 0,9920 | 0,9922 | 0,9925 | 0,9927 | 0,9929 | 0,9931 | 0,9932 | 0,9934 | 0,9936 |
| 2,5 | 0,9938 | 0,9940 | 0,9941 | 0,9943 | 0,9945 | 0,9946 | 0,9948 | 0,9949 | 0,9951 | 0,9952 |
| 2,6 | 0,9953 | 0,9955 | 0,9956 | 0,9957 | 0,9959 | 0,9960 | 0,9961 | 0,9962 | 0,9963 | 0,9964 |
| 2,7 | 0,9965 | 0,9966 | 0,9967 | 0,9968 | 0,9969 | 0,9970 | 0,9971 | 0,9972 | 0,9973 | 0,9974 |
| 2,8 | 0,9974 | 0,9975 | 0,9976 | 0,9977 | 0,9977 | 0,9978 | 0,9979 | 0,9979 | 0,9980 | 0,9981 |
| 2,9 | 0,9981 | 0,9982 | 0,9982 | 0,9983 | 0,9984 | 0,9984 | 0,9985 | 0,9985 | 0,9986 | 0,9986 |
| 3,0 | 0,99865 | 0,99869 | 0,99874 | 0,99878 | 0,99882 | 0,99886 | 0,99889 | 0,99893 | 0,99896 | 0,99900 |
| 3,1 | 0,99903 | 0,99906 | 0,99910 | 0,99913 | 0,99916 | 0,99918 | 0,99921 | 0,99924 | 0,99926 | 0,99929 |
| 3,2 | 0,99931 | 0,99934 | 0,99936 | 0,99938 | 0,99940 | 0,99942 | 0,99944 | 0,99946 | 0,99948 | 0,99950 |
| 3,3 | 0,99952 | 0,99953 | 0,99955 | 0,99957 | 0,99958 | 0,99960 | 0,99961 | 0,99962 | 0,99964 | 0,99965 |
| 3,4 | 0,99966 | 0,99968 | 0,99969 | 0,99970 | 0,99971 | 0,99972 | 0,99973 | 0,99974 | 0,99975 | 0,99976 |
| 3,5 | 0,99977 | 0,99978 | 0,99978 | 0,99979 | 0,99980 | 0,99981 | 0,99981 | 0,99982 | 0,99983 | 0,99983 |
| 3,6 | 0,99984 | 0,99985 | 0,99985 | 0,99986 | 0,99986 | 0,99987 | 0,99987 | 0,99988 | 0,99988 | 0,99989 |
| 3,7 | 0,99989 | 0,99990 | 0,99990 | 0,99990 | 0,99991 | 0,99991 | 0,99992 | 0,99992 | 0,99992 | 0,99992 |
| 3,8 | 0,99993 | 0,99993 | 0,99993 | 0,99994 | 0,99994 | 0,99994 | 0,99994 | 0,99995 | 0,99995 | 0,99995 |
| 3,9 | 0,99995 | 0,99995 | 0,99996 | 0,99996 | 0,99996 | 0,99996 | 0,99996 | 0,99996 | 0,99997 | 0,99997 |
| 4,0 | 0,99997 | 0,99997 | 0,99997 | 0,99997 | 0,99997 | 0,99997 | 0,99998 | 0,99998 | 0,99998 | 0,99998 |

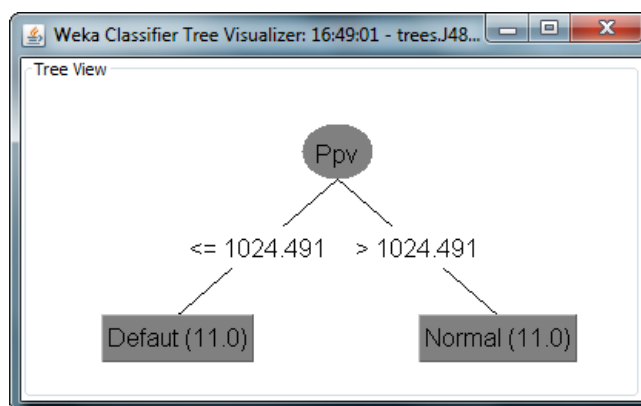
Tableau 2. Table de la loi Normale Centrée Réduite.

- Les données d'entrées pour faire l'apprentissage :

| T (°c) | E (w/m ²) | V _{PV} (V) | I _{PV} (A) | P _{PV} (w) | D |
|--------|-----------------------|---------------------|---------------------|---------------------|--------|
| 24,205 | 368,035 | 244,212 | 4,323 | 1055,803 | Normal |
| 24,14 | 372,285 | 244,159 | 4,364 | 1065,732 | Normal |
| 23,497 | 375,014 | 243,9 | 4,415 | 1076,895 | Normal |
| 22,989 | 379,36 | 243,823 | 4,447 | 1084,446 | Normal |
| 23,553 | 382,728 | 243,969 | 4,499 | 1097,67 | Normal |
| 23,791 | 386,231 | 244,052 | 4,539 | 1107,792 | Normal |
| 23,34 | 389,954 | 243,807 | 4,58 | 1116,809 | Normal |
| 23,733 | 393,495 | 243,656 | 4,624 | 1126,875 | Normal |
| 24,537 | 396,635 | 243,744 | 4,666 | 1137,522 | Normal |
| 23,995 | 400,6 | 243,546 | 4,704 | 1145,668 | Normal |
| 23,691 | 404,063 | 243,217 | 4,751 | 1155,554 | Normal |
| 24,205 | 368,035 | 244,212 | 2,161 | 527,901 | Défaut |
| 24,14 | 372,285 | 244,159 | 2,182 | 532,865 | Défaut |
| 23,497 | 375,014 | 243,9 | 2,207 | 538,447 | Défaut |
| 22,989 | 379,36 | 227,568 | 4,447 | 1012,149 | Défaut |
| 23,553 | 382,728 | 227,704 | 4,499 | 1024,492 | Défaut |
| 23,791 | 386,231 | 211,511 | 4,539 | 960,086 | Défaut |
| 23,34 | 389,954 | 211,299 | 4,58 | 967,901 | Défaut |
| 23,733 | 393,495 | 243,656 | 2,312 | 563,437 | Défaut |
| 24,537 | 396,635 | 243,744 | 0 | 0 | Défaut |
| 23,995 | 400,6 | 243,546 | 0 | 0 | Défaut |
| 23,691 | 404,063 | 243,217 | 0 | 0 | Défaut |

Tableau 2. Partie de l'ensemble d'apprentissage.

On reprend l'exemple de chapitre 3 mais cette fois par WEKA pour montre que les résultats sont les mêmes.



- L'ensemble de Test pour la validation du modèle d'arbre :

| T(°C) | E(w/m²) | Vpv | Ipv | Ppv | D | Prédicte |
|--------------|---------------------------|------------|------------|------------|----------|-----------------|
| 23,675 | 339,198 | 244,671 | 3,985 | 975,064 | Normal | ? |
| 23,562 | 342,924 | 244,635 | 4,022 | 984,142 | Normal | ? |
| 23,939 | 346,686 | 244,516 | 4,067 | 994,471 | Normal | ? |
| 23,37 | 357,694 | 244,224 | 4,196 | 1024,786 | Normal | ? |
| 23,624 | 360,234 | 244,298 | 4,242 | 1036,375 | Normal | ? |
| 23,464 | 364,528 | 244,316 | 4,272 | 1043,812 | Normal | ? |
| 23,675 | 339,198 | 244,671 | 1,992 | 487,532 | Défaut | ? |
| 23,562 | 342,924 | 244,635 | 2,011 | 492,0711 | Défaut | ? |
| 23,939 | 346,686 | 244,516 | 2,033 | 497,235 | Défaut | ? |
| 23,37 | 357,694 | 227,942 | 4,196 | 956,467 | Défaut | ? |
| 23,624 | 360,234 | 211,724 | 4,242 | 898,192 | Défaut | ? |
| 23,464 | 364,528 | 211,74 | 4,272 | 904,637 | Défaut | ? |

Tableau 3. Partie de l'ensemble de Test.

| Les données d'entrée | |
|--|--------------|
| Puissance PV recommandé | 2500-3600 Wp |
| Courant d'entrée max. | 19 A |
| Tension d'entrée Max. à 1000W/m ² et 10°C | 500 V |
| La plage de tension MPP | 150-400 V |
| Temps de connexion au réseau | 50s |
| Les données de sortie | |
| Puissance nominale | 2500 W |
| Puissance Max | 2650 W |
| Rendement Max | 94.3 % |
| Facteur de distorsion | < 3.5 % |
| Courant de sortie nominal | 10.9 A |
| Consommation de nuit | 0.15 W |
| Consommation en fonctionnement | 7 W |
| Tension nominal du réseau | 230 V / 50Hz |

Tableau 4. Caractéristique technique de l'Onduleur IG30 Fronius.