



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 غرداية

Université de Ghardaïa

كلية العلوم والتكنولوجيا

Faculté des Sciences et de la Technologie

Département Automatique et électromécanique

### Mémoire

Pour l'obtention du **diplôme** de **Master**

**Domaine:** Sciences et Technologie

**Filière :** Génie Mécanique

**Spécialité:** Énergie Renouvelable en Mécanique

### Thème

**Génération De Légendes D'images Pour L'état Des Cellules Photovoltaïques  
à l'aide De Techniques D'apprentissage Profond**

Soutenu Publiquement Le 12/06/2025

### Par

Hadj Hibeterrhmane

Devant le jury composé de :

<b>Zeggaoui Abdellah</b>	<b>Professeur</b>	<b>Univ--Ghardaia</b>	<b>Président</b>
<b>Belaid Abdelfetah</b>	<b>Maitre de recherche A</b>	<b>URAER-Ghardaia</b>	<b>Examineur</b>
<b>Arif Toufik</b>	<b>Maitre de recherche A</b>	<b>URAER-Ghardaia</b>	<b>Examineur</b>
<b>Guermoui Mawloud</b>	<b>Directeur de recherche</b>	<b>URAER-Ghardaia</b>	<b>Encadreur</b>
<b>Ferkous Khaled</b>	<b>Maitre de recherche A</b>	<b>Univ-Ghardaia</b>	<b>Co-Encadreur</b>

Année universitaire 2025/2026

# *Dédicaces*

*Je dédie cette réalisation extraordinaire aux personnes les plus chères à mon cœur, à mes parents bien-aimés qui ont été pour moi le refuge chaleureux et un soutien inébranlable, à ceux qui m'ont donné tout ce qu'ils possédaient en amour et en sacrifices pour être la raison de ce que je suis aujourd'hui. À vous deux, je dis : il n'y a pas de mots pour exprimer ma gratitude envers vous, car vous avez semé en moi l'amour de la connaissance et la persévérance, et vous avez continué à me soutenir jusqu'à la dernière étape.*

*À mes chers frères et sœurs, qui ont toujours été à mes côtés dans les bons et les mauvais moments, et qui ont partagé chaque instant de joie et chaque défi, vous m'avez soutenue avec amour et dévouement, votre présence dans ma vie est une bénédiction inestimable.*

*À mes amis qui ont cru en moi depuis le début, qui m'ont encouragée et soutenue moralement à chaque étape, à tous ceux avec qui j'ai partagé de beaux souvenirs, chaque moment d'affection m'a poussée à avancer.*

*À mes chers professeurs, qui ont cultivé en moi l'amour du savoir et ont éclairé mon chemin vers la connaissance, qui m'ont guidée par leur expérience et leurs idées, sans eux, je n'aurais pas pu accomplir cet exploit.*

*Et enfin, à tous ceux qui ont joué un rôle dans ma vie et qui m'ont soutenue par un mot gentil ou un conseil sage, je vous exprime ma profonde gratitude.*

*Je suis reconnaissante envers chaque personne qui a fait partie de mon parcours éducatif. Je demande à Dieu de toujours être à la hauteur de vos attentes et de rester fidèle à vos prières, qui ont été le chemin de cette réussite.*

*Hadj Hibeterrhmane*

# Remerciements

*Je tiens à exprimer ma profonde gratitude à tous ceux qui ont contribué à la réalisation de ce travail.*

*Je remercie tout particulièrement mon encadrant, Monsieur Guermoui Mawloud, pour ses précieuses orientations, ses conseils scientifiques avisés et son soutien constant tout au long de la préparation de cette recherche. Son suivi rigoureux et ses encouragements permanents ont grandement contribué à l'aboutissement de ce travail.*

*J'adresse également mes sincères remerciements à Monsieur l'assistant Ferhous Khaled pour sa contribution efficace, ses remarques constructives et l'aide précieuse qu'il a apportée durant les différentes étapes du travail.*

*Je n'oublie pas de remercier mes collègues et amis pour leur soutien moral et leurs encouragements, ainsi que ma famille pour leur patience et leur appui constant tout au long de mon parcours universitaire.*

*Enfin, je remercie toutes les personnes ayant contribué, de près ou de loin, à la mise à disposition des données et ressources nécessaires à l'achèvement de cette recherche.*

*Hadj Hibeterrhmane*

## Résumé

Dans le cadre de la transition mondiale vers les énergies renouvelables, les cellules photovoltaïques s'imposent comme l'une des principales solutions pour la production d'énergie propre. Cependant, le maintien de leur efficacité et de leur rendement énergétique exige une surveillance précise et continue de leur état, ce qui constitue un défi technique majeur, notamment dans les systèmes de grande envergure ou installés dans des environnements difficiles. Dans ce contexte, l'intelligence artificielle, en particulier les techniques d'apprentissage profond, représente un outil puissant pour le traitement et l'analyse d'un grand volume d'images issues des panneaux solaires. Ce travail porte sur le développement d'un modèle d'intelligence artificielle capable de générer automatiquement des descriptions textuelles (légendes d'images) de l'état des cellules photovoltaïques à partir de leurs images, dans le but de détecter des défauts tels que les fissures, les points chauds ou la corrosion, sans intervention humaine directe. Le modèle proposé repose sur la combinaison de réseaux de neurones convolutifs (CNN) pour l'extraction des caractéristiques visuelles, et de modèles de génération de langage comme les LSTM pour la production de descriptions sémantiquement précises. Cette approche permet d'améliorer l'efficacité de la surveillance et de la maintenance prédictive dans les centrales photovoltaïques, de réduire les coûts d'exploitation, et ouvre de nouvelles perspectives pour l'intégration de l'intelligence artificielle dans le domaine des énergies renouvelables. Les résultats préliminaires démontrent l'efficacité de cette méthode pour générer des descriptions fiables et pertinentes, contribuant ainsi à la transformation numérique de la gestion des infrastructures énergétiques.

**Mot clé :** PV: Photovoltaïques / LSTM : Mémoire à long terme / CNN : Réseau neuronal convolutif / ResNet : Réseau résiduel

## Abstract

Amid the global shift toward renewable energy, photovoltaic (PV) cells have emerged as one of the leading solutions for clean energy production. However, maintaining their efficiency and energy output requires continuous and accurate monitoring, which poses a significant technical challenge, particularly in large-scale systems or those located in harsh environments. In this context, artificial intelligence especially deep learning techniques offers powerful tools for processing and analyzing large volumes of image data captured from solar panels. This work focuses on developing an AI-based model capable of **automatically generating textual descriptions (image captions)** of the condition of PV cells based on their images, with the goal of detecting faults such as cracks, hotspots, or corrosion without direct human intervention. The proposed model combines convolutional neural networks (CNNs) for visual feature extraction with language generation models such as LSTM to produce semantically rich and accurate captions. This approach enhances the efficiency of monitoring and predictive maintenance in photovoltaic installations, reduces operational costs, and opens new avenues for integrating AI into the renewable energy sector. Preliminary results show the effectiveness of this methodology in generating reliable and meaningful descriptions, thereby contributing to the digital transformation of energy infrastructure management.

**Keywords :** PV: Photovoltaïques / LSTM: Long Short-Term Memory / CNN : Convolutional Neural Network / ResNet: Residual Network

## المخلص

في ظل التحول العالمي نحو مصادر الطاقة المتجددة، تُعد الخلايا الكهروضوئية من أبرز الحلول المعتمدة لتوليد الطاقة النظيفة. غير أن الحفاظ على كفاءتها ومردودها الطاقوي يتطلب مراقبة دقيقة ومتواصلة لحالتها، وهو ما يشكل تحدياً تقنياً حقيقياً، خاصة في الأنظمة واسعة النطاق أو تلك التي تعمل في بيئات قاسية وظروف مناخية صعبة. في هذا السياق، يُعد الذكاء الاصطناعي، وبشكل خاص تقنيات التعلم العميق من الأدوات الفعالة لمعالجة وتحليل الكميات الكبيرة من الصور الملتقطة من الألواح الشمسية. ويركز هذا البحث على تطوير نموذج يعتمد على الذكاء الاصطناعي، قادر على توليد أوصاف نصية تلقائية (تسميات أو تعليقات) لحالة الخلايا الكهروضوئية انطلاقاً من صورها، بهدف الكشف عن الأعطال مثل التشققات النقاط الساخنة، أو علامات التآكل، وذلك دون تدخل بشري مباشر. يعتمد النموذج المقترح على الدمج بين الشبكات العصبية الالتفافية (CNN) لاستخلاص الخصائص البصرية، ونماذج توليد اللغة مثل LSTM لإنشاء أوصاف دقيقة ذات دلالة معنوية واضحة. حيث تتيح هذه المنهجية تعزيز فعالية المراقبة والصيانة التنبؤية في محطات الطاقة الشمسية، والحد من التكاليف التشغيلية، كما تفتح آفاقاً جديدة لتطبيقات الذكاء الاصطناعي في قطاع الطاقات المتجددة. وتظهر النتائج الأولية أن هذا النهج فعال في توليد أوصاف موثوقة ودقيقة، مما يُسهم بشكل مباشر في دعم التحول الرقمي في إدارة البنية التحتية الطاقوية.

## الكلمات المفتاحية:

الخلايا الكهروضوئية / الذاكرة طويلة المدى قصيرة المدى / الشبكة العصبية التلافيفية / الشبكة المتبقية

## Table des matières

Introduction générale.....	1
1. Chapitre : Dysfonctionnements Des Cellules Photovoltaïques.....	3
1.1 Introduction.....	4
1.2 Énergie renouvelable.....	4
1.2.1 Énergie solaire.....	5
1.2.2 Énergie solaire thermique.....	5
1.2.3 Solaire Photovoltaïque (PV).....	6
1.3 Les fonctionnement.....	6
1.3.1 Les Matériaux.....	7
1.3.2 La différence entre une cellule photovoltaïque et un panneau solaire.....	10
1.4 L'énergie solaire photovoltaïque en Algérie.....	11
1.4.1 Programme National de Développement des Énergies Renouvelables (PNR).....	11
1.4.2 Usines de fabrication.....	12
1.4.3 Zones clés et projets.....	12
1.5 Technologies avancées.....	14
1.5.1 Capacité actuelle et future.....	14
1.6 Défaillances des cellules photovoltaïques.....	14
1.6.1 Types de Dysfonctionnements.....	14
1.6.2 Méthodes de détection des défauts dans un système photovoltaïque.....	15
1.6.3 Principe de fonctionnement.....	15
1.6.4 Caractéristiques de la caméra thermique.....	16
1.7 Conclusion.....	16
Références bibliographiques.....	17
Chapitre 2 : Techniques d'Apprentissage Profond pour la Génération de Légendes des Défaillances dans les Centrales Photovoltaïques.....	20
2.1 Introduction.....	21
2.2 Apprentissage Profond dans les Systèmes PV.....	22
2.2.1 Définition et concepts de base.....	22
2.2.2 Applications courantes dans le domaine de l'énergie photovoltaïque.....	24
2.3 Le Transfert d'Apprentissage pour l'Extraction de Caractéristiques Visuelles.....	25
2.3.1 Principe du transfert d'apprentissage.....	26
2.3.2 Architectures pré-entraînées couramment utilisées (VGG, ResNet, Inception, etc.).....	27
2.3.3 Adaptation du transfert d'apprentissage aux images PV.....	30
2.4 Décodage Séquentiel : Théorie et Approches.....	31
2.4.1 Introduction aux RNN et LSTM.....	31
2.4.2 Le réseau LSTM : Fonctionnement et avantages.....	31

2.4.3 Le réseau BiLSTM : Motivation et particularités.....	32
2.5 Intégration du Transfert d'Apprentissage et des Décodeurs LSTM/BiLSTM.....	33
2.5.1 Architecture générale du système de génération de légendes.....	34
2.5.2 Flux de données entre l'extracteur de caractéristiques et le décodeur.....	35
2.5.3 Problèmes rencontrés et solutions proposées (vanishing gradient, surapprentissage, etc.).....	35
2.6 Études de Cas et Applications.....	37
2.6.1 Exemples de détection et légendage de défaillances PV.....	38
2.6.2 Analyse comparative des performances (LSTM vs BiLSTM).....	39
2.7 Conclusion.....	39
Références bibliographiques.....	41
Chapitre 3 : Génération automatique de légendes d'images .....	46
3.1 Introduction.....	47
3.2 Intégration Des Modèles Visuels 2d Et Des Modèles De Prédiction Séquentielle Pour La Génération De Légendes D'images.....	47
3.2.1 Extraction des caractéristiques par modèles 2D et apprentissage par transfert.....	48
3.2.2 Modèles séquentiels pour la génération de texte.....	48
3.2.3 Union des deux paradigmes.....	48
3.3 Méthodologie de recherche et développement du modèle.....	49
3.3.1 Niveaux détaillés de l'élaboration du modèle.....	50
3.3.2 Architecture proposée.....	51
3.4 Composante Technique.....	51
3.4.1 Description Du Jeu De Données.....	51
3.4.2 Vue d'ensemble du système / Architecture générale.....	56
3.5 Modèle de génération de légendes.....	57
3.5.1 Architecture du modèle de génération de légendes d'images .....	57
3.5.2 Stratégies d'Apprentissage, Fonctions de Perte et Métriques d'Évaluation en Captioning d'Images.....	58
3.6 Approche de détection des anomalies.....	58
3.6.1 Définition conceptuelle de l'anomalie dans le contexte de la détection.....	58
3.6.2 La méthodologie adoptée pour la détection des anomalies.....	59
3.7 Conclusion.....	61
Références bibliographiques.....	62
Chapitre 4 : Expérimentation et Évaluation des Modèles de Génération de Légendes d'Images.....	64
4.1 Introduction.....	65
4.2 Méthodologie.....	66
4.2.1 Collecte et Préparation des Données.....	66
4.2.2 Architecture du Modèle.....	67
4.3 Evaluation du model de description.....	69

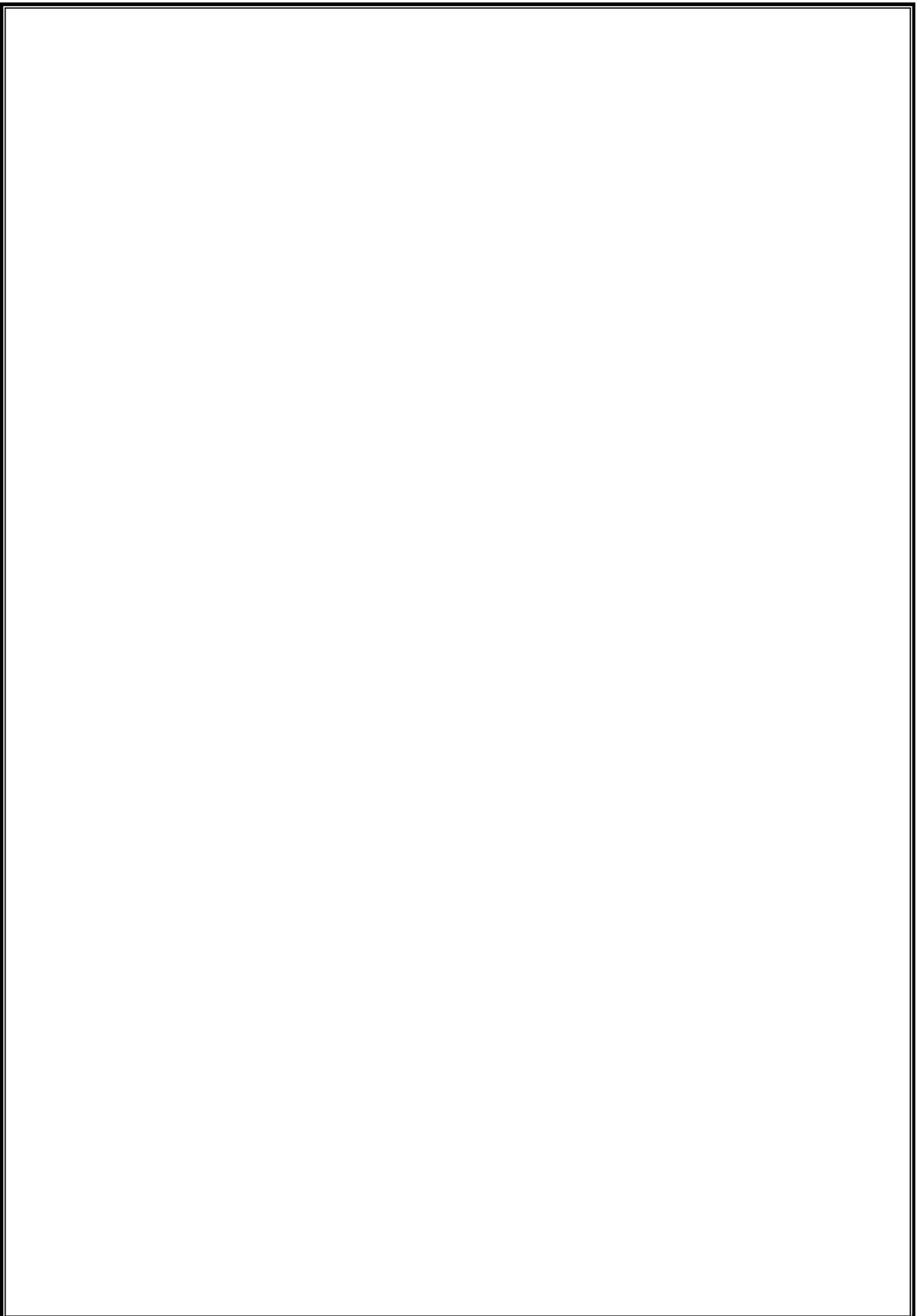
4.3.1 Méthodes d'évaluation en TALN.....	69
4.3.2 Utilisation du score BLEU dans ce travail.....	69
4.4 Résultats et Discussion.....	70
4.4.1 Résultats quantitatifs.....	70
4.4.2 Analyse comparative.....	71
4.4.3 Interprétation linguistique des scores.....	72
4.4.4 Résultats visuels.....	72
4.6 Conclusion.....	79
Conclusion Générale.....	80

## Table des figures

Figure 1.1 : Exploitations Futures Des Energies Renouvelables.....	5
Figure 1.2 : Le Panneau Solaire Thermique .....	5
Figure 1.3 : Le panneau solaire photovoltaïque .....	6
Figure 1.4 : Les Systèmes De Panneaux Photovoltaïques .....	7
Figure 1.5 : Distribution des divers matériaux sur le marché du photovoltaïque.....	7
Figure 1.6 : Le panneau monocristallin .....	8
Figure 1.7 : Panneau solaire poly cristallin 265w .....	9
Figure 1.8 : Ingrédients des cellules solaire .....	9
Figure 1.9 : Cellule photovoltaïque .....	10
Figure 1.10 : Panneau solaire .....	11
Figure 1.11 : Le programme national pour les énergies renouvelables et l'efficacité énergétique .....	12
Figure 1.12 : Un village bientôt alimenté en électricité à 100% solaire .....	13
Figure 1.13: Centrale hybride de Hassi R'mel .....	13
Figure 1.14 : nouvelles centrales solaires .....	14
Figure 1.15 : 5 centrales photovoltaïques ont été mises en œuvre sur une surface de 940 hectares .....	14
Figure 1.16 : Inspection du panneau photovoltaïque à l'aide d'un drone thermique .....	16
Figure 1.17 : La caméra thermique .....	16
Figure 2.1: Neurone formel de Mac Culloch et Pitts .....	23
Figure 2.2 : d'un réseau artificiel.....	24
Figure 2.3 : Explication des systèmes domestiques en réseau.....	24
Figure 2.4 : VGG-Net Architecture .....	27
Figure 2.5 : Architecture du réseau résiduel profond (ResNet) .....	28
Figure 2.6 : Architecture Google Net, propagation directe de droite à gauche .....	29
Figure 3.1 : Architecture encodeur–décodeur pour la génération du langage.....	49
Figure 3.2 : Architecture CNN-LSTM.....	50
Figure 3.3 : Classes de modules.....	51
Figure 3.4 : Répartition des images infrarouges selon les classes d'anomalies des modules photovoltaïques.....	52
Figure 3.5 : Exemples d'images par classe d'anomalie avec leur description cible générée automatiquement.....	55
Figure 4.1 : Organigramme.....	68
Figure 4.2 : Exemples de descriptions générées automatiquement.....	79

## **Table des tableaux**

Tableau 1.1 : Consistance du programme de développement des énergies renouvelables .....	12
Tableau 2.1 : Comparaison des Architectures .....	29
Tableau 3.1 : Classes .....	52
Tableau 4.1 : Résultats comparatifs des scores BLEU pour les modèles ResNet50-LSTM et ResNet50-GRU sur l'ensemble de test.....	70



## Introduction générale

Dans un contexte mondial marqué par une transition accrue vers les énergies renouvelables et par l'aggravation des défis environnementaux liés aux changements climatiques et à l'épuisement des ressources naturelles, le recours à des solutions techniques intelligentes et efficaces devient indispensable pour assurer la durabilité des systèmes de production d'énergie, en particulier celle solaire, reconnue comme l'une des sources d'énergie renouvelable les plus propres et abondantes.

Les cellules photovoltaïques constituent une pierre angulaire dans la production d'électricité propre. Cependant, le maintien de leur efficacité et de leur performance continue nécessite une surveillance rigoureuse et constante de leur état opérationnel, surtout dans les grandes centrales solaires implantées dans des environnements désertiques ou soumis à des conditions climatiques sévères. Dans ce cadre, l'intelligence artificielle, et plus spécifiquement les techniques d'apprentissage profond, offrent des outils prometteurs pour automatiser les processus de surveillance et de détection des défauts sans intervention humaine directe.

À travers ce mémoire, nous proposons une approche technique visant à développer un modèle intelligent capable de générer automatiquement des légendes pour des images des cellules solaires, afin de diagnostiquer leur état et d'identifier les défauts potentiels tels que les fissures, les points chauds ou la corrosion. Ce modèle repose sur la combinaison des réseaux de neurones convolutionnels (CNN) pour l'extraction des caractéristiques visuelles, et des modèles de génération de langage de type LSTM pour produire des descriptions précises et sémantiques.

Le premier chapitre offre un panorama complet des cellules photovoltaïques en expliquant leur principe de fonctionnement, leurs types et composants fondamentaux. Il met l'accent sur les principaux défauts susceptibles de les affecter, comme les fissures, les points chauds et la dégradation thermique, tout en présentant les méthodes de détection couramment utilisées, notamment l'imagerie thermique, dans le but d'améliorer l'efficacité des systèmes solaires et d'assurer une maintenance optimale.

Le deuxième chapitre aborde l'application des techniques d'apprentissage profond pour la génération automatique de descriptions des défauts des cellules photovoltaïques, en s'appuyant sur des réseaux CNN comme ResNet pour extraire les caractéristiques à partir des images, et des modèles LSTM ou BiLSTM pour la production des textes explicatifs. Il souligne également l'importance de l'apprentissage par transfert (Transfer Learning) pour améliorer la performance des modèles, particulièrement en cas de données limitées.

Le troisième chapitre présente la méthodologie de génération automatique de légendes pour les images photovoltaïques, en intégrant le réseau ResNet pour l'extraction des caractéristiques visuelles et le modèle LSTM pour la génération des descriptions linguistiques précises. Cette approche repose sur l'apprentissage profond afin de faire le lien entre vision par ordinateur et traitement du langage naturel, facilitant ainsi l'interprétation intelligente des images pour des applications telles que la surveillance industrielle et la détection automatique des défauts.

Le quatrième chapitre décrit les étapes d'expérimentation des modèles de génération de légendes d'images, en mettant l'accent sur la préparation des données, la construction des modèles basés sur ResNet et LSTM, ainsi que la définition des paramètres d'entraînement. Il traite également de l'évaluation des performances par des métriques standards telles que BLEU et METEOR, et analyse les résultats pour identifier les forces, les limites et les pistes d'amélioration.

Enfin, ce travail propose un système intelligent capable de générer automatiquement des descriptions claires et compréhensibles des défauts des cellules photovoltaïques à partir d'images thermiques, en s'appuyant sur une architecture Encoder-Decoder combinant ResNet50 pour l'extraction des caractéristiques et LSTM/GRU pour la génération textuelle, avec un support multilingue et des résultats prometteurs attestant de l'efficacité de la méthode développée.



Chapitre 1 : Dysfonctionnements Des Cellules Photovoltaïques

## 1.1 Introduction

Les cellules photovoltaïques (PV) sont des composants essentiels des systèmes d'énergie solaire, convertissant la lumière du soleil directement en électricité grâce à l'effet photovoltaïque. Bien que la technologie photovoltaïque soit généralement fiable et durable, les cellules PV sont sujettes à divers dysfonctionnements qui peuvent réduire leur efficacité, leur performance et leur durée de vie globale. Comprendre ces dysfonctionnements est crucial pour le diagnostic, la maintenance et l'amélioration des systèmes d'énergie solaire .

Afin de fournir une source d'énergie propre et durable, l'utilisation des énergies renouvelables augmente chaque jour. L'utilisation de l'énergie solaire a considérablement augmenté ces dernières années. Les solutions au sol et sur toit ont rapidement gagné des parts de marché. De la même manière, l'entretien et le fonctionnement des panneaux solaires doivent être pris en compte. Bien que les systèmes solaires photovoltaïques ne nécessitent aucun entretien, la surveillance du système est nécessaire pour tirer le meilleur parti des installations.

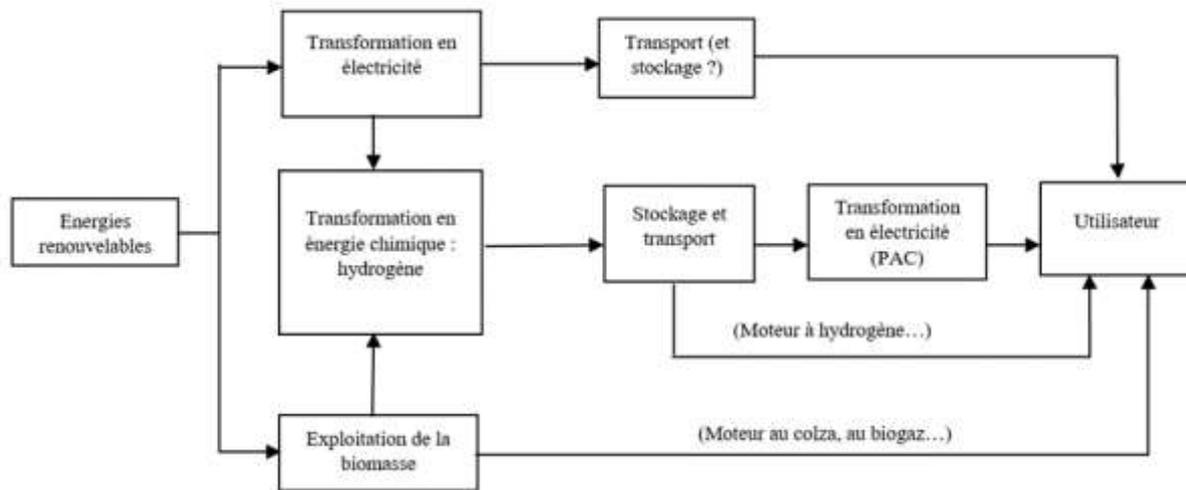
Le rendement du panneau est influencé par un certain nombre de facteurs, notamment l'humidité, la poussière, l'humidité et la température de l'ombre. La surveillance de panneaux individuels dans une grande centrale électrique est une procédure laborieuse. Cependant, toute variable influençant le rendement des panneaux solaires entraînera une résistance intérieure. Par conséquent, une image thermique des surfaces sera en mesure de déterminer rapidement le dysfonctionnement du panneau. Bien qu'il existe plusieurs caméras thermiques sur le marché, il peut s'avérer difficile d'analyser des photos individuelles. En conséquence, le logiciel Python Simulink traite l'image capturée par une caméra thermique pour les différentes procédures d'empilement d'images. Les bordures des modules sont liées au différentiel de température, les points chauds sont donc simples à localiser. seront abordés dans ce chapitre.

## 1.2 Énergie renouvelable

Les énergies renouvelables désignent les énergies issues de sources naturelles qui se régénèrent à un taux supérieur à leur exploitation. Des sources d'énergie renouvelables, telles que la lumière solaire et l'énergie éolienne, sont constamment en cours de renouvellement. Les énergies renouvelables sont multiples et omniprésentes dans notre environnement.

Cependant, les énergies fossiles (charbon, pétrole et gaz) sont des ressources non renouvelables qui nécessitent plusieurs centaines de millions d'années pour se former. Lorsque les combustibles fossiles sont brûlés pour générer de l'énergie, ils engendrent des émissions de gaz à effet de serre nuisibles, comme le dioxyde de carbone. La création d'énergie verte émet nettement moins de gaz à effet de serre que la combustion de combustibles fossiles. Pour lutter contre la crise climatique, il est essentiel de substituer les énergies fossiles, responsables pour l'essentiel des émissions, par des sources d'énergie renouvelables.

Dans la majorité des pays, l'énergie renouvelable est maintenant moins coûteuse et génère trois fois plus d'emplois que les énergies fossiles[1].



**Figure 1.1 :** Exploitations Futures Des Energies Renouvelables.

## 1.2 Énergie solaire

L'énergie solaire désigne la chaleur et la lumière émanant du soleil. On peut utiliser l'énergie solaire pour chauffer et éclairer des bâtiments, ainsi que pour générer de l'électricité.

Dans le passé, les technologies passives ont été utilisées pour exploiter l'énergie solaire. Ces technologies tirent parti de la chaleur et de la lumière du soleil sans recourir à des dispositifs mécaniques ou électriques. Par exemple, en plaçant les structures de manière stratégique et en planifiant certains de leurs composants comme les fenêtres, les saillies et les masses thermiques pour tirer parti de la lumière naturelle pour l'éclairage et le chauffage. Il n'est possible d'exploiter l'énergie solaire qu'en journée, et cela ne peut se faire que si des nuages, des structures ou d'autres éléments ne font pas obstacle à la lumière du soleil.

Deux formes de technologies solaires efficaces, utilisant des dispositifs électriques ou mécaniques, se répandent de plus en plus :

1. Les panneaux ou capteurs solaires thermiques sont utilisés pour chauffer l'eau ou l'air de ventilation dans les bâtiments.
2. La technologie solaire photovoltaïque (PV) utilise des cellules solaires pour convertir directement la lumière du soleil en électricité[2].

### 1.2.1 Énergie solaire thermique

Cette méthode exploite la chaleur solaire afin de produire de l'énergie thermique, utilisable pour le chauffage ou la production d'électricité via des centrales thermodynamiques. Ces installations concentrent les rayons solaires à l'aide de miroirs afin de chauffer un fluide caloporteur, générant ainsi de la vapeur qui fait tourner une turbine [3][4] .



**Figure 1.2 :** Le Panneau Solaire Thermique[5].

### 1.2.2 Solaire Photovoltaïque (PV)

Le mot « photovoltaïque » peut s'appliquer au système physique (l'effet photovoltaïque, conçu par Alexandre Edmond Becquerel en 1839) ou à la technologie qui le relie. L'électricité photovoltaïque est l'énergie produite en convertissant une fraction de rayon solaire en cellule photovoltaïque. De nombreux schémas peuvent amener un photon incident, dans certaines conditions, à déplacer un appareil électronique et à générer un courant électrique. Les cellules photovoltaïques sont reliées au matériau de même couleur, de très grande taille. Les électrons sont exposés à ces matériaux et, lorsqu'ils sont exposés à la lumière, ils font partie d'un circuit ferreux pour produire de l'électricité [6].



**Figure 1.3** : Le panneau solaire photovoltaïque [7].

### 1.3 Le fonctionnement

Le fondement d'une installation solaire photovoltaïque repose sur la transformation de l'énergie lumineuse du soleil en électricité via des cellules photovoltaïques.

Voici une explication détaillée de cette procédure:

#### 1. Captage des rayons solaires

Les panneaux solaires, qui intègrent des cellules photovoltaïques, saisissent la lumière solaire. Ces cellules sont habituellement produites à partir de silicium, un matériau semi-conducteur qui facilite la transformation de l'énergie lumineuse en électricité[8][9].

#### 2. Production d'électricité

Quand les cellules photovoltaïques sont exposées aux photons, des électrons sont libérés du matériau semi-conducteur. Ce mouvement d'électrons génère un courant électrique continu (DC) aux extrémités des cellules [10][11].

#### 3. Conversion en courant alternatif

Le courant continu généré par les panneaux solaires est ensuite dirigé vers des onduleurs. Ces dispositifs transforment le courant continu en courant alternatif (AC), qui est la forme d'électricité utilisée par la majorité des équipements électriques et compatible avec le réseau électrique[8][11].

4. Élévation de la tension

Par la suite, l'électricité est acheminée par des convertisseurs qui augmentent la tension à un état compatible avec le réseau électrique. Cela facilite le transport de l'électricité sur de grandes distances tout en minimisant les pertes[8][11].

5. Injection dans le réseau

Pour finir, l'électricité est introduite dans le réseau électrique à travers une station de livraison. Elle peut être employée directement par les utilisateurs ou conservée dans des batteries pour une exploitation future[8][12].

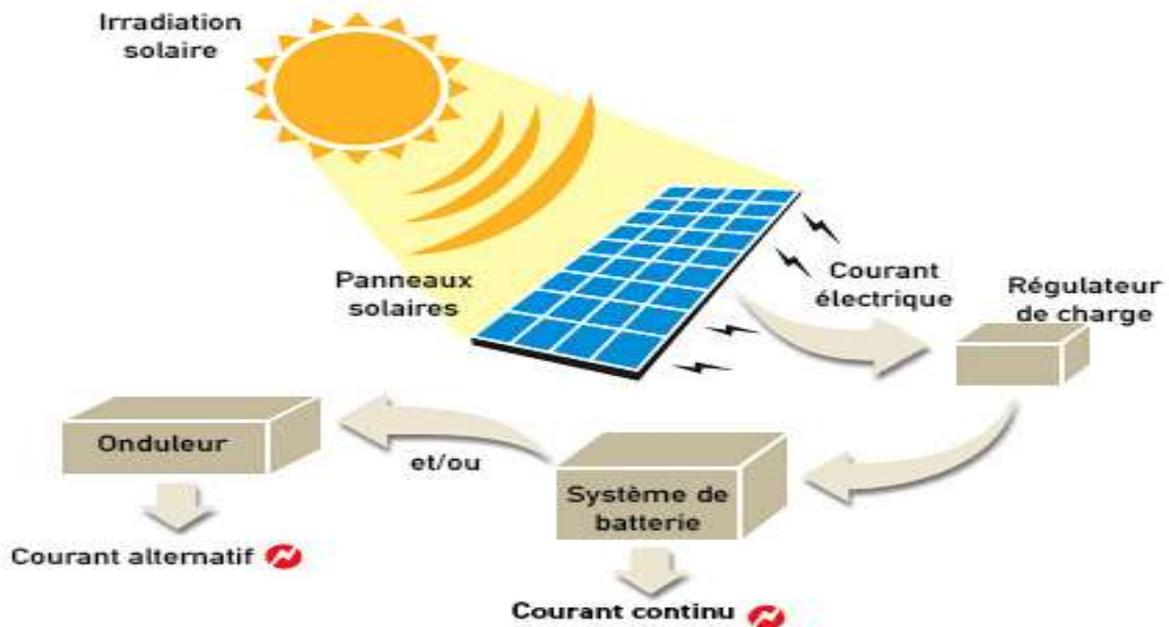


Figure 1.4 : Les Systèmes De Panneaux Photovoltaïques [13].

1.3.1 Les Matériaux

Les cellules photovoltaïques sont principalement constituées de silicium, de verre, de plastique et de liaisons en cuivre et en argent. Le silicium est employé comme matière première dans plus de 90% de la production de panneaux solaires sur le marché photovoltaïque [14][15].

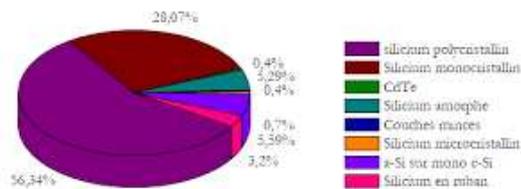


Figure 1.5 : Distribution des divers matériaux sur le marché du photovoltaïque[18].

✓ Silicium :

Grâce à sa capacité à convertir la lumière du soleil en électricité, le silicium est le matériau fondamental des panneaux photovoltaïques.

Il possède à la fois des propriétés photosensibles et semi-conductrices. Le silicium est le second élément le plus répandu sur notre planète, juste après l'oxygène, et il constitue près de 28 % de la croûte terrestre[15-17].

Il existe deux principaux types de panneaux photovoltaïques en silicium [19] :

a) Monocristallins :

Constitués de cellules de silicium monocristallin, dotées d'une structure homogène et d'une pureté remarquable, ils favorisent une conversion plus performante de la lumière solaire en électricité[19].



**Figure 1.6 :** Le panneau monocristallin[19].

b) Polycristallins:

Ces panneaux sont constitués de cellules en silicium polycristallin fondu, ce qui les rend plus économiques à fabriquer par rapport aux monocristallins, même s'ils sont légèrement moins performants en termes d'efficacité énergétique [19].



**Figure 1.7 :** Panneau solaire polycristallin 265w [20] .

D'autres composants comprennent[21]:

**a) Verre trempé:**

Constitue la face côté soleil et offre une protection contre les chocs, les UV et l'humidité.

**b) Enveloppe EVA:**

Une résine chimique adhésive qui protège les cellules photovoltaïques de l'humidité, des champignons et de l'oxydation, assurant l'étanchéité, la conductivité thermique et optique.

**c) Tedlar:**

Un polymère utilisé pour fabriquer la face arrière du panneau solaire, protégeant contre les agressions environnementales.

**d) Châssis en aluminium:**

Offre un support mécanique et une protection supplémentaire.



**Figure 1.8 :** Ingrédients des cellules solaire [21].

### 1.3.2 La différence entre une cellule photovoltaïque et un panneau solaire

La distinction majeure entre une cellule photovoltaïque et un panneau solaire est liée à leur rôle et à leur structure. Une cellule photovoltaïque transforme directement l'énergie lumineuse en électricité par le biais de l'effet photovoltaïque. En revanche, un panneau solaire est constitué d'une combinaison de plusieurs cellules photovoltaïques reliées entre elles, formant ainsi un module de taille supérieure, généralement protégé par une couche de verre et renforcé par un cadre en aluminium pour une meilleure durabilité [22] [23].

Voici les principales distinctions :

#### a. Cellule photovoltaïque :

C'est le composant fondamental qui transforme la lumière en électricité en utilisant des matériaux semi-conducteurs tels que le silicium [22] [23].



Figure 1.9 : Cellule photovoltaïque [23].

#### b. Panneau solaire :

Il s'agit d'un module formé par un regroupement de cellules photovoltaïques, destiné à recueillir l'énergie solaire et à produire de l'électricité sur une plus grande échelle. Les panneaux solaires photovoltaïques constituent une option écologique pour générer de l'électricité. Ils ne produisent aucun gaz à effet de serre et les matériaux employés sont recyclables. En outre, ce type d'équipement assure une longue longévité, car un panneau solaire a habituellement une durée de vie pouvant atteindre jusqu'à 30 ans.

Sur le plan technique, l'expression « panneau solaire » désigne un terme global qui regroupe toutes les formes de modules exploitant l'énergie solaire. Cela inclut les panneaux photovoltaïques qui génèrent de l'électricité, les panneaux thermiques qui produisent de la chaleur, et les panneaux aérovoltaiques qui fusionnent ces deux fonctions. Ainsi, les panneaux photovoltaïques constituent une sous-catégorie particulière des panneaux solaires[22][24].



Figure 1.10 : Panneau solaire [25].

#### 1.4 L'énergie solaire photovoltaïque en Algérie

L'Algérie a mis en œuvre des projets importants pour renforcer son industrie de cellules photovoltaïques, dans le but d'améliorer sa capacité à produire de l'énergie renouvelable. Le pays jouit de l'un des plus importants potentiels solaires au monde, avec une exposition au soleil dépassant les 2000 heures par an et pouvant aller jusqu'à 3900 heures dans les régions montagneuses et le Sahara [26].

##### 1.4.1 Programme National de Développement des Énergies Renouvelables (PNR)

L'Algérie a lancé un programme ambitieux visant à installer 22 000 MW d'électricité d'origine renouvelable d'ici 2030, dont 4 500 MW devaient être installés d'ici 2020. Ce programme souligne l'engagement du pays envers les énergies renouvelables. [28].

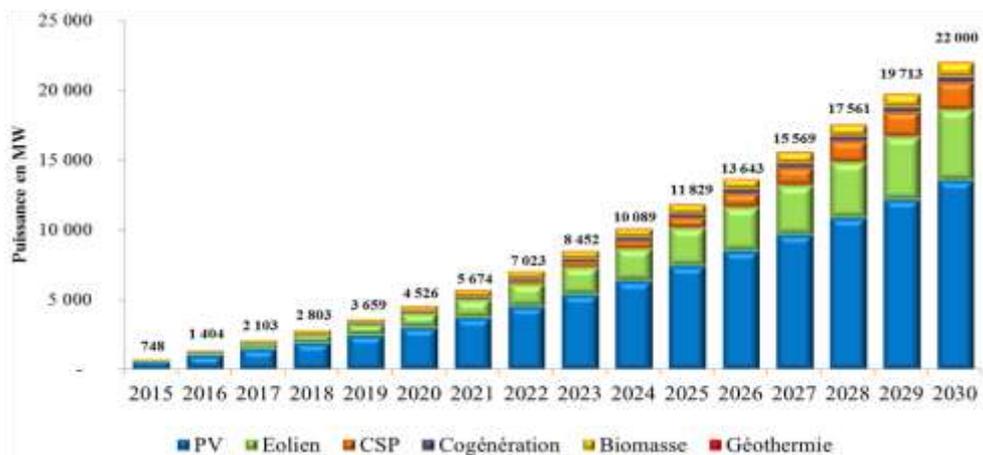


Figure 1.11 : Le programme national pour les énergies renouvelables et l'efficacité énergétique [29]

Unité : MW	1ère Phase 2015-2020	2ème Phase 2021-2030	TOTAL
Photovoltaïque	3 000	10 575	13 575
Eolien	1 010	4 000	5 010
CSP	-	2000	2 000
Cogénération	150	250	400
Biomasse	360	640	1 000
Géothermie	05	10	15
<b>TOTAL</b>	<b>4 525</b>	<b>17 475</b>	<b>22 000</b>

**Tableau 1.1** : Consistance du programme de développement des énergies renouvelables [29][30].

### 1.4.2 Usines de fabrication

En Algérie, plusieurs installations ont été mises en place pour la fabrication de panneaux solaires :

- L'ENIE (société nationale des industries électroniques) a injecté deux milliards de DA dans une nouvelle usine à Sidi Bel Abbès, qui a une capacité de production initiale fixée à 25 MW par an [28].
- Milltech, une entreprise algérienne, a inauguré une installation à Chelghoum El Aid dotée d'une capacité annuelle de 100 MW. Elle utilise des technologies sophistiquées telles que le bi-verre ainsi que des modules polycristallins et monocristallins[31][33].
- Zergoun Green Energy ainsi que d'autres sociétés ont aussi initié la construction d'installations pouvant produire jusqu'à 200 MW, dans le but de satisfaire les exigences du marché local tout en considérant l'opportunité d'exportation[33][34].

### 1.4.3 Zones clés et projets

- **Ilizi**

La wilaya d'Ilizi a été choisie pour accueillir le premier village algérien alimenté à 100 % par l'énergie solaire. Ce choix est motivé par sa situation géographique au cœur du Sahara, son nombre limité de foyers et son éloignement des réseaux d'électricité et de gaz[35].



**Figure 1.12** : Un village bientôt alimenté en électricité à 100% solaire [36].

- **Hassi R'mel (Laghouat)**

Une centrale hybride solaire-gaz de 150 MW est située dans cette localité[37].



**Figure 1.13:** Centrale hybride de Hassi R'mel [38].

- **Adrar**

Cette région bénéficie d'un ensoleillement notable et de vitesses de vent considérables. Il y a une centrale Adrar de 10 MW à Kabertène, dans la région d'Adrar [37][39].



**Figure 1.14 :** nouvelles centrales solaires[40].

- **Oued N'Chou (Ghardaia):**

Dans cet endroit, on a installé une centrale pilote de production d'énergie solaire de 1,1 MWc. [37].



**Figure 1.15 :** 5 centrales photovoltaïques ont été mises en œuvre sur une surface de 940 hectares[41].

- **Régions d'In Salah, Adrar, Timimoune et Bechar:**

L'extension de la liaison électrique entre le Nord et le Sahara facilitera l'implantation de vastes centrales d'énergies renouvelables dans ces zones, ainsi que leur incorporation dans le réseau énergétique national[37].

### **1.5 Technologies avancées**

En Algérie, diverses sortes de modules sont produites dans les usines, comme les technologies MonoPerc, bifaciales et Half-Cut Cell. L'objectif de ces progrès est d'améliorer l'efficacité et la durabilité des panneaux solaires[33] [43].

#### **1.5.1 Capacité actuelle et future**

À la clôture de l'année 2023, l'Algérie avait une capacité photovoltaïque totale de 436,8 MW. On estime qu'elle pourrait atteindre entre 600 et 700 MW d'ici la fin de 2025 en raison de l'accroissement constant des capacités de production.

Ces initiatives indiquent que l'Algérie est résolument impliquée dans l'expansion de son domaine photovoltaïque afin de diminuer sa dépendance vis-à-vis des énergies fossiles et favoriser une transition énergétique pérenne [34][43].

### **1.6 Défaillances des cellules photovoltaïques**

Différents problèmes peuvent affecter l'efficacité et la production d'énergie des cellules photovoltaïques. Ces problèmes peuvent résulter de vices de fabrication, de facteurs environnementaux ou de l'usure [41][42].

#### **1.6.1 Types de Dysfonctionnements**

##### ***1. Fissures et fractures***

Les fissures, qui sont souvent ardues à prévenir, peuvent apparaître à divers moments de la vie du module et peuvent être dues à des contraintes thermiques et mécaniques, ou survenir lors du processus de fabrication. Si une portion de la cellule devient inactive, les micro-fissures peuvent provoquer une diminution de la puissance[43].

## 2. Délamination du revêtement antireflet (ARC)

Des contraintes thermiques et mécaniques peuvent fissurer et séparer le revêtement antireflet, ce qui entraîne des parties de cellules inactives et une perte de puissance [43].

## 3. Défaut de point chaud

Un hotspot est une région de forte résistance entraînant une concentration d'énergie et un échauffement excessif. Les panneaux solaires emploient des diodes de dérivation afin d'éliminer les cellules défectueuses et de sauvegarder le reste du panneau [44].

## 4. Défaut de boîte de jonction

Une défaillance de la boîte de jonction, qui connecte les panneaux solaires, peut affecter les performances globales [44].

## 5. Défaut de module

Un défaut de module se produit lorsque le module du panneau solaire tombe en panne, ce qui entraîne une surchauffe ou une mise hors ligne complète [44].

## 6. Observation des défauts

L'ombrage partiel des panneaux solaires peut réduire la production d'énergie et provoquer un échauffement [44].

## 7. Défaut de diode

Un défaut de diode fournit un chemin de courant dans une cellule ou un panneau défectueux. L'anomalie de la diode est généralement visible dans environ 1/3 du panneau [44].

### 1.6.2 Méthodes de détection des défauts dans un système photovoltaïque

#### Caméras thermiques

L'emploi de caméras thermiques pour surveiller les panneaux solaires comporte plusieurs bénéfices. Sur l'image thermique de haute qualité, les irrégularités sont nettement perceptibles et, à la différence de nombreuses autres techniques, il est possible d'inspecter les panneaux en cours d'exploitation [45].



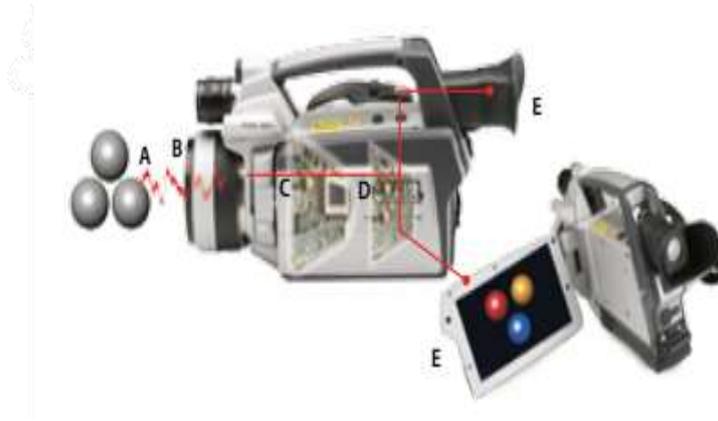
**Figure 1.16 :** Inspection du panneau photovoltaïque à l'aide d'un drone thermique [46].

### 1.6.3 Principe de fonctionnement

L'énergie infrarouge de l'objet (A) est focalisée par l'optique (B) sur un détecteur infrarouge (C). Le détecteur envoie des informations à la partie électronique du capteur (D) qui se charge

du traitement de l'image. L'électronique convertit les données du détecteur en une image (E) pouvant être visualisée dans un viseur, une vidéo ou un écran LCD.

La thermographie consiste à convertir les mesures du rayonnement infrarouge en images radiométriques, permettant de lire les valeurs de température. Par conséquent, chaque pixel de l'image radiométrique est une mesure de température et, à cet effet, les caméras thermiques contiennent des algorithmes complexes [47].



**Figure 1.17:** La caméra thermique[47]

#### 1.6.4 Caractéristiques de la caméra thermique

Rapidité, qualité et rentabilité sont des objectifs primordiaux. Pour les atteindre, il est crucial que les installations industrielles fonctionnent sans interruption : 24 heures sur 24, 365 jours par an. Toute panne ou perte de temps doit être évitée. La gestion préventive d'une usine représente donc une tâche de grande envergure. Si vous aviez la possibilité de détecter à l'avance une défaillance imminente, vous pourriez intervenir au moment le plus opportun. Malheureusement, les soucis les plus sérieux restent souvent indétectables jusqu'à ce qu'ils deviennent critiques. Cependant, grâce aux caméras thermiques qui rendent visibles ce qui l'est moins, il est possible d'anticiper ces défaillances. Sur une image thermique, les anomalies sont rapidement identifiables.

Pour pouvoir produire en continu, de nombreuses usines établissent des programmes de maintenance préventive avec les plus précieux outils de diagnostic sur le marché : les caméras thermiques [47].

#### 1.7 Conclusion

Dans ce chapitre, nous offrons le principe de fonctionnement et la possibilité de cette énergie, et nous nous concentrons en particulier sur la capacité de l'énergie solaire en Algérie car il a d'énormes capacités d'exploiter l'énergie renouvelable, en particulier l'énergie solaire photovoltaïque, en particulier dans le sud partie, et il peut y avoir des dysfonctionnements dans les cellules photovoltaïques qui affectent les systèmes d'énergie solaire de performance.

C'est pourquoi une surveillance régulière et une maintenance appropriée sont nécessaires pour déterminer rapidement ce problème et assurer une production d'énergie optim

## Références bibliographiques

- [1] United Nations. "What is renewable energy?" United Nations, 2020.
- [2] Natural Resources Canada. "Solar Energy." Natural Resources Canada, January 14, 2025.
- [3] Enbridge Gaz Québec. "What is solar energy? An inspiring energy." Enbridge Gaz Québec, 2025.
- [4] Connaissance des Énergies. "Solar Energy: Photovoltaic and Thermal Exploitation." Connaissance des Énergies, July 2024
- [5] Mon Kit Solaire. "Lifespan of a Thermal Solar Panel: What You Need to Know." Mon Kit Solaire, 2025
- [6] Connaissance des Énergies. "Photovoltaic Solar Energy: Operation, Panels, and Power Plants." Connaissance des Énergies, January 2025.
- [7] Batiadvisor. "Photovoltaic Solar Energy: An Economical and Ecological Solution." Batiadvisor, 2025.
- [8] ENGIE Green. "How Does a Solar Power Plant Work?" ENGIE Green, 2025
- [9] Connaissance des Énergies. "Photovoltaic Solar Energy: Operation, Panels, and Power Plants." Connaissance
- [10] Groupe Roy Énergie. "Photovoltaic Power Plant: How It Works." Groupe Roy Énergie, 2025.
- [11] Générale du Solaire. "What is a Photovoltaic Power Plant? And How Does It Work?" Générale du Solaire, May 2, 2023.
- [12] Technique Solaire. "What is a photovoltaic solar power plant?" Technique Solaire, 2025.
- [13] Ferme Solaire. "Understanding the Energy Chain of a Solar Panel." Ferme Solaire, June 21, 2024.
- [14] Connaissance des Énergies. Fiche pédagogique : énergie solaire. (n.d.).
- [15] Libow. Comment fabriquer un panneau solaire photovoltaïque ? (n.d.).
- [16] ENGIE MyPower. Comment sont fabriqués les panneaux solaires ? (n.d.).
- [17] Optimwatt. Quel est le matériau de base du panneau photovoltaïque ? (n.d.).
- [18] Systèmes Solaires. Le baromètre du photovoltaïque. Systèmes Solaires, (154), 41–52. (2003).
- [19] Ref : Guide Panneaux Photovoltaïques. Le panneau monocristallin. (n.d.).
- [20] UPES Énergie. Panneau solaire polycristallin 265W. (n.d.).
- [21] Jade TechnologieComposition d'un panneau solaire. . (n.d.).
- [22] DualSun. Différence entre panneaux photovoltaïques et solaires. (n.d.).
- [23] Techno-Science. Définition : panneau photovoltaïque. (n.d.).
- [24] L'énergie tout compris. Quelle est la différence entre un panneau solaire et un panneau photovoltaïque ? (n.d.).
- [25] SodiShop Guinée. Panneau solaire photovoltaïque 100W YE0038. (n.d.).
- [26] AAPI Algérie. Secteur des énergies nouvelles et énergies renouvelables. (n.d.).
- [27] Université des Sciences et Technologies Houari Boumediene (USTHB). Présentation Master Matériaux EnR. (2023).
- [28] ENIE. Panneaux photovoltaïques produits localement. (n.d.).
- [29] Ministère de l'Énergie Algérien. Programme de développement des énergies renouvelables. (n.d.).
- [30] ResearchGate. Tableau : Programme national des énergies renouvelables. (n.d.).

- [31] Journal Photovoltaïque. Une troisième usine de production photovoltaïque en Algérie. (n.d.).
- [32] MillTech DZ. Présentation de l'entreprise MillTech. (n.d.).
- [33] Algérie Eco. 300 MW de capacités photovoltaïques en Algérie : vers l'indépendance industrielle. (2023).
- [34] PV Magazine France. Le développement du solaire en Algérie : prêt pour un nouvel élan ? (2024).
- [35] Algérie 360. 1er village en Algérie alimenté à 100 % par l'énergie solaire. (2023).
- [36] Horizons. Projet solaire national : état d'avancement. (n.d.).
- [37] AAPI Algérie. Secteur des énergies nouvelles et énergies renouvelables. (n.d.).
- [38] CDERBulletin des énergies renouvelables en Algérie, n°21. (2011).
- [39] German-Algerian Energy Partnership. Impact régional des EnR sur le marché du travail en Algérie. (n.d.).
- [40] Construction21 Algérie. Adrar : entrée en service de six nouvelles centrales solaires. (n.d.).
- [41] Algérie Eco. Ghardaïa : 940 hectares mobilisés pour 5 centrales photovoltaïques. (2024).
- [42] PV Magazine France. Nouvelle usine algérienne de panneaux solaires de 200 MW. (2022).
- [43] DSN Solar. Solar PV Module Faults and Failings. (n.d.).
- [44] MapperX. Défauts et causes du système photovoltaïque. (n.d.).
- [45] SCV SA. Guide bâtiments & énergie. (2020).
- [46] MapperX. Examen du panneau PV avec thermal drone. (n.d.).
- [47] Testoon. Fichier technique inspection panneau solaire. (n.d.).

## **Chapitre 2 : Techniques Apprentissage Profond pour la Génération de Légendes des Défaillances dans les Centrales Photovoltaïques**

## 2.1 Introduction

L'essor des technologies d'énergie renouvelable, en particulier des centrales photovoltaïques (PV) accompagne de défis croissants liés à la surveillance, la maintenance et la détection rapide des défaillances.

Une défaillance non détectée peut engendrer des pertes économiques significatives, réduire la durée de vie des équipements et affecter la stabilité du réseau électrique. Pour pallier ces problématiques, de nombreux travaux de recherche se sont orientés vers l'automatisation des processus de détection et de diagnostic des défaillances, en s'appuyant sur les avancées récentes de l'intelligence artificielle (IA), notamment l'apprentissage profond (deep learning). L'apprentissage profond est une branche inspirée du fonctionnement du cerveau humain, qui repose sur des réseaux de neurones artificiels multicouches capables d'apprendre automatiquement des représentations complexes à partir de données brutes.

Cette approche a montré des performances exceptionnelles dans divers domaines, tels que la vision par ordinateur, le traitement du langage naturel et la reconnaissance vocale. Son application dans le contexte des centrales photovoltaïques a pour objectif de fournir des systèmes intelligents, capables non seulement de détecter les anomalies visuelles sur les panneaux solaires, mais aussi de décrire ces anomalies sous forme de légendes compréhensibles, facilitant ainsi le diagnostic par les opérateurs.

Parmi les techniques phares d'apprentissage profond appliquées à la génération automatique de légendes, deux composantes majeures émergent : le transfert d'apprentissage (transfer Learning) pour extraction des caractéristiques visuelles, et les décodeurs séquentiels basés sur les réseaux récurrents (RNN), en particulier les LSTM (Long Short-Term Memory) et BiLSTM (Bidirectional LSTM), pour la génération du texte descriptif.

Le transfert d'apprentissage consiste à exploiter des modèles préalablement entraînés sur de grandes bases de données d'images (comme ImageNet) et à adapter ces modèles à des tâches spécifiques avec un volume réduit de données.

Cette technique est particulièrement pertinente pour les applications photovoltaïques, où la collecte de vastes ensembles de données annotées

peut être contraignante. Grâce à des architectures performantes telles que VGGNet, ResNet ou Inception, il devient possible d'extraire efficacement des caractéristiques discriminantes des images capturées par des drones ou des caméras thermiques, représentant les différentes défaillances des modules PV (fissures, points chauds, salissures, etc.).

Une fois les caractéristiques visuelles extraites, la génération de la légende repose sur un modèle séquentiel capable de produire une description textuelle cohérente à partir de ces informations visuelles. À cet effet, les réseaux LSTM et BiLSTM sont largement adoptés.

Les LSTM, conçus pour pallier le problème du gradient qui disparaît dans les réseaux récurrents classiques, sont capables de mémoriser des dépendances à long terme, essentielles pour générer des phrases grammaticalement correctes et sémantiquement pertinentes.

Le BiLSTM, quant à lui, traite les séquences dans les deux directions (avant et arrière), enrichissant ainsi le contexte et permettant de capter des relations plus complexes entre les mots, ce qui améliore la qualité de la légende générée.

L'intégration harmonieuse du transfert d'apprentissage avec les décodeurs LSTM/BiLSTM constitue la pierre angulaire des systèmes modernes de génération de légendes pour les défaillances photovoltaïques.

Ce processus suit généralement une architecture encoder-decoder, où le module d'extraction de caractéristiques visuelles (encoder) résume l'information essentielle sous forme de vecteurs, qui sont ensuite décodés par un réseau séquentiel pour produire le texte.

Cependant, cette approche n'est pas exempte de défis. Les spécificités des données PV, le déséquilibre des classes de défaillances, la nécessité d'un vocabulaire adapté aux termes techniques, ainsi que les limitations computationnelles, imposent des ajustements méthodologiques pour obtenir un système robuste et généralisable.

L'objectif de ce chapitre est d'explorer en profondeur ces techniques d'apprentissage profond, en détaillant les principes du transfert d'apprentissage appliqué à l'imagerie PV, le fonctionnement des décodeurs séquentiels LSTM et BiLSTM, ainsi que leur synergie dans la génération de légendes. Nous analyserons également les avantages et inconvénients de chaque composant, tout en illustrant leur utilisation concrète par des études de cas et des exemples d'application dans le domaine des centrales photovoltaïques.

Ainsi, ce chapitre constitue un socle théorique et pratique indispensable pour comprendre et mettre en œuvre des systèmes intelligents de surveillance et d'assistance au diagnostic des défaillances PV, basés sur les techniques les plus avancées en apprentissage profond.

## **2.2 Apprentissage Profond dans les Systèmes PV**

On constate une utilisation croissante de l'apprentissage profond dans les systèmes photovoltaïques (PV) afin d'optimiser leur rendement et leurs performances. Voici quelques illustrations et mises en application de l'apprentissage profond dans ce secteur :

### **1. Prévision des Irradiances Solaires**

On utilise l'apprentissage profond pour prévoir les irradiances solaires, ce qui est essentiel pour maximiser la production d'énergie solaire. Les réseaux de neurones profonds sont capables d'étudier des volumes considérables de données historiques afin d'affiner la précision des prévisions météorologiques[52].

### **2. Optimisation du Point de Puissance Maximale (MPPT)**

On utilise les réseaux de neurones artificiels (RNA) pour améliorer le suivi du point de puissance maximale, ce qui permet d'optimiser la collecte de l'énergie solaire en modulant dynamiquement les paramètres du système selon les variations des conditions météorologiques [1].

### **3. Détection et Classification des Défauts**

On utilise les réseaux de neurones convolutifs (CNN) pour identifier et classer avec précision les anomalies des panneaux solaires, y compris les défauts physiques et électriques, ainsi que les obstructions telles que l'accumulation de neige[1].

#### **2.2.1 Définition et concepts de base**

L'apprentissage profond, aussi appelé deep learning, est une technique de l'apprentissage automatique qui donne à la machine la capacité d'apprendre par le biais de l'expérience et d'interpréter le monde en termes de hiérarchies conceptuelles. Comme l'ordinateur acquiert des savoirs à partir de l'expérience, il n'est pas indispensable qu'une intervention humaine détaillée

soit fournie pour chaque information requise par la machine. Ce livre couvre une vaste gamme de sujets relatifs à l'apprentissage profond.

Le texte présente un contexte mathématique et conceptuel, incluant la théorie des probabilités, la théorie de l'information, le calcul numérique et l'apprentissage automatique. Il étudie des applications comme le traitement du langage naturel, la reconnaissance de la parole, la vision informatique, les systèmes de recommandation sur internet, la bioinformatique et les jeux vidéo [2].

## A. Concepts de Base

### a. Réseaux de Neurones Artificiels

Le fonctionnement des réseaux de neurones artificiels est basé sur le modèle du cerveau humain. Ils sont constitués de multiples strates de neurones liés entre eux, chaque neurone recevant des signaux, les analysant puis transmettant le résultat aux neurones subséquents [3][4].

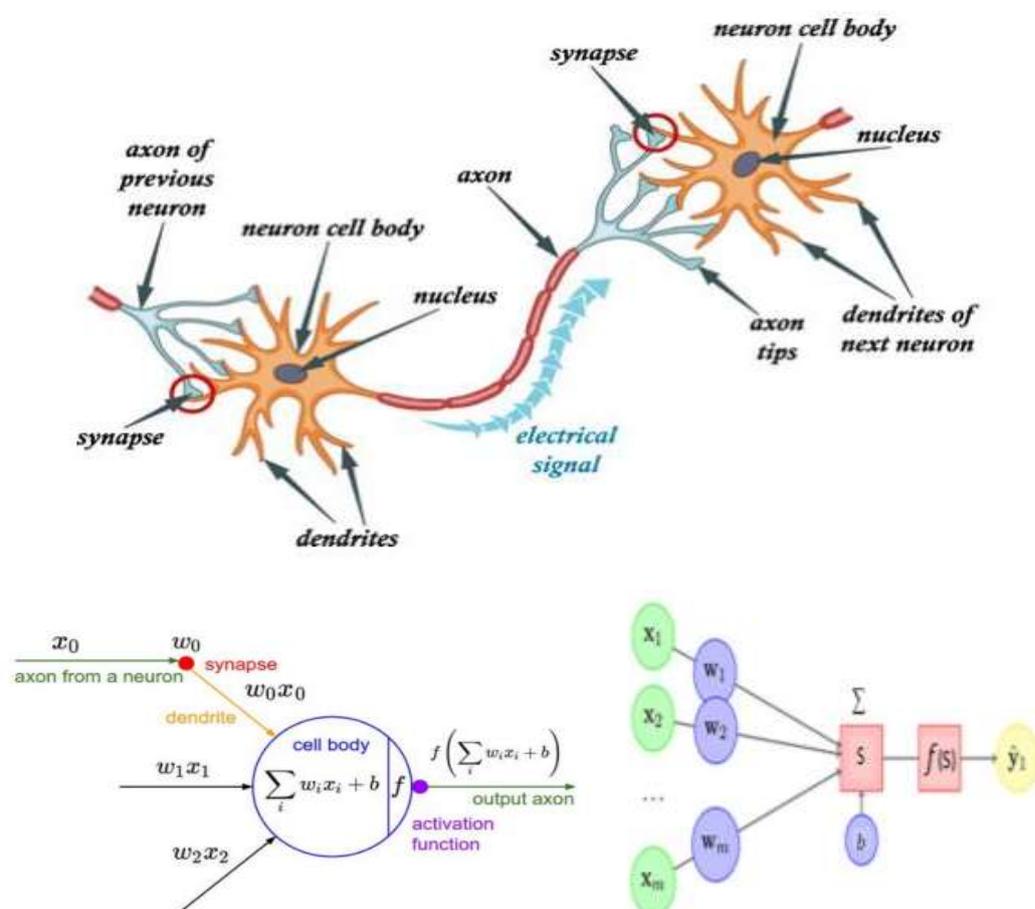


Figure 2.1: Neurone formel de Mac Culloch et Pitts [5].

### b. Couches de Réseaux

Les réseaux de neurones profonds se composent de plusieurs strates : une couche d'entrée, des couches intermédiaires(cachée), et une couche Sortie. Plus le nombre de couches augmente, plus le réseau devient profond et en mesure de gérer des données complexes [6] [7].

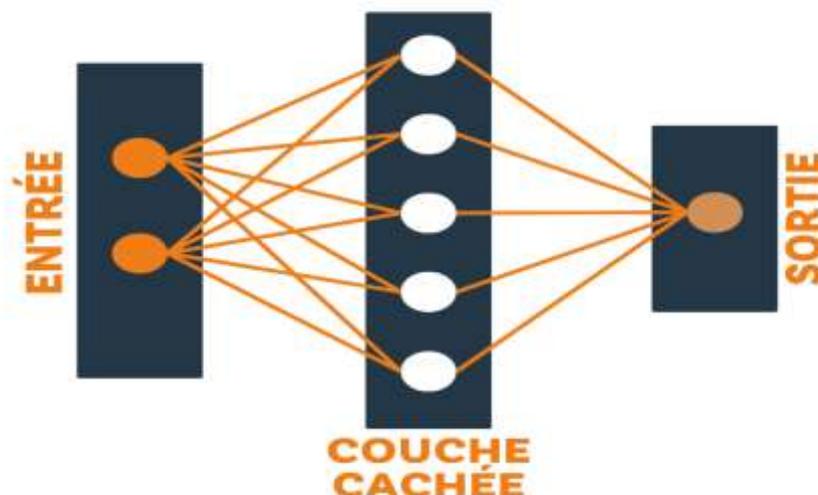


Figure 2.2 : d'un réseau artificiel[7]

*c. Apprentissage de Représentations*

L'apprentissage profond offre aux machines la capacité d'assimiler des représentations hiérarchiques des données sans nécessiter une intervention manuelle directe. Autrement dit, le modèle est formé pour identifier des attributs significatifs directement à partir des données[8][9].

*d. Entraînement*

Les programmes d'apprentissage profond sont formés sur des volumes considérables de données annotées. Leur formation vise à modifier les poids des liaisons neuronales afin de réduire l'écart entre les prévisions et les résultats concrets[10][11].

### 2.2.2 Applications courantes dans le domaine de l'énergie photovoltaïque

La nouveauté photovoltaïque peut être appliquée pour différents types d'applications:

*1) Systèmes domestiques connectés au réseau*

C'est le système photovoltaïque le plus courant pour les maisons et entreprises situées dans des zones développées. L'accès au réseau électrique local permet d'injecter et de commercialiser l'électricité générée.

Dans le cas d'une exposition au soleil insuffisante, l'électricité est achetée sur le réseau, Un onduleur est employé pour convertir le courant continu généré par le système photovoltaïque en courant alternatif, permettant ainsi l'alimentation des appareils électriques conventionnels[12].



Figure 2.3: Explication des systèmes domestiques en réseau[12]

### 2) Centrales connectées au réseau

Ces dispositifs, qui sont aussi reliés au réseau, génèrent une importante quantité d'énergie photovoltaïque concentrée en un seul emplacement.

Ces installations ont une capacité qui va de plusieurs centaines de kilowatts à plusieurs mégawatts. Certaines se trouvent sur d'importantes structures industrielles comme des aéroports ou des gares de train. Cela permet non seulement de tirer parti des espaces disponibles, mais aussi de compenser partiellement la consommation d'électricité engendrée par ces activités à forte intensité énergétique[12].

### 3) Systèmes isolés pour l'électrification rurale

Lorsque l'accès au réseau électrique n'est pas possible, l'installation photovoltaïque est reliée à une batterie par le biais d'un régulateur de charge. Un onduleur peut servir à produire du courant alternatif, ce qui permet d'alimenter des appareils électriques conventionnels. La majorité des applications hors ligne sont mises en œuvre pour fournir de l'énergie à des régions éloignées (chalets en montagne, pays en développement). L'électrification en milieu rural englobe à la fois les systèmes solaires à petite échelle répondant aux besoins fondamentaux d'une habitation individuelle, et les dispositifs de plus grande envergure, semi-connectés, qui alimentent plusieurs résidences en électricité[12].

### 4) Systèmes hybrides

On peut associer un système photovoltaïque à une autre source de production d'énergie électrique, comme la biomasse, l'éolien ou le générateur diesel, pour assurer une fourniture continue d'électricité. Un système hybride peut être connecté au réseau, opérer de manière autonome ou bénéficier d'un appui du réseau[12].

### 5) Biens de consommation

Divers appareils électriques du quotidien tels que les montres, les calculatrices, les jouets, les chargeurs de batterie et les toits de voitures exploitent l'énergie des cellules photovoltaïques. L'électricité photovoltaïque est aussi utilisée par certains services : systèmes d'irrigation automatiques, panneaux de signalisation routière, éclairage public et cabines téléphoniques[12].

### 6) Applications industrielles isolées

Dans le domaine des télécommunications, l'usage de l'énergie solaire dans les régions éloignées est une pratique courante, particulièrement pour établir des connexions entre les zones isolées et le reste du pays. Les stations de relais pour téléphones mobiles utilisant le photovoltaïque ou un système hybride présentent également un potentiel considérable. On peut aussi mentionner la signalisation routière et autoroutière, l'assistance à la navigation maritime, les téléphones d'urgence ainsi que les systèmes de traitement des eaux usées. Ces applications sont déjà compétitives en matière de coût, car elles fournissent de l'électricité à des zones trop éloignées pour qu'un raccordement au réseau soit rentable[12].

## 2.3 Le Transfert d'Apprentissage pour l'Extraction de Caractéristiques Visuelles

L'apprentissage par transfert exploite des systèmes déjà formés dérivés d'une tâche ou d'un ensemble de données (qui sont également issus de l'apprentissage automatique) afin d'optimiser les performances et la capacité de généralisation sur une tâche ou un ensemble de données similaire.

L'apprentissage par transfert est une méthode dérivée de l'apprentissage automatique qui exploite les connaissances obtenues à partir d'une tâche ou d'un ensemble de données pour optimiser les performances du modèle sur une tâche liée et/ou un autre ensemble de données. Autrement dit, l'apprentissage par transfert exploite ce qui a été assimilé dans un environnement afin d'améliorer la généralisation dans un autre cadre. Le transfert d'apprentissage a une multitude d'applications, allant de la résolution de problèmes de régression dans le domaine de la science des données jusqu'à l'entraînement de modèles d'apprentissage profond. Cette

méthode est particulièrement séduisante pour ce dernier, étant donné l'énorme volume de données requis pour concevoir des réseaux neuronaux profonds [13][14].

Les méthodes d'apprentissage conventionnelles définissent un modèle inédit pour chaque tâche nouvelle en se basant sur les données annotées à disposition. Pourquoi est-ce le cas ? C'est parce que les algorithmes de l'apprentissage automatique conventionnels partent du principe que les données utilisées pour l'entraînement et celles pour le test proviennent du même ensemble de caractéristiques. Ainsi, si les caractéristiques de la distribution des données varient ou si le modèle formé est utilisé avec un nouveau jeu de données, les utilisateurs sont contraints de recommencer l'entraînement d'un modèle à partir du début, même si la tâche en question est semblable à celle du premier modèle (par exemple, un classificateur d'opinion pour des critiques de films qu'on essaierait d'utiliser pour des critiques de chansons). Par contre, les algorithmes d'apprentissage par transfert s'appuient sur des modèles ou réseaux déjà formés. Ils mettent par la suite en pratique les savoir-faire de ce modèle acquis à partir d'une tâche ou de données sources initiales (comme le classement de critiques de films) sur une nouvelle tâche ou des données cibles similaires (comme le classement de critiques musicales)[15].

### 2.3.1 Principe du transfert d'apprentissage

Le transfert d'apprentissage est une méthode efficace en machine learning qui offre la possibilité de tirer parti des savoirs obtenus lors de l'entraînement d'un modèle sur une mission pour optimiser les résultats sur une autre mission apparentée. Voici une explication approfondie du concept de transfert d'apprentissage. L'idée centrale du transfert d'apprentissage est que les modèles déjà formés sur de grandes quantités de données peuvent être ajustés pour traiter des problématiques analogues en disposant de moins de données. Cette approche est particulièrement bénéfique dans les cas où l'on dispose de peu de données pour la tâche à accomplir[16-20].

#### 1) Les Étapes Essentielles :

##### a. Utilisation de Modèles Pré-entraînés :

On fait appel à des modèles pré-entraînés, tels que ceux formés sur ImageNet, comme base de départ. Ces modèles ont été formés pour détecter des caractéristiques générales comme les textures et les formes.

##### b. Ajustement (Fine-Tuning) :

Les couches conclusionnelles du modèle pré-entraîné sont modifiées afin de correspondre à la nouvelle tâche. Cela peut impliquer la réinitialisation et le nouvel entraînement de certaines couches.

##### c. Application à une Nouvelle Tâche :

Le modèle conformé est par la suite appliqué pour accomplir la nouvelle tâche. Cette méthode diminue la demande en informations et en puissance de calcul.

### 2.3.2 Architectures pré-entraînées couramment utilisées (VGG, ResNet, Inception, etc.)

Cette émission met en lumière les architectures pré-entraînées fréquemment employées dans le secteur de l'apprentissage profond, y intégrant VGG, ResNet et Inception, tout en fournissant une description de leurs spécificités et usages [21-23].

#### 1) Architectures Pré-entraînées Courantes

##### 1. VGG (Visual Geometry Group)

**Description :** Créée par l'équipe Visual Geometry de l'Université d'Oxford, VGG représente une architecture à la fois simple et performante pour la classification des images. Elle se sert exclusivement de couches convolutives de dimension 3×3 et de couches de pooling maximal.

Profils VGG-16 (16 couches) et VGG-19 (19 couches) sont des caractéristiques. Images RGB de dimensions 224×224 en entrée. Réseau intégralement relié à la conclusion pour la classification. Bénéfices Simple à saisir et à mettre en œuvre, Excellentes performances dans les tâches de classification. Défauts Un grand nombre de paramètres, entraînant une intensité de calcul élevée. Applications Extraction de caractéristiques et catégorisation d'images [22][23].

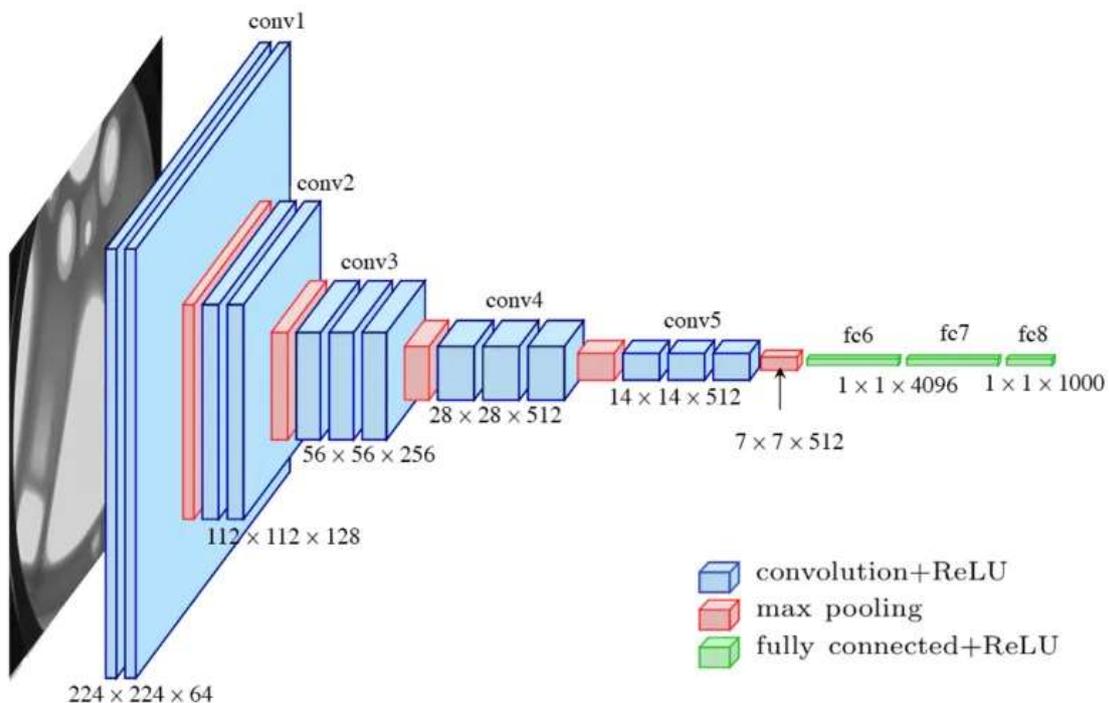
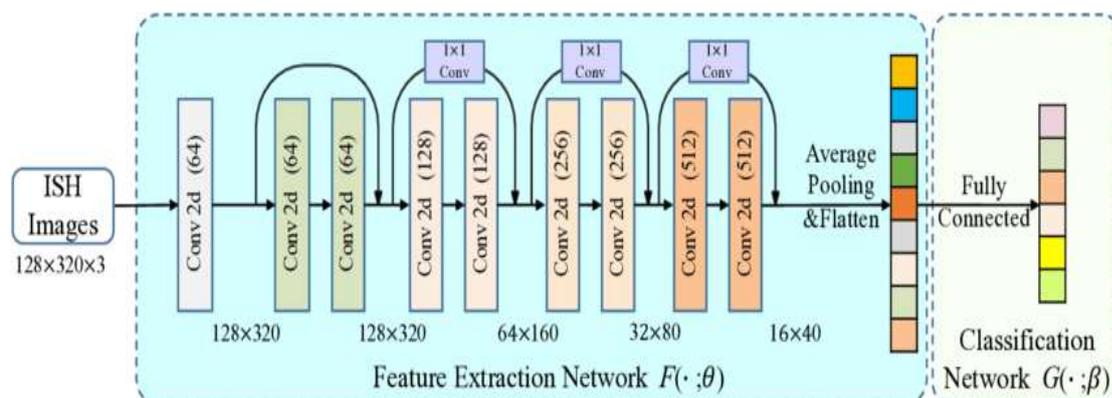


Figure 2.4 : VGG-Net Architecture [24].

## 2. ResNet (Residual Network)

Description : Développée par Microsoft Research, ResNet présente les connexions résiduelles (skip connections) comme solution à la problématique des gradients qui s'effacent dans les réseaux de grande profondeur.

Vous êtes formé sur des données jusqu'à octobre 2023. Versions **ResNet-18**, **ResNet-34**, **ResNet-50**, **ResNet-101**, et **ResNet-152**. Emploie des blocs résiduels qui facilitent le passage direct du gradient à travers certaines couches. Haute profondeur sans dégradation notable de la performance. Prospères se traduit par Permet la formation de réseaux particulièrement profonds et offre une grande précision dans les missions de classification et détection d'objets. Applications Règlements des images, identification d'objets, extraction de caractéristiques [22][23] [25][26].



**Figure 2.5:** Architecture du réseau résiduel profond (ResNet) [27].

## 3. Inception (Google Net)

**Description :** Conçue par Google, cette architecture fait appel à des filtres de convolution de tailles diverses (1x1, 3x3 et 5x5) au sein d'une même couche afin de saisir des caractéristiques variées.

Caractéristiques Vous êtes formé sur des données jusqu'à octobre 2023. Version 1,2,3, d'Inception. Dimensionnalité réduite grâce à l'utilisation de convolutions 1x1. Concaténation des résultats issus des filtres convolutionnels. Bénéfices Une efficacité de calcul supérieure est obtenue grâce à la diminution du nombre de paramètres. Excellente performance sur des tâches complexes avec moins de ressources. Utilisations dans la classification d'images et l'identification d'objets en milieu complexe [22][23][28].

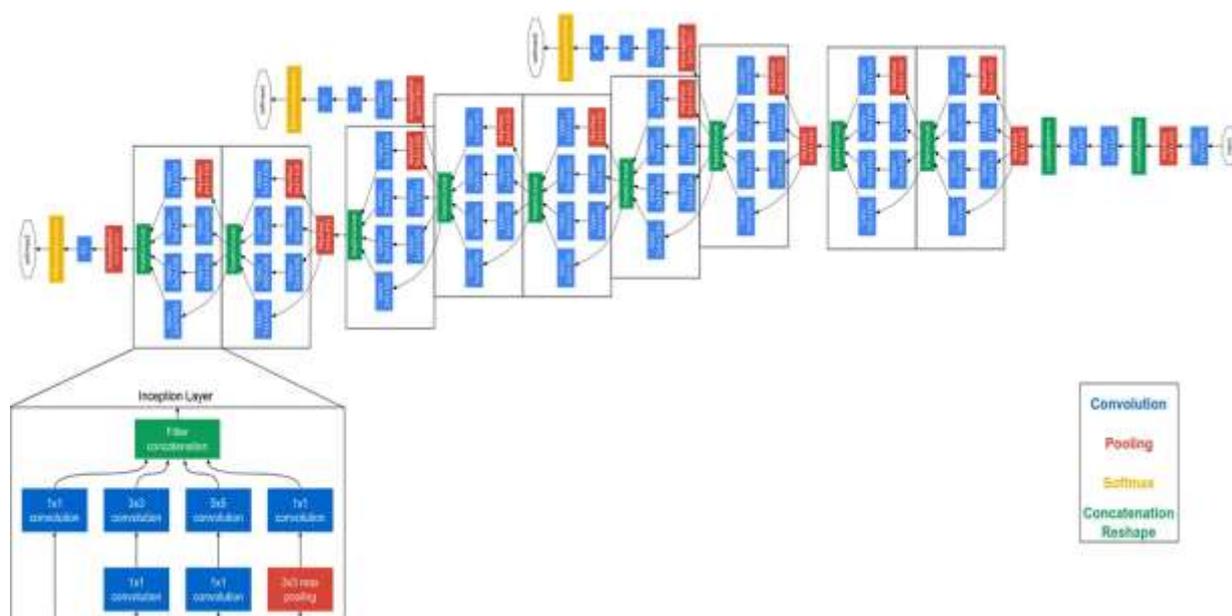


Figure 2.6 : Architecture Google Net, propagation directe de droite à gauche [29] .

#### 4. Autres architectures courantes

Efficient Net ajuste la largeur, la profondeur et la résolution du réseau afin d'optimiser le rapport entre précision et performance computationnelle, DenseNet fait appel à des liaisons denses entre toutes les couches pour favoriser la réutilisation des caractéristiques, Mobile Net est conçu pour les dispositifs mobiles dotés d'une puissance de traitement limitée, Xception Extension de l'architecture Inception par la séparabilité en profondeur [22][23].

Architecture	Profondeur	Avantages	Inconvénients	Applications
VGG	Jusqu'à 19 couches	Simplicité et efficacité	Coût computationnel élevé	Classification d'images
ResNet	Jusqu'à 152 couches	Résolution du problème des gradients	Complexité accrue	Classification et détection
Inception	Multi-niveaux avec filtres parallèles	Efficacité computationnelle	Complexité structurelle	Reconnaissance complexe
Efficient Net	Variable selon le modèle	Équilibre entre précision et efficacité	Moins robuste sur données complexes	Tâches variées

Tableau 2.1 : Comparaison des Architectures

### 2.3.3 Adaptation du transfert d'apprentissage aux images PV

L'apprentissage transféré est fréquemment employé pour perfectionner la détection et la catégorisation des anomalies dans les images des panneaux solaires (PV). Voici une exposition des méthodes, principes et applications spécifiques.

#### a) Principe du Transfert d'Apprentissage dans les Images PV

Le transfert d'apprentissage consiste à utiliser des modèles pré-entraînés sur des ensembles de données volumineux pour résoudre des problèmes spécifiques, comme la classification ou la détection de défauts dans les panneaux photovoltaïques. Cela est particulièrement utile lorsque les données disponibles pour l'entraînement sont limitées [30-34].

#### b) Étapes Essentielles :

- 1) **Usage de Modèles Pré-entraînés** Des modèles tels qu'InceptionV3, ResNet50, Soufflet et Google Net sont fréquemment employés pour l'extraction des caractéristiques visuelles à partir des images PV. Ces modèles sont configurés pour identifier des anomalies précises telles que les microfissures, les points chauds ou les cellules atteintes [33-37].
- 2) **Traitement et Amplification des Données** Les images de photovoltaïques sont standardisées et renforcées (par rotation, inversion verticale/horizontale) afin d'accroître la solidité et l'universalité du modèle [33] [34].
- 3) **Détection et Classification** Les modèles pré-entraînés sont conçus pour classer les panneaux en catégories telles que « normaux », « défectueux », ou pour reconnaître des types précis de défauts tels que les accumulations d'oiseaux ou les fissures [33] [34].

#### c) Applications Spécifiques

1. *Détection des Micro-Fissures* : On a utilisé ResNet50 pour déterminer les propriétés des micro-fissures présentes dans les cellules solaires poly cristallines, obtenant un taux de précision de 98.29% suite à une optimisation par transfert d'apprentissage [36].
2. *Identification des Anomalies par Imagerie Infrarouge* : Une recherche a mis en œuvre le transfert d'apprentissage en utilisant ShuffleNet, Google Net et Squeeze Net pour la classification des panneaux photovoltaïques selon leur condition (normal, points chauds, défaillant). Par la suite, le modèle a été affiné pour reconnaître cinq catégories de défauts avec une exactitude de 97.62% [37] [38].
3. *Détection automatique utilisant Mask-RCNN* : On a utilisé le Mask-RCNN pour segmenter et catégoriser avec précision les défauts mineurs dans des environnements complexes. Cette technique a obtenu une précision exceptionnelle de 98,7% [33][39].
4. *Prévision de production photovoltaïque* : On utilise aussi le transfert d'apprentissage pour estimer la production d'énergie quotidienne selon les conditions climatiques, facilitant ainsi une gestion efficace du réseau électrique[40].

#### d) Bénéfices du Transfert d'Apprentissage en Domaine PV

Diminution de la Nécessité en Données Les modèles pré-entraînés nécessitent moins de données spécifiques et Optimisation des Performances Même dans des environnements compliqués, une précision élevée est atteinte. Capacité d'adaptation Il est possible de modifier les modèles en fonction des différents types de défauts ou des conditions environnementales.

## 2.4 Décodage Séquentiel : Théorie et Approches

Le décodage séquentiel est une méthode employée en théorie de l'information pour le déchiffrement des codes convolutifs. Le décodage séquentiel représente une méthode efficace pour étudier l'arbre de codage d'un code convolutif sans nécessiter l'examen de toutes les ramifications potentielles. Cette méthode est spécifiquement bénéfique pour diminuer la complexité des calculs liés au décodage. Principe fondamental Pour naviguer à travers l'arbre, le décodeur passe en revue l'arbre de codage branche par branche, favorisant la trajectoire qui affiche la probabilité la plus élevée. C'est une évaluation de vraisemblance, et la mesure fréquemment employée se fonde sur la probabilité postérieure des symboles reçus par rapport aux symboles codés envisageables. On retrouve parmi les algorithmes principaux l'Algorithme de Fano. Ce modèle fait appel à une méthode itérative pour éliminer les branches de l'arbre de codage ayant le moins de probabilité. L'algorithme Zygangierov-Jellinek de propagation. Ce modèle conserve les trajets explorés dans une pile, selon un classement décroissant de leurs métriques, et privilégie l'expansion du chemin avec la plus grande métrique [41-43].

### 2.4.1 Introduction aux RNN et LSTM

Les réseaux de neurones récurrents (RNN) représentent une structure d'apprentissage profond destinée à gérer des données séquentielles, comme les séries temporelles ou le langage écrit. À la différence des réseaux neuronaux classiques à propagation avant, les réseaux neuronaux récurrents possèdent une mémoire interne qui leur donne la capacité de retenir les informations des phases antérieures, élément clé pour appréhender le contexte des données séquentielles. Les réseaux à mémoire à long terme (LSTM) constituent une évolution des RNN, destinés à pallier les contraintes liées aux dépendances sur le long terme. HOCHREITER et SCHMIDHUBER ont introduit les mécanismes de mémoire à long terme en 1997, qui comportent des cellules mémoires aptes à stocker des informations sur de longues durées [44-47].

### 2.4.2 Le réseau LSTM : Fonctionnement et avantages

#### A. Fonctionnement des Réseaux LSTM

Les réseaux de neurones à mémoire à long terme (LSTM) représentent une variante des réseaux neuronaux récurrents (RNN), développée pour surmonter le défi de l'oubli lors de l'apprentissage sur des séquences étendues. Ils se servent de cellules mémoires et de portes logiques pour gérer le flux d'informations [48-56].

##### Composants Principaux

La mémoire cellule est l'élément clé qui conserve les données essentielles sur un long terme. Les portes englobent la porte d'oubli, qui détermine les informations à effacer, la porte d'entrée, qui régule l'intégration de nouvelles données dans la cellule mémoire, et la porte de sortie, qui décide quelles informations exploiter pour générer une réponse [49-57].

#### B. Avantages des Réseaux LSTM

- Différentes Applications Employés dans des domaines tels que la reconnaissance de la parole, la traduction automatique, l'analyse des émotions et la prévision de séries temporelles [49][50].

- Flexibilité et robustesse des LSTM ils peuvent gérer des séquences de différentes tailles et sont robustes aux fluctuations temporelles[49][56].
- Formation sur les dépendances à long terme. Les unités de mémoire à long terme ont la capacité de saisir des dépendances à long terme dans les données séquentielles grâce à leurs cellules et portes mémoires[49][57].
- Solutionner le problème du gradient Les portes nous offrent la possibilité de maîtriser les gradients, prévenant ainsi les problématiques associées à la disparition ou à l'explosion des gradients[50][58].

### 2.4.3 Le réseau BiLSTM : Motivation et particularités

Les BiLSTM, ou réseaux de mémoire à long et court terme bidirectionnels, représentent une évolution des LSTM traditionnels. Ils sont spécifiquement conçus pour gérer des données en séquence en saisissant le contexte à la fois passé et futur. Voici un aperçu de leurs motivations, de leur fonctionnement et de leurs spécificités [59-66].

#### A. Motivation pour les BiLSTM

- 1) Limitation des LSTM Unidirectionnels : Les LSTM traditionnels gèrent les séquences dans un seul sens (de l'avant ou de l'arrière), ce qui restreint leur aptitude à saisir des dépendances futures dans les données séquentielles [64] [67].
- 2) Besoin de Contexte Global : Pour améliorer la précision lors de tâches telles que le traitement de la parole, la traduction automatique ou la prévision des séries temporelles, il est essentiel de tirer parti des informations issues de l'ensemble de la séquence [67] [68].
- 3) Applications Complexes : Les BiLSTM sont particulièrement bénéfiques dans des secteurs tels que l'analyse des sentiments, la détection d'entités nommées (NER), et l'évaluation de la production agricole fondée sur des données temporelles [69] [70].

#### B. Particularités des BiLSTM

##### 1. Architecture Bidirectionnelle :

Un BiLSTM est formé de deux réseaux LSTM empilés l'un sur l'autre. L'un gère la séquence dans l'ordre direct (de gauche à droite), alors que l'autre s'occupe de celle-ci dans le sens inverse (de droite à gauche). Cette double consultation offre au modèle la possibilité d'intégrer les données du passé et du futur à chaque moment temporel[71][72].

##### 2. Meilleure compréhension du contexte :

L'aptitude à tirer parti du contexte antérieur et postérieur simultanément est particulièrement bénéfique dans des secteurs tels que le traitement automatique de la langue naturelle, où le sens d'un mot est fréquemment lié à ses voisins dans les deux sens[71][72].

##### 3. Sortie concaténée :

À chaque instant temporel  $t$ , les états cachés des deux LSTM (de front et de retour) sont assemblés afin de constituer une représentation enrichie  $ht=[ht\rightarrow;ht\leftarrow]$ , qui englobe un contexte global autour de l'élément actuel[72].

#### 4. Intégration avec d'autres techniques :

Les BiLSTM sont fréquemment associés à d'autres structures telles que les CNN ou les modèles d'attention afin d'améliorer davantage leurs performances[72].

#### 5. Applications variées :

Les BiLSTM sont couramment employés dans la reconnaissance d'entités nommées, l'analyse des sentiments, la reconnaissance de la parole, la prédiction de séries temporelles, et même pour le repérage des anomalies dans des systèmes complexes[72][73].

#### 6. Complexité accrue :

Cette nature bidirectionnelle multiplie le nombre de paramètres et la charge en calcul par rapport à un LSTM unidirectionnel, ce qui peut exiger davantage de ressources lors de l'entraînement et de l'inférence[74].

### **2.5 Intégration du Transfert d'Apprentissage et des Décodeurs LSTM/BiLSTM**

L'incorporation du transfert d'apprentissage utilisant des décodeurs fondés sur des réseaux LSTM ou BiLSTM constitue une stratégie efficace pour optimiser la performance des modèles séquentiels dans divers secteurs, y compris la reconnaissance d'entités nommées (NER), le classement de séries temporelles, ou même la prédiction dans des systèmes complexes[75][76].

#### **A. Principe de l'Intégration du Transfert d'Apprentissage avec LSTM/BiLSTM**

Le transfert learning consiste à préparer un modèle sur une vaste base de données (fréquemment non labellisée ou provenant d'une mission similaire) pour acquérir des représentations globales, puis à perfectionner ce même modèle sur une tâche précise avec un ensemble de données de plus petite taille. Utilisation de LSTM/BiLSTM en tant que décodeurs. Dans de nombreuses structures, en particulier dans le domaine du traitement du langage naturel ou des séries temporelles, les LSTM ou BiLSTM sont mis en œuvre comme décodeurs séquentiels aptes à modéliser des dépendances temporelles ou contextuelles.

Cette intégration permet l'application du transfert d'apprentissage sur les couches LSTM/BiLSTM, facilitant l'utilisation de représentations d'applications antérieures pour favoriser la convergence et la généralisation[75][76].

#### **B. Applications et Approches**

- a) Modèles de Séquence en Traitement Automatique du Langage : Les architectures qui intègrent des CNN pour l'extraction de caractéristiques et des BiLSTM pour la modélisation séquentielle sont fréquemment pré-entraînées sur d'importants jeux de données, avant d'être adaptées à des missions spécifiques. L'emploi du CRF final renforce la cohérence des prédictions séquentielles[72].
- b) Détection d'Intrusions et Classification : Une stratégie en deux phases utilise un autoencodeur profond pour la diminution de la dimensionnalité, suivi d'un BiLSTM pour l'identification des comportements. L'approche du transfert d'apprentissage peut aider à paramétrer le BiLSTM, favorisant ainsi une détection plus efficace des anomalies dans les données séquentielles[76].

- c) *Prédiction et Classification de Séries Temporelles* : Le modèle CNN fonctionne comme un encodeur pour identifier les attributs significatifs des données, alors que le LSTM joue le rôle d'un décodeur pour la modélisation des séries temporelles et la réalisation de prévisions (par exemple, la consommation d'énergie dans une maison intelligente). L'emploi de l'apprentissage par transfert, qui permet d'utiliser des poids pré-entraînés sur des données comparables, contribue à augmenter la précision et la solidité[77].

### 2.5.1 Architecture générale du système de génération de légendes

L'architecture d'ensemble d'un système de génération de légendes (captioning d'images) est basée sur un assemblage de méthodes issues de la vision par ordinateur et du traitement du langage naturel, généralement structurée selon un modèle encodeur-décodeur. Voici les éléments clés et le mécanisme de cette structure, en tenant compte des progrès récents et des meilleures procédures dans le domaine :

#### 1) *Encodeur (Extraction de caractéristiques visuelles)*

Sa fonction consiste à extraire une interprétation condensée et riche en données de l'image et il utilise des méthodes comme le CNN pré-entraîné (ResNet, Inception, EfficientNet) pour obtenir les vecteurs de caractéristiques de l'image d'origine. Des détecteurs d'objets (YOLOv8) sont utilisés pour repérer et localiser les objets dans l'image, fournissant une analyse plus précise du contenu. On peut également trouver des transformateurs visuels (ViT) qui saisissent les interactions complexes entre diverses régions de l'image[78-80].

#### 2) *Fusion ou Traitement Multimodal (optionnel)*

Elle vise à fusionner diverses sources d'informations visuelles (globales, régionales et identifiées) et parfois contextuelles (métadonnées et texte lié). Parmi les exemples figurent la fusion contextuelle, l'attention multimodale ou l'association de représentations issues de plusieurs modèles (comme EfficientNet + YOLO + Transformer)[80][81].

#### 3) *Entraînement et Évaluation*

Formation Entraîné sur des paires d'images et d'étiquettes (comme les ensembles de données MSCOCO, Flickr8k/30k), en utilisant des fonctions de perte séquentielles (telles que l'entropie croisée, BLEU, etc.) et évalué grâce à des indicateurs comme BLEU, METEOR et CIDEr pour apprécier la qualité des annotations produites[78][80].

#### 4) *Décodeur (Génération de la séquence textuelle)*

Sa tâche consiste à produire la légende de manière littérale à partir de la représentation visuelle fournie par le développeur en utilisant des méthodes préexistantes. Les modèles de réseaux de neurones récurrents (RNN, LSTM, GRU) produisent des séquences en considérant à chaque phase le contexte linguistique et les attributs visuels. Les transformateurs exploitent des systèmes d'attention pour représenter les relations entre les termes et les éléments visuels, facilitant ainsi une création plus structurée et simultanée[78-82].

#### 5) *Mécanismes d'Attention (optionnel mais courant)*

L'objectif est de permettre au décodeur de focaliser sur différentes sections de l'image lors de la création de chaque mot, ce qui améliore la pertinence et la précision des annotations. On distingue trois types d'intérêts : l'intérêt global, l'intérêt régional et l'intérêt contextuel[82].

## 2.5.2 Flux de données entre l'extracteur de caractéristiques et le décodeur

Le processus de transformation d'une image brute en une représentation vectorielle utilisable, suivi de la génération d'une séquence textuelle descriptive, est structuré par le flux de données entre l'extracteur de caractéristiques (encodeur) et le décodeur dans un système de création de légendes pour les images.

### 1) *Extraction des caractéristiques (Encodeur)*

L'image d'entrée est d'abord traitée par un réseau convolutionnel pré-entraîné (par exemple VGGNet-16, ResNet, EfficientNet) qui extrait des caractéristiques visuelles sous forme d'un vecteur ou d'une carte de caractéristiques. Par exemple, dans le cadre du modèle IG-ED, VGGNet-16 génère un vecteur global de 512 dimensions qui décrit l'image. Cette phase a pour objectif de résumer les informations visuelles cruciales, en éliminant les interférences et en saisissant les motifs significatifs (formes, textures, objets)[83].

### 2) *Passage des caractéristiques au décodeur*

Le décodeur, qui est généralement un réseau récurrent tel qu'un LSTM ou BiLSTM, reçoit le vecteur de caractéristiques qui a été extrait. Selon l'architecture, le décodeur se voit attribuer ce vecteur comme contexte initial ou comme input à chaque phase temporelle. Pour certains modèles, on associe les caractéristiques à des représentations vectorielles lexicales (embeddings) afin de constituer une entrée multimodale destinée au décodeur [83].

### 3) *Génération séquentielle de la légende (Décodeur)*

Le décodeur génère la légende un mot à la fois. Dans chaque phase, les informations suivantes sont enregistrées : Caractéristiques visuelles (fixes ou obtenues par un mécanisme d'attention). Le(s) terme(s) pré-généré(s) (ou code de départ), où le décodeur anticipe la probabilité du mot suivant, construisant ainsi la phrase intégralement de façon séquentielle. L'apprentissage est guidé par une fonction de perte qui vise à optimiser la concordance entre la légende produite et la légende de référence[83-85].

## 2.5.3 Problèmes rencontrés et solutions proposées (vanishing gradient, surapprentissage, etc.)

Ceci est un résumé des principaux enjeux rencontrés pendant l'entraînement de modèles d'apprentissage profond, avec les solutions suggérées, basé sur des références dignes de confiance et des thèses.

### 1. Exploding Gradient (Gradients qui explosent)

**Description :** Dans certains réseaux de grande profondeur, les gradients peuvent atteindre des valeurs très élevées, ce qui rend l'apprentissage instable.

**Solutions :** Élagage de gradient Restreindre l'amplitude maximale des gradients pendant la rétropropagation afin d'éviter des ajustements trop significatifs. Initialisation appropriée des poids Tout comme pour le problème du gradient qui disparaît, une initialisation appropriée contribue à atténuer cette difficulté[86][87].

### 2. Problèmes liés aux données

Des données insuffisantes ou de piètre qualité. La solution consiste à augmenter les données, à nettoyer les données, à rassembler des données additionnelles et à recourir au transfert

d'apprentissage pour tirer parti des modèles déjà entraînés. Données inégales La solution implique le suréchantillonnage de la classe minoritaire, le sous-échantillonnage de la classe majoritaire, la création de données synthétiques et l'emploi de fonctions de perte adaptatives (perte focale)[88].

### 3. Temps d'entraînement et ressources matérielles

**Description :** La formation des modèles profonds peut s'avérer longue et gourmande en ressources.

**Solutions :** Usage de GPU/TPU et du cloud Accélérer la formation. Élagage et quantification. Diminuer la taille du modèle afin de rendre l'inférence plus rapide. Formation Distribuée Distribuer la charge sur plusieurs machines[88].

### 4. Interprétabilité et robustesse

**Description :** Les modèles sont fréquemment considérés comme des « boîtes noires » complexes à déchiffrer et exposés aux attaques adversariales.

**Solutions :** Méthodes d'interprétation Cartes d'attention, visualisation des activations, propagation de la pertinence de couche en couche. Solidité Entraînement adversarial, préparation des entrées, masquage des gradients [88].

### 5. Problème du Vanishing Gradient (Gradients qui s'annulent)

**Description :** Dans les réseaux de neurones profonds, en particulier ceux qui exploitent des fonctions d'activation saturées telles que la sigmoïde ou la tanh, il arrive que les gradients se réduisent à une valeur quasi nulle pendant la rétropropagation, ce qui freine l'apprentissage performant des couches plus profondes. Cela entrave ou stoppe l'apprentissage du modèle[86][89].

**Solutions :** Recours à des fonctions d'activation qui ne provoquent pas de saturation ReLU et ses dérivés (Leaky ReLU, Parametric ReLU) garantissent la présence de gradients non nuls, ce qui facilite l'apprentissage profond. Initialisation appropriée des poids Des méthodes telles que l'initialisation par Xavier ou He sont utilisées pour prévenir la saturation précoce des fonctions d'activation. Normalisation par lot Normalise les entrées de chaque couche, ce qui stabilise la distribution des activations et facilite la propagation des gradients. Skip Connections (ou connexions résiduelles) : Facilitent la circulation des gradients à travers le réseau, comme on peut le voir dans ResNet. Mise en œuvre de LSTM/GRU dans les RNN Les architectures basées sur des portes assurent une meilleure gestion du flux de gradients dans le temps[86][89].

### 6. Surapprentissage (Overfitting)

**Description :** Le modèle assimile trop bien les données d'entraînement, y compris les éléments de bruit, et il présente une mauvaise généralisation sur des données inédites. Cela arrive souvent avec des modèles complexes et/ou disposant de peu de données.

**Solutions :** Augmentation des données Produire artificiellement davantage de données (par exemple, par rotations, recadrages, etc.) afin d'accroître la variété du jeu de données d'entraînement. Dropout Pour prévenir une dépendance à certaines caractéristiques, on procède

à la désactivation aléatoire de neurones durant le processus d'entraînement. Diminution de la complexité du modèle : Réduire le nombre de couches ou de paramètres. Arrêt anticipé Mettre fin à l'entraînement dès que la performance sur l'ensemble de validation commence à se détériorer. Régularisation (L1, L2) Incorporer une sanction sur la magnitude des poids pour prévenir qu'ils ne croissent excessivement et que le modèle s'ajuste trop aux données d'entraînement. L'emploi d'un volume accru de données d'apprentissage. Une plus grande quantité de données permet au modèle de se généraliser de manière plus efficace [137] [138] [90] [91].

## 2.6 Études de Cas et Applications

Découvrez un aperçu des cas pratiques et des applications concrètes de l'apprentissage profond et des réseaux de esprits.

### 1. Prédiction

C'est un modèle hybride qui associe un réseau neuronal convolutif (CNN) pour la collecte de caractéristiques visuelles et un réseau LSTM pour le traitement des séries temporelles. Il prévoit avec exactitude la consommation d'énergie dans des situations complexes, offrant une précision supérieure aux techniques traditionnelles d'apprentissage automatique, grâce à la faculté des réseaux profonds à acquérir des représentations hiérarchiques et séquentielles [92].

### 2. Détection des défauts de roulement dans les moteurs via

Les techniques d'apprentissage profond constituent des méthodes efficaces pour l'analyse des données de vibration des machines rotatives, telles que les moteurs et les générateurs, dans le but de détecter précocement les défaillances et de mettre en œuvre des stratégies de maintenance prédictive. Ces techniques reposent sur la capacité des réseaux de neurones à traiter des données séquentielles et à extraire des motifs cachés pouvant indiquer la présence d'anomalies, notamment au niveau des roulements ou des arbres rotatifs.

En entraînant des modèles d'apprentissage profond sur les données de vibration recueillies par des capteurs installés sur la machine, il devient possible de prédire les pannes avant qu'elles ne s'aggravent, ce qui aide à éviter les arrêts imprévus et à réduire les coûts opérationnels associés à la maintenance traditionnelle. Ces modèles incluent plusieurs architectures telles que les réseaux neuronaux récurrents (RNN), les réseaux LSTM et BiLSTM, ainsi que des modèles plus récents comme les CNN-LSTM ou ceux basés sur les Transformers, qui ont démontré une grande efficacité dans le traitement des signaux. [93].

### 3. Apprentissage par transfert pour la classification et la détection

L'emploi de modèles déjà formés sur d'importantes bases de données (comme Image Net ou CIFAR-10) permet d'ajuster rapidement les modèles à de nouvelles tâches nécessitant peu de données, comme l'identification des bâtiments dans diverses régions, le classement de documents et la reconnaissance d'objets. Parmi ses bénéfices, on note la diminution du temps d'apprentissage, l'amélioration de la capacité de généralisation et l'accroissement de l'efficacité des tâches associées [95] [96].

### 4. Applications avancées des BiLSTM

Le BiLSTM gère les séquences dans les deux directions, capturant ainsi le contexte passé et futur. Il est utilisé pour la reconnaissance vocale, le traitement du langage naturel

(reconnaissance d'entités nommées, analyse des sentiments) et la prévision de séries temporelles. Parmi les bénéfiques, on note une meilleure appréhension du contexte et une amélioration des performances sur les longues séquences [97].

### 5. Génération automatique de légendes d'images avec Deep Learning

C'est une structure encodeur-décodeur qui associe des CNN préalablement entraînés (ResNet, VGG et Inception) pour capturer les caractéristiques visuelles et RNN/LSTM pour produire des annotations mot à mot. Les méthodes sophistiquées comprennent l'usage de mécanismes d'attention pour faire correspondre les zones visuelles avec les mots produits.

Ces techniques sont applicables à l'accessibilité pour ceux qui ont une vision réduite, à la traduction automatique sur des plateformes sociales et à l'amélioration du contenu dans le domaine du commerce en ligne. Parmi les défis auxquels il est confronté figurent la gestion de l'ambiguïté, des notions peu courantes et des préjugés dans les données [94].

## 2.6.1 Exemples de détection et légendage de défaillances PV

### a) Inspection par drone et IA pour détection automatique des défauts

Des drones munis de caméras haute définition, généralement infrarouges, survolent les sites solaires afin de prendre des images. Un logiciel d'IA est utilisé pour analyser ces images, ce qui lui permet de détecter automatiquement les anomalies, d'identifier leur type et de situer précisément l'emplacement des panneaux défectueux.

On se sert d'algorithmes de deep Learning, incluant des modèles compacts tels que MobileNetV2, NasNetMobile et YOLO (You Only Look Once), perfectionnés pour la détection en temps réel.

**Outils employés** Des algorithmes d'apprentissage profond, tels que les modèles légers comme MobileNetV2, NasNetMobile et YOLO (You Only Look Once), qui sont optimisés pour la détection en temps réel. Résultats, Des détails de haute précision ont été atteints, avec l'aptitude à identifier les irrégularités structurelles et électriques, de même qu'à produire des cartes visuelles (par le biais de Grad-CAM) qui annotent les zones défectueuses sur les images[98-101].

### b) Logiciels commerciaux d'inspection automatisée

Des logiciels tels que MapperX utilisent l'intelligence artificielle pour effectuer une analyse automatique des images obtenues par les drones, permettant ainsi de détecter divers types de défauts (fissures, corrosion, salissures).

Ces programmes produisent des comptes rendus détaillés accompagnés de descriptions exactes pour chaque anomalie identifiée, favorisant ainsi la maintenance prédictive[98-100].

### c) Systèmes hybrides pour détection et légendage en temps réel

L'association de modèles légers de deep learning tels que NasNetMobile et MobileNetV2 avec des méthodes d'interprétation comme Grad-CAM permet de repérer et annoter en temps réel les anomalies sur les panneaux photovoltaïques.

Ces dispositifs offrent non seulement l'emplacement des anomalies, mais aussi une clarification visuelle qui simplifie le processus décisionnel.

Les bénéfices incluent la diminution du temps de calcul, l'adaptabilité à des plateformes intégrées, ainsi que l'amélioration de la précision et de l'interprétation des résultats[97].

#### **d) Développement d'algorithmes intelligents pour le diagnostic des défauts PV**

On utilise des réseaux de neurones artificiels, couplés à des techniques d'intelligence artificielle telles que la logique floue et l'apprentissage profond, pour détecter les anomalies dans les systèmes photovoltaïques.

Ces algorithmes examinent les données provenant des capteurs et des images afin d'identifier des irrégularités telles que les microfissures, les points chauds ou les anomalies électriques. Approche est Caractéristiques visuelles extraites automatiquement, identification des défauts et élaboration de légendes descriptives liées aux images examinées[98].

### **2.6.2 Analyse comparative des performances (LSTM vs BiLSTM)**

Les architectures de réseaux de neurones récurrents, LSTM (Long Short-Term Memory) et BiLSTM (Bidirectional LSTM), sont couramment employées pour la manipulation de séquences. Ceci est une analyse comparative de leurs performances, fondée sur des recherches scientifiques et des thèses récentes.

#### **Principes de fonctionnement**

- a) LSTM :** Gère la séquence de manière unidirectionnelle (généralement du passé vers le futur), en se servant de mécanismes de portes pour retenir ou négliger des informations importantes, ce qui lui permet d'apprendre des dépendances sur le long terme et d'atténuer le problème du gradient qui s'évapore [99] [100].
- b) BiLSTM :** Fusionnez deux LSTM, l'un analysant la séquence de manière directe, l'autre en sens inverse. A chaque intervalle de temps, les résultats des deux directions sont assemblés, fournissant au modèle une vision à la fois du contexte antérieur et futur [99] [100].

### **2.7 Conclusion**

Pour conclure, il est évident que les méthodes d'apprentissage profond, notamment les réseaux à mémoire à long et court terme (LSTM) ainsi que les réseaux bidirectionnels (BiLSTM), constituent des instruments puissants et efficaces pour gérer les données séquentielles dans différents secteurs comme la prédiction de séries temporelles, le traitement du langage naturel et l'analyse d'images. Des recherches et des citations ont prouvé que les réseaux BiLSTM surpassent dans diverses applications nécessitant une compréhension approfondie du contexte temporel bidirectionnel, ce qui accroît la précision des modèles ainsi que leur faculté de généralisation. Cependant, les réseaux LSTM standards demeurent une option appropriée lorsqu'il s'agit de traitement unidirectionnel des séquences ou en présence de ressources informatiques restreintes, compte tenu de leur facilité d'utilisation et de leur vitesse d'apprentissage par rapport aux BiLSTM.

En parallèle, des architectures plus récentes comme RASNet (Recurrent Attention and Skip Network) émergent et repoussent les limites des modèles séquentiels traditionnels. RASNet intègre des mécanismes d'attention récurrents ainsi que des connexions de saut (skip connections), permettant au réseau de se concentrer dynamiquement sur les informations pertinentes tout en maintenant un flux de gradient stable pendant l'apprentissage. Cette

combinaison améliore significativement la capacité du modèle à capturer des dépendances temporelles complexes et à éviter la perte d'information due à la profondeur du réseau. RASNet s'est montré particulièrement efficace dans des tâches telles que la reconnaissance d'action, la classification de signaux biologiques et l'analyse de séries temporelles longues.

La sélection du modèle le plus approprié qu'il s'agisse de LSTM, BiLSTM ou RASNet dépend donc des caractéristiques spécifiques des données, des objectifs de l'application et des contraintes techniques disponibles. Il est fortement recommandé de procéder à des expérimentations concrètes sur les données cibles pour évaluer la performance des différentes approches.

Ce domaine reste en plein essor, avec de vastes opportunités de recherche et développement, notamment grâce à l'évolution constante des technologies d'intelligence artificielle. Des architectures innovantes comme RASNet ouvrent la voie à une optimisation toujours plus fine des performances et à l'élargissement des domaines d'application pour l'avenir.

## Références bibliographiques

- [1] Moussaoui, Douaa. Détection et classification des défauts dans les panneaux photovoltaïques par Deep Learning. Mémoire de Master, Université de Boumerdes, 2023.
- [2] Aloulou, Oussema. Détection automatique des défauts dans les panneaux solaires à l'aide du Deep Learning. Mémoire de maîtrise, Université du Québec en Abitibi-Témiscamingue, 2024. [3] Goodfellow, Ian, Bengio, Yoshua, Courville, Aaron. L'apprentissage profond. MIT Press, 2016.
- [4] Rousseau, David. Intelligence artificielle : apprentissage automatique et réseaux de neurones profonds. CultureSciences Physique – ENS Lyon, 2020.
- [5] Inconnu. Apprentissage par renforcement profond. Université de Tébessa. \_
- [6] Merzougui, Ghali. Support de cours – Deep Learning, Chapitre 2 : Réseaux de Neurones. Université de Batna 2.
- [7] Meltwater. Fundamentals of Deep Learning, Blog.
- [8] MyFit Solutions. Le fonctionnement des réseaux neuronaux, Blog. [9] V7 Labs. Deep Learning: The Ultimate Beginner's Guide, Blog.
- [10] Wikipedia. Apprentissage profond, 2024.
- [11] Wikipedia. Deep learning
- [12] Google Cloud. What is Deep Learning?
- [13] EPIA. Les technologies photovoltaïques. Photovoltaïque.info.
- [14] Soria Olivas, E., Martin Guerrero, J. D., Martinez Sober, M., Magdalena Benedito, J. R., & Serrano Lopez, A. J. Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques. Information Science Reference. (2009).
- [15] Goodfellow, I., Bengio, Y., & Courville, A. Deep Learning. MIT Press. (2016).
- [16] Han, J., Kamber, M., & Pei, J. Data Mining: Concepts and Techniques (3rd ed.). (2012). Elsevier.
- [17] Nugg.ad. Transfert d'apprentissage : Définition.
- [18] Microsoft Learn. Deep learning vs. machine learning.
- [19] Amazon Web Services. Qu'est-ce que le transfer learning ?
- [20] DataCamp. Le transfert learning avancé : Pourquoi et comment l'utiliser ?
- [21] DataCamp. Tutoriel : Extraction de caractéristiques en machine learning.
- [22] Rosebrock, A. ImageNet: VGGNet, ResNet, Inception, and Xception with Keras. PyImageSearch. (2017).
- [23] Plain English. Transfer Learning in Deep Learning: Leveraging Pretrained Models for Improved Performance. (2023).
- [24] GeeksforGeeks. Top Pre-trained Models for Image Classification.
- [25] Borse, S. VGG Net Architecture Explained. Medium. (2020).

- [26] Rosebrock, A. (2017). ImageNet: VGGNet, ResNet, Inception, and Xception with Keras. PyImageSearch.
- [27] DigitalOcean. Popular Deep Learning Architectures: ResNet, InceptionV3, SqueezeNet.
- [28] ResearchGate. The architecture of our deep residual network (ResNet) in the precise stage prediction.
- [29] He, K., Zhang, X., Ren, S., & Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778. (2016).
- [30] Samin, S. A., Sulaiman, N., & Ghazali, N. B. Detection of Faults in Solar Photovoltaic Panels Using CNN Techniques. Progress in Engineering Application and Technology (PEAT), 3(2), 153–159. (2022).
- [31] Zyout, I., & Oataweh, H. Detection of PV Solar Panel Surface Defects Using Image Processing and Machine Learning Techniques. Semantic Scholar. (2021).
- [32] PubMed. Solar Panel Fault Detection Using Deep Learning Approaches: A Review. (2021).
- [33] Sharma, P., et al. Solar Panel Defect Detection Using Deep Learning Techniques: A Comparative Review. Frontiers in Energy Research. (2024).
- [34] Al-Khayyat, N. T., & Hasan, A. H. Improving Image Captioning Using Transfer Learning and Hybrid Deep Learning Models. International Journal of Heat and Technology, 41(1), 279–285. (2023).
- [35] Samin, S. A., Sulaiman, N., & Ghazali, N. B. Detection of Faults in Solar Photovoltaic Panels Using CNN Techniques. Progress in Engineering Application and Technology (PEAT), 3(2), 153–159. (2022).
- [36] Zhang, Y., et al. A Novel CNN-BiLSTM Model for Automated Image Captioning Tasks. Sensors, 23(8), 3850. (2023).
- [37] PubMed. Solar Panel Fault Detection Using Deep Learning Approaches: A Review. (2021).
- [38] Salehinejad, H., et al. A Deep Learning Approach for Visual Captioning with Enhanced Attention Mechanisms. Computers in Biology and Medicine, 136, 104703. (2021).
- [39] Chatterjee, A., et al. Image Captioning With Transformers: A Performance Analysis. PeerJ Computer Science, 10, e2148. (2024).
- [40] Ref : European Geosciences Union (EGU). EGU General Assembly 2024 Abstract: Solar Panel Monitoring and Fault Detection Using AI. (2024).
- [41] Mermet, P., & Morel, G. Analyse des séries temporelles par les modèles autorégressifs: principes et applications. GRETSI, 4(6), 2411–2418. (1987).
- [43] Haccoun, D. Reconnaissance de Formes par Réseaux de Neurones à Rétropropagation. École Polytechnique de Montréal – Département de Génie Électrique. (1984).
- [44] Guibert, N., & Malbos, P. Approche non linéaire en reconnaissance de formes : application aux séries temporelles. GRETSI, 3(3), 1602–1610. (1986).
- [45] GeeksforGeeks. Introduction to Recurrent Neural Network (RNN). (2023).
- [46] Amazon Web Services (AWS). What is a Recurrent Neural Network (RNN)?. (2023).
- [98] GeeksforGeeks. Deep Learning – Introduction to Long Short-Term Memory (LSTM). (2023).
- [47] Metaschool. LSTM (Long Short-Term Memory): A Beginner-Friendly Guide. (2023).
- [48] Noury, A. Pense-bête : Réseaux de Neurones Récurrents (RNN, LSTM, GRU). Stanford University - CS230.

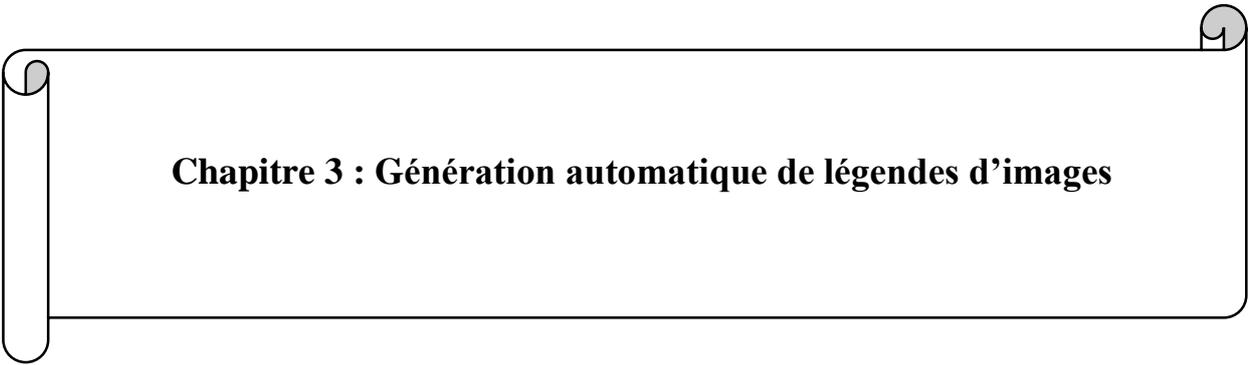
- [49] Coursera. LSTM Neural Network Explained: Architecture and Use Cases. (2024).
- [50] Olah, C. Understanding LSTM Networks. (2015).
- [51] DIDA Learning. What is an LSTM Neural Network?. (2023).
- [52] Simplilearn. LSTM Explained: Long Short-Term Memory Networks. (2023).
- [53] MLArchive. Understanding Long Short-Term Memory (LSTM) Networks in Deep Learning. (2023).
- [54] Interview Kickstart. Long Short-Term Memory Explained: LSTM Architecture & Applications. (2023).
- [55] MathWorks. LSTM Networks. (2023).
- [56] Noury, A. Pense-bête : Apprentissage profond. Stanford University - CS229.
- [57] SerpAI. Bidirectional LSTM: Concepts, Use Cases & Architecture. (2023).
- [58] Rajeswari, S., & Ganesan, R. Improved Fault Classification Using Bi-LSTM Deep Learning Model. *Journal of Electrical Systems*, 18(1), 1–12. (2022).
- [59] Olanrewaju, R. F., et al. A Bidirectional LSTM Model for Weather Forecasting. *Research Square*, 1–15. (2022).
- [60] Kim, M., & Moon, Y. S. BiLSTM model based on multivariate time series data. *Semantic Scholar*. (2021).
- [61] Baeldung. Bidirectional vs Unidirectional LSTM – Key Differences. (2023).
- [62] Baeldung. Bidirectional vs Unidirectional LSTM – Key Differences. (2023).
- [63] Liu, Z., et al. BiLSTM-Based Model for Plant Disease Detection Using Leaf Images. *Frontiers in Plant Science*, 15, 1500499. (2024).
- [64] Papers With Code. BiLSTM – Papers and Code. (2024).
- [65] T1che, A. Bidirectional LSTM (BiLSTM) in Deep Learning A Powerful Sequential Model. *LinkedIn*.(2023).
- [66] Schneppat, R. Bidirectional Long Short-Term Memory (BiLSTM).(2022).
- [119] Rajeswari, S., & Ganesan, R. Improved Fault Classification Using Bi-LSTM Deep Learning Model. *Journal of Electrical Systems*, 18(1), 1–12. (2022).
- [67] Liu, Z., et al. BiLSTM-Based Model for Plant Disease Detection Using Leaf Images. *Frontiers in Plant Science*, 15, 1500499. (2024).
- [68] Datafranca. Réseau récurrent bidirectionnel à longue mémoire court terme (BiLSTM).
- [69] Cherfaoui, N. Utilisation de BiLSTM pour la classification des séries temporelles multivariées. *Mémoire de Master. Université du Québec à Montréal*. (2023).
- [70] Redjati, R. Mémoire final: Réseaux neuronaux bidirectionnels pour la classification d'images. *Université d'Annaba*. (2024).
- [71] MathWorks. bilstmLayer. (2024).
- [72] Ref : Delpech, Q. Apprentissage de représentations croisées texte-image avec des réseaux neuronaux. *JEP-TALN-RECITAL 2019*, 15. (2019).
- [73] Chikr, M., & Abid, F. Image Captioning using Deep Learning: A Review. *Open Journal of Artificial Intelligence*, 4(1), 1–10. (2024).

- [74] Djaballah, M. A. Application des réseaux neuronaux bidirectionnels. Mémoire de Master. Université de Guelma.
- [75] ExpertBeacon. A Comprehensive Guide to Image Captioning: Generating Textual Descriptions from Visual Data. (2023).
- [76] 33rd Square. Solving an Image Captioning Task Using Deep Learning. (2023).
- [77] Kumar, A., & Singh, P. A Deep Learning-Based Image Captioning System Using BiLSTM and Attention Mechanism. CMC-Computers, Materials & Continua, 81(3), 3345–3360. (2024).
- [78] Singh, Y., & Ghosh, S. Image Captioning using CNN and Bi-LSTM with Attention Mechanism. Semantic Scholar. (2021).[132] Benitez, M. Image Captions with Deep Learning: State of the Art Architectures. Towards Data Science.(2020).
- [79] Yuan, H., & Zhang, K. Attention-Based Bi-LSTM Model for Image Captioning. International Journal of Computer Science, 51(9), 1–10. (2024).
- [80] Saeed, Tanzeel. "Image Captioning Using Deep Learning Techniques." International Journal of Novel Research and Development (IJNRD), vol. 9, no. 9, 2024.
- [81] Wu, Xiang, et al. "DIFNet: Boosting Visual Information Flow for Image Captioning." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022.
- [82] Malik, Ajit Jaokar. "Vanishing Gradient Problem in Deep Learning with Examples." Vitalflux, 2023.
- [83] EasyAI. "The Problem of Overfitting and How to Resolve It." EasyAI Tech Blog, 2022.
- [84] Data Science Society. "Challenges and Solutions in Training Deep Learning Models." DataScienceSociety.net, 2022.
- [85] KDnuggets. "Vanishing Gradient Problem: Definition and Solutions." KDnuggets, 2022.
- [86] GeeksForGeeks. "What Are the Possible Approaches to Fixing Overfitting on a CNN?" GeeksForGeeks, 2021.
- [87] OpenGenus. "Problems in Deep Learning." OpenGenus IQ, 2023.
- [88] Djaballah, Mohammed Abdelmoumen. "Détection des défauts dans les panneaux photovoltaïques à l'aide du deep learning." Mémoire de master, Université 8 mai 1945 Guelma, 2023.
- [89] Redjati, Rachid. "Détection automatique des anomalies dans les images infrarouges des panneaux photovoltaïques." Mémoire de master, Université Badji Mokhtar – Annaba, 2024.
- [90] FasterCapital. "Exploiter le Deep Learning pour la génération de légendes d'images." FasterCapital, 2023.
- [91] Esri. "Improve a Deep Learning Model with Transfer Learning." Learn ArcGIS, 2022.
- [92] Begnan, Aristide Médard. "Conception et réalisation d'un système de reconnaissance d'images basé sur l'apprentissage profond." Mémoire de maîtrise, Université du Québec à Trois-Rivières (UQTR), 2022.
- [93] T1che, Ahmed. "Bidirectional LSTM (BiLSTM): A Powerful Sequential Model." LinkedIn Pulse, 2023.
- [94] MapperX. "Détection des défauts des panneaux solaires." MapperX, 2023.
- [95] Dynae. "Maintenance prédictive et surveillance des panneaux photovoltaïques." Dynae Actualités, 2023.
- [96] MapperX. "Analyse et détection des défauts des panneaux photovoltaïques." MapperX, 2023.
- [97] Mehadjbia, Abdelhak. "Détection automatique des défauts dans les panneaux photovoltaïques." Mémoire de maîtrise, Université du Québec en Abitibi-Témiscamingue (UQAT), 2025.

[98] Charef, Walid. "Système de surveillance thermique pour les installations photovoltaïques." Mémoire de master, Université d'El Oued, 2023.

[99] Serp AI. "Bidirectional LSTM (BiLSTM) – Understanding the Architecture." Serp.ai, 2022.

[100] GeeksForGeeks. "Difference between a Bidirectional LSTM and an LSTM." GeeksForGeeks, 2021.



## **Chapitre 3 : Génération automatique de légendes d'images**

### 3.1 Introduction

L'un des défis majeurs de l'intelligence artificielle réside dans la compréhension sémantique et la représentation du contenu visuel, avec pour ambition de rapprocher la perception machine de celle de l'humain. Les systèmes intelligents tentent ainsi de convertir des informations visuelles complexes en descriptions linguistiques précises, cohérentes et contextuellement adaptées. Dans ce cadre, le domaine de la génération de légendes d'images (Image Captioning) s'impose comme un carrefour stratégique entre la vision par ordinateur (Computer Vision) et le traitement automatique du langage naturel (Natural Language Processing, NLP).

L'objectif principal de cette discipline est de produire automatiquement des descriptions textuelles détaillées des objets, scènes et relations présentes dans une image, tout en préservant la justesse sémantique et la fluidité linguistique. Cette tâche s'avère particulièrement utile dans des domaines où le traitement visuel manuel est complexe ou coûteux, comme la surveillance par drone, le diagnostic médical, ou l'inspection de panneaux solaires.

Au cours de la dernière décennie, le domaine a connu des avancées considérables grâce aux progrès des modèles d'apprentissage profond. D'une part, les réseaux de neurones convolutifs (CNN), tels que VGG16, ResNet ou Inception, ont démontré une remarquable capacité à extraire automatiquement des caractéristiques visuelles riches et discriminantes. D'autre part, les modèles séquentiels tels que les réseaux de neurones récurrents (RNN) et en particulier les variantes LSTM (Long Short-Term Memory) et GRU (Gated Recurrent Unit) se sont imposés pour modéliser efficacement les séquences linguistiques, en tenant compte du contexte et des dépendances temporelles.

Plus récemment, les architectures Transformer, initialement développées pour le NLP, ont été intégrées avec succès dans les systèmes de génération de légendes, grâce à leur capacité à capturer les relations globales dans les séquences et à permettre une parallélisation efficace de l'apprentissage.

Ainsi, la génération automatique de légendes représente un pont technologique entre l'analyse visuelle et la production linguistique, où une image devient le point de départ d'un processus cognitif artificiel simulant l'interprétation humaine. Cette tâche consistant à générer une phrase descriptive à partir d'une image donnée s'impose aujourd'hui comme une composante essentielle dans de nombreuses applications intelligentes et autonomes.

### 3.2 Intégration Des Modèles Visuels 2d Et Des Modèles De Prédiction Séquentielle Pour La Génération De Légendes D'images

La génération automatique de légendes d'images (image captioning) constitue un défi multidisciplinaire à la croisée de la vision par ordinateur et du traitement du langage naturel. Elle nécessite la capacité de comprendre le contenu sémantique d'une image et de produire une description textuelle cohérente. Les architectures modernes adoptent une approche en deux étapes : (1) l'extraction de caractéristiques visuelles via des modèles d'apprentissage profond 2D, et (2) la génération de séquences textuelles à l'aide de modèles de prévision tels que LSTM, GRU, BiLSTM ou Transformer.

### 3.2.1 Extraction des caractéristiques par modèles 2D et apprentissage par transfert

Les réseaux de neurones convolutifs (CNN) pré-entraînés sur de grands corpus d'images comme ImageNet, tels que VGG16, ResNet50, InceptionV3 ou EfficientNet, se sont avérés très efficaces pour extraire des caractéristiques spatiales et sémantiques à partir d'images. Dans le cadre de l'apprentissage par transfert (transfer learning), ces modèles sont utilisés soit comme extracteurs de caractéristiques figés, soit en étant partiellement ré-entraînés sur le domaine cible. Ils permettent ainsi de convertir une image en un vecteur dense ou une carte de caractéristiques représentative de son contenu visuel.

Cette stratégie est particulièrement avantageuse dans les domaines où les jeux de données sont limités ou spécialisés (comme les images infrarouges ou thermiques), car elle réduit le coût computationnel et améliore la robustesse du modèle.

### 3.2.2 Modèles séquentiels pour la génération de texte

Une fois les caractéristiques visuelles extraites, la tâche devient un problème de modélisation séquentielle : il s'agit de prédire une séquence de mots à partir du contexte visuel fourni. Plusieurs architectures récurrentes ou à base d'attention sont utilisées comme décodeurs :

- **LSTM (Long Short-Term Memory)** : Ce modèle gère efficacement les dépendances à long terme, ce qui le rend particulièrement adapté à la génération de séquences textuelles fluides et cohérentes.
- **GRU (Gated Recurrent Unit)** : Variante simplifiée du LSTM, le GRU offre une performance similaire tout en étant plus léger et rapide à entraîner.
- **BiLSTM (Bidirectional LSTM)** : En traitant les séquences dans les deux directions (avant et arrière), le BiLSTM capte un contexte plus riche. Il est généralement utilisé dans la phase d'encodage, mais peut également contribuer à des architectures hybrides.
- **Transformer** : Les modèles de type Transformer, reposant entièrement sur des mécanismes d'attention multi-tête, surpassent les modèles récurrents dans de nombreuses tâches. Ils permettent une parallélisation efficace et capturent mieux les dépendances globales au sein des séquences générées.

### 3.2.3 Union des deux paradigmes

L'union entre les modèles 2D (pour la compréhension visuelle) et les modèles séquentiels (pour la génération de langage) constitue une architecture dite encodeur-décodeur puissante pour la génération de légendes. Le CNN agit comme un encodeur visuel, tandis que le LSTM, GRU ou Transformer sert de décodeur linguistique. Pendant l'entraînement, le décodeur apprend à transformer les caractéristiques visuelles en une séquence de mots, en maximisant la probabilité de la légende correcte.

Dans la pratique, les architectures CNN-LSTM et CNN-Transformer sont devenues des références (**Figure 3.25**). Cette synergie permet au système de « voir » l'image via l'encodeur 2D, puis de la « décrire » par une phrase via le décodeur séquentiel. Les mécanismes d'attention, notamment dans les Transformers, améliorent encore la performance en permettant au modèle de se focaliser dynamiquement sur les zones pertinentes de l'image au fil de la génération des mots.

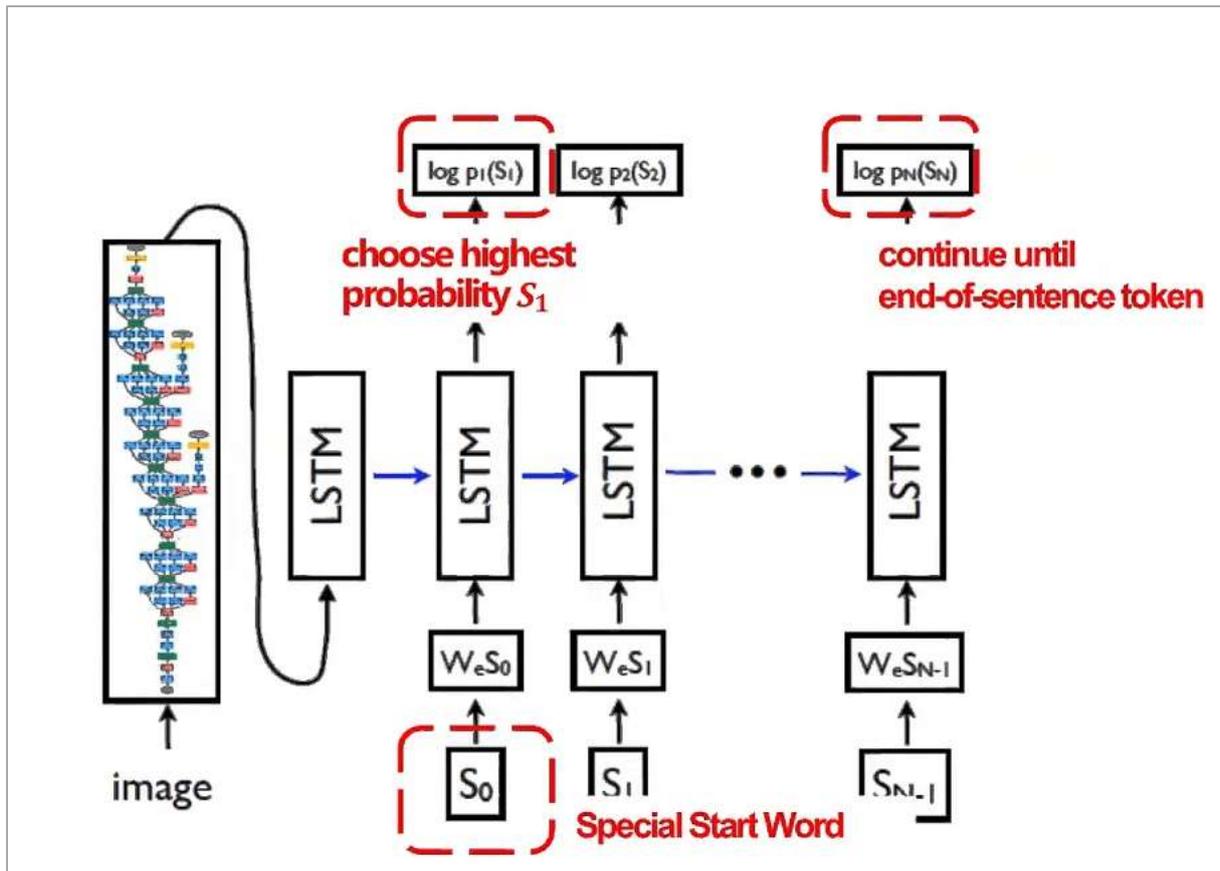


Figure 3.1 : Architecture encodeur décodeur pour la génération du langage.

### 3.3 Méthodologie de recherche et développement du modèle

Ce travail s'appuie sur l'application d'un modèle unifié pour la création automatique de légendes d'images, en combinant diverses techniques de pointe issues de l'apprentissage profond. Cette méthode s'organise en deux phases majeures :

✓ **Phase d'extraction des caractéristiques visuelles en utilisant ResNet :**

Pour cette phase, le réseau ResNet est employé pour l'extraction des caractéristiques à partir des images. On insère l'image dans un modèle ResNet déjà entraîné, qui analyse l'image à travers diverses couches du réseau dans le but d'en dégager des caractéristiques abstraites. Avec l'architecture de « connexions de saut » (skip connections) de ResNet, le modèle est capable de gérer des couches plus profondes et complexes, renforçant ainsi sa capacité à extraire précisément les caractéristiques visuelles. Ces attributs sont indispensables pour produire une description exacte de l'image [1] [2].

✓ **Phase de création de légendes avec LSTM :**

Après l'extraction des caractéristiques, celles-ci sont transmises au réseau LSTM afin de produire le texte. LSTM, une forme de réseau neuronal récurrent (RNN), a la capacité de traiter des séquences temporelles de données, comme par exemple la création de phrases.

Ce modèle est formé pour assimiler les contextes textuels sur une longue durée, ce qui lui permet de produire des descriptions cohérentes et naturelles.

LSTM est formé sur un jeu de données comprenant des images accompagnées de légendes de référence afin d'assimiler les liens entre les images et les termes [1] [2].

### 3.3.1 Niveaux détaillés de l'élaboration du modèle

Prétraitement des données : On utilise des jeux de données comme MSCOCO ou Flickr8k, qui renferment des images accompagnées de légendes écrites, pour l'entraînement.

On segmente les données en groupes d'apprentissage et de vérification afin d'évaluer la performance du modèle.

- **Entraînement via ResNet :** Le modèle ResNet est formé sur l'ensemble des images afin d'obtenir les représentations numériques (features). Le modèle pré-entraîné a des poids fixes (c'est-à-dire qu'ils ne sont pas modifiés pendant l'entraînement) et l'attention se porte sur l'optimisation des couches de filtrage et de regroupement.
- **Apprentissage avec LSTM :** Une fois que les représentations d'images ont été obtenues grâce à ResNet, elles sont transmises au réseau LSTM pour la création de légendes. Durant la formation, une multitude de légendes de référence est employée pour enseigner au réseau comment produire des textes pertinents et cohérents.
- **Évaluation du modèle :** Le modèle est évalué en utilisant des indicateurs tels que BLEU, METEOR, et CIDER, qui servent à comparer la qualité des légendes produites avec celle des légendes de référence [1].

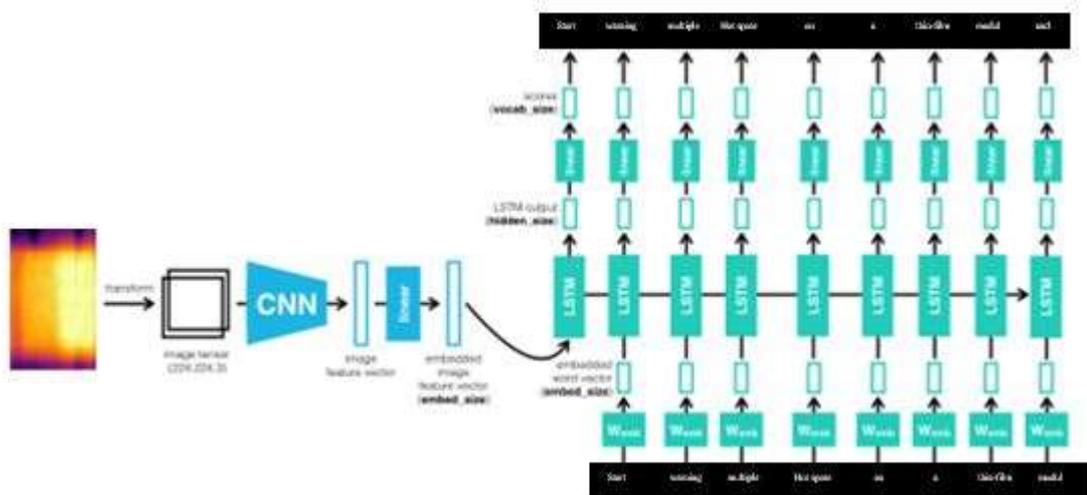


Figure 3.2 : Architecture CNN-LSTM.

### 3.3.2 Architecture proposée

**Encodeur (ResNet)** : un CNN pré-entraîné (ResNet50) dont la couche de classification finale est substituée par une couche linéaire qui projette les caractéristiques obtenues dans un espace d'embedding correspondant à celui des mots.

- **Réseau LSTM** : un réseau LSTM qui utilise les embeddings des mots antérieurs ainsi que le vecteur des caractéristiques de l'image comme entrées, afin de prédire le mot suivant dans la légende.
- **Mécanisme d'attention (facultatif)** : quelques modèles intègrent un mécanisme d'attention « soft attention » qui autorise le décodeur à mettre l'accent sur diverses sections de l'image lors de la création de chaque mot, ce qui accroît la pertinence des descriptions [3].

### 3.4 Composante Technique

#### 3.4.1 Description Du Jeu De Données

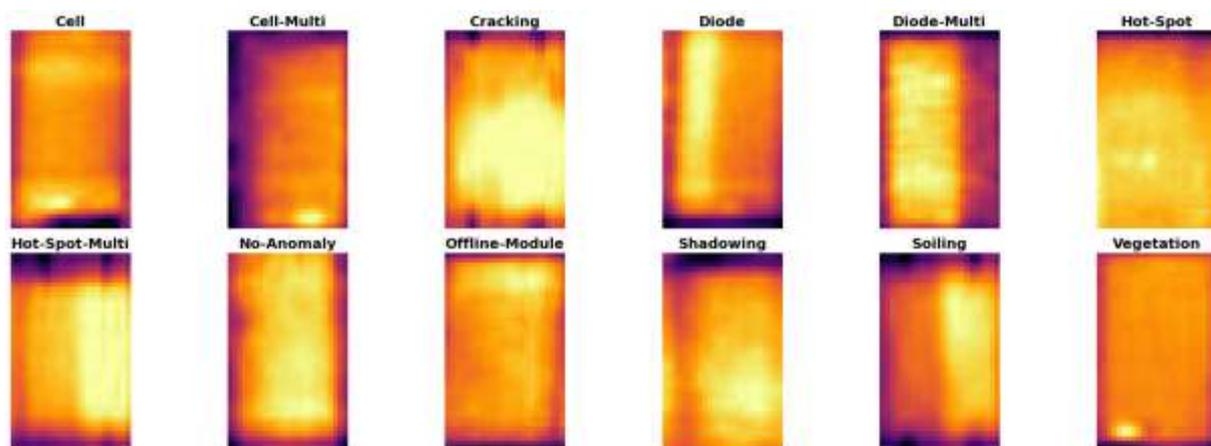
Le jeu de données se compose de 20 000 images infrarouges, chacune ayant une résolution de 24 par 40 pixels. Douze classes de modules solaires y sont définies, dont 11 classes représentant différentes anomalies, et une classe représentant l'absence d'anomalie (**Figure 3.3**).

Le **tableau 3.1** énumère chaque classe accompagnée d'une description. Ce jeu de données comprend de véritables anomalies uniques de modules solaires.

Les données ont été agrégées par l'équipe de Raptor Maps et recueillies par des avions pilotés et des systèmes aériens sans pilote, équipés de capteurs d'imagerie infrarouge à moyenne ou longue longueur d'onde (3 à 13,5  $\mu\text{m}$ ) ainsi que de capteurs du spectre visible[4].

La résolution des images varie de 3,0 à 15,0 cm/pixel. Les anomalies ont été extraites au niveau de chaque module individuel et classées par type.

Les images correspondantes dans le spectre visible ont été utilisées lors de la classification afin d'en améliorer la précision.



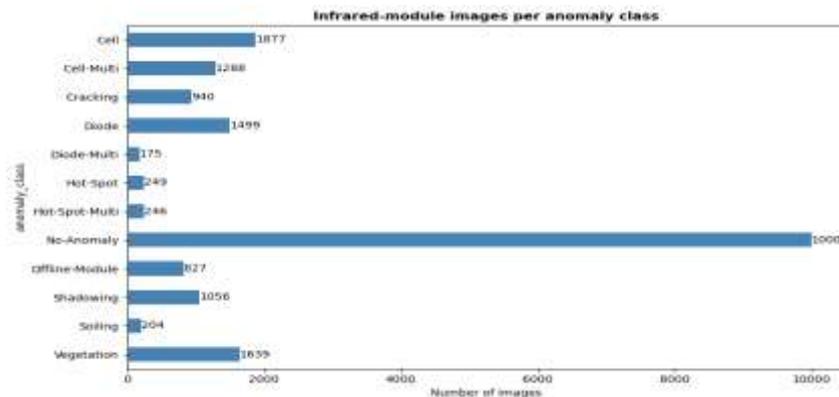
**Figure 3.3** : Classes de modules.

Class Name	Images	Description
Cell	1,877	Hot spot occurring with square geometry in single cell.
Cell-Multi	1,288	Hot spots occurring with square geometry in multiple cells.
Cracking	941	Module anomaly caused by cracking on module surface.
Hot-Spot	251	Hot spot on a thin film module.
Hot-Spot-Multi	247	Multiple hot spots on a thin film module.
Shadowing	1056	Sunlight obstructed by vegetation, man-made structures, or adjacent rows.
Diode	1,499	Activated bypass diode, typically 1/3 of module.
Diode-Multi	175	Multiple activated bypass diodes, typically affecting 2/3 of module.
Vegetation	1,639	Panels blocked by vegetation.
Soiling	205	Dirt, dust, or other debris on surface of module.
Offline-Module	828	Entire module is heated.
No-Anomaly	10,000	Nominal solar module.

**Tableau 3.1 : Classes**

La figure 3.4, illustre la distribution du nombre d'images infrarouges par classe d'anomalie dans le jeu de données *InfraredSolarModules*. Sur un total de 20 000 images, 10 000 représentent des modules sans anomalie ("No-Anomaly"), tandis que les 10 000 restantes sont réparties entre 11 classes d'anomalies distinctes, telles que les défauts de cellules, les diodes défectueuses, les zones chaudes (hot-spots), le masquage (shadowing) et la végétation.

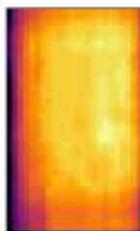
On observe un déséquilibre significatif dans les classes, la classe "Cell" étant la plus représentée parmi les anomalies avec 1877 images, tandis que "Diode-Multi" et "Soiling" sont parmi les moins fréquentes avec respectivement 175 et 204 images. Cette disparité souligne la nécessité de techniques de pondération ou d'échantillonnage dans les tâches de classification automatisée.



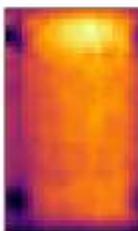
**Figure 3.4 : Répartition des images infrarouges selon les classes d'anomalies des modules photovoltaïques.**

**La Figure 3.5** présente chaque classe d'anomalie identifiée dans le jeu de données *InfraredSolarModules*, accompagnée d'un exemple d'image représentative. Pour chacune de ces classes, nous visons à générer automatiquement une **description textuelle pertinente et spécifique** en utilisant notre modèle de génération de légendes basé sur l'apprentissage profond. Ces descriptions permettront d'associer à chaque image infrarouge un texte explicatif décrivant visuellement et sémantiquement l'anomalie détectée (par exemple : « présence de points chauds sur plusieurs cellules », « ombrage partiel dû à la végétation », ou « cellule déconnectée »). Cette approche a pour objectif de faciliter l'interprétation des inspections thermographiques, notamment dans les environnements où le tri manuel des anomalies est long et coûteux. Le modèle proposé extrait les caractéristiques visuelles des images par un encodeur convolutionnel, puis produit des légendes en langage naturel via un décodeur séquentiel.

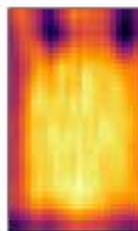
**No-Anomaly**



nominal photovoltaic solar cell



nominal photovoltaic solar cell



nominal photovoltaic solar cell



nominal photovoltaic solar cell

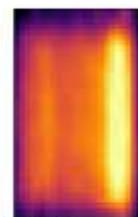
**Hot-Spot-Multi**



multiple hot spots on a thin-film module



warning multiple hot spots on a thin-film module

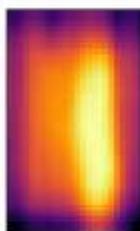


warning multiple hot spots on a thin-film module



warning multiple hot spots on a thin-film module

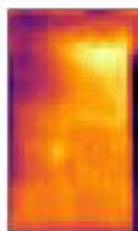
**Hot-Spot**



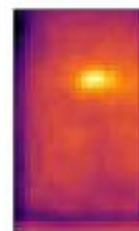
hot spot on a thin-film module



warning hot spot on a thin-film module



warning hot spot on a thin-film module

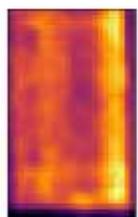


warning hot spot on a thin-film module

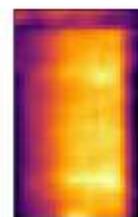
**Diode-Multi**



multiple activated by-pass diodes affecting 2 over 3 of module



warning multiple activated by-pass diodes typically affecting 2 over 3 of module



warning multiple activated by-pass diodes typically affecting 2 over 3 of module



warning multiple activated by-pass diodes typically affecting 2 over 3 of module

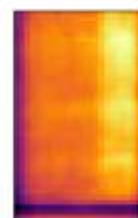
**Diode**



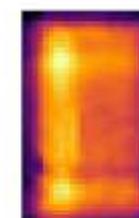
activated by-pass diode typically 1 over 3 of the module



warning activated by-pass diode typically 1 over 3 of the module



warning activated by-pass diode typically 1 over 3 of the module



warning activated by-pass diode typically 1 over 3 of the module

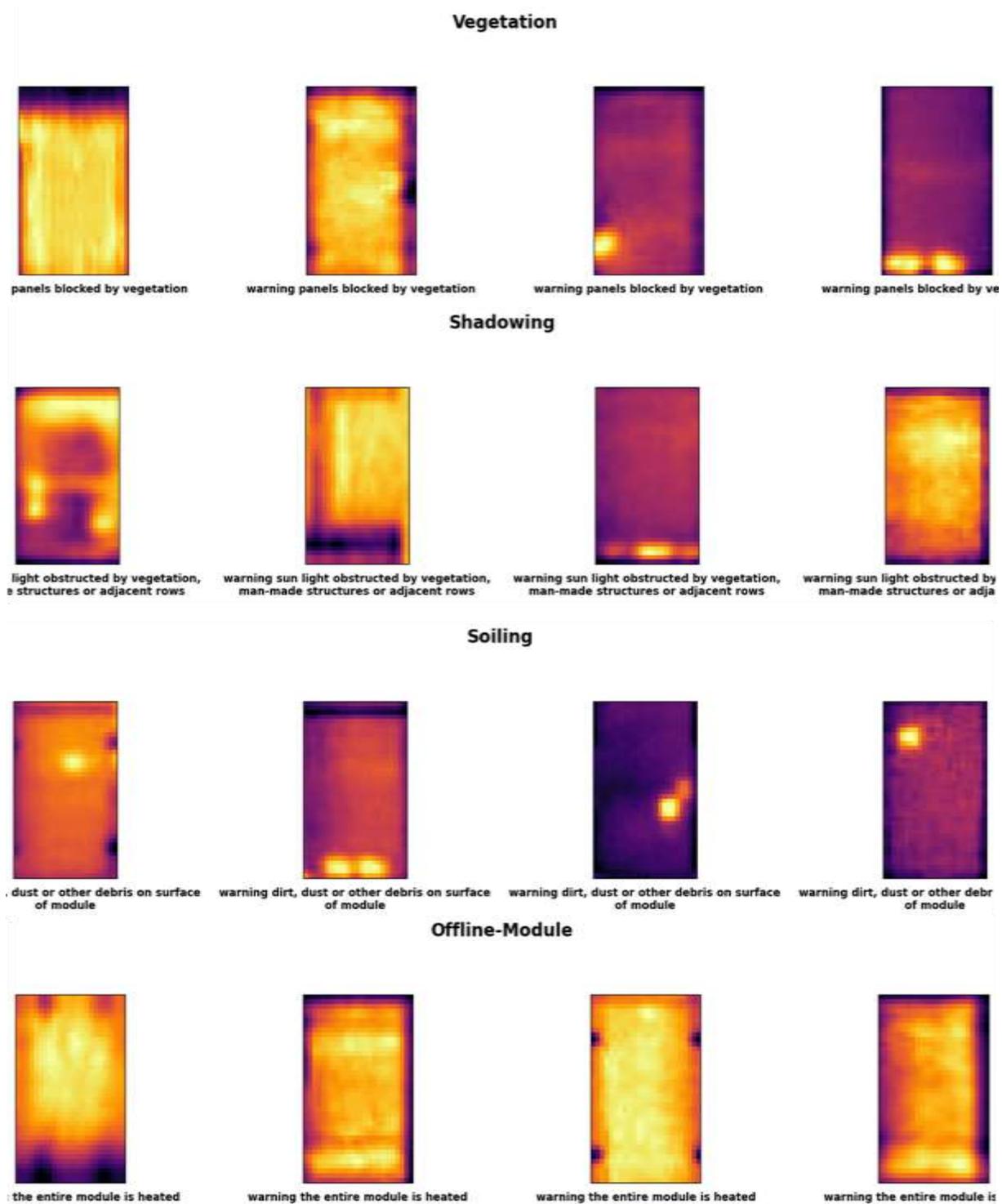


Figure 3.5 : Exemples d'images par classe d'anomalie avec leur description cible générée automatiquement

### 3.4.2 Vue d'ensemble du système / Architecture générale

L'architecture proposée repose sur un modèle de type Encodeur–Décodeur, intégrant un réseau convolutionnel profond (ResNet50Plus) et un réseau LSTM (Long Short-Term Memory), dans une structure qui combine les avantages des deux approches.

ResNet50Plus agit comme un extracteur de caractéristiques visuelles, exploitant sa capacité à analyser les détails spatiaux des images, tandis que le LSTM se charge de générer une description textuelle en modélisant les dépendances séquentielles à long terme entre les mots.

Ce système est conçu pour analyser automatiquement une image donnée et produire une légende descriptive cohérente et en langage naturel, établissant ainsi un lien efficace entre la vision par ordinateur et le traitement du langage naturel.

#### 1. Architecture LSTM :

Le réseau LSTM (Long Short-Term Memory) est une version améliorée des RNN, spécifiquement élaborée pour traiter les dépendances à long terme dans les séquences, en esquivant le souci du gradient qui se réduit à néant (vanishing gradient). La cellule mémoire, qui garde l'information pertinente au fil du temps via un système de portes ( gating ), constitue le cœur de l'unité centrale.

##### a. Porte de l'oubli (Forget Gate $f_t$ )

La Porte D'oubli (forget gate) joue un rôle essentiel dans la gestion de la mémoire à long terme de la cellule LSTM. Elle décide quelles informations de la mémoire précédente doivent être effacées ou conservées, en se basant sur l'entrée actuelle  $x_t$  et l'état caché précédent  $h_{t-1}$ . Concrètement, une fonction d'activation sigmoïde génère des valeurs comprises entre 0 et 1 pour chaque élément de la mémoire : une valeur proche de 0 signifie que l'information correspondante doit être oubliée, tandis qu'une valeur proche de 1 indique qu'elle doit être conservée. Ce mécanisme permet au modèle d'éliminer les informations devenues inutiles tout en préservant celles qui restent pertinentes pour la tâche à accomplir.

##### b. Porte d'entrée (Input Gate $i_t$ )

La Porte D'entrée (input gate) joue un rôle crucial dans le fonctionnement de la cellule LSTM. Elle détermine quelles nouvelles informations issues de l'entrée actuelle  $x_t$  et de l'état caché précédent  $h_{t-1}$  doivent être intégrées dans la mémoire existante. Son fonctionnement repose sur deux étapes principales : tout d'abord, une fonction d'activation (généralement une sigmoïde) est appliquée pour décider de l'importance de chaque élément de l'entrée ; ensuite, une suggestion de mise à jour est générée, souvent à l'aide d'une tangente hyperbolique, qui propose le nouveau contenu à ajouter à la mémoire. Ces mécanismes permettent de contrôler avec finesse l'apprentissage et la rétention d'informations pertinentes à chaque étape temporelle.

##### c. Porte de sortie (Output Gate $o_t$ )

La Porte De Sortie (output gate, notée  $o_t$ ) est responsable de la régulation de l'information transmise par la cellule LSTM à l'étape suivante. Elle détermine quelles parties de la mémoire actuelle  $c_t$  doivent être produites comme nouvel état caché  $h_t$ . En d'autres termes, cette porte contrôle ce que la cellule "divulgue" à l'extérieur, influençant ainsi les prédictions futures et l'apprentissage continu.

Elle agit en combinant l'entrée actuelle  $x_t$  et l'état caché précédent  $h_{t-1}$ , passe par une fonction d'activation sigmoïde, puis applique une fonction tanh sur la cellule mémoire pour moduler les valeurs transmises, assurant une sortie utile et pertinente[5-7].

## 2. Architecture du modèle hybride (ResNet50Plus + LSTM)

Dans des architectures hybrides telles que RASNET50PLUS LSTM, un CNN de profondeur (ici ResNet50Plus, une version optimisée du ResNet50) est employé pour extraire des caractéristiques complexes en termes d'espace à partir d'images ou de données structurées.

Ces caractéristiques sont par la suite transmises à un LSTM qui saisit les dépendances séquentielles ou temporelles.

Le réseau ResNet50Plus est utilisé en tant que couche d'extraction de caractéristiques, transformant chaque image ou frame en un vecteur de caractéristiques représentant les informations discriminantes de l'image. Ces vecteurs sont ensuite organisés sous forme de séquences temporelles, telles que des frames successives d'une vidéo ou des séries d'images capturées à des intervalles réguliers.

Ces séquences sont ensuite transmises à une unité LSTM (Long Short-Term Memory), chargée d'apprendre les relations temporelles entre les vecteurs. Le LSTM est particulièrement adapté à ce type de données grâce à son mécanisme de portes (oubli, entrée, sortie), qui lui permet de conserver ou d'ignorer certaines informations au fil du temps, en fonction de leur pertinence.

En pratique, les couches CNN sont souvent encapsulées dans une couche appelée TimeDistributed, qui applique la même couche CNN à chaque élément de la séquence de manière indépendante. Cela permet de préserver la structure temporelle de la séquence avant de la transmettre au LSTM pour un traitement plus approfondi [8][9].

### 3.5 Modèle de génération de légendes

#### 3.5.1 Architecture du modèle de génération de légendes d'images

L'architecture classique des modèles de génération automatique de descriptions d'images repose sur une approche encodeur-décodeur, combinant un réseau de neurones convolutionnel (CNN) pour l'extraction des caractéristiques visuelles, et un réseau de neurones séquentiel, généralement de type LSTM (Long Short-Term Memory), pour la génération du texte descriptif [10-12].

- **Encodeur (CNN)**

Un modèle CNN pré-entraîné, tel que VGG16 ou ResNet, est utilisé pour extraire un vecteur de caractéristiques visuelles à partir de l'image. Les couches entièrement connectées sont généralement supprimées, ne conservant que les couches convolutionnelles afin d'obtenir un vecteur fixe ou une carte de caractéristiques spatiale [166].

- **Décodeur (LSTM)**

Le vecteur de caractéristiques produit par le CNN est injecté dans un décodeur basé sur un LSTM, qui se charge de générer la légende de manière progressive. À chaque étape temporelle, le LSTM prédit le mot suivant en fonction de son état interne, du mot précédemment généré et du contexte visuel associé à l'image. Pour améliorer la pertinence sémantique de la légende, un mécanisme d'attention peut être intégré, permettant au modèle de se focaliser dynamiquement sur différentes régions de l'image au cours de la génération [10] [11] [13].

### **3.5.2 Stratégies d'Apprentissage, Fonctions de Perte et Métriques d'Évaluation en Captioning d'Images**

Les modèles de génération de légendes d'images (image captioning) sont entraînés pour prendre une image en entrée et produire une description textuelle correspondante [14] [15].

#### **1. Apprentissage supervisé :**

Le modèle est formé à partir d'un jeu de données contenant des paires image-légende. Il apprend ainsi à associer une image à un texte descriptif approprié.

#### **2. Extraction de caractéristiques avec les CNN :**

Des réseaux convolutifs pré-entraînés comme ResNet ou VGG sont utilisés pour extraire les caractéristiques visuelles de l'image. Ces caractéristiques représentent un résumé visuel utilisé pour la génération de texte.

#### **3. Génération du texte avec LSTM ou Transformers :**

Une fois les caractéristiques extraites, des architectures comme LSTM (Long Short-Term Memory) ou Transformers sont utilisées pour générer la légende mot par mot. Souvent, un mécanisme d'attention (attention) est intégré pour permettre au modèle de se concentrer sur différentes zones de l'image lors de la génération de chaque mot.

#### **4. Optimisation du modèle :**

L'entraînement utilise l'algorithme Adam, connu pour son efficacité dans la mise à jour des poids. En complément, un planificateur du taux d'apprentissage (learning rate scheduler) ajuste ce taux au cours du temps pour améliorer la performance du modèle.

#### **5. Fonction de perte Entropie croisée :**

Pendant l'apprentissage, les mots prédits sont comparés aux mots réels dans la légende de référence. La fonction d'entropie croisée est utilisée pour mesurer l'écart entre les prédictions et les véritables mots, et le modèle cherche à minimiser cette différence afin d'améliorer la qualité des descriptions générées.

### **3.6 Approche de détection des anomalies**

#### **3.6.1 Définition conceptuelle de l'anomalie dans le contexte de la détection**

Dans le domaine de la science des données et de l'apprentissage automatique, une anomalie désigne une observation, un événement ou un comportement qui s'écarte de manière significative des schémas réguliers ou attendus. Autrement dit, il s'agit d'une déviation statistique ou comportementale importante par rapport à un modèle appris ou à une norme historique. Bien que les anomalies soient rares en termes de fréquence, elles peuvent avoir une importance critique, car elles signalent souvent des incidents tels que des fraudes, des défaillances techniques, des erreurs ou des attaques informatiques. Par exemple, dans un réseau informatique, une anomalie peut se manifester par un trafic inhabituel, une tentative de connexion suspecte ou un transfert de données anormal. Dans le contexte de la vidéosurveillance, elle peut correspondre à un événement rare comme un comportement agressif, une intrusion ou un accident. En général, la détection d'anomalies vise à identifier ces événements atypiques pour permettre une réaction rapide, une surveillance renforcée ou une prise de décision automatisée [14].

### 3.6.2 La méthodologie adoptée pour la détection des anomalies

La détection des anomalies constitue un outil fondamental dans l'analyse des données, visant à identifier les comportements ou états inhabituels en les comparant à un modèle normal préalablement appris. Cette approche revêt une importance croissante dans divers domaines, notamment celui des systèmes photovoltaïques (PV), où elle peut être utilisée pour détecter les pannes ou les baisses de performance, telles qu'une diminution inattendue de la production d'énergie ou des variations soudaines de tension ou de courant. Les recherches et revues scientifiques récentes ont montré que l'intégration des techniques d'intelligence artificielle et d'apprentissage automatique dans les modèles de détection des anomalies améliore la capacité des systèmes à surveiller les performances et à optimiser la maintenance préventive, contribuant ainsi à accroître la fiabilité et l'efficacité des installations solaires.

- **Les éléments fondamentaux de l'approche de détection des anomalies**

Dans les systèmes photovoltaïques, le maintien d'une performance optimale est essentiel pour garantir une production efficace d'électricité à partir de l'énergie solaire. Cependant, ces systèmes peuvent être sujets à des pannes ou à des anomalies entraînant une baisse de rendement ou un arrêt complet. Ainsi, les méthodologies de détection des anomalies constituent des outils fondamentaux pour assurer la continuité du fonctionnement et identifier les problèmes dès leur apparition.

#### A. Méthodologies utilisées pour la détection des anomalies dans les systèmes photovoltaïques

##### 1. Observation visuelle et techniques thermographiques

L'inspection visuelle constitue une méthode simple mais essentielle pour l'identification de défauts manifestes dans les installations photovoltaïques. Elle permet notamment de repérer des anomalies telles que la corrosion des composants, la casse du verre frontal des modules ou encore la détérioration des câbles électriques. Cependant, certains défauts plus subtils, comme les points chauds (*hot spots*) ou les surchauffes d'origine électrique, ne sont pas détectables à l'œil nu. Dans ce cas, des techniques avancées telles que la **thermographie infrarouge** ou l'**électroluminescence** sont mises en œuvre. Ces approches reposent sur l'utilisation de caméras thermiques ou d'équipements spécialisés capables de mettre en évidence les **signatures thermiques anormales** associées à des dysfonctionnements internes au sein des modules ou des câblages du système PV [15].

##### 2. Analyse des caractéristiques électriques (courbe I-V)

L'analyse de la courbe courant-tension (I-V) constitue l'une des méthodes les plus répandues pour la détection et la localisation des défauts dans la partie continue (DC) des générateurs photovoltaïques. En comparant la courbe mesurée à celle attendue dans des conditions normales, il est possible d'identifier diverses anomalies telles que des défauts de connexion, des effets d'ombrage partiel ou une dégradation des cellules photovoltaïques.

Cette approche s'appuie souvent sur des modèles mathématiques représentatifs du comportement du générateur PV. Des algorithmes numériques, tels que la méthode de Newton-

Raphson, sont utilisés pour résoudre ces modèles et simuler le comportement électrique idéal, facilitant ainsi la détection des écarts caractéristiques des défauts [16] [17].

### **3. Modélisation et diagnostic paramétrique**

La modélisation paramétrique repose sur des méthodes analytiques permettant d'estimer, à partir de mesures électriques, les paramètres fondamentaux des modules PV tels que la résistance série, la résistance parallèle ou encore le courant de court-circuit.

L'estimation de ces paramètres peut être effectuée à l'aide de techniques statistiques comme le maximum de vraisemblance. L'analyse des résidus (écarts entre les mesures réelles et les valeurs modélisées) permet ensuite de détecter, localiser et quantifier les dégradations affectant les performances des modules [18].

### **2. Approches basées sur l'apprentissage automatique**

Les techniques d'intelligence artificielle, et en particulier les réseaux de neurones artificiels (ANN), connaissent un essor important dans le domaine du diagnostic des systèmes photovoltaïques. Ces modèles sont capables d'apprendre à partir de grandes quantités de données historiques ou simulées, afin de reconnaître automatiquement les signatures associées à différents types de défauts. Par rapport aux approches classiques fondées sur des seuils fixes, les méthodes d'apprentissage automatique offrent généralement une meilleure robustesse, précision, et capacité de généralisation face à des conditions de fonctionnement variables [19].

### **5. Méthode de diagnostic par comparaison entre sous-champs**

Dans les systèmes photovoltaïques comportant plusieurs sous-champs aux orientations ou inclinaisons différentes, une méthode efficace consiste à comparer les performances relatives de ces sous-ensembles. Cela se fait en mesurant des grandeurs électriques (courant, tension, puissance) à différents moments, puis en transposant les mesures d'un sous-champ de référence selon son orientation et inclinaison, afin de générer des valeurs de référence pour les autres sous-champs. La comparaison entre ces références et les mesures réelles permet de détecter rapidement des anomalies. Cette méthode présente l'avantage d'être peu dépendante des capteurs d'irradiance et s'adapte bien aux variations météorologiques, grâce à l'utilisation de modèles de transposition adaptés [20].

### **6. Détection d'anomalies à partir des séries temporelles de production**

L'analyse des séries temporelles de production énergétique constitue une autre approche puissante pour la détection des anomalies. Des algorithmes spécifiques examinent l'évolution de la production électrique au fil du temps afin d'identifier des comportements inhabituels, tels qu'une absence de production momentanée ou prolongée pendant la journée, des effets d'ombrage (le matin ou en fin de journée), ou encore une orientation sous-optimale des modules. Ces techniques s'appuient sur des métriques de performance qui permettent de quantifier la sévérité des anomalies et se montrent robustes face aux variations des conditions climatiques [121].

### **3.7 Conclusion**

Au terme de ce chapitre, nous avons présenté une approche conceptuelle détaillée pour la conception d'un modèle intelligent basé sur des techniques d'apprentissage profond, visant à générer automatiquement des descriptions textuelles de l'état des cellules photovoltaïques à partir d'images thermiques ou visuelles. Ce modèle repose sur une intégration entre un extracteur de caractéristiques visuelles fondé sur le réseau ResNet et un générateur de texte séquentiel de type LSTM, formant ainsi une architecture hybride combinant vision par ordinateur et traitement du langage naturel.

Ce travail a permis de formaliser la méthodologie de développement envisagée, incluant la préparation et l'analyse des données, la conception de l'architecture du système, ainsi que la définition des paramètres d'apprentissage et des critères d'évaluation. Les défis potentiels liés à la rareté des données annotées, à la diversité des défauts et à la complexité de la génération linguistique ont également été discutés.

Bien que la mise en œuvre pratique du modèle et sa phase d'expérimentation n'aient pas encore été achevées, cette étude constitue une base solide pour les étapes à venir. Les perspectives futures incluent la validation du modèle sur des données réelles, ainsi que l'intégration de modèles plus récents tels que les Transformers ou CLIP, en vue d'améliorer la précision des descriptions et la capacité du système à généraliser dans des contextes industriels concrets.

## Références bibliographiques

- [1] Susmiti. "Image Caption Generation: A Detailed Comparative Study of CNN-LSTM vs ResNet-GRU." Medium, 2022.
- [2] Satya Mallick. "Image Captioning using Deep Learning." LearnOpenCV, 2021.
- [3] Zhang, Yuan, Mingjin Chen, Weiguo Fan, et al. "Automatic Image Captioning Based on ResNet50 and LSTM with Soft Attention." Journal of Healthcare Engineering, vol. 2020, Article ID 8909458, 2020.
- [4] Millendorf, Matthew, Edward Obropta, and Nikhil Vadhavkar. "Infrared solar module dataset for anomaly detection." Proc. Int. Conf. Learn. Represent. 2020.).
- [5] Djabri, Mohammed. "Génération automatique de descriptions textuelles d'images par apprentissage profond." Mémoire de maîtrise, Université du Québec à Montréal (UQAM), 2020.
- [6] DataScientest. "Long Short-Term Memory (LSTM) : tout savoir sur ce type de réseau de neurones." DataScientest.com, 2022.
- [7] Satya Mallick. "Image Captioning using Deep Learning." LearnOpenCV, 2021.
- [8] Jason Brownlee. "CNN Long Short-Term Memory Networks." Machine Learning Mastery, 2017.
- [9] Asif Iqbal Khan, Muhammad Usama, et al. "A Combined Deep CNN-LSTM Network for the Detection of Novel Coronavirus (COVID-19) from X-ray Images." medRxiv preprint, 2020.
- [10] Soh, Moses. "Learning CNN-LSTM Architectures for Image Caption Generation." CS224d Final Project Report, Department of Computer Science, Stanford University, 2015.
- [11] Pravallika, V., Uday Kiran, V., Rahul, B., Neelima, N., Rishi Patnaik, G., & Sreejyothshna Ankam, Dr. (2025). Deep learning-based image captioning: A hybrid CNN-LSTM approach. International Journal of Research Publication and Reviews, 6(4), 2459–2463.
- [12] Stha, A. "Image Caption Generator using CNN and LSTM." GitHub repository, 2021.
- [13] Osman, Asmaa A. E., Mohamed A. Wahby Shalaby, Mona M. Soliman, and Khaled M. Elsayed. "Novel concept-based image captioning models using LSTM and multi-encoder transformer architecture." Scientific Reports, vol. 14, Article no. 20762, 2024.
- [168] Daniella. "Image Captioning: AI that describes your images." Innovatiana, January 14, 2025.
- [169] Oukil, Kawtar, Houda Mabrouk, and M. Mohammed Reda. "Image Captioning." Presentation, Faculty of Sciences Ain Chock, Department of Mathematics and Computer Science, 2024 [169] Foorthis, Ralph. "On the nature and types of anomalies: A review of deviations in data." International Journal of Data Science and Analytics, vol. 12, no. 4, 2021, pp. 297–331
- [14] Lebert, Nicolas, Cécile Miquel, and Jérémy Sarantou. "Methods for detecting electrical malfunctions in photovoltaic installations." Agence Qualité Construction, June 2019.
- [15] Bun, Long. "Detection and Localization of Faults in a Photovoltaic System." Master's thesis, Mouloud Mammeri University of Tizi-Ouzou, 2020.
- [16] Zaiour, Meriem. "Detection and Localization of Faults in a Photovoltaic System." Master's thesis, Badji Mokhtar University – Annaba, 2020.
- [17] Ayang, Albert. "Diagnostic of a photovoltaic storage system using parametric estimation and ADRC control, integrated into a standalone cogeneration power plant." PhD thesis, Université du Québec à Chicoutimi, 2020.

[18] Toubal Maamar, Alla Eddine, Samir Ladjouzi, Rachid Taleb, and Yacine Kacemi. "Detection and classification of faults for a PV generator: Comparative study between thresholding method and neural networks." *Journal of Renewable Energies*, vol. 21, no. 1, 2018, pp. 45–53.

[19] Pichenot, Grégoire, and Imene Khaled. "Method for diagnosing a photovoltaic system." French patent FR3028114A1, published May 6, 2016

[20] Branco, Pedro Miguel Mayer. "Anomaly detection in photovoltaic systems." Master's thesis, NOVA Information Management School, Universidade Nova de Lisboa, 2019

A decorative frame resembling a scroll, with a vertical bar on the left side and rounded corners. The text is centered within this frame.

## **Chapitre 4 :Expérimentation et Évaluation des Modèles de Génération de Légendes d'Images**

## 4.1 Introduction

Le développement rapide des énergies renouvelables, notamment le photovoltaïque, s'accompagne d'un besoin croissant en systèmes intelligents capables de surveiller, diagnostiquer et améliorer en continu la performance des installations. Dans ce contexte, l'inspection automatisée des panneaux solaires représente un enjeu technologique majeur, visant à assurer une maintenance proactive, à réduire les pertes d'énergie et à prolonger la durée de vie des systèmes. L'inspection thermique, réalisée à l'aide de caméras infrarouges embarquées sur des drones, permet de détecter rapidement les anomalies de température révélatrices de défauts physiques ou électriques au niveau des cellules photovoltaïques. Cependant, l'interprétation manuelle des images thermiques est une tâche laborieuse, sujette à erreurs, et peu évolutive dans un cadre industriel à grande échelle.

Face à cette problématique, ce travail s'inscrit dans la volonté de développer un système automatisé basé sur l'intelligence artificielle, capable non seulement de détecter des anomalies à partir d'images thermiques, mais également de **générer automatiquement une description textuelle précise et multilingue** de l'état observé. C'est dans cette optique que s'insère la génération automatique de légendes d'images, également connue sous le nom d'Image Captioning. Il s'agit d'un domaine à l'intersection du traitement d'image et du traitement automatique du langage naturel (TALN), consistant à produire une phrase descriptive à partir du contenu visuel d'une image.

Le présent chapitre constitue la partie expérimentale de notre projet et vise à mettre en œuvre et à évaluer deux architectures de modèles d'Image Captioning appliquées spécifiquement aux images thermiques issues de panneaux photovoltaïques. La première architecture repose sur la combinaison du réseau neuronal convolutif **ResNet50** pour l'extraction des caractéristiques visuelles, avec un réseau **LSTM (Long Short-Term Memory)** pour la génération de la description textuelle. La seconde architecture remplace le LSTM par un **GRU (Gated Recurrent Unit)**, un modèle récurrent plus léger en termes de paramètres.

Ces deux types de réseaux – LSTM et GRU – appartiennent à la famille des réseaux de neurones récurrents (RNN), qui sont conçus pour modéliser les dépendances temporelles ou séquentielles dans les données. Le **LSTM** se distingue par sa capacité à capturer des dépendances à long terme grâce à une architecture composée de cellules mémoires et de trois portes (entrée, oubli, sortie), ce qui le rend performant mais coûteux en ressources. Le **GRU**, quant à lui, utilise une structure plus simple à deux portes (mise à jour et réinitialisation), ce qui réduit les temps d'apprentissage tout en conservant des performances élevées dans de nombreux cas. Le choix de comparer ces deux approches répond à une question centrale : faut-il privilégier la précision linguistique ou la légèreté computationnelle dans un système embarqué de description automatique ?

Les données utilisées dans notre expérimentation ont été rigoureusement préparées. Un jeu de 2000 images thermiques a été collecté, couvrant une diversité de situations, allant des panneaux en bon état aux défauts typiques comme les cellules défectueuses ou les diodes de dérivation activées. Ces images ont été divisées selon une répartition classique en **80 % pour l'apprentissage** et **20 % pour les tests**, assurant ainsi un bon compromis entre l'efficacité de l'entraînement et la représentativité des cas réels dans l'évaluation.

L'entraînement des modèles a été réalisé en tenant compte de plusieurs paramètres critiques tels que la taille des lots (*batch size*), le taux d'apprentissage (*learning rate*), la fonction de

perte (*loss function*) ainsi que le nombre d'époques. Ces paramètres ont été optimisés pour équilibrer la rapidité de convergence, la stabilité de l'apprentissage et la généralisation du modèle. De plus, des techniques de traduction automatique ont été intégrées pour générer des descriptions multilingues (anglais, français, arabe), renforçant ainsi la polyvalence du système dans un contexte international.

Enfin, pour évaluer objectivement les performances linguistiques des modèles, nous nous sommes appuyés sur la **métrique BLEU (Bilingual Evaluation Understudy)**, largement reconnue dans le domaine de la génération de textes. Ce score mesure la similarité entre les descriptions générées par le modèle et des descriptions de référence, en analysant la concordance des mots et des expressions à plusieurs niveaux de granularité (unigrammes, bigrammes, etc.).

Ce chapitre détaille l'ensemble du protocole expérimental : description des données, présentation des architectures, configuration de l'entraînement, métriques d'évaluation et analyse des résultats. Il constitue une contribution concrète à la mise en œuvre de systèmes intelligents de diagnostic visuel assisté par IA dans le secteur de l'énergie photovoltaïque.

### 4.2 Méthodologie

Cette section décrit de manière détaillée le processus de mise en œuvre du système de génération automatique de légendes à partir d'images thermiques de panneaux photovoltaïques (Figure 4.1). Elle couvre la collecte des données, leur préparation, ainsi que la conception de l'architecture des modèles testés. L'objectif principal est de construire un pipeline cohérent qui transforme une image brute en une description textuelle pertinente et précise, adaptée à un usage multilingue dans un contexte industriel réel.

#### 4.2.1 Collecte et Préparation des Données

La première étape de tout projet d'apprentissage profond consiste à constituer un ensemble de données représentatif et de haute qualité. Dans le cadre de ce projet, les données sont constituées de **2000 images thermiques** de panneaux solaires, capturées dans un environnement réel à l'aide de **drones équipés de caméras infrarouges de haute précision**. Ce choix technologique a permis de collecter des données à grande échelle, rapidement et sans interruption du fonctionnement des installations.

Les images thermiques contiennent des informations cruciales sur la **distribution de température** à la surface des modules photovoltaïques. Les variations anormales de température sont souvent le symptôme de **défauts internes** : cellules défailtantes, points chauds, connexions rompues ou dysfonctionnements des diodes de dérivation (by-pass). Ces images offrent donc une source d'information riche pour la détection automatique de pannes.

Afin de garantir la validité de l'apprentissage supervisé, chaque image a été **annotée manuellement** avec une description technique de l'état observé. Ces descriptions ont été élaborées par des experts ou extraites de bases de connaissances industrielles, et elles ont été validées pour garantir leur exactitude. Un exemple typique de description est :

*"Warning: activated by-pass diodes typically affecting 2 over 3 of the PV module."*

Chaque annotation est ensuite traduite automatiquement en français et en arabe à l'aide de bibliothèques Python telles que **googletrans** ou **DeepL API**, afin de constituer une base de données multilingue. Cette approche favorise l'adaptabilité du système dans des régions géographiques variées et pour des utilisateurs de différents niveaux techniques.

La base de données a été **divisée en deux ensembles distincts** :

- **Ensemble d'apprentissage (Train Set) : 80 %**, soit 1600 images, utilisé pour entraîner le modèle.
- **Ensemble de test (Test Set) : 20 %**, soit 400 images, destiné à évaluer la capacité de généralisation du modèle.

Avant l'entraînement, les images ont subi une série d'opérations de **prétraitement** :

- Redimensionnement à une taille standard (80x80 pixels pour ResNet).
- Normalisation des valeurs de pixels.
- Éventuel filtrage des bruits ou artefacts thermiques.

Ces étapes sont essentielles pour garantir l'homogénéité des entrées, réduire la variance et faciliter la convergence de l'apprentissage.

### 4.2.2 Architecture du Modèle

L'architecture retenue pour ce projet s'appuie sur une approche **Encoder-Decoder**, largement utilisée dans les systèmes de génération de texte à partir d'images. Elle est composée de deux modules fondamentaux :

1. **L'encodeur visuel (Visual Encoder)**
2. **Le décodeur linguistique (Text Decoder)**

#### 1. Extraction de caractéristiques visuelles – ResNet50

Le module encodeur est basé sur le réseau neuronal **ResNet50**, un modèle convolutionnel pré-entraîné sur ImageNet, capable de capturer des caractéristiques visuelles complexes à différents niveaux de profondeur. ResNet50 comprend **50 couches** et utilise des **connexions résiduelles** qui permettent de résoudre le problème du gradient évanescent et d'accélérer la convergence.

Dans notre pipeline, ResNet50 prend en entrée une image thermique prétraitée et produit un **vecteur de caractéristiques de haute dimension** (typiquement de taille 2048), représentant les aspects visuels dominants de l'image (formes, textures, contrastes thermiques). Ce vecteur sert ensuite de base pour la génération linguistique.

#### 2. Génération de texte – LSTM ou GRU

Deux variantes ont été testées pour le module de décodage :

##### **LSTM (Long Short-Term Memory)**

Le LSTM est un réseau neuronal récurrent (RNN) spécialement conçu pour modéliser les dépendances à long terme dans les séquences. Il intègre trois portes :

- **Porte d'entrée** : contrôle quelles informations du vecteur d'entrée sont pertinentes.
- **Porte d'oubli** : décide quelles informations passées doivent être ignorées.
- **Porte de sortie** : génère la sortie à chaque étape de la séquence.

Cette architecture le rend particulièrement apte à générer des descriptions longues, grammaticalement correctes, et sémantiquement cohérentes.

### GRU (Gated Recurrent Unit)

Le GRU est une version simplifiée du LSTM, avec uniquement deux portes :

- **Porte de mise à jour**
- **Porte de réinitialisation**

Il consomme moins de ressources, converge plus rapidement, et conserve des performances compétitives dans la majorité des tâches de génération. Il est donc particulièrement adapté pour des déploiements en temps réel ou sur des appareils embarqués.

### 3. Fonctionnement global

Le vecteur de caractéristiques extrait par ResNet50 est injecté dans le modèle LSTM ou GRU, qui génère séquentiellement une phrase descriptive mot par mot. Ce processus repose sur un dictionnaire préétabli (vocabulaire) et utilise des techniques comme l'**Embedding** des mots et la **fonction Softmax** pour choisir le mot le plus probable à chaque étape.

La **fonction de perte** utilisée est la **Cross-Entropy**, mesurant l'écart entre la séquence prédite et la séquence de référence. L'optimisation est assurée par l'algorithme **Adam**, avec un taux d'apprentissage ajusté automatiquement.

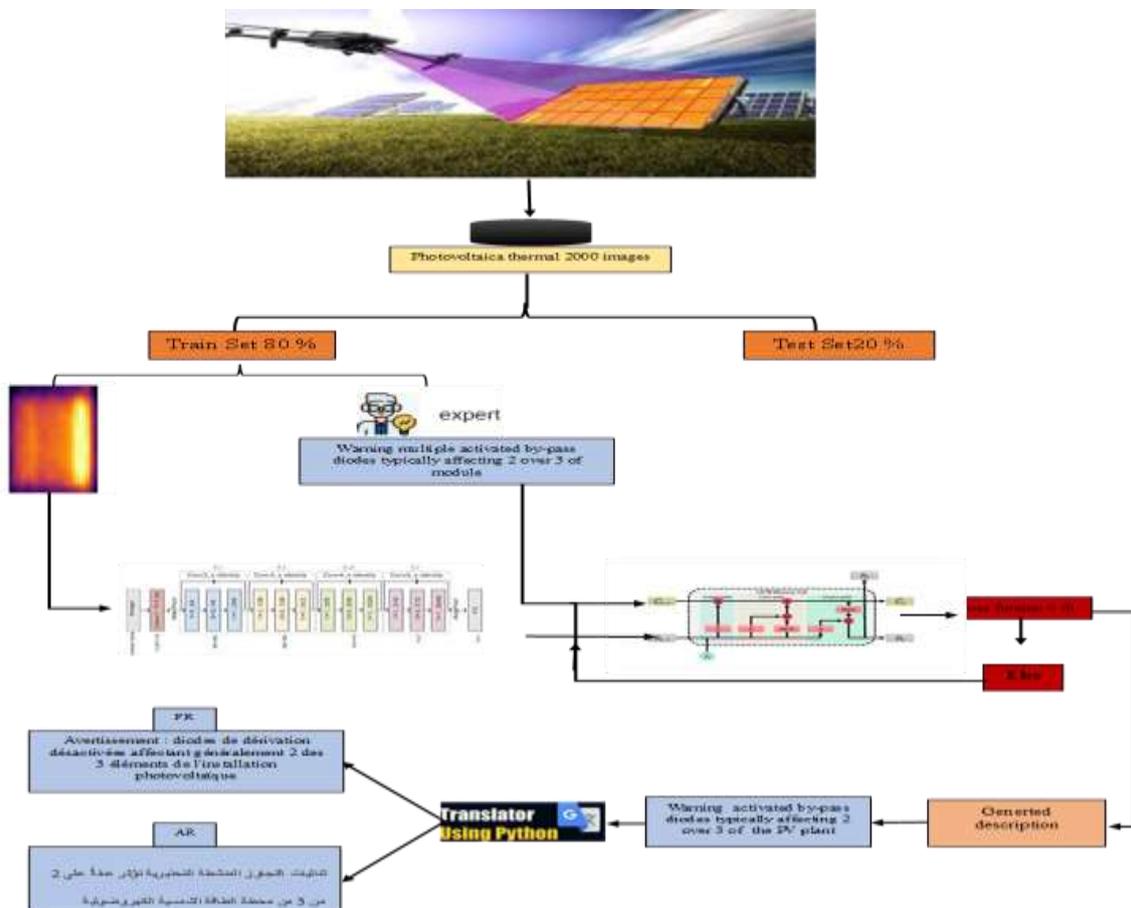


Figure 4.1: Organigramme

### 4.3 Evaluation du model de description

L'évaluation des performances dans les systèmes de génération automatique de texte, tels que les modèles de description d'images (Image Captioning), constitue un défi fondamental en traitement automatique du langage naturel (TALN). Contrairement aux tâches de classification, l'évaluation de la génération de texte ne peut se réduire à une simple exactitude binaire, car plusieurs réponses peuvent être linguistiquement valides pour une même entrée visuelle.

#### 4.3.1 Méthodes d'évaluation en TALN

Plusieurs métriques ont été proposées pour quantifier la qualité des textes générés par les modèles neuronaux :

- **BLEU (Bilingual Evaluation Understudy)** : mesure la précision des n-grammes dans la sortie du modèle par rapport à des phrases de référence. C'est l'une des premières métriques utilisées dans la traduction automatique et l'une des plus répandues.
- **ROUGE (Recall-Oriented Understudy for Gisting Evaluation)** : se concentre davantage sur le rappel des mots et des séquences, en particulier utilisée dans les tâches de résumé automatique. Elle évalue la couverture des unités lexicales importantes présentes dans les références.
- **METEOR** : prend en compte les correspondances flexibles entre les mots, y compris les synonymes et les variantes morphologiques. Elle s'appuie sur des ressources linguistiques comme WordNet.
- **CIDEr (Consensus-based Image Description Evaluation)** : conçue spécifiquement pour l'évaluation des descriptions d'images. Elle attribue un poids plus élevé aux n-grammes fréquemment observés dans plusieurs descriptions de référence.

Chacune de ces métriques présente des avantages et des limites. Par exemple, **BLEU est strictement basé sur la correspondance exacte des mots**, ce qui peut ignorer des variations légitimes du langage. En revanche, des métriques comme METEOR ou CIDEr tiennent compte de la diversité linguistique.

#### 4.3.2 Utilisation du score BLEU dans ce travail

Dans le cadre de ce mémoire, nous avons choisi d'utiliser la **métrique BLEU** pour évaluer la performance des modèles de génération de descriptions (ResNet50-LSTM et ResNet50-GRU). Ce choix est motivé par la robustesse et la simplicité de BLEU, ainsi que sa large adoption dans la littérature scientifique liée à la génération automatique de texte.

Le **score BLEU** mesure le degré de similarité entre la phrase générée par le modèle et une ou plusieurs phrases de référence, sur la base des **n-grammes** (séquences de  $n$  mots consécutifs). Il s'agit donc d'une métrique de **précision multigranulaire**, qui pénalise à la fois les erreurs lexicales et les erreurs de structure.

La formule de calcul du score BLEU est la suivante :

$$BLEU = \exp \left( \sum_{\{n=1\}}^N w_n \log p_n \right) \times BP \quad (4.1)$$

Avec :

- **P<sub>n</sub>** : la précision des n-grams de rang n,
- **W<sub>n</sub>** : le poids attribué à chaque rang (souvent égal),
- **N** : le nombre maximal de n-grams considérés (souvent N = 4),
- **BP** : le **Brevity Penalty**, une pénalité appliquée si la phrase générée est trop courte par rapport à la référence.

Cette formule permet de capturer plusieurs dimensions de la qualité d'un texte :

BLEU-1 : évalue la précision des mots isolés (unigrammes)

BLEU-2 : évalue la cooccurrence de paires de mots (bigrammes)

BLEU-3 et BLEU-4 : introduisent la dimension syntaxique et la cohérence des segments plus longs

En utilisant plusieurs niveaux de BLEU (du 1 au 4), nous avons pu analyser la qualité linguistique de manière progressive, allant de la correspondance lexicale simple à la fluidité syntaxique globale. Dans les sections suivantes, nous présentons les résultats numériques obtenus pour chaque modèle testé, suivis d'une analyse comparative illustrée par des graphiques et exemples concrets de descriptions générées.

#### 4.4 Résultats et Discussion

Cette section présente les résultats obtenus après l'entraînement et l'évaluation des deux modèles de génération de descriptions : ResNet50-LSTM et ResNet50-GRU. Les performances ont été mesurées à l'aide du BLEU Score, décomposé en quatre niveaux (BLEU-1 à BLEU-4) afin de mieux cerner la qualité linguistique des descriptions générées à différents niveaux de granularité.

##### 4.4.1 Résultats quantitatifs

Le tableau ci-dessous (tableau 4.4) synthétise les scores BLEU obtenus pour chaque architecture :

Model	Rasnet50-LSTM	Rasnet50-GRU
BLEU-S1	0.769706	0.759639
BLEU-S2	0.753116	0.750207
BLEU-S3	0.741402	0.738911
BLEU-S4	0.723224	0.722981

**Tableau 4.1** : Résultats comparatifs des scores BLEU pour les modèles ResNet50-LSTM et ResNet50-GRU sur l'ensemble de test

Ces résultats indiquent que les deux modèles étudiés – ResNet50-LSTM et ResNet50-GRU – présentent des performances remarquables en matière de génération de descriptions automatiques à partir d'images thermiques. Les scores BLEU obtenus sont tous supérieurs à 0.72, ce qui traduit une excellente qualité linguistique dans les phrases générées, tant au niveau du choix des mots (précision lexicale) que de la structure grammaticale (cohérence syntaxique). Plus précisément, on observe que le modèle ResNet50-LSTM atteint un score BLEU-1 de 0.7697, ce qui indique une correspondance très élevée entre les mots générés et ceux des descriptions de référence. Ce niveau de performance témoigne d'une forte capacité du modèle à sélectionner un vocabulaire pertinent et adapté au contenu visuel analysé. En comparaison, ResNet50-GRU obtient un score BLEU-1 de 0.7596, légèrement inférieur, mais tout de même très élevé, confirmant sa robustesse.

En progressant vers des n-grammes plus longs (BLEU-2 à BLEU-4), qui évaluent la fluidité des séquences de mots et la cohérence contextuelle, les scores décroissent de manière modérée, comme attendu. Cela s'explique par la nature plus stricte de l'évaluation : il ne suffit plus de choisir les bons mots, mais aussi de les ordonner correctement pour former des phrases grammaticalement correctes. Le modèle LSTM conserve un avantage stable à chaque niveau :

- BLEU-2 : **0.7531** (LSTM) vs. **0.7502** (GRU)
- BLEU-3 : **0.7414** (LSTM) vs. **0.7389** (GRU)
- BLEU-4 : **0.7232** (LSTM) vs. **0.7230** (GRU)

Bien que l'écart entre les deux modèles soit faible (de l'ordre de 0.002 à 0.01), il met en évidence la **capacité supérieure du LSTM à capturer les relations à long terme entre les mots**, ce qui est essentiel pour produire des descriptions fluides et grammaticalement structurées. Ce constat est particulièrement significatif dans les cas où les descriptions techniques incluent des séquences plus complexes, des syntagmes spécialisés, ou des formulations conditionnelles fréquentes dans le langage industriel.

Toutefois, **le modèle GRU ne doit pas être sous-estimé**. Malgré une structure plus simple et une vitesse d'apprentissage supérieure, il parvient à atteindre des performances très proches de celles du LSTM, en particulier au niveau BLEU-4, où l'écart devient quasiment négligeable (**0.723224 vs 0.722981**). Cela suggère que, dans des contextes où les ressources computationnelles sont limitées, ou dans des environnements embarqués, **GRU constitue une alternative pragmatique et efficace**, surtout si la génération linguistique est soumise à des contraintes de temps réel.

Enfin, il est important de souligner que l'évaluation par BLEU Score reste une **approche quantitative**, et qu'elle doit idéalement être complétée par une **analyse qualitative** des phrases générées, ainsi que par des retours d'utilisateurs finaux sur la lisibilité, la précision et la pertinence des descriptions. Cette analyse sera abordée dans les sections suivantes, notamment à travers des exemples concrets issus du jeu de test.

### 4.4.2 Analyse comparative

- Le modèle ResNet50-LSTM montre une légère supériorité par rapport à ResNet50-GRU à tous les niveaux du BLEU Score. Cette différence, bien que marginale, peut être attribuée à la structure plus complexe du LSTM, qui lui permet de mieux capturer les dépendances à long terme dans les séquences textuelles. Cela se reflète particulièrement dans les scores BLEU-3 et BLEU-4, qui impliquent des trigrammes et quadrigrammes, indicateurs d'une meilleure cohérence syntaxique.

- À l'inverse, le modèle ResNet50-GRU, tout en étant légèrement moins précis, présente l'avantage d'une architecture plus simple et plus rapide à entraîner. Dans des contextes où les ressources computationnelles sont limitées ou où une réactivité en temps réel est requise (ex. : drones, systèmes embarqués), le GRU peut représenter un compromis efficace entre performance et efficacité.

#### 4.4.3 Interprétation linguistique des scores

Les scores BLEU sont plus qu'un simple indicateur numérique : ils reflètent la capacité du modèle à produire des descriptions :

- BLEU-1 élevé ( $\geq 0.75$ ) : indique une bonne précision au niveau du choix des mots, c'est-à-dire que le vocabulaire utilisé est pertinent.
- BLEU-2 et BLEU-3 élevés : montrent une bonne construction syntaxique de phrases courtes à moyennes.
- BLEU-4 élevé ( $\geq 0.72$ ) : suggère que les modèles réussissent à générer des phrases non seulement correctes lexicalement, mais aussi fluides et grammaticalement structurées.

#### 4.4.4 Résultats visuels

Dans la classe "**No-anomaly**" Les images montrent un ensemble de cellules solaires photovoltaïques dans leur état nominal, c'est-à-dire l'état de référence idéal dans lequel la cellule est supposée fonctionner avec une efficacité maximale, sans aucun type de défaut ni de dysfonctionnement, qu'il soit visible ou fonctionnel.

Ces images sont classées dans la catégorie "No-anomaly" (sans anomalie), ce qui indique que la structure des cellules est intacte sur le plan physique et que les composants qui les constituent tels que les surfaces en verre, les couches actives et les connexions électriques ne présentent ni fissures, ni fractures, ni déformations de couleur pouvant indiquer des défauts de fabrication ou une dégradation due au temps ou aux conditions environnementales.

Les cellules solaires nominales se caractérisent par une distribution régulière et symétrique des sous-cellules à l'intérieur du panneau. Les barres conductrices (busbars) sont clairement visibles, et aucune tache sombre ni zone de réflexion irrégulière n'est présente, ce qui témoigne d'un bon écoulement du courant électrique et de l'absence de problèmes liés à l'absorption de la lumière ou à la conversion photoélectrique.

Cette uniformité visuelle et technique reflète l'état de fonctionnement optimal de la cellule, lequel sert de référence pour comparer les autres cellules lors des processus d'inspection et d'évaluation.

D'un point de vue pratique, ces images sont utilisées comme données de "Ground Truth" (vérité terrain) dans les ensembles de données exploités dans les projets d'intelligence artificielle, en particulier ceux dédiés au développement de modèles d'apprentissage profond (Deep Learning) pour le diagnostic des performances des cellules solaires.

Ces cellules nominales servent de modèles de référence permettant d'entraîner les algorithmes à distinguer entre les états sains et les états défectueux, ce qui contribue à améliorer la précision de la détection automatique des anomalies, à réduire le besoin d'inspection manuelle et, en fin

de compte, à renforcer la fiabilité des systèmes photovoltaïques dans les environnements industriels ou sur le terrain.

Dans la classe "**végétation**", quatre images liées au même cas apparaissent, dont trois portent la même description suivante :

GT : panels blocked by vegetation

EN : panels blocked by vegetation

FR : panneaux bloqués par végétation

AR : لوحات محجوبة بالنباتات

Cette description indique clairement que les panneaux solaires sur ces images sont partiellement ou totalement recouverts par de la végétation telle que des herbes, des arbustes ou des branches d'arbres, ce qui empêche la lumière du soleil d'atteindre pleinement la surface des panneaux et affecte ainsi négativement leur efficacité dans la production d'énergie électrique.

Cependant, la quatrième image, bien qu'elle appartienne à la même classe "végétation", présente une description notablement différente :

GT : panels blocked by vegetation

EN : nominal photovoltaic solar cell

FR : cellule solaire photovoltaïque nominale

AR : الخلية الشمسية الكهروضوئية الاسمية

Ici, il y a une contradiction dans la description, car le GT et l'EN indiquent que les panneaux sont bloqués par la végétation, tandis que les descriptions en anglais, français et arabe indiquent que l'image montre une cellule solaire dans son état nominal, c'est-à-dire dans son état naturel et sain.

Explication de la comparaison : Dans les première, deuxième et quatrième images, la description est précise et cohérente : les panneaux sont couverts par la végétation, ce qui indique la présence d'un obstacle naturel affectant la performance des panneaux. Cette situation cause généralement une diminution de la capacité des panneaux à capter les rayons du soleil et à les convertir en électricité.

En revanche, pour la troisième image, il semble y avoir une erreur d'étiquetage ou un chevauchement des données. Il n'est pas logique qu'un panneau soit à la fois bloqué par la végétation et classé comme étant en état nominal parfait. L'état nominal signifie que la cellule fonctionne à une efficacité optimale sans aucun obstacle ni défaut, ce qui contredit la présence de végétation recouvrant les panneaux.

Dans la classe "**Offline-Module**" Les images montrent un échauffement global du module photovoltaïque, où l'on observe que l'ensemble du module est chauffé de manière uniforme. Cette situation est classée dans la catégorie « Offline-Module », ce qui signifie que le module

est hors service ou déconnecté du réseau. Dans ce contexte, le fait que tout le module soit chauffé indique une anomalie pouvant être causée par une coupure dans le circuit électrique ou une interruption du flux de courant, entraînant une accumulation de chaleur provenant du rayonnement solaire sans conversion efficace en énergie électrique.

Ce type d'anomalie se manifeste clairement dans les images thermiques, où l'on détecte des températures élevées et homogènes à la surface du module. Cela contraste avec le fonctionnement normal, où la chaleur est répartie de manière variable selon l'activité photovoltaïque. En fonctionnement optimal, la majorité de l'énergie solaire est transformée en électricité ; dans ce cas précis, l'énergie thermique s'accumule en raison d'un arrêt du processus de conversion.

Ce cas constitue une référence précieuse dans les ensembles de données de type « Ground Truth », utilisés pour l'entraînement des modèles d'intelligence artificielle. Il permet aux systèmes d'apprentissage de reconnaître les modules « inactifs » ou « morts », c'est-à-dire ceux qui, bien qu'exposés au soleil, ne produisent pas d'électricité. L'identification de ce type de défaillance améliore la surveillance intelligente, facilite la détection rapide des unités inefficaces, et contribue ainsi à optimiser la maintenance et la performance globale des installations photovoltaïques.

Dans la Class "**Cracking**" Quatre images illustrent une anomalie causée par des fissures à la surface du module solaire. Pour les trois premières images, les descriptions sont identiques et cohérentes, comme suit :

GT : module anomaly caused by cracking on the module surface

EN : module anomaly caused by cracking on the module surface

FR : anomalie du module causée par la fissuration sur la surface du module

AR : الشذوذ الناجم عن التكسير على سطح الوحدة النمطية :

Cette description indique la présence de fissures visibles à la surface des cellules ou du panneau solaire, ce qui peut affecter négativement les performances du module, en raison d'une possible infiltration d'humidité, d'une mauvaise conductivité électrique ou d'une accélération de la dégradation de la cellule. Ce type de défaut est généralement causé par des contraintes thermiques, une pression mécanique ou des défauts de fabrication.

Quant à la quatrième image, bien qu'elle soit classée dans la même catégorie "Cracking", sa description est différente :

GT : module anomaly caused by cracking on the module surface

EN : multiple hot spots on a thin film module

FR : plusieurs points chauds sur un module à couches minces

AR : يقع ساخنة متعددة على وحدة فيلم رقيقة :

Dans ce cas, il existe une contradiction évidente entre les champs :

Le champ GT indique la présence de fissures,

alors que les descriptions EN / FR / AR évoquent la présence de plusieurs points chauds, un autre type d'anomalie causée par une élévation anormale de la température à certains endroits du module.

Dans les trois premières images, la description est précise et cohérente avec la catégorie "Cracking", car les cellules solaires présentent des défauts causés par des fissures clairement visibles qui altèrent leurs performances.

En revanche, dans la quatrième image, il est probable qu'il y ait une erreur de classification ou une confusion dans les annotations, car il est incohérent de classer un module comme fissuré alors que la description fait référence à des "points chauds" – un problème thermique, et non mécanique.

Dans la classe "**Hot-Spot**" Quatre images illustrent un défaut de type "point chaud" dans les modules photovoltaïques. Pour les trois premières images, les descriptions sont identiques et cohérentes, comme suit :

GT : hot spot on a thin film module

EN : hot spot on a thin film module

FR : point chaud sur un module à couches minces

AR : بقعة ساخنة على وحدة فيلم رقيقة

Cette description fait référence à une zone spécifique sur la surface du module où la température augmente anormalement. Ce phénomène, connu sous le nom de "point chaud" (Hot Spot), survient généralement en raison d'un dysfonctionnement d'une ou plusieurs cellules dans la série, entraînant une accumulation de chaleur à cet endroit. Ce type d'anomalie provoque une baisse de performance du panneau, et peut endommager de manière permanente la cellule ou la couche mince.

Quant à la quatrième image, bien qu'elle soit classée également dans la catégorie "Hot-Spot", sa description est très différente :

GT : hot spot on a thin film module

EN : multiple activated pass pass pass diode typically affecting 1111

FR : diode de pass de pass de pass à passes activées multiples affectant généralement 1111

AR : تمرير متعدد المنشطات على 1111 عادة ما تؤثر الصمام الثنائي بتمرير

Dans ce cas, un conflit évident apparaît entre les champs :

Alors que le champ GT continue d'indiquer un "point chaud sur un module à couches minces", les descriptions EN / FR / AR semblent incohérentes ou confuses, et font plutôt référence à un

dysfonctionnement lié aux diodes de dérivation activées (bypass diodes), qui sont généralement utilisées pour protéger les cellules contre les dommages causés par les points chauds.

Dans la classe "**Diode-Multi**", qui concerne les cas d'activation multiple des diodes de dérivation (Bypass Diodes) dans les modules solaires, les quatre images documentent un dysfonctionnement particulier lié à cette activation.

Pour les première, deuxième et quatrième images :

Les descriptions sont identiques et cohérentes à travers les langues (GT, EN, FR, AR).

Elles indiquent une activation multiple des diodes de dérivation, affectant typiquement deux tiers du module (2 sur 3).

Cela reflète une défaillance réelle du module, puisque ces diodes sont utilisées pour contourner les cellules endommagées ou ombragées, donc leur activation signifie qu'une partie du module ne fonctionne pas efficacement.

Cette description est précise et conforme à la classe "Diode-Multi", qui se concentre sur les défauts liés à l'activation de plusieurs diodes.

Pour la troisième image Bien qu'elle appartienne à la même classe, ses descriptions diffèrent :

Le champ GT indique toujours une activation multiple des diodes (signalant un défaut).

En revanche, les descriptions EN, FR et AR indiquent une cellule photovoltaïque en état nominal, c'est-à-dire fonctionnant normalement et sans défaut.

Il s'agit d'une contradiction claire, car une cellule en parfait état ne peut pas simultanément présenter une activation multiple des diodes de dérivation.

Cela suggère une erreur probable dans la classification de la troisième image ou dans la saisie des descriptions, puisqu'une image d'une cellule saine ne devrait pas être incluse dans une classe dédiée aux défauts.

Les images 1, 2 et 4 ont des descriptions cohérentes et logiques, reflétant un dysfonctionnement réel dû à l'activation multiple des diodes de dérivation.

L'image 3 présente un désaccord entre la description GT, qui indique un défaut, et les autres descriptions qui indiquent un état nominal, suggérant une erreur de classification ou de description.

Il est donc nécessaire de revoir cette troisième image pour corriger sa classification ou ajuster ses descriptions afin qu'elles correspondent à l'état réel du module.

Cette classe "**Diode**" présente des images documentant l'activation d'une diode de dérivation (Bypass Diode) au sein du module solaire, où la diode activée concerne généralement environ un tiers du module.

Les descriptions dans différentes langues indiquent qu'une des diodes de dérivation a été activée, ce qui permet de contourner une partie des cellules du panneau, affectant typiquement environ un tiers du module solaire. L'activation de ces diodes se produit lorsqu'il y a un problème ou une ombre sur cette partie du panneau ; les diodes protègent alors le module contre d'éventuels dommages, tels que l'accumulation de chaleur ou la perte de performance.

Dans la classe "**Cell-Multi**", les images documentent l'apparition de « points chauds » ayant une forme géométrique carrée, mais il existe une contradiction dans les descriptions entre les langues :

Dans la description GT, il est indiqué que les points chauds apparaissent sur plusieurs cellules avec une géométrie carrée.

En revanche, dans les descriptions EN et FR, il est mentionné que le point chaud se trouve sur une seule cellule.

La description en arabe fait référence à la présence d'un point chaud, mais reste peu claire et imprécise.

Cette disparité dans les descriptions suggère une possible erreur ou un dysfonctionnement dans la nomenclature ou la classification, car la description de plusieurs cellules ne correspond pas à celle d'une seule cellule. Il est donc nécessaire de procéder à une revue attentive des données afin de clarifier la situation réelle.

Le nombre d'images par classe joue un rôle crucial dans la performance des modèles d'apprentissage profond. Lorsque certaines classes, comme "No-Anomaly", disposent de milliers d'images (par exemple 10 000), tandis que d'autres, telles que "Diode-Multi", n'en possèdent que quelques centaines (par exemple 175), un déséquilibre se crée dans les données. Ce déséquilibre peut entraîner un biais du modèle en faveur des classes majoritaires, car elles sont plus souvent représentées durant l'apprentissage, ce qui permet au modèle de mieux les reconnaître. En revanche, les classes peu représentées risquent de ne pas être suffisamment apprises, ce qui se traduit par une performance réduite lors de leur classification. En outre, un grand nombre d'images permet au modèle de capturer une diversité plus riche des cas possibles au sein d'une classe, favorisant ainsi une meilleure généralisation. À l'inverse, les classes avec peu d'images sont plus exposées au risque de sur-apprentissage (overfitting), car le modèle peut s'adapter de manière excessive aux rares exemples disponibles. Cela affecte aussi la fiabilité des évaluations, car les performances sur les classes peu fournies peuvent varier fortement. Pour atténuer ces effets, plusieurs approches sont possibles : le rééchantillonnage des données pour équilibrer les classes, la pondération de la fonction de perte afin de donner plus d'importance aux classes rares, ou encore l'augmentation de données, qui consiste à générer artificiellement des exemples supplémentaires pour enrichir les classes sous-représentées.

On dit que le nombre d'images influence directement la capacité du modèle à distinguer avec précision entre les catégories, car les catégories avec un grand nombre d'exemples forment mieux le modèle, tandis que les catégories avec peu d'exemples nécessitent une attention particulière pour éviter les biais et la dégradation des performances.



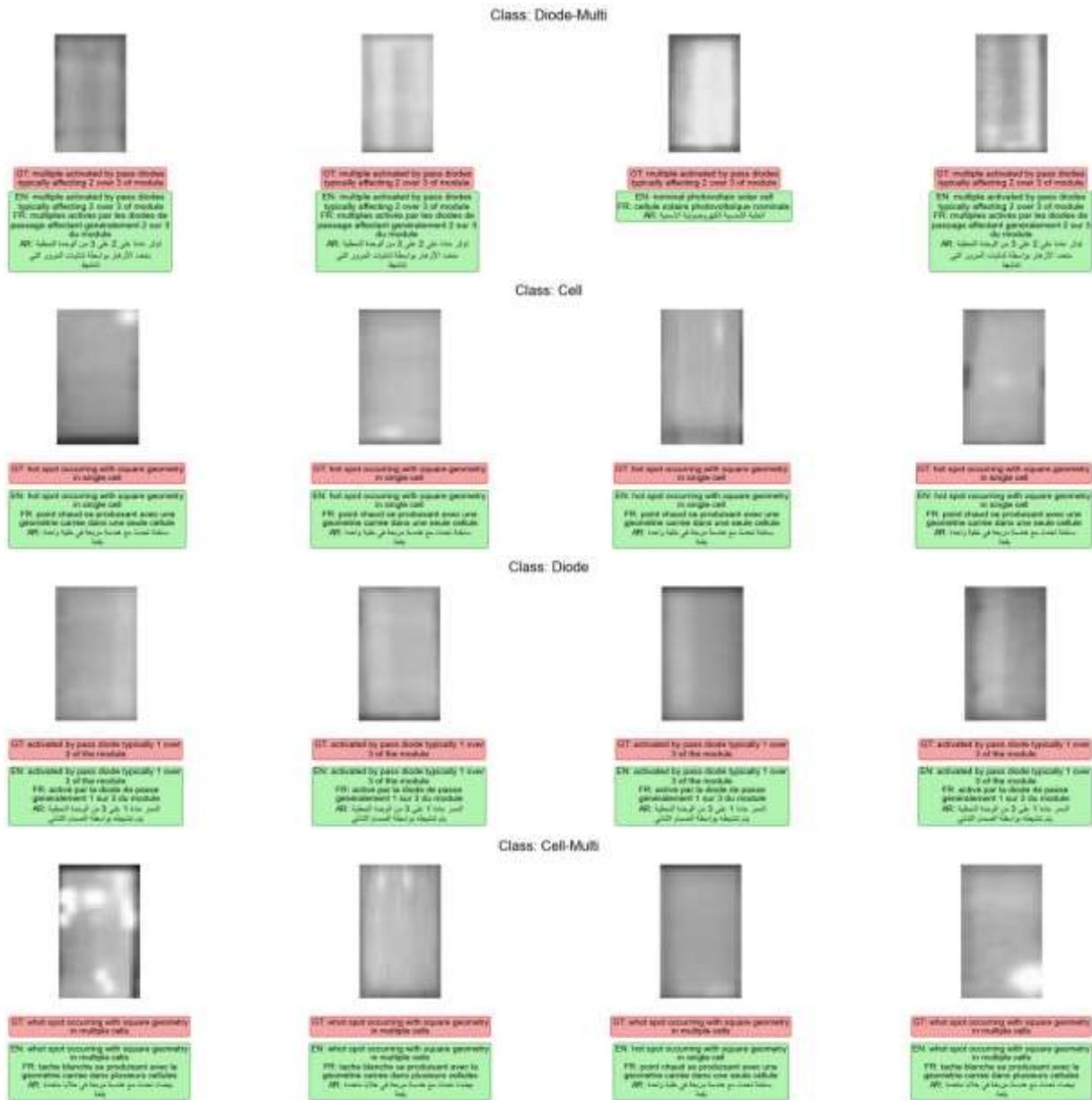


Figure 4.2 : Exemples de descriptions générées automatiquement

#### 4.6 Conclusion

Dans ce chapitre, nous avons présenté et analysé les performances de deux modèles de génération de descriptions pour des images thermiques de panneaux photovoltaïques, en combinant ResNet50 avec des réseaux LSTM et GRU. Les résultats expérimentaux, évalués à l'aide du BLEU Score, ont montré une qualité élevée des descriptions générées, avec une légère supériorité du modèle LSTM sur le GRU en termes de précision linguistique. La génération multilingue renforce également la portée pratique du système. Ces résultats confirment la faisabilité d'un système intelligent de diagnostic visuel, ouvrant la voie à des perspectives d'amélioration via des architectures plus avancées comme les Transformers.

## Conclusion Générale

Ce mémoire de Master s'est inscrit dans une problématique contemporaine à fort impact environnemental, économique et technologique : l'amélioration des systèmes de maintenance et de diagnostic dans les installations photovoltaïques à l'aide de l'intelligence artificielle. À travers une démarche rigoureuse, combinant vision par ordinateur, traitement automatique du langage naturel (TALN) et apprentissage profond, nous avons proposé, implémenté et évalué un système intelligent de génération automatique de descriptions pour des images thermiques de panneaux solaires.

La contribution principale de ce travail réside dans la conception d'un système automatisé et multilingue capable de transformer des données visuelles complexes – en l'occurrence des images infrarouges – en descriptions textuelles compréhensibles par un opérateur humain. Ce système s'appuie sur une architecture de type Encoder-Decoder, où le réseau ResNet50 extrait les caractéristiques visuelles, et où deux variantes de décodeurs (LSTM et GRU) sont comparées pour la génération textuelle.

La mise en œuvre méthodologique a couvert toutes les étapes du cycle de vie d'un projet en deep learning : collecte et annotation des données (2000 images thermiques acquises par drone), division du corpus en ensembles d'apprentissage et de test, prétraitement des données, définition des paramètres d'entraînement (batch size, learning rate, epochs), et enfin évaluation des performances. Nous avons également intégré un module de traduction automatique permettant d'obtenir des descriptions en anglais, français et arabe, ce qui confère au système une dimension pratique et adaptable à des contextes multiculturels ou internationaux.

Les résultats expérimentaux obtenus sont particulièrement encourageants : les deux modèles évalués (ResNet50-LSTM et ResNet50-GRU) ont généré des descriptions avec un score BLEU supérieur à 0.72 à tous les niveaux, ce qui témoigne d'une bonne adéquation entre les phrases produites et les descriptions de référence. Le modèle LSTM a montré une légère supériorité, notamment pour les séquences longues (BLEU-3 et BLEU-4), traduisant une meilleure capacité à structurer syntaxiquement les phrases. En revanche, le modèle GRU, plus léger, se distingue par son efficacité computationnelle, le rendant attractif pour des déploiements sur des systèmes embarqués ou à faible puissance.

Au-delà de ces performances quantitatives, ce travail met en évidence plusieurs avantages scientifiques et pratiques :

- Il démontre la faisabilité de combiner vision thermique et génération de texte pour automatiser des tâches de diagnostic visuel dans le domaine de l'énergie renouvelable.
- Il offre une approche évolutive, modulaire, et transposable à d'autres types d'infrastructures (centrales solaires, réseaux électriques, etc.).
- Il apporte une valeur ajoutée sur le plan industriel, en facilitant la maintenance prédictive, en réduisant les coûts d'inspection manuelle, et en améliorant la réactivité en cas de panne.

Malgré ces avancées, le projet comporte certaines limites :

- L'évaluation repose principalement sur la métrique BLEU, qui mesure la similarité lexicale mais ne tient pas pleinement compte du sens ou de la fluidité sémantique.
- Le corpus d'apprentissage, bien que riche, reste limité à une catégorie spécifique d'images thermiques (modules photovoltaïques en contexte fixe).
- Le système ne gère pas encore les cas ambigus, flous ou fortement bruités, qui pourraient apparaître dans des conditions météorologiques défavorables.

Pour surmonter ces limites et ouvrir des perspectives d'amélioration, plusieurs pistes peuvent être envisagées :

- Intégrer d'autres métriques d'évaluation plus sensibles à la qualité sémantique (METEOR, ROUGE, CIDEr).
- Élaborer un système hybride combinant génération de texte et classification des défauts pour fournir un double diagnostic (visuel + textuel).

- Étendre la base de données à d'autres environnements photovoltaïques (urbains, agricoles, désertiques) pour améliorer la généralisation.
- Explorer des architectures plus avancées comme les Transformers ou les Vision-Language Models (tels que ViLT, BLIP ou Flamingo) qui unifient le traitement du texte et de l'image dans un seul modèle.
- Enfin, envisager le déploiement réel du système sous forme d'application mobile ou de module intégré dans un drone autonome, permettant une interaction directe sur le terrain.

En somme, ce travail constitue une preuve de concept solide pour l'automatisation de la description sémantique d'images dans le secteur énergétique. Il illustre la puissance des modèles de deep learning dans des tâches complexes, à la frontière entre l'interprétation visuelle et la génération de langage. Les résultats obtenus ouvrent la voie à des développements futurs riches et multidisciplinaires, à la croisée de l'intelligence artificielle, de la linguistique computationnelle, de l'ingénierie énergétique et de la robotique appliquée.



غرداية في : 30./06./ 2025

## شهادة ترخيص بالتصحيح والايداء:

انا الاستاذ زقاوي عبد الله

بصفتي المشرف المسؤول عن تصحيح مذكرة تخرج (ليسانس/ماستر/دكتورا) المعنونة ب:

Génération De Légendes D'images Pour L'état Des Cellules Photovoltaïques à l'aide De Techniques D'apprentissage Profond

من انجاز الطالب (الطالبة):

Hadj Hibeterrhmane

التي نوقشت/قيمت بتاريخ 12/06/2025

اشهد ان الطالب/الطالبة قد قام /قاموا بالتعديلات والتصحيحات المطلوبة من طرف لجنة المناقشة وقد تم التحقق من ذلك من طرفنا أعضاء لجنة المناقشة وقد استوفت جميع الشروط المطلوبة .

مصادقة رئيس القسم

رئيس قسم الآلية والكهروميكانيك

عزواي محمد



امضاء المسؤول عن التصحيح