

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche
Scientifique



UNIVERSITÉ CHAHID HAMMA LAKHDAR D'EL-OUED
Faculté des Sciences Exactes
Département d'Informatique

THÈSE

Pour l'obtention du grade de

DOCTEUR 3^{ème} CYCLE LMD EN
INFORMATIQUE

Option : système d'informations interopérables

Présenté par : M^{me} SETTOU Tarablesse

Titre

Optimisation de la recherche d'image
cartographique par le contenu lié à un
phénomène géographique

Soutenu le : 08/07/2021

Devant le jury composé de :

Président : Dr Lejdel Brahim Université d'El-Oued
Rapporteur : Professeur Kholadi Mohamed-Khireddine Université d'El-Oued
Co-Rapporteur : Dr. Ben Ali Abdelkamel Université d'El-Oued
Examineur : Professeur Kazar Okba Université de Biskra
Examineur : Dr Bouzenada Mourad Université Constantine2 de Constantine
Examineur : Professeur Rezeg Khaled Université de Biskra

Année universitaire 2020 – 2021

Dédicaces

Avant tous, je veux rendre grâce à Dieu le tout puissant de m'avoir donné le courage et la patience pour réaliser ce travail malgré toutes les difficultés rencontrées.

Je dédie ce modeste travail à :

A ma très chère mère pour sa tendresse et son amour inconditionnel et sa présence à mon côté dans les moments difficiles.

A mon très chère père pour sa patience et son encouragement et sa présence à mon côté dans les moments difficiles.

A mon très cher mari pour son aide inconditionnel, son encouragement et sa présence à mon côté dans les moments difficiles.

A mes chères sœurs et mes chers frères pour leurs conseils et leur encouragement.

A mes chers grands parents.

A mes grandes familles Settou qui ont été toujours derrière moi

A mes autres familles Kertiou pour leur encouragement

A mon cher enseignant Dr. Benali Abdelkamel pour son aide et ses conseils et son encouragement durant ce travail

A tous ceux que j'aime,

A tous mes amis.

Remerciements

Je remercie Allah le tout puissant, qui nous a donné le courage et la patience pour terminer ce travail.

En premier lieu, je tiens à exprimer ma profonde gratitude à mon directeur de thèse Prof. Kholladi Mohamed-Khireddine qui m'a aidé et conseillé durant ce travail.

Je remercie vivement l'enseignant Dr. Benali Abdelkamel qui m'a coencadré durant cette thèse, pour sa disponibilité, son soutien, son aide précieuse, ses conseils, ses orientations tant précieux et ses remarques constructives.

Mes remerciements vont également aux membres de jury qui m'ont fait l'honneur d'accepter de lire et d'évaluer ma thèse et de fournir des critiques pertinentes.

Je suis aussi reconnaissant à tous ceux qui ont collaboré à notre formation en particulier les enseignants du département d'Informatique de la faculté des Sciences exactes de l'université Chahid Hamma Lakhdar d'El-Oued.

Je remercie également tous ceux qui m'ont soutenu, encouragé et qui ont participé de près ou de loin à l'élaboration de ce travail.

Résumé

L'un des problèmes cruciaux après le tremblement de terre est de savoir comment détecter et identifier rapidement et précisément les régions endommagées ? Plusieurs méthodes automatisées ont été développées pour analyser les images de télédétection (RS) afin de pouvoir procéder à la classification des dommages sismiques.

La performance de la classification des dommages dépend principalement de la puissance de représentations des caractéristiques prises en compte. Bien que les caractéristiques de bas niveau (provenant d'experts) puissent atteindre des performances satisfaisantes dans une certaine mesure, le gain de performances est faible et ne se généralise pas toujours bien.

Récemment, le réseau de neurones convolutif (CNN) a démontré sa capacité à dériver des représentations de caractéristiques plus puissantes que les caractéristiques de bas niveau dans de nombreux domaines.

Notre contribution principale réalisée dans cette thèse est l'évaluation des représentations hybrides de caractéristiques dérivées à partir de plusieurs modèles pré-entraînés de CNN pour la classification des dommages sismiques. De plus, dans cette étude, contrairement aux études précédentes, nous explorons la possibilité d'exploiter la combinaison des vecteurs de caractéristiques extraites à la fois de deux premières couches entièrement connectées (FC1 et FC2) de chaque modèle CNN utilisé.

Nous avons validé nos propositions sur deux grands ensembles de données, y compris des images fortement variées en termes de caractéristiques de la scène, de conditions d'éclairage et de caractéristiques des images, capturées à partir de différents événements sismiques dans plusieurs régions géographiques. Des expériences intensives ont montré que notre proposition peut améliorer considérablement les performances.

Mots Clés : Télédétection, Classification des dommages sismiques, Représentation des caractéristiques hybrides, CNN pré-entraîné et Apprentissage en profondeur.

Abstract

One of the crucial problems after earthquakes is how to quickly and accurately detect and identify damaged areas? Several automated methods have been developed to analyze remote sensing (RS) images in order to be able to classify seismic damage.

The performance of damage classification is mainly depending on powerful learning feature representations. Though the hand-crafted features can achieve satisfactory performance to some extent, the performance gain is small and does not always generalize well.

Recently, the convolutional neural network (CNN) has demonstrated its capability of deriving more powerful feature representations than hand-crafted features in many domains.

Our main contribution made in this thesis is the investigation of hybrid feature representations derived from several pre-trained CNN models for earthquake damage classification. Also, in this study, in contrast to previous works, we explore the combination of feature representations extracted from the last two fully-connected layers of a particular CNN model.

We validated our proposals on two large datasets, including images highly varying in scene characteristics, lighting conditions, and image characteristics, captured from different earthquake events and several geographic locations. Extensive experiments showed that our proposals can improve significantly the performance.

Keywords: Remote sensing, Earthquake damage classification, Hybrid feature representation, Pre-trained CNN and Deep learning.

المخلص

تتمثل إحدى المشكلات الحاسمة بعد الزلازل في كيفية اكتشاف وتحديد المناطق المتضررة بسرعة وبدقة؟ تم تطوير العديد من الطرق الآلية لتحليل صور الاستشعار عن بعد (RS) لغرض التمكن من تصنيف الأضرار الزلزالية.

يعتمد أداء تصنيف الضرر بشكل أساسي على قوة تمثيلات ميزات التعلم التي يتم أخذها في الاعتبار. على الرغم من أن الميزات المصنوعة يدويًا يمكن أن تحقق أداءً مرضياً إلى حد ما ، إلا أن مكاسب الأداء صغيرة ولا يتم تعميمها بشكل جيد دائماً.

في الآونة الأخيرة ، أثبتت الشبكة العصبية التلافيفية (CNN) قدرتها على اشتقاق تمثيلات ميزات أكثر قوة من الميزات المصنوعة يدويًا في العديد من المجالات.

مساهمتنا الرئيسية المنجزة في هذه الأطروحة هي تقييم التمثيلات الهجينة للسّمات المستمدة من العديد من نماذج CNN المدربة مسبقاً لتصنيف الأضرار الزلزالية. علاوة على ذلك في هذه الدراسة ، على عكس الدراسات السابقة ، نستكشف إمكانية استغلال مجموعة تمثيلات الميزات المستخرجة من كلا الطبقتين المتصلتين بالكامل (FC1 et FC2) لكل نموذج CNN مستخدم.

لقد تحققنا من كفاءة وأداء مقترحاتنا في مجموعتين كبيرتين من البيانات، تحتوي على صور شديدة التباين في خصائص المشهد وظروف الإضاءة وخصائص الصورة التي تم التقاطها من أحداث زلزالية مختلفة في مناطق جغرافية متعددة. أظهرت التجارب المكثفة أن اقتراحنا يمكن أن يحسن الأداء بشكل ملحوظ.

الكلمات المفتاحية : الاستشعار عن بعد ، تصنيف الضرر الزلزالي ، تمثيل الخصائص الهجينة ، شبكة CNN المدربة مسبقاً و التعلم العميق.

Table des matières

Dédicaces	i
Remerciements	ii
Résumé	iii
Abstract	iv
المخلص	v
Table des matières	vi
Liste des figures	xii
Liste des tableaux	vii
Liste d'abréviations	viii
Introduction Générale.....	1
Chapitre I. Gestion des catastrophes naturelles par des données de télédétection.	4
1. Introduction	5
2. Qu'est-ce qu'une catastrophe ?	5
2.1 Les catastrophes naturelles	5
2.2 Les catastrophes technologiques	6
3. Cycle de gestion des catastrophes	6
3.1 Atténuation.....	7
3.2 Préparation	7
3.3 Intervention (ou réaction)	8
3.4 Rétablissement	8
4. Gestion des catastrophes naturelles	8
4.1 La recherche sur le terrain.....	8
4.2 Les technologies de télédétection	9
5. Télédétection	9
5.1 Domaines d'application de télédétection	9
6. Images de télédétection	10
6.1 Résolution	11
6.1.1 Résolution spatiale.....	11
6.1.2 Résolution spectrale.....	11
6.1.3 Résolution temporelle.....	11
6.1.4 Résolution radiométrique	12
6.2 Plateforme d'acquisition	12
6.2.1 Plateformes aériennes	12

6.2.2	Plateformes satellitaires	12
6.3	Types de données de télédétection.....	13
6.3.1	Données optiques.....	13
6.3.2	Données radar	13
6.3.3	Données LiDAR	13
7.	Techniques de gestion des catastrophes naturelles	14
7.1	Interprétation visuelle	14
7.2	Techniques Multi-temporelles	14
7.3	Techniques Mono-temporelles.....	15
8.	Conclusion.....	15
Chapitre II .Extraction de caractéristiques et classification		16
1.	Introduction	17
2.	Caractéristiques de bas niveaux	17
2.1	Approches globales	18
2.1.1	Descripteurs de couleur	18
2.1.1.1	Histogramme de couleurs	19
2.1.1.2	Moments de couleur.....	19
2.1.1.3	Cohérence spatiale	19
2.1.1.4	Couleurs dominantes.....	20
2.1.1.5	Corrélogrammes.....	20
2.1.2	Descripteurs des textures	20
2.1.2.1	Matrice de cooccurrence de niveaux de gris "GLCM"	21
2.1.2.2	Filtres de Gabor.....	21
2.1.2.3	Transformée en ondelettes	22
2.1.2.4	Motif de texture binaire local "LBP"	22
2.1.2.5	Histogramme de gradient orienté "HOG"	23
2.1.3	Descripteurs de forme.....	23
2.2	Approches locales	23
2.2.1	Détecteur de coins de Harris.....	24
2.2.2	Transformation de caractéristiques visuelles invariante à l'échelle "SIFT"...	24
2.2.3	SIFT dense "DSIFT"	25
2.2.4	Accélérées de caractéristiques visuelles robustes "SURF"	25
3.	Caractéristiques de niveau intermédiaire	25
3.1	Représentation par sac de mots visuels "BOW"	27
3.2	Vecteur de descripteurs agrégés localement "VLAD"	27

3.3	Vecteur de Fisher "FV"	27
3.4	Codage linéaire à contraintes de localité "LLC"	28
4.	Caractéristiques de plus hauts niveaux.....	28
5.	Classification.....	29
5.1	Machines à vecteurs de support "SVM"	29
5.2	Méthode des k-plus proches voisins "KNN"	31
5.3	Arbre de décision "DT".....	31
5.4	Réseaux de neurones artificiels "ANN"	32
5.5	Réseaux bayésiens "BN".....	32
5.6	Sélection et adaptation du vecteur de support "SVSA"	33
6.	Conclusion.....	34
Chapitre III .Apprentissage en profondeur "Deep learning"		35
1.	Introduction	36
2.	Apprentissage en profondeur	36
2.1	La relation entre l'apprentissage automatique "ML" et l'apprentissage en profondeur "DL"	36
2.2	Réseau de neurones "Neural Networks "	37
2.2.1	Sigmoid.....	39
2.2.2	Tanh.....	40
2.2.3	ReLU	40
2.3	L'architecture de DL.....	41
3.	Les domaines d'application de DL.....	42
3.1	Colorisation automatique des images en noir et blanc.....	42
3.2	Ajout automatique de sons aux films muets	42
3.3	Traduction automatique	43
3.4	Génération d'écriture manuscrite automatique.....	43
3.5	Génération automatique de texte	43
3.6	Reconnaissance d'images	43
3.7	Les applications de télédétection	44
3.8	Génération automatique de légendes d'images.....	44
3.9	Robotique et véhicules autonomes.....	44
3.10	Prédire les séismes	44
4.	Les approches de DL.....	45
4.1	L'auto-encodeur superposé.....	45
4.2	Les machines Boltzmann restreintes.....	46

4.2.1	Les réseaux de croyances profondes	47
4.2.2	Les machines Boltzmann profondes.....	48
4.3	Les réseaux de neurones convolutifs	49
5.	L'architecture de réseau de neurone convolutif "CNN"	50
5.1	Couche d'entrée	51
5.2	Couches convolutionnelles	51
5.3	Couches de correction "ReLU".....	53
5.4	Couche de regroupement	53
5.5	Couche entièrement connectée et couche de sortie de classification.....	53
5.6	Couches auxiliaires	54
6.	Les scénarios de classification basée sur CNN	55
6.1	Apprentissage de CNN à partir de zéro	56
6.2	Réglage d'un modèle pré-entraîné	56
6.3	Les modèle pré-entraîné sans réglage	57
7.	Les modèles de CNN pré-entraîné	57
7.1	LeNet-5	57
7.2	AlexNet.....	58
7.3	ZF Net	59
7.4	Le réseau VGG-16	60
7.5	Le réseau VGG-19	61
7.6	GoogLeNet.....	62
7.7	ResNet.....	63
8.	Conclusion.....	63
Chapitre IV.Gestion des catastrophes naturelles par les algorithmes de ML et de DL : état de l'art		64
1.	Introduction	65
2.	Gestion des séismes.....	65
2.1	Catastrophes des tremblements de terre.....	65
2.2	Gestion des séismes par télédétection.....	67
2.3	Gestion des séismes par les algorithmes de ML et de DL	70
2.3.1	Gestion des séismes par les algorithmes de ML.....	70
2.3.2	Gestion des séismes par les algorithmes de DL	72
3.	Gestion des inondations	75
3.1	Catastrophes des inondations	75
3.2	Gestion des inondations par télédétection.....	77

3.3	Gestion des inondations par les algorithmes de ML et de DL.....	80
3.3.1	Gestion des inondations par les algorithmes de ML.....	80
3.3.2	Gestion des inondations par les algorithmes de DL	81
4.	Gestion des incendies	83
4.1	Catastrophes des incendies	83
4.2	Gestion des incendies par télédétection	84
4.3	Gestion des séismes par les algorithmes de ML et de DL	85
4.3.1	Gestion des séismes par les algorithmes de ML.....	85
4.3.2	Gestion des séismes par les algorithmes de DL	87
5.	Gestion des glissements de terrain	88
5.1	Catastrophes des glissements de terrain.....	88
5.2	Gestion des glissements de terrain par des données de télédétection	90
5.3	Gestion des glissements de terrain par les algorithmes de ML et de DL.....	91
5.3.1	Gestion des glissements de terrain par les algorithmes de ML	91
5.3.2	Gestion des glissements de terrain par les algorithmes de DL.....	93
6.	Conclusion.....	94

Chapitre V.Optimisation de la classification de dommages via des représentations hybrides de caractéristiques de DL dérivées d’images aériennes post-séisme..... 95

1.	Introduction	96
2.	Approche proposée.....	96
2.1	La phase d’extraction des caractéristiques.....	97
2.1.1	Les caractéristiques extraites à partir d’un modèle CNN particulier.....	98
2.1.1.1	Les caractéristiques extraites à partir d’AlexNet	98
2.1.1.2	Les caractéristiques extraites à partir de VGG-Net	100
2.1.1.3	Les caractéristiques extraites à partir de GoogLeNet	104
2.1.2	La représentation hybride proposée.....	111
2.2	La phase de classification des dommages.....	111
2.2.1	Apprentissage	112
2.2.2	Test	112
3.	Conception expérimentale.....	112
3.1	Description de la base d’images utilisée	113
3.2	Environnement de travail et réglage des paramètres	114
3.3	Méthodes étudiées.....	115
3.4	Mesure de la performance et validation statistique.....	116
4.	Résultats et discussion.....	116

4.1	Expérience 1: Combinaison des caractéristiques de FC1 et FC2	116
4.2	Expérience 2: Combinaison de caractéristiques de différents modèles CNN.....	118
4.3	Comparaison de l'approche proposée avec la littérature	121
5.	Conclusion.....	125
	Conclusion générale	127
	Références	129

Liste des figures

Chapitre I : Gestion des catastrophes naturelles par des données de télédétection

Figure I.1 : Cycle de gestion des catastrophes. 7

Chapitre II : Extraction de caractéristiques et classification

Figure II.1 : Méthodes basées sur les caractéristiques de niveau intermédiaire. 26

Figure II.2 : Illustration de l'hyperplan séparant les points de données "descripteurs" de deux classes. Les vecteurs de support et la marge sont marqués..... 30

Chapitre III : Apprentissage en profondeur "Deep learning"

Figure III.1 : Relation entre l'intelligence artificielle, l'apprentissage automatique et l'apprentissage en profondeur. 37

Figure III. 2 : L'extraction des caractéristiques dans la ML et DL..... 38

Figure III. 3 : Exemple de neurone artificiel..... 39

Figure III.4 : Gauche : un réseau de neurones à 2 couches (une couche cachée de 4 neurones (ou unités) et une couche de sortie avec 2 neurones) et trois entrées. Droite : Un réseau de neurones à 3 couches avec trois entrées, deux couches cachées de 4 neurones chacune et une couche de sortie. Notez que dans les deux cas, il existe des connexions (synapses) entre les neurones d'une couche à l'autre, mais pas à l'intérieur d'une couche. 39

Figure III.5 : Les fonctions d'activation..... 40

Figure III.6 : Comparaison entre les architectures d'ANN et DNN..... 42

Figure III.7: Diagramme schématique des AE..... 45

Figure III.8 : L'architecture de SAE. 46

Figure III.9 : L'architecture de RBM. 48

Figure III.10 : L'architecture de DBN "a" et de DBM "b". 48

Figure III.11 : L'architecture de CNN. 49

Figure III.12 : (a) Exemple de réseau entièrement connecté sans interaction clairsemée et (b) exemple de l'interaction clairsemée dans CNN. 51

Figure III.13 : Exemple illustrant l'opération des couches convolutionnelles..... 52

Figure III.14 : Exemple illustre l'opération de zero-padding dans la couche convolutionnelle. 53

Figure III.15 : Exemple illustrant la méthode de regroupement. 54

Figure III.16 : Couche entièrement connectée et couche de sortie de classification. 55

Figure III.17 : Exemple illustrant l'application de décrochage : l'exemple de gauche du réseau entièrement connecté avec deux couches cachées, et l'exemple de droite du réseau après l'application de décrochage. 55

Figure III.18 : L'architecture globale de LeNet-5 [50]. 58

Figure III.19: L'architecture globale d'AlexNet. Elle est composée de cinq couches de convolution (C1-C5) et de trois couches de FC (FC1-FC3) [49]. 59

Figure III.20 : L'architecture globale de VGG-16. Elle est composée de cinq blocs de couches de convolution (C#1-C#5) et de trois couches de FC (FC1-FC3). 61

Figure III.21 : L'architecture globale de VGG-19. 61

Figure III.22 : L'architecture globale de GoogLeNet. 63

Chapitre V : Optimisation de la classification de dommages via des représentations hybrides de caractéristiques de DL dérivées d'images aériennes post-séisme.

Figure V.1 : Schéma général de notre architecture proposée pour la classification des dommages. 97

Figure V.2 : Les caractéristiques extraites à partir d'AlexNet. 99

Figure V.3 : Les caractéristiques extraites à partir de VGG-16 et VGG-19. 103

Figure V.4 : Les caractéristiques extraites à partir de GoogLeNet. 110

Figure V.5 : La représentation hybride proposée. 111

Figure V.6 : Phase de classification. 112

Figure V.7 : Exemples des échantillons de chaque type de base d'images, représentant des échantillons non endommagés (côté gauche) et endommagés (côté droit) des ensembles d'images capturées par l'avion (échantillons du haut) et d'UAV (échantillons du bas). 114

Figure V.8 : Comparaisons multiples de CNN utilisés en fonction du taux de succès. Chaque rectangle contient quatre lignes (partant des cercles de symboles) représentant les rangs moyens et leurs intervalles pour les quatre modèles CNN utilisés. Les trois rectangles du haut concernent la base de données aérienne et les trois rectangles du bas concernent la base de données d'UAV; les colonnes de rectangles de gauche à droite sont respectivement pour représenter les caractéristiques de FC1, FC2, FC1 & FC2. 118

Figure V.9 : Distribution des taux de succès pour la méthode Hybrid-1 et ses composants CNN. Chaque rectangle contient cinq diagrammes en boîte représentant la distribution des taux de succès de la classification des dommages en utilisant les vecteurs de caractéristiques de FC1, FC2 ou FC1 & FC2 pour Hybrid-1 et chacun de ses composants AlexNet, VGG-16, VGG-19 et GoogLeNet. Les trois rectangles du haut concernent la base de données aérienne et les trois rectangles du bas concernent la base de données d'UAV. 119

Figure V.10 : Distribution des taux de succès pour la méthode Hybrid-2 et ses composants CNN. Chaque rectangle contient quatre diagrammes en boîte représentant la distribution des taux de succès de la classification des dommages en utilisant les vecteurs de caractéristiques de FC1, FC2 ou FC1 & FC2 pour Hybrid-2 et chacun de ses composants AlexNet, VGG-16 et VGG-19. Les trois rectangles du haut concernent la base de données aérienne et les trois rectangles du bas concernent la base de données d'UAV. 120

Figure V.11 : Distribution des taux de succès pour la méthode Hybrid-3 et ses composants CNN. Chaque rectangle contient trois diagrammes en boîte représentant la distribution des taux de succès de la classification des dommages en utilisant les vecteurs de caractéristiques de FC1, FC2 ou FC1 & FC2 pour Hybrid-3 et chacun de ses composants VGG-16 et VGG-19. Les trois rectangles du haut concernent la base de données aérienne et les trois rectangles du bas concernent la base de données d'UAV. 120

- Figure V.12 : Comparaisons multiples des vecteurs des trois méthodes hybrides étudiés. Chaque rectangle contient trois lignes (partant des cercles de symboles) représentant les rangs moyens et leurs intervalles pour les trois méthodes hybrides. Les trois rectangles du haut concernent la base de données aérienne et les trois rectangles du bas concernent la base de données d'UAV. 121
- Figure V.13 : Comparaison de notre proposition avec les méthodes de la littérature. 124

Liste des tableaux

Chapitre V : Optimisation de la classification de dommages via des représentations hybrides de caractéristiques de DL dérivées d'images aériennes post-séisme

Tableau V.1 : L'architecture détaillée d'AlexNet à partir de Matlab.	98
Tableau V.2 : L'architecture détaillée de VGG-16 à partir de Matlab.	100
Tableau V.3 : L'architecture détaillée de VGG-19 à partir de Matlab.	101
Tableau V.4 : L'architecture détaillée de GoogLeNet à partir de Matlab.....	104
Tableau V.5 : Description des échantillons d'apprentissage et de test dérivés d'images de deux plateformes pour différentes régions géographiques.....	113
Tableau V.6 : Les caractéristiques étudiées de chaque hybridation.....	116
Tableau V.7 : Performances de la classification des dommages sismiques (TS) en utilisant les vecteurs de caractéristiques de FC1, FC2 et FC1&FC2 de quatre modèles CNN pré-entraînés.	117
Tableau V.8 : Comparaison (en termes de TS) entre les caractéristiques de CNN avec et sans transformation ReLU.....	118
Tableau V.9 : Performances de classification de dommage (TS) fournies basé sur les représentations hybrides.	119
Tableau V.10 : Caractéristiques de chaque base d'images utilisée dans les travaux de la littérature.....	123
Tableau V. 11 : Comparaison de notre approche avec les méthodes de la littérature (en termes de précision de résultats de classification).	124

Liste d'abréviations

WHO: The **W**orld **H**ealth **O**rganization
SIG: Geographic Information System
SAR: Synthetic Aperture Radar, radars à synthèse d'ouverture
LiDAR: Light Detection And Ranging
MLA: Machine Learning Algorithm
RGB: Red, Green, Blue
HSV: Hue, Saturation, Value
GLCM: Gray-Level Co-occurrence Matrix
LBP: Local Binary Pattern
HOG: Histogram of Oriented Gradient
SIFT: Scale-Invariant Feature Transform
DSIFT: Dense SIFT
SURF: Speeded Up Robust Features
BOW: Bag of Visual Words
VLAD: Vector of Locally Aggregated Descriptor
FV: The Fisher Vector
LLC: Locality-constrained Linear Coding
DBN: Deep Belief Nets
DBM: Deep Boltzmann Machines
SAE: Stacked Auto Encoder
CNN: Convolutional Neural Networks
SVM: Support Vector Machines
KNN: K-Nearest-Neighbors
DT: Decision Trees
ANN: Artificial Neural Network
BN: Bayesian Network
SVSA: Support Vector Selection and Adaptation
AI: Artificial Intelligence
ML: Machine Learning
NN: Neural Networks
DL: Deep Learning
ReLU: Rectified Linear Unit
DNN: Deep Neural Networks
GPU: Graphics Processing Unit
UAV: Unmanned Aerial Vehicle

Introduction Générale

Au cours des dernières décennies, il y a eu une augmentation considérable des catastrophes naturelles et de leurs effets sur la population et l'économie. Les événements indésirables, tels que cyclones, tornades, inondations, éruptions volcaniques, tremblements de terre, tsunamis, etc., se transforment en catastrophe lorsqu'ils frappent une zone de population vulnérable. Les catastrophes sismiques sont les catastrophes naturelles les plus fréquentes et les plus dangereuses qui causent d'énormes dégâts dans le monde entier. Selon l'United States Geological Survey, environ 50 tremblements de terre se produisent chaque jour dans le monde. Les tremblements de terre provoquent souvent des dommages environnementaux et structurels importants et des pertes économiques estimées à des milliards de dollars. En effet, avec l'augmentation rapide de l'urbanisation, les tremblements de terre ont causé plus de 60% de tous les décès dus à des catastrophes naturelles. Selon la National Aeronautics and Space Administration (NASA), plus de 10 000 personnes dans le monde sont mortes de tremblements de terre au cours des cinq dernières années.

Après le tremblement de terre, la détection et l'identification rapide et précise des dommages aident à réduire les dommages causés par la catastrophe sismique et fournissent également des informations importantes telles que l'emplacement, le taux et le type de dommage pour les efforts d'intervention, d'urgence et de reconstruction. Traditionnellement, l'enquête sur le terrain dans les zones touchées avait été menée par de nombreux chercheurs et les spécialistes pour détecter les régions endommagées et étudier la distribution des dégâts. Cependant, l'identification de la région endommagée par l'enquête sur le terrain est coûteuse, laborieuse et prend beaucoup de temps et est fortement influencée par l'expérience, les compétences et le point de vue des inspecteurs. Ce qui a rendu difficile l'obtention d'informations précises et opportunes sur les tremblements de terre, en particulier dans les régions inaccessibles. La technologie de télédétection offre la possibilité d'obtenir des données post-séisme plus rapidement et moins cher. Elle est donc une source appropriée et plus efficace pour l'évaluation des dégâts sismiques et fournir des informations rapides et fiables aux décideurs, aux intervenants en cas d'urgence.

Au cours des dernières décennies, plusieurs types de données de télédétections tels que l'optique, le LiDAR et le radar (SAR) ont été largement exploités pour la détection rapide de dommages dans de vastes zones après les tremblements de terre. En effet, les capteurs optiques fournissent des images à très haute résolution de plus d'un mètre via de nombreuses plateformes telles que les satellites, les avions et les drones (UAV). Les avions et les UAVs sont reconnus comme la source la plus appropriée pour faciliter le processus de détection des dommages, car ces images représentent la surface du sol telle qu'elle apparaît à l'œil humain, ce qui permet de faire des mesures précises.

La détection des dommages sismiques à partir des données de télédétection peut être manuelle ou à l'aide de procédures d'analyse d'image automatisées. Bien que les méthodes manuelles, ou les méthodes d'interprétation visuelle, soient actuellement la méthode la plus précise car elles peuvent donner de bons résultats pour détecter les dommages et aussi distinguer différents types et sévérités de dommages des images de télédétection ; mais elles nécessitent beaucoup de temps de traitement, ont besoin d'experts qualifiés et conduisent à un

travail difficile et laborieux. Pour surmonter ces inconvénients, plusieurs méthodes automatisées ont été développées dans le but de détecter les régions endommagées par des indices de dommages visibles de l'extérieur tels que des tas de gravats, des débris des éléments cassés et une fragmentation dans une région de l'image. Ces méthodes sont classées en deux groupes :

- (1) Les techniques multi-temps qui comparent et analysent les changements significatifs entre les images avant et après le séisme ;
- (2) Et les techniques mono-temporelles qui utilisent seulement les ensembles de données de télédétection collectées après le tremblement de terre.

Les techniques mono-temporelles sont plus rapides et plus adaptées que les techniques multi-temporelles pour certains raisons telles que les données proches du pré-événement sont difficiles à obtenir dans certaines situations, la paire homogène de données pré- et post-événement est difficile à obtenir, l'enregistrement de haute précision est délicat et prend du temps. Les méthodes mono-temporelles sont principalement basées sur des approches d'apprentissage supervisé conçues pour être efficaces dans la classification des dommages. Ces approches se composent de deux modules principaux qui sont le module d'extraction de caractéristiques et le module de classification.

Principalement, la performance de ces approches dépend de la puissance des méthodes d'extraction de caractéristiques. Bien que les méthodes basées sur des caractéristiques de bas niveau se sont prouvées performantes, mais l'amélioration de performance de ces caractéristiques est faible et ne se généralise pas bien. Parce que les images de télédétection sont très complexes et très diverses dans leurs caractéristiques, la stratégie de représentation de caractéristiques joue un rôle important pour améliorer la performance de ces caractéristiques. Bien que les stratégies de représentation, comme la représentation par sac de mots visuels (BOW), aient considérablement amélioré la performance de la classification, des problèmes liés à la généralisation persistent. Parce qu'ils ne se généralisent souvent pas bien lorsque l'emplacement géographique n'est pas visible et aussi lorsque les images diffèrent fortement par leurs caractéristiques telles que l'échelle, la vue et la résolution de la caméra, la plateforme de capture et la complexité de la scène.

Récemment, CNN a démontré qu'il était capable de dériver automatiquement des représentations de caractéristiques puissantes à partir des images de télédétection elles-mêmes pour la classification des dommages sismiques, contrairement aux représentations globales conventionnelles, y compris celles qui utilisent les méthodes de représentation comme BOW, qui ont besoin d'une quantité considérable de compétences en ingénierie et une expertise de domaine.

Notre contribution principale dans ce travail est l'évaluation d'une approche hybride d'extraction de caractéristiques pour la classification des dommages sismiques. Cette approche hybride est basée sur des représentations profondes des caractéristiques extraites par plusieurs modèles pré-entraînés de CNN. À notre connaissance, aucune approche similaire n'a été proposée dans la littérature. De plus, contrairement aux études précédentes, nous explorerons dans ce travail la possibilité d'exploiter la combinaison des vecteurs de caractéristiques extraites à la fois de deux premières couches entièrement connectées (FC1 et

FC2) de chaque modèle CNN utilisé. Le but de notre proposition est d'améliorer la performance de classification et d'examiner la capacité de généralisation. Pour la validation, nous avons utilisé de grands ensembles de données plus récemment utilisés dans [49], avec des images aériennes capturées à partir de deux plateformes différentes (avions et UAV), à partir de différents événements sismiques dans plusieurs régions géographiques. Également, ces images sont fortement variées en termes de caractéristique de la scène, de conditions d'éclairage et de caractéristiques des images telles que l'échelle de l'image, la vue et la résolution de la caméra, et la complexité de la scène.

Organisation du mémoire

Ce travail s'articule autour de cinq chapitres :

- Le premier chapitre présente les différentes notions de base de gestion des catastrophes à partir des données de télédétection. D'abord, nous présenterons les types et le cycle de gestion des catastrophes géographiques. Puis, nous expliquerons les concepts de base pour les technologies et les applications de télédétection. Enfin, nous portons notre attention sur les différentes techniques de gestion des catastrophes par des données de télédétection, parmi elles, ceux qui appliquent principalement des algorithmes d'apprentissage automatique.
- Le deuxième chapitre présente un état de l'art sur les algorithmes d'apprentissage automatique (ML) qui fonctionnent en deux phases complémentaires, extraction des caractéristiques et classification. Nous présenterons les différentes méthodes et les techniques connues dans la littérature pour chaque phase. Plus précisément, nous expliquerons les méthodes d'extraction des caractéristiques basées sur les caractéristiques de bas niveaux et sur les caractéristiques de niveaux intermédiaire.
- Le troisième chapitre présente un état de l'art sur l'apprentissage en profondeur (DL), où nous expliquerons le passage de l'apprentissage automatique à l'apprentissage en profondeur, puis nous présenterons les applications et les différentes approches de l'apprentissage en profondeur. En plus, nous détaillerons l'architecture de réseau CNN et ses modèles.
- Le quatrième chapitre présente les travaux antérieurs qui sont basés sur les méthodes de ML et les méthodes de DL en particulier, pour la gestion des catastrophes naturelles à partir des données de télédétection.
- Le cinquième chapitre est consacré à la présentation de notre contribution pour la classification des dommages sismiques basés sur les données de télédétection aériennes, ainsi que les résultats de validation obtenus. Nous commencerons par donner l'architecture globale du système proposé. Ensuite, nous détaillerons en étudiant séparément chacun de ses deux phases principales. Et enfin, nous présenterons notre protocole expérimental, puis analyserons et interpréterons les résultats correspondants.

Enfin, nous terminerons cette thèse par une conclusion générale dans laquelle nous récapitulerons les résultats obtenus par notre contribution, puis nous proposerons des perspectives d'avenir.

Chapitre I

Gestion des catastrophes naturelles par des données de télédétection

À travers ce chapitre, nous présenterons les concepts de base des catastrophes, des technologies de télédétection et décrirons les différentes techniques de gestion des catastrophes en utilisant des données de télédétection.

1. Introduction

Les catastrophes ont un impact dévastateur sur la vie humaine et causent un degré élevé de dégâts. Chaque année, les catastrophes fortes comme les tremblements de terre et les inondations causent des milliers de dégâts et de victimes. La superficie, le taux et le type de dommage sont des informations essentielles pour les opérations de sauvetage, de secours humanitaires et de reconstruction dans la zone sinistrée.

La rapidité et la précision de l'évaluation des dégâts sont essentielles pour le sauvetage et l'estimation des pertes économiques. Des études antérieures ont montré que l'utilisation des données de télédétection pour l'évaluation des dégâts causés par les catastrophes est très importante et donne des informations rapides et essentielles pour le sauvetage et les secours.

Dans ce chapitre, nous présenterons les éléments de base nécessaires à la gestion des catastrophes à partir des données de télédétection. Nous commençons par présenter les types et le cycle de gestion des catastrophes géographiques. Puis, nous expliquerons les concepts de base pour les technologies et les applications de télédétection. Et enfin, nous concentrerons notre attention sur les différentes techniques de gestion des catastrophes par des données de télédétection.

2. Qu'est-ce qu'une catastrophe ?

De nombreuses définitions ont été données aux catastrophes. L'Organisation mondiale de la Santé "The World Health Organization, WHO" définit une catastrophe comme "un phénomène écologique soudain d'une ampleur suffisante nécessitant une assistance extérieure" [1]. Il a été également défini comme tout événement, se produisant généralement soudainement, conduit à la perturbation de la vie quotidienne, ce qui donc perturbe sérieusement le travail de la société et cela peut causer de nombreuses blessures, des morts et d'importantes pertes matérielles [2].

Les catastrophes sont généralement divisées en deux types, se distinguant par leurs causes [2] :

- Les catastrophes naturelles;
- Et les catastrophes technologiques causées par l'homme.

2.1 Les catastrophes naturelles

La catastrophe naturelle est une catastrophe qui résulte d'un événement naturel se produisant à la suite d'une action des forces naturelles et ont tendance à être acceptée comme étant regrettable, mais inévitable [2]. Les catastrophes naturelles concernent des événements climatiques et géologiques [1].

Dans les dernières années, les catastrophes naturelles ont fait 100 000 victimes coûtant plus de 140 milliards de dollars américains [1]. Les catastrophes naturelles sont peut-être les catastrophes les plus "inattendues" et coûteuses globalement en termes de perte de vies humaines et de ressources. Les phénomènes comme : les séismes, les éruptions volcaniques,

les tsunamis, les glissements de terrain, les inondations, les tempêtes, les cyclones, les orages, etc., sont tous des risques naturels qui tuent des milliers de personnes et causent des milliards de dollars de dégâts en terme d'habitat et de propriétés chaque année [3].

2.2 Les catastrophes technologiques

Les catastrophes technologiques ou d'origine humaine sont les catastrophes résultant de certaines activités humaines, telles que les meurtres, les explosions chimiques, les déversements de pétrole, les accidents nucléaires, les accidents de transport, les accidents maritimes, ferroviaires et aériens, les marées noires, les déforestations, les grands incendies de forêts, la guerre et les attaques délibérées, etc., sont en hausse, plus particulièrement dans les pays en voie de développement où les critères en matière de sécurité industrielle et leurs pratiques sont insuffisants [3].

Les catastrophes liées aux activités humaines sont imprévisibles, peuvent s'étendre sur plusieurs zones géographiques, peuvent être inévitables et peuvent avoir des dommages physiques limités mais aussi un effet à long terme. Comme pour les catastrophes naturelles, ils sont également beaucoup plus difficiles à gérer pour la communauté et à accepter pour les victimes [1].

3. Cycle de gestion des catastrophes

Le cycle de gestion des catastrophes illustre le processus en cours par lequel les gouvernements, les entreprises et la société civile, réagissent pendant et immédiatement après une catastrophe et prennent des décisions pour planifier et réduire l'impact des catastrophes [4].

La gestion des catastrophes peut être considérée comme un processus cyclique. Le cycle de gestion des catastrophes est constitué par une série de phases reliées étroitement les unes aux autres : atténuation, préparation, intervention et rétablissement comme sur la figure I.1 [5]. Les phases se déroulent en principe de manière séquentielle mais en réalité deux ou plusieurs phases peuvent se produire simultanément. Une phase du cycle ne doit pas nécessairement s'achever pour laisser la place à une autre [2].

La prise des actions appropriées durant chaque phase entraîne une meilleure préparation, les meilleures alertes, une vulnérabilité réduite et/ou la prévention des catastrophes à venir. Le cycle complet de gestion des catastrophes inclut l'élaboration de politiques et de plans publics, qui modifient les causes des catastrophes ou en atténuent les effets sur les personnes, les biens et les infrastructures [2] et [4].

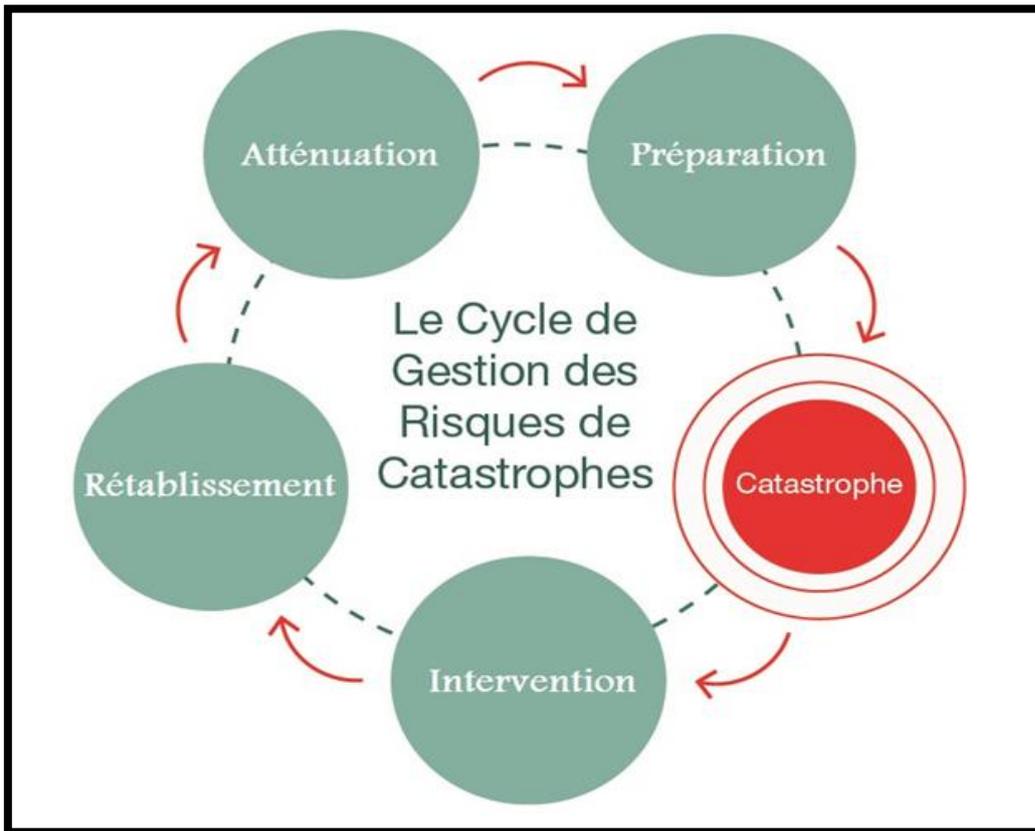


Figure I.1 : Cycle de gestion des catastrophes.

3.1 Atténuation

Le mot "Atténuation" (ou en anglais *Mitigation*) désigne l'ensemble des actions, des lignes directrices et des mesures utilisées pour minimiser les conséquences, réduire les impacts d'une catastrophe et contrôler les risques résultants [6].

L'atténuation se déroule avant ou entre événements pour réduire les impacts. L'implémentation de processus d'atténuation peut être considérée comme une partie du processus de récupération si elle est appliquée après une catastrophe [5].

Le processus engage l'identification des risques de catastrophe, codes du bâtiment et zonage, l'analyse de la vulnérabilité, zonage révisé et gestion de l'utilisation des sols, sensibilisation et éducation du public, et d'autres efforts pour faire participer la collectivité à un événement catastrophique [2].

3.2 Préparation

La phase de préparation, "ou en anglais (*Preparedness*)", se concentre sur le développement des plans bien longtemps avant la situation de catastrophe, pour confirmer que les procédures et les ressources nécessaires pour faire face à une catastrophe sont en place pour venir rapidement en aide aux personnes touchées, réduire la confusion et restaurer l'ordre public aussi rapidement que possible suite à une éventuelle catastrophe [2] et [6].

Chapitre I Gestion des catastrophes naturelles par des données de télédétection

Parmi les activités de planification : exercices de préparation à la situation d'urgence/formation, la préparation des forces de sauvetage (police, ambulance, pompiers, etc.), systèmes d'alerte, promotion d'autosuffisance communale avant et durant les urgences [6]. Les activités doivent être conçues de manière à garantir qu'en cas de catastrophe, le personnel approprié soit en mesure de fournir la meilleure réponse possible [5].

3.3 Intervention (ou réaction)

L'intervention, (ou en anglais *Response*) est la phase aiguë et critique durant ou immédiatement après une catastrophe. La réaction est définie comme les mesures initiales prises au moment de l'événement pour minimiser les dangers que génère une catastrophe [5], [6].

Les activités d'intervention adressent les menaces immédiates présentées par la catastrophe, notamment l'évaluation des dégâts, le sauvetage de vies humaines, la satisfaction des besoins humanitaires (nourritures, abris, vêtements, santé publique et sécurité), le nettoyage et le début de la distribution des ressources [6].

3.4 Rétablissement

La quatrième phase est le rétablissement (ou en anglais *Recovery*), qui peut être trouvé sur le nom de réhabilitation, de récupération et de reprise. Le rétablissement consiste à rétablir le fonctionnement de la collectivité [5]. La reprise faudrait remettre la zone touchée dans un état similaire ou meilleur qu'avant la catastrophe, notamment : la restauration et l'amélioration, les habitations temporaires, les subventions, les soins médicaux, les moyens de subsistance et les conditions de vie des communautés touchées par des catastrophes [2].

4. Gestion des catastrophes naturelles

Durant les dernières années, l'occurrence des catastrophes naturelles a considérablement augmenté, surtout celles qui sont liées à la météorologie comme les inondations, les tempêtes, les cyclones, les sécheresses et les températures extrêmes, les séismes, et touchent globalement plus du tiers de la population mondiale. La gestion des catastrophes naturelles est l'un des domaines les plus importants qui vise à prendre des décisions cruciales pour les opérations de sauvetage, l'estimation des pertes économiques et de reconstruction des zones affectées. Les informations nécessaires pour les opérations de sauvetage et de reconstruction peuvent être obtenues par deux méthodes principales [5] :

- La recherche sur le terrain;
- Et les technologies de télédétection.

4.1 La recherche sur le terrain

La recherche sur le terrain (en anglais *field survey* ou *field inspection*) dépend d'un certain nombre de visites dans la zone affectée (terrain) par une équipe de personnes bien formée et qualifiée qui identifie et enregistre des informations importantes afin de prendre des décisions décisives pour le sauvetage et l'assistance [7]. Les méthodes de recherche sur le terrain permettent de fournir des informations très détaillées et précises sur le niveau et le

degré des dommages causé par la catastrophe étudiée. Mais cette approche nécessite beaucoup de travail et de temps et elle est fortement influencée par l'expérience, les compétences et la perspective des inspecteurs. De plus, l'accès à certaines zones peut être impossible en raison de restrictions d'ordre politiques ou pratiques [8] et [9].

4.2 Les technologies de télédétection

Les méthodes de recherche sur le terrain sont très limitées et fondamentalement trop lentes, coûteuses et statiques [10]. La technologie de télédétection en anglais (*Remote sensing technologies*) est la plus prédominante méthode pour fournir des données permettant de réaliser des évaluations de la gestion des catastrophes [9].

Les technologies de télédétection consistent à utiliser des images de télédétection pour surveiller une vaste surface de la terre dans les cas normaux et dans les cas avec catastrophes naturelles. Cette méthode permet d'avoir un faible coût et une évaluation rapide des dégâts sans contact avec la région affectée, ce qui réduit le niveau de risque pour les spécialistes [11].

5. Télédétection

Le terme de "Télédétection" (ou en anglais *Remote sensing*) désigne la science, et jusqu'à un certain point, l'art, qui permet d'acquérir des informations sur la surface de la terre sans être en contact physique avec elle. Cela veut dire que l'acquisition se fait à distance, à l'aide d'un dispositif d'enregistrement de capteurs radiométriques à bord des plateformes aériennes ou spatiales qui n'a aucun contact physique ou intime avec la surface de la Terre (un objet ou un phénomène) [2] et [12].

La télédétection est la principale méthode d'acquisition de l'image d'observation de la terre. Elle offre la possibilité d'observer et de collecter rapidement des données sur de grandes surfaces [12].

5.1 Domaines d'application de télédétection

La télédétection est l'un des techniques les plus répandues dans les différentes recherches. Les domaines d'application des télédétections sont nombreux que variés, telle qu'ils peuvent servir aussi dans beaucoup de domaines de gestion des catastrophes [13]. Voici quelques domaines d'application importants dans lesquels la télédétection est utile [14] et [15] :

- **Systèmes d'informations géographiques** : sont des systèmes d'information, de matériels, de logiciels, et de processus capables d'intégrer, saisir, rassembler, stocker, mettre en formes, analyser, partager et exploser des informations par référence géographiques, contribuant notamment à la gestion de l'espace [2] et [12]. Au sens plus général, la technologie du SIG est un outil peut permettre à son utilisateur d'avoir des informations interactives, d'analyser les informations spatiales, de mettre en formes les données, de cartographier et de présenter les résultats de toutes ces opérations [2].
- **Agriculture** : pour l'identification du type de culture, la surveillance de l'état des cultures, l'humidité du sol mesure, travail du sol et identification des résidus de récolte, cartographie des types de culture, l'évolution des terres agricoles, analyse de pentes, ... ;

Chapitre I Gestion des catastrophes naturelles par des données de télédétection

- **Protection civile** : la gestion et la prévention des catastrophes, la cartographie de dunes de sable, l'évaluation et surveillance environnementales, la couverture rapide en cas d'urgence, pollution, ... ;
- **Forêts** : pour la cartographie des coupes à blanc et linéaires, l'estimation de la biomasse, les espèces identification et cartographie des cicatrices de feu, l'évolution des forêts, surveillance d'incendie, contrôle des feux de forêt, ... ;
- **Planification urbaine** : cadastre, voirie, les géomètres, le changement urbain, les instituts géographiques et cartographiques nationaux, les organismes spécialisés dans la récolte de données, le développement des infrastructures, les plans d'occupation des sols, la classification d'occupation du sol, l'identification de la désertification, le suivi des changements d'occupation du sol, ... ;
- **Géologie** : pour la cartographie géologique, la prospection minière, la cartographie structurelle et analyse de terrain, l'optimisation de parcours de distribution, planification, ... ;
- **Océanographie** : pour faire la vision synoptique de vastes régions qu'il est impossible d'obtenir par les moyens traditionnels (bateaux), mesure de la température de surface de l'océan, l'identification de la glace de mer, la mesure du champ côtier éolien et la mesure de la pente des vagues, surveillance des océans et des côtes, caractéristiques de l'océan, couleur de l'océan, détection de déversement d'hydrocarbures érosion, sédimentation, ... ;
- **Transport** : pour la planification des transports urbains, l'optimisation d'itinéraires, la gestion des réseaux routiers et des transports publics, l'évaluation de l'état des routes rurales, ... ;
- **Marketing** : pour la localisation des clients, la recherche d'implantation de sites ou de bâtiments, l'optimisation de parcours de distribution, la planification, ... ;
- **Sécurité de l'eau** : pour la gestion des réseaux d'eau et d'assainissement, la surveillance maritime, la conservation des lacs et des rivières, ... ;
- **Cartographie** : pour la cartographie thématique, les calculs topographiques, ... ;
- **Expédition** : pour la détection et la classification des navires, ... ;
- **Tourisme** : pour la gestion des infrastructures, les itinéraires touristiques, ... ;
- **Hydrologie** : pour la surveiller les régions humides et la couverture de neige, ... ;
- **Télécoms** : implantation d'antennes pour les téléphones mobiles, ... ;
- **Militaire** : surveillance militaire, défense, ... ;
- **L'étude de l'atmosphère** : études météorologiques et climatiques, ... ;
- **Etc.**

6. Images de télédétection

L'image de télédétection est une source de donnée idéale et précise sur l'occupation du sol et sur les phénomènes qui se déroulent à la surface de la terre. Ce type d'image est considéré comme la représentation graphique des mesures capturées par des capteurs placés à bord des plateformes autour de la terre. La plupart des images de télédétection sont composées d'une matrice d'éléments d'image, ou un ensemble de pixels, qui représentent la valeur de l'intensité du rayonnement réfléchi par l'objet observé dans la gamme de longueur d'ondes, auxquelles le capteur est sensible [16].

Car les images de télédétection constituent des sources d'informations primordiales et importantes pour les divers domaines d'applications comme la gestion des ressources agricoles et environnementales, exploration minérale, évaluation rapide des dégâts sur de vastes zones pour des interventions rapides, prévision météo, il est nécessaire déterminer quels types de données sont appropriés à telle application, prendre en compte les caractéristiques des images de télédétection telles que la résolution, la plateforme d'acquisition et la qualité des données [5].

6.1 Résolution

La résolution des images de télédétection est exprimée par la taille de la zone de la surface de la terre couverte par un pixel. Chaque pixel de l'image correspond à une partie de la surface de la terre. Pour certains dispositifs de télédétection, la distance entre la cible observée et la plateforme est un facteur important pour déterminer la précision des données obtenues et la zone observée par le capteur. Sur une plateforme très proche de la cible, le capteur surveille généralement une zone plus petite mais avec beaucoup de précision. Alors qu'il existe quatre aspects différents de résolution [17] :

- La résolution spatiale ;
- La résolution spectrale ;
- La résolution temporelle ;
- Et la résolution radiométrique.

6.1.1 Résolution spatiale

La résolution spatiale est une mesure de la grosseur ou de la finesse d'une grille de raster. La résolution spatiale correspondant à la surface élémentaire d'échantillonnage observée instantanément par le capteur. Cette surface élémentaire correspond au pixel [16].

La capacité à distinguer deux objets voisins différents est donc dépend de la résolution spatiale du capteur. Dans les images à haute résolution, on peut discerner de petits objets lorsque la résolution spatiale augmente ; l'image contient plus de détails. Les détails sont précieux pour certaines applications, mais ils sont également coûteux. Les capteurs de l'armée, par exemple, sont conçus pour déceler le plus de détails possibles et ont donc une très haute résolution. En règle générale, plus la résolution augmente, plus la superficie totale visible par le capteur diminue [17].

6.1.2 Résolution spectrale

La résolution spectrale décrit la capacité d'un capteur à discerner des bandes de longueurs d'ondes différentes [16]. La distinction entre deux objets est plus facile lorsqu'un système de télédétection à une grande résolution spectrale [17].

6.1.3 Résolution temporelle

L'intervalle de survol est défini comme le temps que met un satellite à compléter un cycle orbital, qui est habituellement de quelques jours [17]. La résolution temporelle ou la répétitivité correspond à cet intervalle de survol afin de capter la même scène, c'est-à-dire,

capter exactement la même image au même angle. La résolution temporelle effective du satellite dépend donc de divers facteurs, tels que la capacité du satellite et de ses capteurs, et le mode de manœuvre du satellite [16].

6.1.4 Résolution radiométrique

Le dernier aspect de la résolution est la résolution radiométrique, qui désigne la capacité d'un capteur, dans une bande spectrale définie, à distinguer des signaux électromagnétiques d'énergies différentes. Elle peut être exprimée par le rapport signal/bruit [16].

6.2 Plateforme d'acquisition

Un capteur installé sur une plateforme distante de la surface ou de la cible observée peut enregistrer correctement l'énergie réfléchie ou émise par celle-ci. On distingue deux types différents de plateformes aériennes et satellitaires.

6.2.1 Plateformes aériennes

Généralement, la photographie aérienne est une image prise des airs. Les photos aériennes sont prises à la verticale à l'aide d'une caméra spécialisée de prise de vues hautement précis, installée à bord d'un aéronef équipé pour de tels travaux. Ces types de photos rapprochés sont faits généralement à l'aide d'un drone, d'un hélicoptère, d'un avion, d'une montgolfière ou d'un ballon radiocommandé [18].

Les photos aériennes sont acquises à une résolution spatiale de l'ordre du décimètre, ce qui permet de faire des mesures précises. Ces mesures sont utilisées en géologie, en surveillance, en foresterie, en cartographie et dans bien d'autres domaines. Elles fournissent une source importante qui permet d'identifier des bâtiments, l'infrastructure de transport, l'état des dégâts, les types des dommages après les catastrophes, etc. ; mais elles ne peuvent pas couvrir une vaste zone avec un temps d'acquisition et garantir la répétitivité des prises de vue de la même scène. Ces images peuvent être obtenues à tout moment et dans tous les cas, cependant, ne peuvent pas être toujours disponibles pour un certain nombre de raisons [19].

- L'avion et l'équipement de levé aérien ne sont toujours pas disponibles.
- Les coûts sont élevés ; des ressources financières limitées pourraient empêcher l'affrètement d'un avion.
- Les permissions de vol pourraient être refusées par les autorités compétentes,
- La météo va perturber le plan de vol, des ciels nuageux peuvent empêcher le processus d'acquisition d'images.

6.2.2 Plateformes satellitaires

Les plateformes satellitaires sont des satellites artificiels en orbite autour de la terre sous forme elliptique. Ces plateformes peuvent fournir plusieurs types d'images qui sont numériquement transmis, habituellement appelés images satellites, ou images satellitaires [18].

Les images satellitaires sont très riches en informations de toutes sortes et ont l'avantage d'être capables d'observer une grande surface en même temps. Elles sont des

Chapitre I Gestion des catastrophes naturelles par des données de télédétection

sources de données idéales pour analyser divers aspects du territoire, créer et mettre à jour des cartes, évaluer rapidement des dégâts sur de vastes régions pour des actions de réponses rapides, etc. ; mais les limites des images satellitaires doivent être prises en compte, telles que [5] :

- La plupart des capteurs satellites ont des résolutions limitées ;
- Les satellites ne peuvent pas collecter simultanément des données d'images à plusieurs endroits ;
- Les satellites apparaissent à un endroit particulier sur la terre toujours au même moment de la journée ;
- Et les capteurs optiques par satellite ont des restrictions liées à la météo.

6.3 Types de données de télédétection

La plupart des données utilisées dans les recherches peuvent être classées en trois catégories : (1) données optiques, (2) données radar, et (3) données LiDAR [9].

6.3.1 Données optiques

Les données de télédétection optiques peuvent être acquises à la fois par des capteurs satellitaires et aériens. Ces données sont couramment utilisées dans la surveillance, l'évaluation des catastrophes, l'identification et la caractérisation des écosystèmes, car elles fournissent des informations intéressantes sur l'occupation et l'utilisation des terres, et permettent par la même occasion d'identifier et de caractériser différents éléments du paysage à des échelles différentes. Les données optiques sont faciles à interpréter puisqu'elles représentent la surface du sol telle qu'elle apparaît à l'œil humain. Les capteurs optiques fournissent un fonctionnement plus facile, moins complexe, plus fiable et sont moins coûteux [5] et [9].

6.3.2 Données radar

Le grand avantage de radar, également appelés radars à synthèse d'ouverture (ou en anglais *Synthetic Aperture Radar, SAR*), est qu'il peut être utilisé indépendamment de la lumière du soleil et des conditions météorologiques et pourrait être utile, en particulier dans des zones à nébulosité persistante [5].

L'imagerie radar présente un intérêt particulier lorsque la couverture nuageuse empêche l'acquisition d'images optiques, elle peut être la seule donnée disponible en cas de mauvaises conditions météorologiques [19]. Les principaux problèmes sont les caractéristiques spectrales inhabituelles des capteurs radar qui sont très différentes du système visuel humain et donc difficilement interprétables et aussi la résolution spatiale très limitée [5].

6.3.3 Données LiDAR

LiDAR est un acronyme de détection et de télémétrie par ondes lumineuses (en anglais *Light Detection And Ranging*), qu'utilise un laser pour mesurer la distance entre le capteur et les points [20]. Les systèmes LiDAR permettent une acquisition rapide et étendue de données de hauteur précises qui peut être utilisées pour identifier certaines mesures très précises qui ne peuvent pas être identifiées à l'aide d'images optiques ou radar. Mais, l'acquisition de

données LiDAR pendant la pluie ou le brouillard à travers les nuages n'est pas possible, car LiDAR est très sensible aux aérosols et aux particules nuageuses. Aussi beaucoup de régions n'ont pas de couverture LiDAR car la technologie LiDAR est relativement récente [5].

7. Techniques de gestion des catastrophes naturelles

Des nombreuses approches pour la surveillance et l'estimation des dégâts ont été proposées dans la littérature. Elles peuvent être classées en trois groupes [20] :

- Interprétation visuelle ;
- Techniques Multi-temporelles ;
- Et techniques Mono-temporelles.

7.1 Interprétation visuelle

L'approche d'interprétation visuelle est employée pour détecter et évaluer les dégâts en utilisant les différents types d'images de télédétection et des données SIG, mais dépend de nombreux outils auxiliaires, comme par exemple MapInfo, ArcGIS, InfoMap, ArcView, ERDAS, etc. L'interprétation visuelle est la plus utilisée pour l'estimation des dégâts, car c'est la méthode la plus précise actuellement [20].

Malheureusement, l'interprétation visuelle prend beaucoup de temps, nécessite des opérateurs qualifiés, génère une lourde charge de travail et la difficulté à prendre en charge des critères d'interprétation cohérents dans le cas où un grand nombre de photo-interprètes travaillent en parallèle pour interpréter de vastes zones en un peu de temps [5]. Cela le rend impropre à l'évaluation rapide des dégâts sur une vaste zone [20]. L'interprétation visuelle peut être encore plus compliquée à des résolutions spatiales supérieures à une mètre, si elles sont accompagnées d'une formation d'ombre, de différences de capteurs et de conditions solaires entre les images pré- et post-événements [21].

7.2 Techniques Multi-temporelles

Les images de télédétection acquises avant et après les catastrophes sont utilisées pour déterminer les déformations qu'ont eues lieu pendant la catastrophe. Les approches multi-temporelles comparent et analysent les changements significatifs entre les images pré- et post-catastrophes [22].

En fonction de la disponibilité et de la qualité des données, il est généralement possible d'obtenir un meilleur taux de réussite lorsque des données pré- et post-événements sont utilisées. Par rapport aux approches utilisant uniquement des données post-événements, des résultats plus précis peuvent être obtenus par ces approches [20].

Les approches multi-temporelles possèdent une principale limitation dans plusieurs villes, en particulier dans les pays en développement : il est difficile d'obtenir des références pré-événements, le couple homogène d'images pré- et post-événements est difficile à identifier, et l'enregistrement de haute précision est délicat et prend énormément de temps [23]. Même lorsque les données pré-événements sont disponibles, il peut exister un problème pour obtenir des résultats fiables et précis parce que les grandes différences entre les données

pré- et post-événements, en termes de résolution d'images, d'angles de vue, de couleurs, de spectres, de capteurs, etc. De plus, les difficultés sont amplifiées par un grand écart temporel [20].

7.3 Techniques Mono-temporelles

Les approches mono-temporelles détectent les dégâts en utilisant un seul ensemble des données de télédétection collectées après la catastrophe [23].

Dans ce cas, le problème de détection de dégâts peut être considéré comme un problème de classification, s'applique principalement aux algorithmes d'apprentissage automatique supervisé (en anglais *Supervised Machine Learning Algorithms*), pour la reconnaissance visuelle des éléments endommagés, et alors elles sont directement liées à la résolution d'image. Les performances de ces approches dépendent principalement de la puissance des méthodes d'extraction des caractéristiques [22].

Ces approches sont plus rapides et plus appropriées que les approches multi-temporelles car elles utilisent uniquement des données post-événements. Mais l'identification précise des dégâts sans savoir ce qui existait avant la catastrophe est très difficile [23].

8. Conclusion

Nous avons présenté, au long de ce chapitre, les différentes notions de base de gestion des catastrophes à partir des données de télédétection. Où nous avons commencé par présenter les types et le cycle de gestion des catastrophes géographiques. Puis, nous avons expliqué les concepts de base pour les technologies et les applications de télédétection. Enfin, nous avons porté notre attention sur les différentes techniques de gestion des catastrophes par des données de télédétection.

Dans le prochain chapitre, nous allons présenter un état de l'art sur les algorithmes d'apprentissage automatique. Dans un premier temps, nous décrirons en détail les deux phases principales de ces algorithmes. Puis, dans un deuxième temps, nous présenterons les différentes approches existantes pour chaque phase depuis plusieurs années.

Chapitre II

Extraction de caractéristiques et classification

À travers ce chapitre, nous ferons un état de l'art sur les algorithmes d'apprentissage automatique qui permettent un fonctionnement en deux phases complémentaires, extraction des caractéristiques et classification. Où nous présenterons les différentes méthodes et les techniques connues dans la boîte à outils de la littérature pour chaque phase.

1. Introduction

Les algorithmes d'apprentissage automatique jouent un rôle important dans les divers domaines, en particulier dans le domaine de télédétection. Ils ont été appliqués avec succès dans plusieurs applications, comme par exemple, la recherche d'image par le contenu, la reconnaissance de visage, la reconnaissance de parole, la classification des scènes naturelles, la gestion des catastrophes, etc. Le défi consiste à remplacer l'interprétation visuelle des images, réalisée par des photo-interprètes experts, par des techniques de classification automatique ou semi-automatique. Généralement, ces approches consistent en deux modules principaux : module d'extraction de caractéristiques et module de classification.

Généralement, la phase d'extraction des caractéristiques est une étape essentielle et cruciale dans la classification, car la performance des classifications dépend en grande partie du choix des descripteurs employés et est relative à la qualité des techniques d'extraction des descripteurs utilisés. Le descripteur choisi doit être le plus représentatif, le plus discriminatoire et le plus rapide en termes de temps de calcul. Cette phase consiste à extraire automatiquement des diverses caractéristiques d'image pour obtenir les descripteurs, qu'est appelé aussi un vecteur de caractéristiques ou signature (en anglais *feature*). Un bon descripteur doit permettre de décrire le contenu d'image avec certaines variations, en prenant en compte le changement d'illumination, la variation de l'échelle, la translation et rotation, le changement de point de vue, etc. Le choix des caractéristiques extraites est souvent guidé par la volonté d'invariance ou de robustesse par rapport aux changements d'illumination, à des transformations de l'image, ou aux deux à la fois. Beaucoup de recherches se font sur les descripteurs pour pouvoir les utiliser dans la reconnaissance et l'interprétation des images. On peut généralement diviser les méthodes d'extraction des caractéristiques existantes en trois catégories principales en fonction des caractéristiques extraites : méthodes basées sur les caractéristiques de bas niveaux, méthodes basées sur les caractéristiques de mid- niveaux ou niveaux intermédiaires, et méthodes basées sur les caractéristiques de plus hauts niveaux.

Dans le chapitre précédent, nous avons vu les différentes notions de base de gestion des catastrophes à partir des données de télédétection. En plus, nous avons abordé les différentes techniques de gestion des catastrophes par des données de télédétection, parmi elles, ceux qui s'appliquent principalement des algorithmes d'apprentissage automatique.

Dans ce chapitre, nous ferons un état de l'art sur les algorithmes d'apprentissage automatique qui peuvent fonctionner en deux phases complémentaires, extraction des caractéristiques et classification. Où nous présenterons les différentes méthodes et les techniques connues dans la boîte à outils de la littérature pour chaque phase. Plus précisément, nous expliquerons les méthodes d'extraction des caractéristiques basées sur les caractéristiques de bas niveaux et sur les caractéristiques de niveaux intermédiaire.

2. Caractéristiques de bas niveaux

Les méthodes basées sur les caractéristiques de bas niveaux ou les caractéristiques visuelles (en anglais *handcrafted feature methods*) s'intéressent à l'information contenue dans l'image au niveau du pixel, telles que la couleur, la texture, la forme, des informations

spatiales et spectrales, ou leurs combinaisons, qui constituent la caractéristique principale d'une image et portent donc des informations utiles pour la classification [26]. On peut utiliser des descripteurs caractérisant la totalité de l'image ou plusieurs descripteurs locaux caractérisant chacun une partie de l'image. Donc, il y a principalement deux approches pour les caractéristiques qui peuvent être extraites [25] :

- Des descripteurs globaux ;
- Et des descripteurs locaux.

2.1 Approches globales

Dans ce cas, il s'agit de fournir des observations sur la totalité de l'image. L'extraction de descripteurs globaux permet d'extraire un seul descripteur, qui décrit la totalité de l'image. Cela les rend robustes au bruit [24].

Ces caractéristiques permettent de réduire le nombre de calculs nécessaires, la taille de la base de données ainsi que le coût de la classification. Cependant, l'approche globale ne permet pas de distinguer des parties de l'image ; ils ne distinguent pas, par exemple, les objets dans l'image, sauf dans le cas où l'image ne contient qu'un seul objet dans un fond uni [25]. La couleur, la texture et la forme sont des caractéristiques nécessaires pour la description du contenu globale d'images. Elles jouent un rôle très important dans la perception visuelle humaine [27].

2.1.1 Descripteurs de couleur

Les descripteurs de couleurs sont des caractéristiques riches d'information visuelle et très utilisées pour la représentation des images. Le descripteur de couleur est très souvent le premier descripteur qui est employé pour la description des images. Plusieurs travaux ont déjà prouvé qu'il est un descripteur efficace [25].

Une couleur est généralement représentée au moyen de triplets numériques permettant de coder l'intensité de ces composantes. Ces composantes définissent un espace de couleurs. Il existe plusieurs espaces de couleurs qu'ont chacun certaines caractéristiques intéressantes, comme par exemple les espaces colorimétriques définis selon des propriétés comme RGB pour Red, Green, Blue, et ceux basés sur la perception humaine des couleurs telles que HSV pour Hue, Saturation, Value ou HSB pour Hue Saturation Brightness [24]. Si on change l'espace colorimétrique, on trouve qu'avec la même méthode il peut émerger des informations différentes de l'image. Alors, avant de sélectionner le descripteur de couleur approprié, l'espace de couleur doit être déterminé d'abord [28].

Il existe nombreuses possibilités d'attributs pour caractériser la couleur. Les descripteurs de couleurs le plus utilisé dans la littérature sont [28] :

- Histogramme de couleurs ;
- Moments de couleur ;
- Cohérence spatiale ;
- Couleurs dominantes ;
- Et corrélogrammes.

2.1.1.1 Histogramme de couleurs

L'histogramme de couleurs est le descripteur de couleurs le plus courant et le plus rencontré dans la littérature [26]. L'histogramme de couleur d'une image représente la distribution des intensités de pixels. Chaque valeur d'intensité possible est associée par le nombre d'occurrences dans l'image [29]. Il existe plusieurs catégories d'histogrammes. Elles peuvent être classées par exemple selon l'espace de couleur considéré lors du calcul tels que l'histogramme RGB, l'histogramme HSV, etc., qui sont associés respectivement aux espaces de couleurs RGB et HSV [25].

L'histogramme de couleur est facile et rapide à calculer, et robuste à la rotation et à la translation. Cependant, l'utilisation d'histogrammes de couleurs pose des problèmes, notamment : ils ne possèdent pas d'informations spatiales sur les positions des couleurs donc par conséquent il est très difficile de distinguer les images avec les mêmes couleurs mais avec une distribution de couleurs différente. Aussi, le descripteur d'histogramme de couleur est également sensible à de petits changements de luminosité, ce qui devient problématique pour les images acquises dans des conditions différentes [26].

2.1.1.2 Moments de couleur

Le descripteur d'histogramme de couleur utilise la distribution complète de couleur. Par conséquent, il est à de grandes tailles, On doit stocker de nombreuses données. Au lieu de calculer la distribution complète de couleur, on calcule les descripteurs des moments de couleur, appelées aussi les moments statistiques, qui consiste à calculer seulement des caractéristiques dominantes de couleur telles que l'espérance, la variance et d'autres moments sur chaque composante de couleur. La littérature a prouvé que les méthodes des moments de couleur utilisées sont plus rapides et donnent des résultats plus efficaces que les méthodes d'histogrammes [28].

2.1.1.3 Cohérence spatiale

Les descripteurs de cohérence spatiale ont pour but de combler en partie, l'absence d'information spatio-colorimétrique de l'image dans les deux descripteurs précédents. Ce descripteur est calculé pour chaque classe de couleur identifiée. La cohérence spatiale de couleur c est alors donnée par le rapport $\frac{H_I(c)}{H(c)}$, où $H_I(c)$ représente l'histogramme de connexité de couleur c , et $H(c)$ représente l'histogramme de couleur c [28]. Un histogramme de connexité est calculé par :

$$H_I(c) = \sum_{i=0}^{x=I} \sum_{j=0}^{x=I} \delta(I(i,j), c) \alpha(i,j) \quad (2.1)$$

Où I est l'image segmentée de taille (X, Y) , c 'est la couleur du pixel (i, j) , δ est le symbole de Kronecker et $\alpha(i, j)$ définie par:

$$\alpha(i, j) = \begin{cases} 1 & \text{si } \forall k, k' \in (-w, w) I(i+k, j+k') + I(i, j) \\ 0 & \text{sinon} \end{cases} \quad (2.2)$$

La fenêtre $(2W + 1) * (2W + 1)$ représente le degré de compacité souhaité.

2.1.1.4 Couleurs dominantes

Dans le cas d'histogramme, il existe une grande partie de l'espace mémoire qu'est utilisée inutilement, comme par exemple les histogrammes à plusieurs dimensions sont creux, c'est-à-dire, que la majorité des cellules ne comptent aucun pixel. Les descripteurs par couleurs dominantes, aussi appelée DCD (pour l'anglais : *Dominant Color Descriptor*), permettent de résoudre ce problème. Ces caractéristiques ne stockent pas les cellules vides, elles ne stockent que les couleurs, qui appartiennent à l'image [28].

2.1.1.5 Corrélogrammes

L'histogramme et la couleur dominante donnent une bonne estimation de densité de couleur d'une image, mais ils perdent la distribution spatiale des couleurs. Ces descripteurs souvent ne donnent pas des caractéristiques discriminatives, ils peuvent donner le même descripteur pour deux images qui ont exactement le même nombre de pixels de chaque couleur, mais uniquement leur répartition spatiale change [28].

Ce problème peut être adouci, si l'on considère, pour chaque couleur, non plus le nombre d'occurrences des pixels ayant cette couleur, mais les probabilités pour les différentes couleurs qu'elles soient présentes à différentes distances. Ces histogrammes optimisés sont appelés des corrélogrammes. Les corrélogrammes semblent donner de très bons résultats. Cependant, l'implémentation des corrélogrammes est très délicate et très coûteuse en temps de calcul [28].

2.1.2 Descripteurs des textures

Le descripteur de texture est une caractéristique fondamentale des images. Il est plus en plus utilisé, car il concerne un élément important de la vision humaine et il permet de pallier certains problèmes posés par le descripteur de couleur, notamment lorsque les distributions de couleurs sont très proches. Cependant, il n'existe pas de définition formelle de texture et il n'en existe pas non plus de représentation officielle [28]. Une texture peut être caractérisée par la répétition d'un motif ou de quelques éléments, il peut être vu comme un ensemble de pixels spatialement agencés selon un certain nombre de relations spatiales, ainsi créant une région homogène et invariante par translation [24]. La texture peut être également définie par : "Une texture est un champ de l'image qui apparaît comme un domaine cohérent et homogène, c'est-à-dire formant un tout pour un observateur" [30].

De nombreuses approches ont été proposées pour la caractérisation de texture. Parmi les plus connues, on peut citer :

- Matrice de cooccurrence de niveaux de gris ou GLCM ;
- Transformée en ondelettes ;
- Filtres de Gabor ;
- Motif de texture binaire local ou LBP ;
- Et histogramme de gradient orienté ou HOG.

2.1.2.1 GLCM

La matrice de cooccurrence de niveaux de gris (en anglais *Gray-Level Cooccurrence Matrix*) est un descripteur important de texture et couramment utilisé pour la description des textures d'une image. Cette approche est l'une des méthodes statistiques basées sur des évaluations quantitatives de la distribution de niveaux de gris, qui ont un principe simple basé sur l'examen de la relation spatiale entre les pixels et ses voisins [29]. La GLCM peut être interprétée par la texture d'une image comme la régularité d'apparition de chaque paire de niveaux de gris selon une distance donnée dans l'image [28].

D'abord, il faut mettre l'image à l'échelle de gris pour trouver les n valeurs d'intensité. À partir de ces valeurs on crée une matrice de cooccurrence C de dimension $n \times n$, qui décrit le nombre d'apparence de chaque couple de pixels à un repère d'un vecteur d'orientation θ Où $\theta = (0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4})$ [30].

$$C = \begin{bmatrix} p(1,1) & \cdots & p(1,n) \\ \vdots & \ddots & \vdots \\ p(n,1) & \cdots & p(n,n) \end{bmatrix} \quad (2.3)$$

Où chaque élément $p(i,j)$ est défini par une distance d , une direction θ et le nombre de couples de valeurs (i,j) séparés par une distance d dans la direction θ présents dans l'image [30]. A partir du GLCM calculé, on peut décrire quelques statistiques, telle qu'il y a 14 caractéristiques statistiques extraites à partir de cette matrice. Les caractéristiques les plus appropriées et largement utilisées sont l'homogénéité, la corrélation, l'énergie, l'entropie, le contraste et le moment des différences inverses [28].

Les descripteurs de couleur donnent un histogramme identique des images qu'ont les mêmes valeurs de luminance, seul l'arrangement spatial de ces valeurs les distingue. Les matrices de cooccurrence ont l'avantage de discriminer des textures qui ont des histogrammes identiques. Elle compte parmi les méthodes les plus utilisées telles que ses applications en télédétection sont nombreuses. Cependant, le GLCM ne peuvent pas caractériser tous les types de texture et l'utilisation de cette approche est peu pratique, car elle nécessite la définition de plusieurs paramètres : choix des distances séparant les couples, des directions d'analyse, de la taille de la fenêtre d'analyse et des indices pour interpréter la matrice [30].

2.1.2.2 Filtres de Gabor

Les descripteurs de filtres de Gabor, appelés aussi ondelettes de Gabor (ou en anglais *Gabor wavelets descriptors*), sont largement utilisés en vision par ordinateur, pour extraire les caractéristiques de textures à partir des images. Le filtre de Gabor est l'un des méthodes fréquentielles, qui s'appuient sur la transformée de Fourier et reposent sur l'analyse d'une fonction de densité spectrale dans un domaine fréquentiel [25].

Généralement, les paramètres de texture sont déterminés en calculant la moyenne et l'écart-type des niveaux de gris de l'image filtrée par Gabor. En fait, un ensemble de valeurs de moyenne et d'écart-type qui sera calculé égal au nombre d'échelles multiplié par le nombre d'orientations utilisées. Ces valeurs calculées seront regroupées dans un seul vecteur de caractéristiques [28].

Les descripteurs de filtre de Gabor sont obtenus en convolant l'image avec un ensemble de filtres de Gabor, qui est parfois appelé la banque de filtre de Gabor. Cette méthode permet d'extraire de l'image des informations pertinentes, à la fois en espace et en fréquence, relatives à la texture [9]. Plusieurs recherches montrent que ce descripteur est très efficace, robuste, et facile à calculer, telles qu'ils simulent de manière convenable le système visuel humain en reconnaissance de textures [25].

2.1.2.3 Transformée en ondelettes

La description de texture à base d'ondelettes est très utilisée en vision par ordinateur, pour extraire les caractéristiques de textures à partir des images. Ce descripteur est une méthode fréquentielle [25] et à la base de plusieurs analyses de texture, telles que les filtres de Haar [28].

La transformée en ondelettes est similaire la transformée de Fourier, mais elle se distingue par le fait qu'il permet une représentation dans l'espace des différentes fréquences contenues dans un signal original. La transformée en ondelettes $F(a, b)$ d'un signal continu $f(t)$ peut être exprimée avec l'équation suivante [30] :

$$F(a, b) = \int_{-\infty}^{+\infty} f(t) \frac{1}{\sqrt{a}} \varphi^* \left(\frac{t-b}{a} \right) dt \quad (2.4)$$

Où a représente le facteur de dilatation, b le facteur de translation et φ^* le complexe conjugué de la fonction analysante φ . La valeur de la transformée en ondelettes $F(a, b)$ correspond à la projection du signal original sur la fonction d'analyse $\varphi^* \left(\frac{t-b}{a} \right)$ et une ondelette fournit une représentation du signal à toutes les échelles a [30].

Pour calculer une transformée en ondelettes, on doit réaliser un produit de convolution du signal avec une paire de filtres, au lieu de calculer le produit scalaire de l'ondelette avec le signal. La méthode texture de transformée en ondelettes la plus couramment employée est la transformée de Haar [28]. Le filtre de Haar est fréquemment employé pour obtenir la description texture d'une image.

2.1.2.4 LBP

Le motif binaire local (ou en anglais : *Local Binary Pattern*) est connu comme un descripteur de texture vraiment puissant et plus populaire, qui donne une représentation de texture robuste et invariante aux transformations géométriques telles que la translation, la rotation et le changement d'échelle. Il donne un excellent résultat dans plusieurs applications, notamment, la reconnaissance des visages, la texture, la classification d'image, etc. [31].

C'est une des méthodes d'analyse géométrique, qui permettent de caractériser la texture en définissant les primitives et les règles d'arrangement qui les relient. D'abord, il faut mettre l'image en niveaux de gris. Le principe général du LBP est de comparer le niveau de luminance d'un pixel avec les niveaux de ses voisins. Cela rend donc compte d'une information relative à des motifs réguliers dans l'image. Certaines microstructures ou zones d'intérêts tels que des coins ou des bords peuvent être détectées par le LBP [29].

2.1.2.5 HOG

L'histogramme de gradient orienté (ou en anglais HOG pour *Histogram of Oriented Gradient*) est une méthode d'extraction de caractéristiques vraiment intéressante et couramment utilisée en vision par ordinateur. L'idée essentielle de HOG est de rassembler les variations de dégradé d'une image donnée [31].

Ces descripteurs peuvent être obtenus en divisant l'image en petites blocs adjacentes, appelées cellules, puis des histogrammes de gradient orienté dérivés pour les pixels de chaque cellule sont concaténés pour former un vecteur de caractéristiques. Cela donne la représentation globale d'image [9]. Puisque le descripteur HOG opère sur les cellules localisées, la méthode maintient l'invariance à des transformations géométriques et photométriques [31].

2.1.3 Descripteurs de forme

Au même titre que pour le descripteur de couleur et de texture, le descripteur de forme est une caractéristique visuelle très importante pour décrire le contenu d'une image. Ce descripteur est généralement une description très riche d'un objet ou son contour [28]. Il permet, comme leur nom l'indique, de présenter une information pertinente sur le contenu de l'image et précisément sur la forme [25].

De nombreuses techniques ont été proposées pour représenter une forme. Ces techniques peuvent être distinguées en deux catégories : les descripteurs basés sur les régions, qui peuvent considérer une région par rapport aux caractéristiques des pixels que cette région contient et les descripteurs basés sur les contours ou les frontières qui peuvent décrire une région au moyen des pixels situés sur son contour [24]. En cours d'extraction des descripteurs de forme, on fait appel aux techniques de segmentation d'images. La segmentation est une technique de traitement d'images au niveau de pixel. Pour cela, un sous module de segmentation a été élaboré pour délimiter les frontières d'objets (les formes). Cela rend l'extraction du descripteur de forme une tâche difficile [27].

2.2 Approches locales

La seconde approche est l'approche locale. Elle caractérise un ensemble de régions significatives de l'image, qui sont riches en termes de contenu de l'information locale. Ces descripteurs de bas niveau ont été utilisés avec succès dans plusieurs applications, car elles absorbent mieux certaines variations, comme par exemple les transformations affines et les changements d'illumination, et leur extraction est plus simple, mais ils sont coûteux [24].

Ces descripteurs peuvent être des régions de l'image obtenues soit par segmentation de l'image entière ou par description des points d'intérêt. Ces méthodes consistent à deux phases successives : ils visent tout d'abord à détecter les zones importantes de l'image, puis à calculer le descripteur de chaque zone détectée [25]. L'idée de base de première phase est qu'une image peut être décrite par un petit nombre de pixels significatifs (points saillants). Par exemple, les pixels correspondant aux bords et aux coins contiennent les informations les plus significatives par rapport aux pixels des régions homogènes. Le but de deuxième phase est d'extraire le descripteur d'entité locale pour chaque point saillant d'image donné [9]. Plusieurs méthodes ont été proposées comme un détecteur et un descripteur des points d'intérêts, telle que [32] :

- Détecteur de coins de Harris ;
- Transformation de caractéristiques visuelles invariante à l'échelle (ou SIFT) ;
- SIFT dense (ou DSIFT) ;
- Et accélérées de caractéristiques visuelles robustes (ou SURF).

2.2.1 Détecteur de coins de Harris

Le détecteur de Harris et Stephens ou le détecteur de Harris est une approche de détection de coins et elle a été inventée par Harris et Stephens. Cette approche permet la description locale d'une région d'intérêt dans l'image en termes de bords et de coins, telle qu'elle permet de détection des croisements de ligne par exemple. Ces points peuvent être reconnus en considérant seulement une petite fenêtre autour du point, puis en déplaçant la fenêtre dans toutes les directions. Pour ces points, le changement d'intensité devrait être essentiel et très important. Cette détection est basée sur l'analyse de gradients calculés localement ou dans une région d'intérêt [29].

2.2.2 SIFT

David Lowe a proposé et développé un descripteur robuste appelé SIFT (pour l'anglais *Scale-Invariant Feature Transform*), que l'on peut traduire par "transformation des caractéristiques visuelles invariante à l'échelle" permettant de transformer le contenu visuel d'une image en ensemble de vecteurs de caractéristiques, qui sont invariants à la translation, à l'échelle, à la rotation et partiellement invariante aux changements d'illumination et de projection affine et 3D [25].

SIFT est une méthode de caractérisation plus populaire pour l'extraction des points d'intérêts et couramment utilisée dans le domaine de la vision par ordinateur pour la reconnaissance d'objets, la reconnaissance de visages, le suivi de mouvement, l'indexation multimédia, etc. [28].

Cette méthode permet à la fois de détecter et de caractériser les points d'intérêts dans l'image de la famille des histogrammes de gradients orientés, c'est-à-dire dans chaque point particulier de l'image, un descripteur de dimension 128 a été extraire par la méthode SIFT. Ce descripteur se présente sous forme d'un histogramme grossier des orientations des gradients contenus dans son voisinage [24].

Le SIFT comporte généralement quatre étapes, à savoir : la détection d'extrema, la localisation des points d'intérêts, le calcul de l'orientation et du descripteur du point d'intérêt [26]. En sélectionnant 16 blocs autour du point d'intérêt de taille chacune 4×4 pixels, dans chaque bloc un histogramme d'orientation de gradients est calculé telle que chaque histogramme est de taille 8 bins qui représentent les 8 orientations principales entre 0 et 360 degrés : $\{0^\circ, 45^\circ, 90^\circ, 135^\circ, 180^\circ, 225^\circ, 270^\circ, 315^\circ\}$. Les vecteurs de $4 \times 4 \times 8 = 128$ dimensions sont donc le vecteur représentatif du descripteur SIFT [25].

2.2.3 DSIFT

Le SIFT dense (en anglais *Dense SIFT* ou *DSIFT*) est une variante de SIFT avec des descripteurs extraits à de multiples échelles. Ce descripteur est à peu près équivalent au calcul des SIFTs. Il est calculé à des points sur une grille régulière avec une échelle et une orientation fixées. SIFT identifie les points d'intérêts à l'aide de la fonction de différences de gaussiennes avant de calculer l'histogramme des gradients orientés pour décrire ces points d'intérêt, mais DSIFT n'identifie pas les points d'intérêts, il divise simplement l'image en cellules avant de calculer l'histogramme des gradients orientés pour la description, mais tous les deux produisent des vecteurs de caractéristiques de 128 dimensions. Le principal avantage de l'utilisation de DSIFT par rapport à SIFT est la rapidité. Ces descripteurs sont également invariants par rotation et souvent utilisés pour la catégorisation d'objet. Un SIFT dense génère plus de descripteurs à chaque emplacement et à l'échelle d'une image, ce qui par conséquence augmente la précision de la reconnaissance [32].

2.2.4 SURF

Le descripteur SURF (en anglais *Speeded Up Robust Features*) que l'on peut traduire par "Accélérées de caractéristiques visuelles robustes", est aussi un algorithme qui permet à la fois la détection de points d'intérêts d'une image et la description associée à celle-ci, présenté pour la première fois par Herber Bay en 2006 [24].

Ce descripteur approxime et obtient même de meilleurs résultats que le descripteur SIFT. Il est également plus performant en termes de rapidité. SURF est un détecteur de points d'intérêts robuste aux rotations et aux changements d'échelles. Il est utilisé dans le domaine de vision par ordinateur, pour des tâches de détection d'objet ou de reconstruction 3D. Le calcul de SURF utilise des détecteurs Hessienne rapides "approximation des ondelettes de Haar", pour détecter les structures de types blobs qui sont localisés aux maxima du déterminant de la matrice Hessienne [32].

3. Caractéristiques de niveau intermédiaire

En général, les méthodes basées sur des caractéristiques de bas niveau telles que les histogrammes de couleur, GLCM, HOG, SIFT, etc., ont prouvé qu'elles sont performantes. Mais l'amélioration des performances de ces caractéristiques est faible et ont également avec une mauvaise généralisation, particulièrement dans le cas où les images sont complexes et ont des caractéristiques très variables en termes de l'éclairage, la rotation et les changements d'échelle, etc. Ces caractéristiques peuvent constituer une première représentation mais peu robustes aux variations de luminosité, échelle, occlusion, rotation, etc. Pour remédier à ces

limitations, l'utilisation des stratégies de représentation des caractéristiques joue un rôle crucial et très important dans l'amélioration des descripteurs obtenus. Plusieurs recherches ont montré que la performance des caractéristiques de bas niveaux s'est améliorée lorsqu'on utilise des stratégies de représentation de caractéristiques pour donner des caractéristiques plus représentatives et plus discriminatoires. Ces caractéristiques sont invariantes à la translation, à l'échelle, à la rotation, etc. [31].

Contrairement à la méthode basée sur les caractéristiques de bas niveaux, les méthodes basées sur les caractéristiques de niveau intermédiaire ou de mid- niveau construisent une représentation d'image en codant des descripteurs locaux de bas niveau comme sur la figure II.1. Dans ce cas, ces méthodes utilisent des techniques d'apprentissage non supervisées (par exemple K-Means) qui visent à apprendre un ensemble des caractéristiques de base niveaux, par exemple SIFT, SURF, etc., pour obtenir une représentation globale des images [26]. Les méthodes de représentation des caractéristiques les plus utilisées dans la littérature sont :

- La représentation par sac de mots visuels (BOW) ;
- Le vecteur de descripteurs agrégés localement (VLAD) ;
- Le vecteur de Fisher (FV) ;
- Et le codage linéaire à contraintes de localité (LLC).

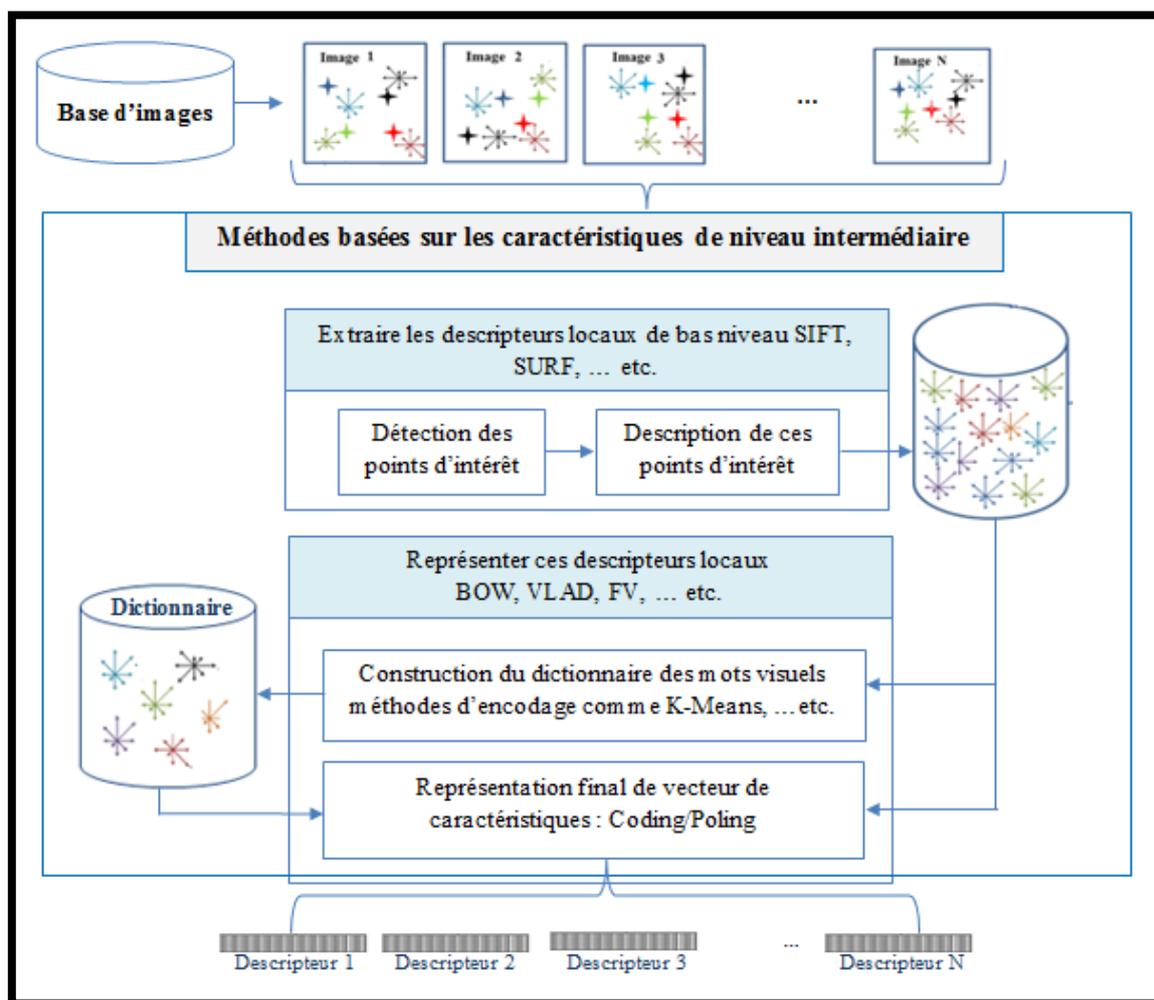


Figure II.1 : Méthodes basées sur les caractéristiques de niveau intermédiaire.

3.1 BOW

La représentation par sac de mots visuels (en anglais *bag of visual words*) est un modèle très populaire pour représenter les images dans la littérature de la vision par ordinateur. BOW décrit généralement une image en utilisant un ensemble de mots visuels appelé vocabulaire visuel ou dictionnaire de mots. Les mots visuels sont souvent produits en regroupant des descripteurs d'image locaux pour former un dictionnaire. Toutes les images sont finalement représentées par un histogramme basé à ce dictionnaire [31].

Le processus de construction l'histogramme de BOW est généralement assuré par trois étapes : l'extraction des caractéristiques visuelles, la construction d'un dictionnaire de mots visuels et la représentation du BOW descripteur. Premièrement, des points d'intérêts sont détectés et décrits à partir des approches proposées pour extraire des caractéristiques de bas niveaux locales. Ensuite, les mots visuels de dictionnaire peuvent être générés en regroupant ces descripteurs locaux de toute l'image à l'aide de méthodes d'encodage de caractéristiques. De nombreuses méthodes d'encodage de caractéristiques ont été rapportées pour la construction d'un dictionnaire de mots visuel, comme la méthode K-moyennes (*K-Means*). Une fois le dictionnaire des mots visuels est construit, les images sont représentées par un histogramme où les cases sont des mots visuels et les poids sont leurs occurrences dans l'image [9].

3.2 VLAD

La méthode VLAD (en anglais *Vector of Locally Aggregated Descriptor*) est une autre approche qui vise à représenter l'image à partir des descripteurs locaux extraits par les méthodes de bas niveaux. Elle consiste à représenter l'image par l'agrégation des sommes des différences entre les descripteurs locaux de celle-ci et les mots visuels de dictionnaire extrait. Comme la méthode BOW, cette méthode construit aussi le dictionnaire en utilisant des descripteurs locaux. Premièrement les points d'intérêts sont détectés et caractérisés. Une méthode non supervisée apprend ces descripteurs pour générer les mots visuels (centre de chaque cluster). Pour cela, l'algorithme K-moyennes est sûrement la solution la plus évidente. Contrairement à la méthode BOW, la méthode VLAD consiste non pas à créer un histogramme mais à stocker les écarts entre les descripteurs et les centres représentant les mots visuels. Lors du codage de chaque descripteur vers son voisin le plus proche dans le dictionnaire, les différences entre eux dans chaque dimension sont cumulées, ce qui donne un vecteur de caractéristiques d'image de dimension $K \times F$, telle que K est le nombre de mots dans dictionnaire et F est la dimension de chaque descripteur local, (128 dans le cas de descripteur SIFT) [33].

3.3 FV

Le vecteur de Fisher (en anglais *The Fisher Vector*) utilise un cadre puissant adopté dans le contexte de la classification appelé noyau de Fisher ou Fisher Kernel comme un outil efficace pour coder les descripteurs locaux d'images en une représentation vectorielle de taille fixe. L'idée de base de cette méthode est la modélisation de l'espace de description et ensuite la déviation du modèle propre à une image par rapport à ce modèle général. Le modèle gaussien de mélange ou (The Gaussian Mixture Model ou GMM) est généralement utilisé

comme un modèle génératif et pourrait être compris comme un "vocabulaire visuel probabiliste". Cependant que BOW compte les occurrences de mots visuels et ne prend donc en compte que des statistiques d'ordre 0, alors que FV offre une représentation plus complète en codant des statistiques d'ordre supérieur "premier et éventuellement deuxième ordre" liées à la distribution des descripteurs. FV permet également une représentation plus efficace, car il faut moins de mots visuels pour atteindre une performance donnée. Cependant, la représentation vectorielle obtenue à l'aide de BOW est généralement assez rare, alors que celle obtenue à l'aide du noyau de Fisher est presque dense. Cela pose des problèmes de stockage et d'entrée/sortie qui ont été résolus à l'aide de techniques de réduction de dimension [33].

3.4 LLC

Le codage linéaire à contraintes de localité (en anglais *Locality-constrained Linear Coding*) est une généralisation de la méthode BOW. La méthode LLC est un schéma de codage efficace qui est adapté des méthodes utilisées en codage. Elle utilise les contraintes de localité pour coder chaque descripteur local dans son système de coordonnées local en modifiant les contraintes de parcelle. La dernière caractéristique peut être générée en regroupant au maximum les coordonnées projetées avec la même taille de dictionnaire. La méthode LLC décompose le calcul du descripteur en deux étapes : la première étape est le codage (ou Coding) qui consiste à coder chaque descripteur du sac en appliquant généralement un traitement non-linéaire sur les descripteurs pour les projeter dans un espace de plus grande dimension, et la deuxième étape est l'agrégation (ou Pooling) qui consiste à rassembler l'ensemble des descripteurs codés en un unique descripteur. Cette approche donne de bonnes performances en classification [33].

4. Caractéristiques de plus hauts niveaux

Les caractéristiques de plus hauts niveaux sont des caractéristiques basées sur les caractéristiques d'apprentissage en profondeur. Durant ces dernières années, plusieurs travaux dans la littérature ont montré que la performance de ces caractéristiques est bien meilleure que les caractéristiques de bas niveaux et aussi de niveaux intermédiaires, car ces deux derniers requièrent un effort considérable de compétences en ingénierie et d'expertise de domaine. Cela rend le processus d'extraction des caractéristiques lourd, complexe, très peu fiable et dépend de l'expérience des experts de domaine telle qu'il doit être, et détermine des descripteurs plus représentatifs et plus discriminatoires pour un problème donné. Les chercheurs ont visé de remplacer ces caractéristiques créées manuellement par des caractéristiques automatiquement apprises à partir de contenu de l'image (image raw) à l'aide d'une technique d'apprentissage en profondeur via des réseaux neuronaux à architecture profonde [26].

Les modèles d'apprentissage en profondeur composés de plusieurs couches de traitement peuvent apprendre des représentations plus puissantes des données avec plusieurs niveaux d'abstraction. En outre, les méthodes d'apprentissage des caractéristiques approfondies se sont également révélées très utiles pour découvrir des structures complexes et des informations discriminantes cachées dans des données de grande dimension. Les caractéristiques des couches supérieures du réseau neuronal profond présentent des propriétés

sémantiques d'abstraction. Il existe actuellement un certain nombre de modèles d'apprentissage en profondeur, qui ont démontré une capacité de représentation des caractéristiques robustes et efficaces dans nombreux applications tels que les réseaux de croyances profondes "DBN", les machines de Boltzmann profondes "DBM", l'auto-codeur superposé "SAE", les réseaux de neurones à convolution "CNN", etc., [26]. Ces méthodes seront vues en détails ultérieurement dans le chapitre suivant.

5. Classification

Le module de classification consiste à utiliser une méthode de classification pour déterminer à quelle classe appartient une entrée de données basée sur les descripteurs de la phase précédente. Elle se divise en deux sous phases : la phase d'apprentissage et la phase de test. La phase d'apprentissage consiste à initialiser la base des modèles qui est composée des descripteurs et des classes correspondantes. La phase de test ou de décision consiste à assigner et à attribuer une classe pour chaque nouvel exemple donné dont la classe est inconnue. Ces techniques se distinguent par leur caractère supervisé et non supervisé selon l'information disponible a priori sur les observations "descripteur". La classification est dite supervisée lorsque les différentes classes des observations sont connues au préalable. Ce type de classification est préférable et plus utilisée pour la classification des images. Cependant, dans le cas où l'on n'a aucune connaissance a priori sur l'appartenance des observations aux classes, on parle de classification non supervisée ou de "clustering" [23].

Pour faire la classification, il existe plusieurs méthodes :

- La machines à vecteurs de support ou SVM ;
- La méthode des k-plus proches voisins ou KNN ;
- L'arbre de décision ou DT ;
- Les réseaux de neurones artificiels ou ANN ;
- Les réseaux bayésiens ou BN ;
- Sélection et adaptation du vecteur de support ou SVSA ;
- etc.

5.1 SVM

L'algorithme des machines à vecteurs de support ou SVM (pour l'anglais *Support Vector Machine*) est considéré comme l'une des méthodes les plus importantes et les plus pratiques pour résoudre les problèmes liés à la classification, à l'apprentissage et à la prédiction. SVM est une méthode d'apprentissage automatique basée sur la théorie de l'apprentissage statistique qui utilise la séparation linéaire. Cette méthode a été très activement utilisée avec succès et a souvent donné les meilleurs résultats dans beaucoup d'applications [34].

La SVM a été développé dans les années 90 par Vapnik [36]. Initialement il est développé comme un algorithme de classification binaire supervisée. L'algorithme sous sa forme initiale revient à chercher une frontière de décision linéaire entre deux classes, mais ce modèle peut considérablement être enrichi en se projetant dans un autre espace permettant d'augmenter la séparabilité des données dans le cas non linéairement séparable. L'originalité

principale de ces méthodes consiste à utiliser efficacement les descripteurs étiquetés pour produire une fonction de décision qui permet de maximiser la marge entre deux classes de données. Soit un problème caractérisé par un ensemble de N observations : $\{(x_i, y_i)\}_{i=1\dots n}$, où le descripteur x_i est étiqueté par le class y_i , dans le cas de classification binaire $y_i = \{-1, +1\}$, et dans un modèle linéaire, on a $f(x) = w \cdot x + b$ où :

$$y = \text{signe}(w \cdot x + b), \text{ qui signifie : } \begin{cases} \text{si } w \cdot x + b > 0 \text{ alors } y = +1 \\ \text{si } w \cdot x + b < 0 \text{ alors } y = -1 \end{cases} \quad (2.5)$$

La marge est la distance minimale entre les échantillons de l'ensemble d'apprentissage et la frontière de décision. Les vecteurs de support, qui donnent son nom au modèle, sont les points de données les plus proches de la surface de décision et déterminent donc la marge. L'hyperplan, la marge et les vecteurs de support sont illustrés dans la figure II.2. Il est évident qu'il existe une multitude d'hyperplans valides mais la propriété remarquable de SVM est que cet hyperplan doit être optimal. Dans ce cas, parmi les hyperplans valides, celui qui passe au milieu des points des deux classes d'exemples est appelé hyperplan optimal.

Parmi les modèles de SVM, on trouve les cas linéairement séparables et les cas non linéairement séparables. Les premiers cas sont les plus simples de SVM car ils permettent de trouver facilement le classificateur linéaire, où tous les points associés aux données peuvent être séparés correctement par une frontière linéaire. L'hyperplan séparateur ou frontière de décision est donc illustré par l'équation $w \cdot x + b = 0$. Le cas non linéairement séparable est le cas où tous les points associés aux données peuvent être séparés correctement par des frontières non linéaires. Pour surmonter les inconvénients des cas non linéairement séparables, l'idée de SVM est de changer l'espace des données. Dans ce cas, il se trouve que l'espace ne suffit pas pour classer les entrées alors on passe donc dans un espace de grande dimension. La transformation non linéaire des données peut permettre une séparation linéaire des exemples dans un nouvel espace.

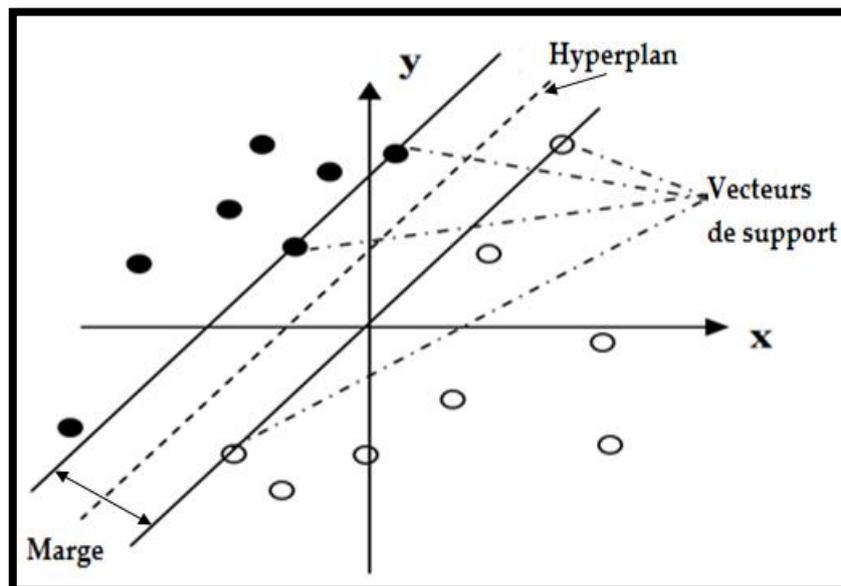


Figure II.2 : Illustration de l'hyperplan séparant les points de données "descripteurs" de deux classes. Les vecteurs de support et la marge sont marqués.

On va donc avoir un changement de dimension et on peut alors appliquer le même algorithme dans ce nouvel espace, ce qui se traduit par une frontière de décision non linéaire dans l'espace initial. Pratiquement, quelques noyaux traduisent la répartition des exemples dans cet espace à un espace de grande dimension. On peut citer : linéaire, polynomial, gaussien et laplacien [35]. Le principal avantage de SVM réside dans sa capacité à traiter une grande variété de problèmes de classification, notamment des problèmes de grande dimension et non séparables linéairement. Les fondements mathématiques de SVM sont solides. Dans ce cas, la décision de SVM est rapide, car les exemples de tests sont non comparés pas avec tous les exemples d'apprentissage, mais juste avec les supports vecteurs. Mais, ils ont un temps de calcul élevé lors d'une régularisation des paramètres de la fonction noyau et ils nécessitent un certain nombre de paramètres clés pour être correctement défini pour obtenir d'excellents résultats de classification [36].

5.2 KNN

L'algorithme de k plus proches voisins ou KNN (pour l'anglais *K-Nearest-Neighbors*) est une méthode de classification supervisée et non paramétrique. KNN est largement utilisé en classification d'une manière générale et en segmentation d'images en particulier. Cette méthode de décision est liée à la notion de "proximité" ou "ressemblance". Elle est une méthode d'apprentissage à base d'instances telle qu'elle ne nécessite pas de phase d'apprentissage. L'idée de cette technique repose sur un principe simple et intuitif de regroupement d'individus en fonction de leur voisinage. Le voisin le plus proche est mesuré par rapport à la valeur k qui définit le nombre de voisins les plus proches devant être examinés pour décrire la classe d'un nouveau point de données. La forme acoustique à classer est comparée à d'autres déjà classées, et on lui affecte la classe la plus représentée parmi les k plus proches. Cette notion de voisinage est quantifiée par une mesure de similarité. La mesure de similarité la plus utilisée est la distance euclidienne. Les avantages de KNN sont la simplicité et la transparence, des données d'entraînement robustes à bruyantes, faciles à comprendre et à implémenter. Mais les inconvénients de KNN sont la complexité des calculs, la limitation de la mémoire, des performances d'exécution médiocres pour de grands ensembles d'apprentissage et des attributs non pertinents peuvent poser problème [35].

5.3 DT

Les algorithmes d'arbre de décision ou DT (pour l'anglais *Decision trees*) sont les algorithmes les plus couramment utilisés dans la classification. DT est un mécanisme transparent qui permet aux utilisateurs de suivre facilement une arborescence pour voir comment la décision est prise. L'objectif principal de DT est de produire un modèle qui calcule la valeur d'une variable demandée en fonction de nombreuses variables d'entrée [35].

Les DTs sont des arbres qui classifient les instances en les triant en fonction des valeurs des entités. Chaque nœud d'un DT représente un descripteur dans une instance à classer et chaque branche représente une valeur que le nœud peut prendre. Les instances sont classées à partir du nœud racine et triées en fonction de leurs descripteurs. Les différentes décisions possibles sont situées aux feuilles de l'arbre, et sont atteints en fonction de décisions prises à chaque étape [34]. DT a l'avantage d'être lisible et rapide à exécuter telle qu'il fournit une technique de modélisation facilement compréhensible et simplifie également le processus de

classification. Cependant, il ne convient pas pour un petit ensemble de données, un temps de traitement élevé, pas de recherche de suivi en arrière, incapable de gérer les valeurs manquantes et pas d'optimisation globale [35].

5.4 ANN

Les réseaux de neurones artificiels ou ANN (pour l'anglais *Artificial Neural Network*) sont les algorithmes les plus couramment utilisés pour résoudre des problèmes de classification, de prédiction, de reconnaissance des formes et d'optimisation. Depuis une dizaine d'années, l'utilisation de ces méthodes s'est développée dans de nombreux domaines comme les sciences économiques, l'écologie et l'environnement, biologie, médecine, etc. La conception d'ANN est à l'origine schématiquement inspirée du fonctionnement des neurones biologiques. Comme le cerveau humain, Ces réseaux peuvent apprendre par expérience. ANN dépend de trois aspects fondamentaux : les entrées, les fonctions d'activation de l'unité et l'architecture du réseau et le poids de chaque connexion d'entrée. Étant donné que les deux premiers aspects sont fixes, le comportement de l'ANN est défini par les valeurs actuelles des poids. Les poids du réseau à entraîner sont initialement définis sur des valeurs aléatoires, puis des instances de l'ensemble d'apprentissage sont exposées de manière répétée au réseau. Les valeurs pour l'entrée d'une instance sont placées sur les unités d'entrée et la sortie du réseau est comparée à la sortie souhaitée pour cette instance. Ensuite, tous les poids du filet sont légèrement ajustés dans la direction qui rapprocherait les valeurs de sortie du filet des valeurs de la sortie souhaitée. Les avantages des méthodes neuronales sont : puissance d'approximation, robustesse pour des tâches difficiles et parallélisme dans le traitement des données. Cependant, les réseaux de neurones ont des problèmes et des limites propres à eux. En plus de temps d'apprentissage qui est lent, l'inconvénient majeur des méthodes neuronales réside dans la détermination de la topologie du réseau de neurones à utiliser : nombre de couches cachées, nombre de neurones dans chaque couche cachée, la stratégie de connectivité, etc. [34].

5.5 BN

Les réseaux bayésiens (ou BN pour l'anglais *Bayesian Network*) sont le représentant le plus connus des algorithmes d'apprentissage statistique. Ils ont été développés au début des 1980 pour essayer de résoudre certains problèmes d'abduction et de prédiction, courants en intelligence artificielle [36]. Un réseau bayésien est un modèle probabiliste graphique entre un ensemble de variables (caractéristiques) permettant de calculer des probabilités conditionnelles apportant des solutions à différentes sortes de problématiques, notamment, diagnostic médical, diagnostic industriel, analyse de risques, détection de spams, exploitation du retour d'expérience, modélisation et simulation de systèmes complexes, etc. [36].

La structure de BN est simple, est un graphe acyclique dirigé (ou DAG pour l'anglais *Directed Acyclic Graph*) qui ne contient pas de boucle, dans lequel les nœuds représentent des variables aléatoires, et les arcs reliant ces dernières sont rattachées à des probabilités conditionnelles, tandis que l'absence d'arcs code les indépendances conditionnelles. De plus, une caractéristique "nœud" est conditionnellement indépendant de ses descendants étant donné ses parents. L'intérêt particulier de ces types de réseaux est de tenir compte simultanément de connaissances a priori d'experts "dans le graphe" et de

l'expérience contenue dans les données. En règle générale, l'apprentissage de celle-ci peut être divisé en deux sous-tâches : d'abord, l'apprentissage de la structure DAG du réseau, puis la détermination de ses paramètres. Les paramètres probabilistes sont codés dans un ensemble de tables, une pour chaque variable, sous la forme de distributions conditionnelles locales d'une variable, compte tenu de ses parents. Cette méthode de nature statistique, suppose que l'échantillon à analyser est composé de N observations tirées, $x = \{x_1, x_2, \dots, x_N\}$ réparties en M classes (C_1, C_2, \dots, C_M) avec leur probabilité a priori $P(C_i)$, probabilité d'avoir la classe C_i , calculée à partir des fréquences d'occurrence des exemples de cette classe. Par application du théorème de Bayes, il est possible de déterminer la probabilité pour qu'une observation x_i appartienne à la classe C_i . Cette probabilité est appelée aussi probabilité a posteriori de la classe C_i . Le rôle de la théorie Bayésienne est de fournir une fonction de décision qui minimise le coût moyen d'erreur par décision prise. Pour une probabilité d'erreur minimum, il suffit de construire un système qui à chaque x associe la classe dont la densité en x est la plus forte, elle nécessite donc la connaissance de la probabilité conditionnelle d'appartenance à une classe C_i donnée, soit $P(x/C_i)$ [35].

La caractéristique la plus intéressante des BN, comparée aux arbres de décision ou aux réseaux de neurones, est très certainement la possibilité de prendre en compte des informations préalables sur un problème donné, en termes de relations structurelles entre ses caractéristiques. L'un des problèmes des classificateurs de BN est qu'il nécessite généralement la discrétion d'attributs continus. Le processus de conversion d'attribut continu en attribut discret a introduit des problèmes de classification. Un autre problème des classificateurs de BN est qu'ils ne conviennent pas aux ensembles de données comportant de nombreuses caractéristiques [34].

Les avantages de BN sont : des propriétés de lissage, des modifications mineures du modèle de réseau bayésien n'influencent pas le fonctionnement du système, applicabilité flexible et le traitement des données manquantes. BN a la capacité de compléter les données manquantes en assimilant toutes les opportunités des valeurs manquantes [35].

5.6 SVSA

L'algorithme de sélection et adaptation de vecteur de support (en anglais *support vector selection and adaptation*) est une nouvelle méthode de classification supervisée qui s'applique aux données linéairement ou non linéairement séparables en termes de certains vecteurs de référence générés par le traitement des vecteurs de support obtenus à partir du SVM linéaire. Il présente une bonne performance dans la classification des données linéairement séparables et également non linéairement séparables. La méthode comprend deux étapes appelées sélection et adaptation. Dans ces deux étapes, une fois que les vecteurs de support sont obtenus par un SVM linéaire, certains d'entre eux sont rejetés et d'autres sont sélectionnés et adaptés pour devenir les vecteurs de référence. La classification est ensuite effectuée à l'aide de la méthode KNN avec les vecteurs de références. À l'étape de la sélection, tous les vecteurs de support sont classés en ce qui concerne l'ensemble de données d'apprentissage basé sur la distance euclidienne minimale. Les vecteurs de support mal classés sont ignorés de l'ensemble des vecteurs de support et les vecteurs restants sont appelés vecteurs de référence. Dans la phase d'adaptation, les vecteurs de références sont mis à jour à

l'aide de l'algorithme d'apprentissage de la quantification vectorielle. À la fin de la phase d'adaptation, les vecteurs de référence adaptés sont utilisés pour la classification des ensembles d'apprentissage et de test. À cette fin, la méthode KNN est appliquée pour classer les échantillons par rapport aux vecteurs de référence. Les distances euclidiennes entre le vecteur d'entrée et les vecteurs de référence sont calculés et la classification est effectuée sur la base de la classe de majorité des vecteurs de référence K les plus proches. L'algorithme est conçu pour la classification binaire. Pour classer un jeu de données multi-classes, l'algorithme peut être généralisé en utilisant l'approche un-contre-un [37].

6. Conclusion

Nous avons présenté, tout au long de ce chapitre, l'état de l'art sur les algorithmes d'apprentissage automatique qui fonctionnent par deux phases complémentaires, l'extraction des caractéristiques et la classification. En plus, nous avons abordé les deux différentes phases d'apprentissage automatique. Où nous avons vu les différentes méthodes et les techniques connus dans la boîte à outils de la littérature pour chaque phase, la phase d'extraction des caractéristiques et ensuite la phase de classification. Nous avons expliqué les méthodes d'extraction des caractéristiques basées sur les caractéristiques de bas niveaux et sur les caractéristiques de niveaux intermédiaire.

Dans le prochain chapitre, nous allons présenter en détail les différentes méthodes basées sur les caractéristiques de plus hauts niveaux, qui sont des caractéristiques basées sur l'apprentissage en profondeur. En plus, nous allons présenter un état de l'art sur les algorithmes d'apprentissage en profondeur.

Chapitre III

Apprentissage en profondeur "Deep learning"

À travers ce chapitre, nous présenterons un état de l'art sur l'apprentissage en profondeur. En premier temps, nous allons expliquer le passage de l'apprentissage automatique à l'apprentissage en profondeur. Puis, nous allons présenter les applications et les différentes approches de l'apprentissage en profondeur. En plus, nous allons expliquer l'architecture de réseau CNN et ses différents modèles en détail.

1. Introduction

Dans le chapitre précédent, nous avons vu les méthodes d'apprentissage automatiques qui se fonctionnent en deux phases complémentaires, l'extraction des caractéristiques et la classification. Nous avons aussi présenté les approches de chaque phase.

Ce chapitre présentera un état de l'art sur l'apprentissage en profondeur, où nous allons expliquer le passage de l'apprentissage automatique à l'apprentissage en profondeur, puis nous allons présenter les applications et les différentes approches de l'apprentissage en profondeur. En plus, nous allons expliquer l'architecture de réseau CNN et ses différents modèles en détail.

2. Apprentissage en profondeur

L'apprentissage en profondeur ou l'apprentissage profond, plus précisément "apprentissage approfondi", en anglais "*Deep learning* ou DL, également appelée *Deep structured learning et Hierarchical learning*", est un domaine de recherche émergent au sein de la communauté de l'apprentissage automatique et de l'intelligence artificielle. Au cours des dernières années, l'apprentissage en profondeur a attiré beaucoup d'attention en raison de ses grandes performances dans divers domaines. DL est un ensemble des méthodes qui apprennent automatiquement les représentations hiérarchiques dans les architectures profondes en vue de la classification. L'objectif est de découvrir des caractéristiques plus abstraites dans les niveaux les plus élevés de la représentation, en utilisant des réseaux de neurones qui séparent facilement les divers facteurs explicatifs des données [38].

2.1 La relation entre ML et DL

L'intelligence artificielle ou AI "en anglais (*Artificial intelligence*)", vise à rendre la machine intelligente en simulant l'intelligence humaine pour résoudre des problèmes à forte complexité logique ou algorithmique. L'apprentissage automatique ou ML, est une branche majeure de l'intelligence artificielle. Elle se définit comme la capacité d'une machine à apprendre la prise de décision à partir d'observations et vise à utiliser l'intelligence artificielle à travers l'apprentissage des données. Dans la ML, un programme analyse un ensemble de données afin d'obtenir des modèles qui permettront de tirer des conclusions sur de nouvelles données. D'autre part DL, à son tour, est un sous-ensemble de ML qui peut résoudre le problème en utilisant des réseaux de neurones, "en anglais (*neural networks* ou NN)", est une nouvelle frontière de l'AI où les caractéristiques les plus représentatives et les plus discriminantes sont apprises de manière hiérarchique comme sur la figure III.1 [39].

La plupart des algorithmes de MLs sont conçus à des descripteurs et qui sont associées à des classificateurs dans une structure et une architecture peu profonde, comme SVM, ANN, DT, KNN, etc. [40]. En outre, ces descripteurs sont extraites à partir d'une méthode qui est déterminée par un expert, "en anglais (*hand-engineering features*)", pour extraire des caractéristiques basées sur le contenu spatial, spectral, texturé, morphologique, etc. [40].

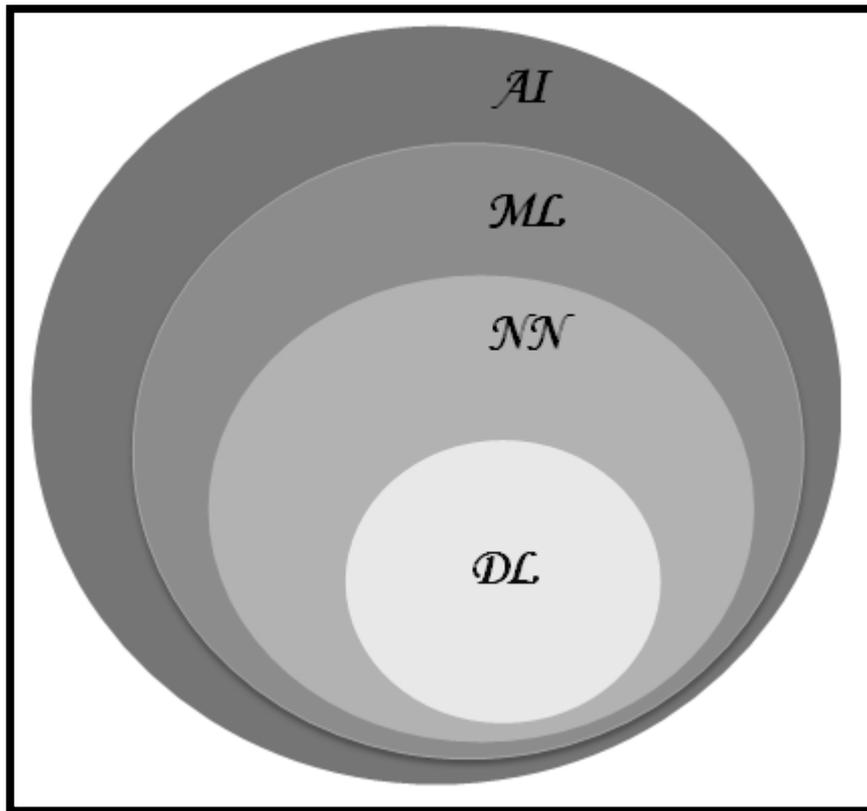


Figure III.1 : Relation entre l'intelligence artificielle, l'apprentissage automatique et l'apprentissage en profondeur.

Comme décrit en détail dans le chapitre précédent, les algorithmes de ML impliquent généralement deux étapes distinctes, mais complémentaires, pour l'extraction des caractéristiques puis les classifier. L'extraction de caractéristiques est mise en œuvre par des opérateurs spécifiques sur l'image entière ou des parties locales de l'image pour transformer l'espace de caractéristiques spectrales d'origine en représentations abstraites pouvant être facilement séparées par un classificateur. Les performances de tout classificateur utilisé sont fortement affectées par les transformations utilisées et les caractéristiques spatiales qu'en obtient. Malgré que les algorithmes de ML sont performantes, mais l'étape d'extraction des caractéristiques est difficile, prend du temps et nécessite une expertise du domaine. Par contre des algorithmes de ML, dans les algorithmes de DL, l'extraction des caractéristiques est réalisée automatiquement par l'algorithme elle-même. La figure III.2 illustre l'extraction des caractéristiques dans la ML et DL. L'apprentissage profond est basé sur ce qui a été appelé des réseaux de neurones artificiels, composés de milliers d'unités, les "neurones", qui effectuent chacune de petites opérations simples. Les résultats d'une première couche de "neurones" servent d'entrée aux calculs d'une deuxième couche et ainsi de suite [40].

2.2 Réseau de neurones "Neural Networks "

L'entraînement des machines à penser comme des humains passe en partie par l'utilisation de réseaux de neurones. Les réseaux de neurones ou NN sont une série d'algorithmes modélisés selon le comportement du cerveau humain. Comme le cerveau peut reconnaître des modèles et nous aider à classer des informations, les réseaux de neurones font de même pour les machines.

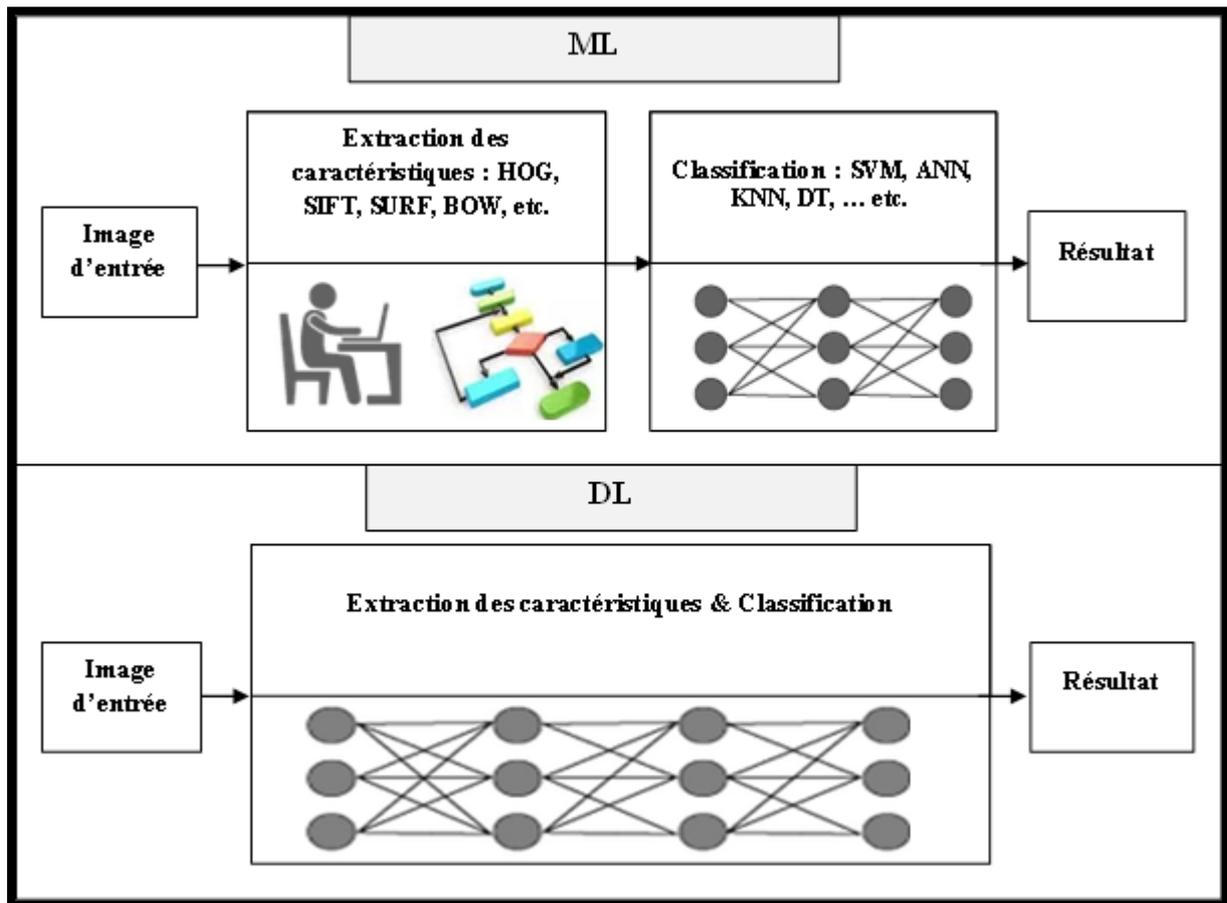


Figure III. 2 : L'extraction des caractéristiques dans la ML et DL.

Le cerveau essaie constamment de comprendre les informations qu'il traite, et pour ce faire, il assigne des éléments à des catégories. Lorsque nous rencontrons de nouvelle chose, nous essayons de le comparer à un élément connu pour nous aider à le comprendre et à lui donner un sens. Les réseaux de neurones font la même chose pour les machines. Les neurones artificiels s'inspirent des neurones biologiques et tentent de formuler le fonctionnement de neurones biologiques sous une forme informatique. Un neurone artificiel à un nombre fini d'entrées avec des poids qui leur sont associés et une fonction d'activation. La sortie du neurone est le résultat de la fonction d'activation appliquée à la somme pondérée des entrées. La figure III.3 est un exemple de neurone artificiel. Les neurones artificiels sont connectés les uns aux autres pour former des réseaux de neurones artificiels "ANN". Le grand potentiel des ANN est le traitement à très rapide fourni dans une vaste mise en œuvre parallèle. Ces réseaux sont composés généralement de trois types de couches : couche d'entrée ou "input layer", couche cachée ou "hidden layer" et couche de sortie ou "output layer".

Les neurones de la couche d'entrée correspondent aux caractéristiques des données d'entrée qui sont représentées par un vecteur de caractéristique. Les résultats de la tâche assignée au réseau sont représentés par les neurones de couche de sortie. Les couches intermédiaires entre l'entrée et la sortie sont les couches cachées. Dans ces réseaux, chaque couche utilisant la sortie de la couche précédente comme entrée, où les données sont transférées de la couche d'entrée aux nœuds de la couche cachée et aux nœuds de sortie.

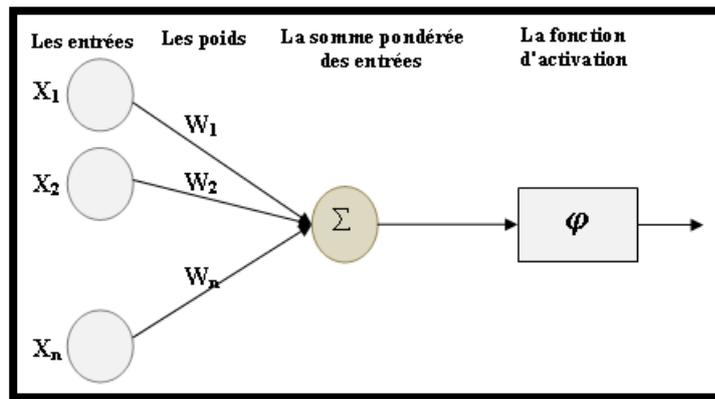


Figure III. 3 : Exemple de neurone artificiel.

Pour les réseaux de neurones réguliers, le type de couche le plus courant est la couche complètement connectée dans laquelle les neurones situés entre deux couches adjacentes sont complètement connectés par paires, mais les neurones d'une seule couche ne partagent aucune connexion. Vous trouverez dans la figure III.4 deux exemples d'architectures de réseau neuronal utilisant une pile de couches complètement connectées [39].

Les fonctions d'activation, sont également appelées les fonctions de transfert, transforment la somme pondérée des entrées qui entrent dans les neurones artificiels. Ces fonctions doivent être non linéaires pour coder des modèles complexes de données. Les fonctions d'activation les plus populaires sont [41] :

- Sigmoid ;
- Tanh ;
- Et ReLU.

2.2.1 Sigmoid

Sigmoid est également appelée fonction logistique, est une fonction d'activation non-linéaire utilisée pour représenter une distribution de probabilité sur une variable binaire. L'équation mathématique et le courbe de sigmoïde sont illustrées dans la figure III.5 à gauche.

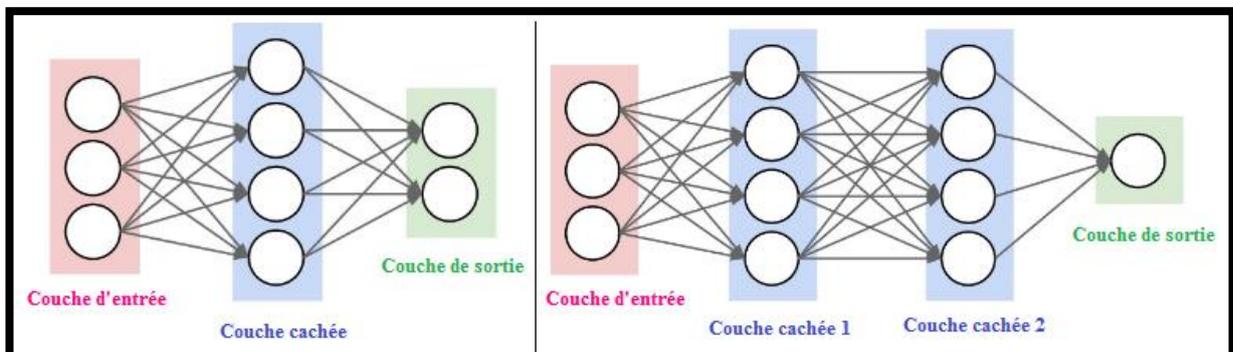


Figure III.4 : Gauche : un réseau de neurones à 2 couches (une couche cachée de 4 neurones (ou unités) et une couche de sortie avec 2 neurones) et trois entrées. Droite : Un réseau de neurones à 3 couches avec trois entrées, deux couches cachées de 4 neurones chacune et une couche de sortie. Notez que dans les deux cas, il existe des connexions (synapses) entre les neurones d'une couche à l'autre, mais pas à l'intérieur d'une couche.

La fonction sigmoïde prend un nombre réel et le réduit dans l'intervalle entre 0 et 1. La beauté d'un exposant réside dans le fait que la valeur n'atteint jamais 0 ni ne dépasse 1 dans ces équations, en particulier, les grands nombres négatifs deviennent 0 et les grands nombres positifs deviennent 1. Il est facile à utiliser et il présente toutes les propriétés intéressantes des fonctions d'activation, notamment, non linéarité, continuellement différentiable : cela signifie que nous pouvons trouver la pente de la courbe sigmoïde en deux points quelconques, qui possède une plage de sortie fixe. Mais, il soulève un problème de gradients de disparition, car les valeurs de fonction ont tendance à réagir beaucoup moins aux changements des valeurs d'entrées. Aussi, sa sortie n'est pas centrée à zéro. Cela fait que les mises à jour du gradient vont trop loin dans des directions différentes, $0 < \text{sortie} < 1$, ce qui rend l'optimisation plus difficile. Elles saturent, éliminent les gradients et ont une convergence lente. Car elle comprend la fonction exponentielle, elles sont coûteuse en terme de calcul [41].

2.2.2 Tanh

Tanh est aussi une fonction non linéaire et également connue sous le nom de "tangente hyperbolique". L'équation mathématique et le courbe de Tanh sont illustrées dans la figure III.5 au milieu. Cette fonction est similaire à Sigmoid, mais la fonction Tanh produit un résultat dans l'intervalle $[-1, 1]$. En terme général, la fonction Tanh est préférable à la fonction Sigmoid car elle est centrée sur zéro. L'optimisation est plus facile et le gradient est plus fort pour le Tanh que pour le Sigmoid, où les dérivés sont plus raides. Il souffre également de problème de gradient disparaissant tel qu'il sature et tue les gradients [41].

2.2.3 ReLU

Unité de Rectification Linéaire ou ReLU, "en anglais (*Rectified Linear Unit*)", est la fonction d'activation la plus utilisée. Elle est devenue très populaire ces dernières années, car il a été constaté qu'elle accélérât considérablement la convergence de la descente de gradient stochastique par rapport aux fonctions Sigmoid et Tanh en raison de sa forme linéaire non saturante. En d'autres termes, l'activation est simplement limitée à zéro. Cette fonction permet de ne transmettre que les valeurs maximales lors de la propagation frontale. Comme indiqué dans la figure III.5 à droite, si l'entrée est négative la sortie est 0 et si elle est positive alors la sortie est elle-même.

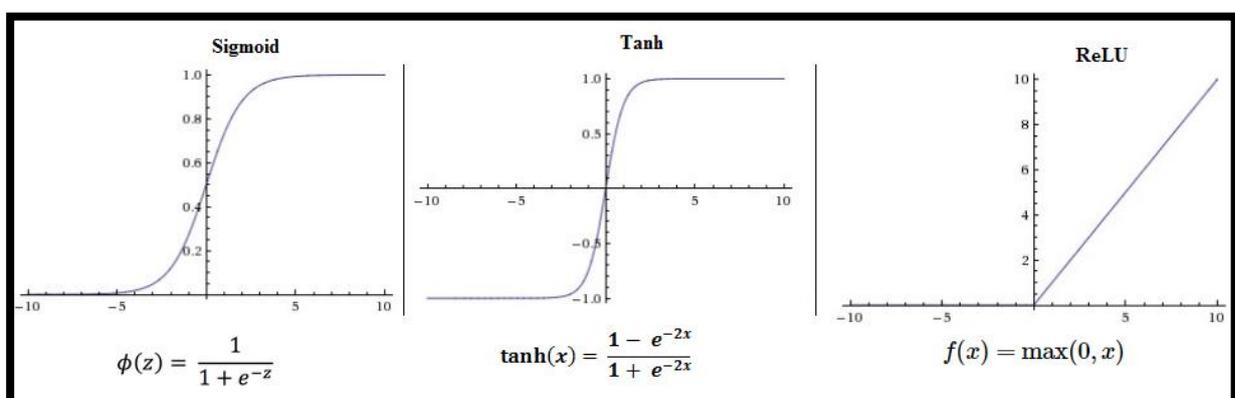


Figure III.5 : Les fonctions d'activation.

En fait, il ne souffre pas de la disparition du gradient ni de son explosion. Comparé aux neurones Tanh et Sigmoides qui impliquent des opérations coûteuses et exponentielles, le ReLU peut être mis en œuvre simplement en effectuant un seuil à zéro de la matrice d'activations. C'est efficace et facile pour le calcul, en raison de sa fonctionnalité ; il n'active pas tous les neurones en même temps. Mais cette fonction n'est pas parfaite. Si la valeur d'entrée est négative, le neurone reste inactif, ainsi les poids ne sont pas mis à jour et le réseau n'apprend pas. En d'autres termes, pour les activations dans la région " $x < 0$ " de ReLU, le gradient sera 0, à cause duquel les poids ne seront pas ajustés pendant la descente. Cela signifie que les neurones qui entrent dans cet état cesseront de répondre aux variations d'erreur/entrée "simplement parce que le gradient est égal à 0, rien ne change". C'est ce qu'on appelle le problème de la mort ReLU [41].

2.3 L'architecture de DL

L'apprentissage en profondeur ou DL est simplement appelé un réseaux de neurones profonds, en anglais "*Deep Neural Networks*" ou DNN, est une extension d'ANN avec plusieurs couches cachées entre les couches d'entrée et de sortie, d'où le terme profond, pour pouvoir apprendre un niveau d'abstraction plus élevé. Comme les architectures d'ANN, les architectures de DNN sont aussi composés de trois types de couches, une couche d'entrée, des couches cachées et une couche de sortie. Cependant, dans les cas d'ANN, Les neurones de la couche d'entrée sont les caractéristiques des données d'entrée, par contre, les neurones de couche d'entrée de DNN représentent aux données brutes ou "Raw data" en anglais. Le nombre de couches cachées définit la profondeur de l'architecture, où se trouve la bonne manipulation mathématique pour transformer l'entrée en sortie. Les données sont exprimées sous forme de caractéristiques représentées en couches cachées de l'architecture DNN. Un réseaux de neurones profonds peut fournir une représentation multi-niveaux des données comme sur la figure III.6. En d'autres termes, l'architecture DNN peut apprendre des caractéristiques de bas niveaux dans les premières couches cachées, puis les couches intermédiaires représentent les caractéristiques de niveaux intermédiaires et un niveau d'abstraction plus élevé dans les dernier couches. Les architectures profondes incluent de nombreuses variantes. Chaque architecture a rencontré le succès dans des domaines spécifiques. L'apprentissage du DNN peut prendre un temps variable, en fonction de la taille de données et de la puissance d'exécution. Généralement, les DNN nécessitent une grand quantité de données et, en raison de la profondeur du réseau, ils effectuent des calculs puissants. Pour cela, une machine puissante est recommandée dans ce cas pour une exécution rapide, telle que les architectures de processeurs parallèles, l'unité de traitement graphique "*Graphics Processing Unit* ou GPU", etc. [38].

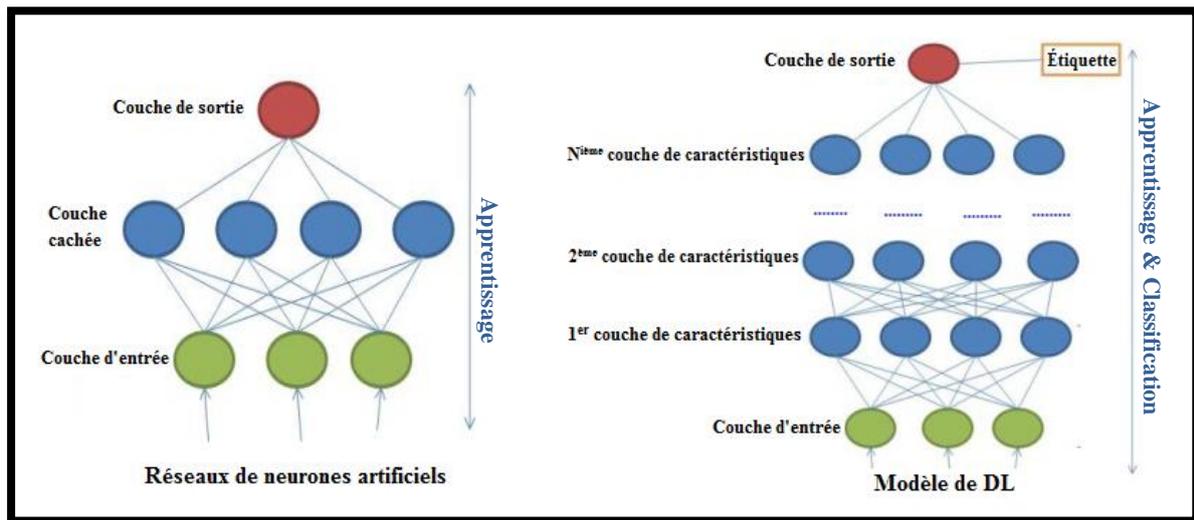


Figure III.6 : Comparaison entre les architectures d'ANN et DNN.

3. Les domaines d'application de DL

DL est l'une des technologies la plus en vogue, où elle change notre façon de voir des technologies d'IA. Récemment, elle a réalisé des progrès considérables et des grands succès dans plusieurs domaines, non seulement dans les tâches de vision par ordinateur, telles que la détection de cible, la reconnaissance visuelle et la robotique, mais aussi dans de nombreuses autres applications pratiques. Les domaines d'application de DL sont aussi nombreux que variés. Voici quelques domaines d'applications importantes dans lesquels le DL a apporté des améliorations considérables et réalisé des résultats très performants [42] :

3.1 Colorisation automatique des images en noir et blanc

La colorisation des images est le problème de l'ajout de couleur, noir et blanc, aux photographies. Traditionnellement, car c'était une tâche très difficile, cela se faisait à la main avec des efforts humains. DL peut être utilisé dans ce cas pour utiliser les objets et leur contexte dans la photographie pour colorer l'image, un peu comme un opérateur humain pourrait aborder le problème avec un exploit visuel et très impressionnant. De manière impressionnante, la même approche peut être utilisée pour coloriser des images fixes de films en noir et blanc [40].

3.2 Ajout automatique de sons aux films muets

Dans cette tâche, le système doit synthétiser des sons pour correspondre à une vidéo silencieuse. Le système est formé à l'aide de 1000 exemples de vidéos avec le son d'une baguette de percussion frappant différentes surfaces et créant différents sons [40]. Un modèle d'apprentissage approfondi associe les images vidéo à une base de données de sons préenregistrés pour sélectionner le son à reproduire qui correspond le mieux à ce qui se passe dans la scène. Le système a ensuite été évalué à l'aide d'un système de test de Turing, dans lequel les humains devaient déterminer quelles vidéos comportaient le son réel ou le son artificiel [43].

3.3 Traduction automatique

Dans cette tâche, la traduction automatique des mots, des expressions ou des phrases dans une langue à d'une autre langue. La traduction automatique existe depuis longtemps, mais DL permet d'obtenir les meilleurs résultats dans deux domaines spécifiques : traduction automatique de texte, et traduction automatique d'images. La traduction de texte peut être effectuée sans aucun prétraitement de la séquence, ce qui permet à l'algorithme d'apprendre les dépendances entre les mots et de les mapper dans une nouvelle langue [43]. Dans les domaines de traduction automatique d'images, DL sont utilisés pour identifier les images comportant des lettres et leur emplacement dans la scène. Une fois identifiés, ils peuvent être convertis en texte et traduits ; puis l'image est recréée avec le texte traduit. Ceci est souvent appelé traduction visuelle instantanée [40].

3.4 Génération d'écriture manuscrite automatique

C'est une tâche où, à partir d'un corpus d'exemples manuscrits, générer une nouvelle écriture pour un mot ou une phrase donnée. L'écriture manuscrite est fournie sous la forme d'une séquence de coordonnées utilisée par un stylo lors de la création des échantillons d'écriture manuscrite [40]. À partir de ce corpus, la relation entre le mouvement du stylo et les lettres est apprise et de nouveaux exemples peuvent être générés par rapport à cela. Ce qui est fascinant, c'est que différents styles peuvent être appris, puis imités [43].

3.5 Génération automatique de texte

C'est une tâche intéressante, où un corpus de texte est appris et à partir de ce modèle, un nouveau texte est généré, mot par mot ou caractère par caractère [43]. Le modèle est capable d'apprendre comment épeler, ponctuer, former des phrases et même capturer le style du texte dans le corpus. Les méthodes de DL sont utilisées pour apprendre la relation entre les éléments des séquences de chaînes en entrée, puis pour générer du texte. Le modèle est capable d'apprendre comment épeler, ponctuer, former des phrases et même capturer le style du texte dans le corpus [40].

3.6 Reconnaissance d'images

Un autre domaine populaire en ce qui concerne le DL est la reconnaissance d'images. Son objectif est de reconnaître et d'identifier les personnes et les objets dans les images, ainsi que de comprendre le contenu et le contexte. La reconnaissance d'images est déjà utilisée dans plusieurs secteurs tels que les jeux, les médias sociaux, la vente au détail, le tourisme, etc. Cette tâche nécessite de classer les objets d'une photographie dans un ensemble d'objets précédemment connus. Une variante plus complexe de cette tâche, appelée détection d'objets, consiste à identifier spécifiquement un ou plusieurs objets dans la scène de la photo et à dessiner un cadre autour d'eux. En plus, les méthodes de DL jouent également un rôle important dans le traitement des images de télédétection, la classification des scènes de télédétection haute résolution, gestion des catastrophes comme les séismes, les éruptions volcaniques, les tsunamis, les glissements de terrain, les inondations, les cyclones, etc. [43].

3.7 Les applications de télédétection

D'autres importants domaines en ce qui concerne le DL sont les domaines d'application de télédétection. Les méthodes de DL jouent un rôle très important de nombreuses applications, où elles utilisent des données de télédétection, notamment la classification de la couverture terrestre et de l'utilisation des terres, l'analyse de séries chronologiques, la détection de changements, les prévisions météorologiques, la détection de changement pour déterminer les changements, qui peuvent être dus à des phénomènes naturels tels que les séismes, les éruptions volcaniques, les tsunamis, les glissements de terrain, les inondations, les cyclones, etc. La détection et l'identification automatisées d'objets et de cibles pour les applications militaires, etc. [44].

3.8 Génération automatique de légendes d'images

Les légendes automatiques des images est la tâche pour laquelle le système doit générer des annotations décrivant le contenu de l'image. En 2014, les algorithmes d'apprentissage en profondeur ont connu une explosion, obtenant des résultats très impressionnants sur ce problème, tirant parti du travail des meilleurs modèles pour la classification et la détection d'objet sur des photographies [43]. Une fois que vous pouvez détecter des objets sur des photographies et générer des étiquettes pour ces objets, vous pouvez voir que l'étape suivante consiste à transformer ces étiquettes en description de phrase cohérente [40].

3.9 Robotique et véhicules autonomes

Pour que les robots autonomes puissent prendre leurs propres décisions intelligentes, fiables et utiles, nous devons leur donner la capacité de comprendre leur environnement. La compréhension de la scène avec le DL peut fournir cette connaissance, comme par exemple les voitures autonomes. Les entreprises qui construisent de tels types de services d'aide à la conduite, ainsi que des voitures autonomes, telles que Google, doivent apprendre à un ordinateur à maîtriser certaines parties essentielles de la conduite à l'aide de systèmes de capteurs numériques au lieu de l'esprit humain. Pour ce faire, les entreprises commencent généralement par former des algorithmes utilisant une grande quantité de données. Nous pouvons imaginer comment un enfant apprend grâce à des expériences constantes et à la réplication. Ces nouveaux services pourraient fournir des modèles commerciaux inattendus aux entreprises [43].

3.10 Prédire les séismes

Les scientifiques de Harvard ont utilisé DL pour apprendre un ordinateur à effectuer des calculs viscoélastiques. Ces sont les calculs utilisés dans les prédictions de séismes. Jusqu'à leur présentation, ces calculs étaient très gourmands en ordinateur, mais cette application de DL a amélioré le temps de calcul de 50 %. En ce qui concerne le calcul des tremblements de terre, le temps est important et cette amélioration peut être essentielle pour sauver des vies [43].

Il existe également de nombreux autres domaines d'application pour DL tels que : les jeux automatiques, la publicité, la reconnaissance de la parole, la reconnaissance faciale, le traitement du langage naturel, etc.

4. Les approches de DL

L'apprentissage en profondeur offre une vision différente de l'apprentissage de caractéristiques et de leur représentations, où des caractéristiques robustes, abstraites et invariantes sont apprises, hiérarchiquement, à partir de données brutes (par exemple, pixels d'image). Cela constitue un avantage important par rapport aux méthodes de bas niveaux et de niveaux intermédiaire décrites dans le chapitre précédent. Plusieurs modèles de DL ont été proposés dans la littérature, dont les plus utilisées sont [42] :

- L'auto-encodeur superposé "SAE" ;
- Les machines Boltzmann profondes "DBM" ;
- Les réseaux de croyances profondes "DBN" ;
- Et les réseaux de neurones convolutifs "CNN".

4.1 L'auto-encodeur superposé

Un auto-encodeur, AE, est simplement défini comme un ANN qui vise à apprendre une représentation cachée dans le but de reconstruire ses entrées, où les entrées du modèle sont similaires aux sorties. Plus spécifiquement, il est formé pour coder l'entrée dans une représentation afin que l'entrée puisse être reconstruite à partir de cette représentation telle que le nombre de nœuds dans la couche d'entrée et le nombre de nœuds dans la couche de sortie sont identiques. En d'autres termes, c'est un algorithme d'apprentissage non supervisé utilisé pour coder efficacement l'ensemble de données dans le but de réduire la dimensionnalité ou de supprimer le bruit des données. L'AE mappe l'entrée via une fonction d'encodeur f pour générer une représentation interne "latente", où code H , c'est-à-dire $H = f(x)$. L'auto-encodeur possède également une fonction de décodeur, g qui mappe H sur la sortie \hat{x} . Dans le processus de codage, l'AE convertit d'abord le vecteur d'entrée x en une représentation cachée H à l'aide d'une matrice de pondération W ; ensuite, dans le processus de décodage, l'AE renvoie H au format d'origine pour obtenir \hat{x} avec une autre matrice de pondération W' . Théoriquement, W' devrait être la transposée de W . Le schéma de principe de ce modèle est présenté à la figure III.7 [44].

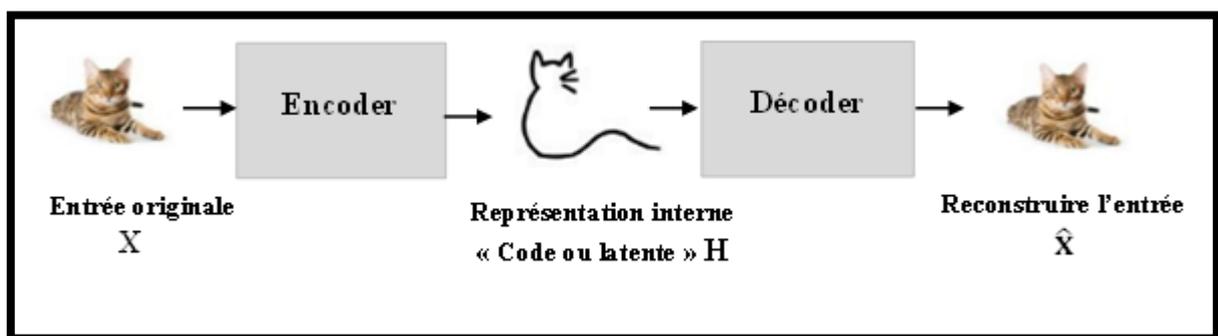


Figure III.7: Diagramme schématique des AE.

Auto-encodeurs empilés, superposés ou profonds, "en anglais (*Stacked Auto-Encoder* ou *SAE*)", peuvent être formés à partir plusieurs couches généralement placées symétriquement dans les parties codeur et décodeur comme sur la figure III.8. L'apprentissage de SAE peut être fait couche par couche. Les caractéristiques apprises d'une couche sont introduites dans la couche suivante, de sorte qu'un réseau avec une couche cachée est d'abord entraîné, puis un réseau avec deux couches cachées est entraîné, etc. À chaque étape, l'ancien réseau avec $k - 1$ couches cachées est créé et une k ème couche cachée supplémentaire est ajoutée, qui prend en entrée la couche cachée précédente $k - 1$ qui a été entraîné. Une fonction de perte L mesure à quelle distance l'AE peut reconstruire la sortie, où L est une fonction de x et $\hat{x} = g(f(x))$. Après l'apprentissage, les activations de la couche cachée sont utilisées comme des caractéristiques apprises et la couche supérieure peut être ignorée. Les activations de la couche supérieure peuvent être traitées comme des descripteur et introduites dans un classificateur approprié, tel que SVM, DT, etc. [38]. Un SAE est défini comme "sous-complet", si la couche de codage a une dimensionnalité inférieure à celle de l'entrée. Dans ce cas, le réseau est obligé de prendre une représentation plus compacte. Par contre, un SAE est défini comme "sur-complet", si la couche de codage a une dimensionnalité supérieure à celle de l'entrée. Dans ce cas le risque, d'apprendre des caractéristiques sans intérêt, existe pour cela certaines restrictions peuvent être appliquées [45].

4.2 Les machines Boltzmann restreintes

Les machines Boltzmann restreintes, "en anglais (*Restricted Boltzmann Machines* ou *RBM*)", sont très utilisées dans les réseaux d'apprentissage en profondeur en raison de leur importance historique et de leur simplicité relative. Les machines RBM ont été utilisées pour générer des modèles stochastiques d'ANN pouvant apprendre de la distribution des probabilités en fonction de leurs ensembles d'entrées. Ces approches peuvent être utilisées dans plusieurs applications, notamment la réduction de dimensionnalité, la classification, l'apprentissage des caractéristiques, etc. Généralement, les RBM sont utilisées comme des méthodes d'extraction des caractéristiques dans le processus d'apprentissage aux tâches de classification [45].

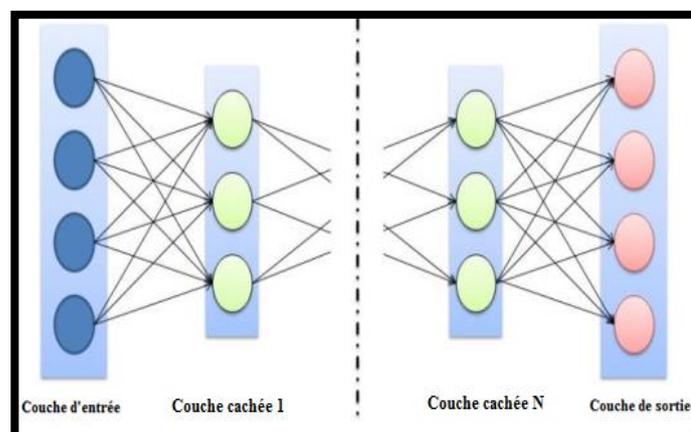


Figure III.8 : L'architecture de SAE.

RBM est un réseau neuronal non supervisé, non dirigé, à deux couches constitué d'une couche visible, ou d'une couche d'entrée, $x = (x_1, x_2, \dots, x_N)$, et d'une couche cachée, $h = (h_1, h_2, \dots, h_N)$ qui porte la représentation de la caractéristique latente de ces ensembles d'entrées. Comme le montre la figure III.9, les neurones sont restreints pour former un graphe bipartite dans un RBM. On peut voir qu'il existe une connexion complète entre les unités visibles et les unités cachées, alors qu'aucune connexion n'existe entre les unités de la même couche, et cette restriction donne son nom à la RBM. Les neurones de couches cachées prennent des unités visibles comme entrées. Les neurones visibles prennent vecteurs d'entrée binaires comme entrée initiale et ensuite des probabilités de couches cachées. Dans la phase d'apprentissage, un échantillonnage de Gibbs, Gibbs Sampling, est effectué et équivaut à calculer une distribution de probabilité. Dans le premier passage, les probabilités de couche cachée h sont calculées à partir des entrées x . Ensuite, dans le deuxième passage, ces valeurs redirigées au couche visible, et reviennent à la couche cachée pour obtenir x' et h' . Les poids sont mis à jour en utilisant les différences dans les produits extérieurs des activations cachées et visibles entre le premier et le deuxième passages. Pour approcher le modèle optimal, un grand nombre de passages est nécessaire, donc cette approche fournit une inférence proche, mais fonctionne bien dans la pratique. Après l'apprentissage, les activations de la couche cachée d'un RBM peuvent être utilisées comme descripteurs à prendre en compte. En utilisant la RBM comme modules d'apprentissage, deux approches fondamentales d'apprentissage profonds ont été proposées: le réseau de croyances profondes "DBN" et la machine Boltzmann profonde "DBM" [38].

4.2.1 Les réseaux de croyances profondes

Bien qu'un RBM puisse être capable d'apprendre des descripteurs à partir d'un ensemble de données d'entrée simples, son contenu peut être limité. Pour cette raison, ils sont généralement empilés pour former un réseau de niveau supérieur. Les réseaux de croyances profondes ou DBN, "en anglais (*Deep Belief Networks*)", sont capables de construire des représentations de plus en plus abstraites à l'aide de leurs multiples couches cachées. Le DBN est un modèle graphique génératif composé de plusieurs couches cachées de variables stochastiques ou latentes, où ils ont des connexions non dirigées uniquement aux deux couches supérieures et des connexions dirigées aux couches inférieures [45]. Plus précisément, un DBN peut être défini comme une pile de RBM, dans lesquelles chaque couche RBM communique à la fois avec les précédentes et les couches suivantes, alors qu'aucune connexion n'existe entre les nœuds de la même couche. À l'exception des première et dernière couches, chaque couche d'un DBN a un double rôle : elle sert de couche cachée aux nœuds dans la couche précédente et de couche d'entrée, ou visible à la couche suivante. Les nœuds d'une couche cachée représentent des descripteurs qui capturent les corrélations présentées dans les données d'entrées. La figure III.10 présente un DBN à trois couches, où un DBN est une pile de RBM. Il est également possible d'empiler n'importe quel type de réseaux simples et non supervisés tels que des AE. Les DBN sont utilisés pour reconnaître, regrouper et générer des images, des séquences vidéo, de capture de mouvement des données et de reconnaissance de la parole. Comparés aux ANN, les DBN sont plus efficaces, en particulier s'ils s'appliquent à des problèmes de données non étiquetées [44].

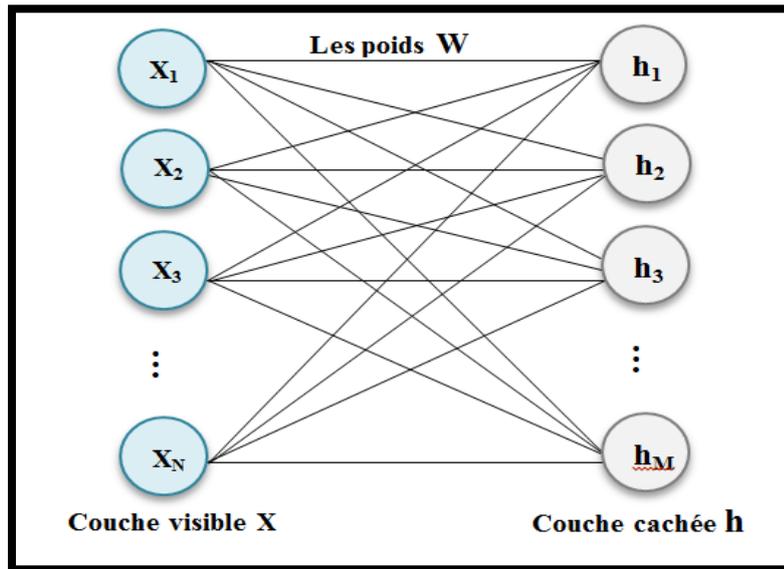


Figure III.9 : L'architecture de RBM.

4.2.2 Les machines Boltzmann profondes

Les machines Boltzmann profondes, "en anglais (*Deep Boltzmann Machines* ou DBM)", est un modèle graphique génératif probabiliste, non supervisé, composé de plusieurs couches cachées de variables stochastiques avec des connexions entièrement non dirigées entre différentes couches. Similairement aux DBN, les DBM peuvent être compris comme une série de RBM empilées les unes sur les autres, utilisées pour identifier les caractéristiques latentes présentées dans les données d'entrées. La principale différence entre les DBM et les DBN réside dans le fait que DBN ont été introduits sous forme de modèles graphiques mixtes dirigés/non dirigés, alors que les DBM sont complètement non dirigés. C'est-à-dire que les nœuds cachés sont regroupés dans une hiérarchie de couches, tels qu'il existe une connectivité complète entre les couches voisines, mais pas de connectivité à l'intérieur des couches ou entre des couches non voisines comme sur la figure III.10. Une DBM permet d'apprendre des représentations de haut niveau des caractéristiques, par rapport aux réseaux de neurones artificiels.

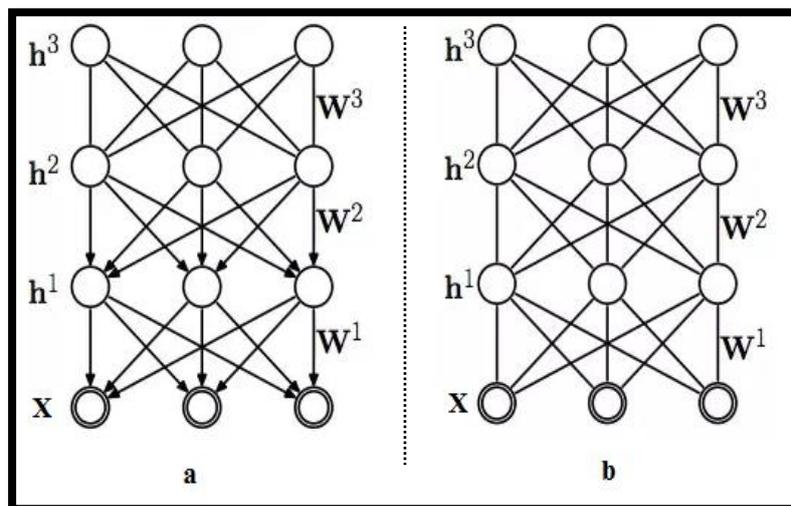


Figure III.10 : L'architecture de DBN "a" et de DBM "b".

Les DBM ont été utilisés dans différentes applications telles que la reconnaissance d'image ou de texte, la reconnaissance d'expression faciale, la reconnaissance de modèles 3D et l'identification de personnes audio-visuelle [46].

4.3 Les réseaux de neurones convolutifs

Les réseaux de neurones convolutifs, "en anglais (*Convolutional Neural Networks*, CNN ou ConvNets)", ont déjà été reconnus comme l'approche d'apprentissage en profondeur la plus efficace, la plus réussie et la plus utilisée pour l'analyse d'images en raison de leurs performances remarquables. Ils ont été appliqués avec un grand succès dans plusieurs domaines de la vision par ordinateur [44]. Actuellement, les CNN ont déjà été appliqués avec succès à la reconnaissance de l'écriture manuscrite, à la détection des visages, à la reconnaissance du comportement, à la reconnaissance de la parole, aux systèmes de recommandation, à la classification des images, à la détection et à la classification d'objets dans une image, à la reconnaissance visuelle, à la recherche d'images et l'annotation de scènes, etc. [53]. CNN est très similaire à un réseau de neurones multi-couches, mais l'architecture d'un CNN est composée de plusieurs couches cachées. Ces modèles sont conçus pour imiter le comportement d'un cortex visuel. Il comprend plusieurs couches de hiérarchie, certaines pour les représentations de caractéristiques et d'autres en tant que type de réseaux de neurones classiques pour la classification comme sur la figure III.11.

Plus précisément, la couche d'entrée est l'image brute, image raw, qui est fournie sous la forme d'une matrice de pixels. La première partie des couches cachées représente des couches spéciales appelées couches convolutives, ou convolution, et des couches de regroupement, ou de mise en commun, qui permettent au réseau de coder certaines caractéristiques d'images et qui fonctionnent comme un extracteur de caractéristiques des images [45]. Ces couches sont ensuite généralement suivies des couches soit totalement connectées, soit partiellement connectées, qui effectuent une classification ou une estimation de probabilité de classe.

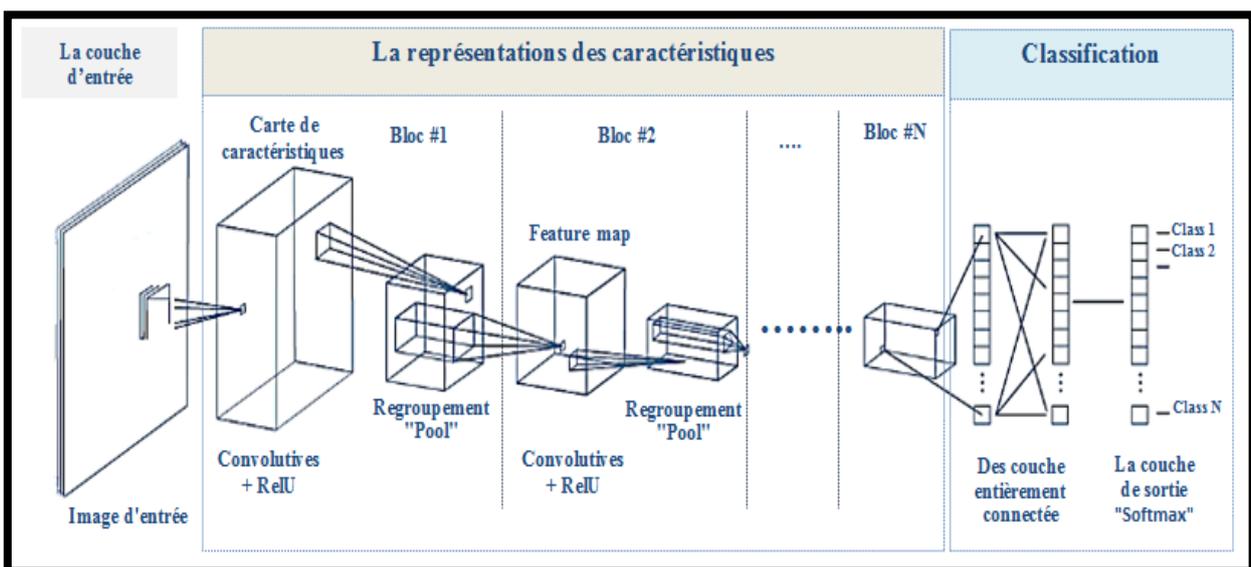


Figure III.11 : L'architecture de CNN.

Le profondeur de CNN de convolution force le réseau d'apprendre des représentations plus abstraites et plus discriminantes. Plus la couche est profonde, plus les caractéristiques apprises sont abstraites et complexes. Les premières couches extraient des caractéristiques de bas niveau telles que les bords, les gouttes ou les blobs et les gradients orientés, tandis que les couches plus profondes capturent les caractéristiques de haut niveau. Cependant, il est difficile de former un CNN profond puissant avec de petits ensembles de données dans la pratique [49].

Trois facteurs jouent un rôle important dans le processus d'apprentissage d'un CNN : l'interaction clairsemée, le partage de paramètres et la représentation équivariante, "ou en anglais (*sparse interaction, sparse connectivity, parameter sharing et equivariant representation*)", [47]. À la différence des NN traditionnels, où les relations entre les nœuds d'entrée et de sortie sont dérivées de la multiplication matricielle, les CNN réduisent la charge du calcul avec une interaction creuse ou clairsemée, où les filtres sont rendus plus petits que les entrées et utilisés pour toute l'image. Plus précisément, ANN calcule les activations pour chaque couche en reliant chaque sortie avec chaque entrée. Les réseaux CNN ont des interactions clairsemées en rendant le filtre plus petit que l'entrée.

Cela rend CNN n'a besoin que de peu de paramètres de stockage et le calcul de la sortie n'a besoin que de peu d'opérations [48]. La figure III.12 illustre les deux cas : (a) exemple de réseau entièrement connecté sans interaction clairsemée, où h_3 est calculé par multiplication de matrice complète, et (b) exemple de l'interaction clairsemée dans CNN avec filtre de taille 3, déplacé avec un stride de 1, où h_3 ne dépend que de x_2, x_3, x_4 . À la différence des NN traditionnels, où chaque élément de matrice de poids est utilisé exactement une seule fois pour calculer la sortie d'une couche, dans les réseaux CNN, les paramètres sont liés, où le poids appliqué à une entrée est lié à la valeur d'un poids appliqué dans un autre endroit. Le même noyau est utilisé dans toute l'image. Par conséquent, au lieu d'apprendre un paramètre pour chaque emplacement, seul un ensemble de paramètres est à prendre en compte. Plus précisément, l'idée de base du partage de paramètres est qu'au lieu d'apprendre un ensemble séparé de paramètres à chaque emplacement, nous n'avons besoin que de prendre un ensemble, ce qui implique une meilleure performance du CNN. Le partage de paramètres est également doté le CNN d'une propriété attrayante appelée équivariance, ce qui signifie que chaque fois que l'entrée change, la sortie change de la même manière. Par conséquent, moins de paramètres sont nécessaires pour CNN par rapport aux autres algorithmes NN traditionnels, ce qui entraîne une réduction de la mémoire et une amélioration de l'efficacité. Les poids partagés et la connectivité fragmentée rendent les CNN idéaux pour le traitement et la classification des images [47].

5. L'architecture de CNN

Comme le montre la figure III.11, les CNN se composent d'une couche d'entrée et de sortie et de plusieurs couches cachées entre l'entrée et la sortie. Ce réseau se compose généralement de plusieurs blocs de couches et chaque bloc est composé de plusieurs couches, notamment les couches de convolution, les couches de correction par l'application d'une fonction "Relu", et les couches de regroupement, qui sont suivies par des couches entièrement connectées.

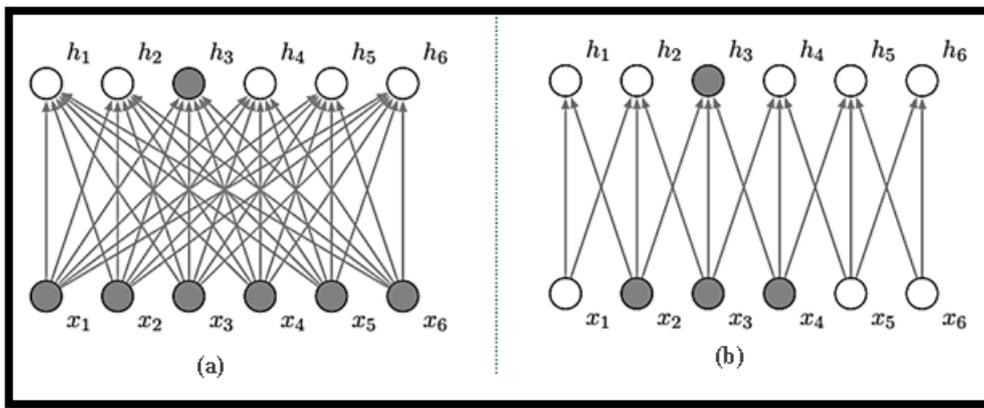


Figure III.12 : (a) Exemple de réseau entièrement connecté sans interaction clairsemée et (b) exemple de l'interaction clairsemée dans CNN.

5.1 Couche d'entrée

La couche d'entrée, "en anglais (Input Layer)", est représentée l'image brute, ou l'image *raw* en anglais. Elle est fournie sous la forme d'une matrice de pixels. De plus, les images peuvent être importées, en tant qu'entrées brutes, directement sur la grille. Cela évite la procédure d'extraction des caractéristiques dans l'apprentissage automatique traditionnel dans laquelle les caractéristiques pertinentes doivent être extraites manuellement. L'apprentissage en profondeur par CNN utilise des images brutes comme entrée pour apprendre certaines caractéristiques. La couche d'entrée définit et spécifie la taille des images d'entrée d'un CNN. La taille des images correspond à la largeur, la hauteur et le nombre de canaux de couleur de ces images. Les images sont représentées à deux dimensions $2D$ pour les images aux niveaux de gris car le nombre de canaux est 1, et à trois dimensions $3D$ pour les images couleurs car le nombre de canaux est 3, pour représenter les couleurs fondamentales en mode RGB, HSV, etc. [41].

5.2 Couches convolutionnelles

Le nom réseau CNN, le réseau de neurone convolutif, indique que le réseau utilise des opérations de convolution dans les couches convolutionnelles. Les couches convolutionnelles ou les couches convolutives, "en anglais (*Convolution Layer*)", est une opération linéaire qui permet de gérer des entrées de tailles variables. Cette Couche est considérée la plus importante pour l'extraction des caractéristiques. Dans cette couche, on applique un produit de convolution qui sert à extraire des caractères spécifiques de l'image traitée. La somme de produit de convolution génère la sortie de cette couche qu'on appelle la carte des caractéristiques, en anglais "feature maps". Généralement, chaque couche de convolution a un certain nombre de filtres, "noyaux, champ réceptif ou poids", qui produisent le même nombre de cartes des caractéristiques en faisant glisser les filtres sur les cartes des caractéristiques de la couche précédente [41].

D'une manière précise, avant tout, il effectue des multiplications élément par élément, c'est-à-dire un produit scalaire, entre une sous-matrice d'une matrice d'entrée et un filtre. Les valeurs de poids initiales d'un filtre sont généralement générées d'une façon aléatoire. Les valeurs de biais peuvent être définies de nombreuses façons en fonction de la configuration

des réseaux. Ces deux valeurs sont ajustées lors de l'entraînement de réseau CNN. La taille d'une sous-matrice est toujours égale à un filtre, mais un filtre est toujours plus petit que la matrice d'image d'entrée. Ensuite, les valeurs multipliées sont additionnées et un biais est ajouté à ces valeurs additionnées. Le filtre a toujours le même nombre de canaux que l'image d'entrée. Des filtres de convolution 2D sont utilisés pour analyser les images en niveaux de gris, tandis que des filtres de convolution 3D sont utilisés pour analyser les images en couleur RGB.

La figure III.13 illustre l'opération de convolution. L'un des avantages de la convolution est qu'elle réduit la taille des données d'entrée, ce qui réduit les coûts de calcul. Un hyper-paramètre supplémentaire de la couche est la foulée ou "stride". Le stride définit le nombre de colonnes et de lignes "pixels" du filtre glissant à la fois sur la largeur et sur la hauteur de la matrice en entrée, c'est-à-dire la distance entre deux positions consécutives du filtre qui consiste à déplacer le noyau d'une sous-matrice à l'autre ; il saute certains pixels. Une taille de stride plus grande mène à moins d'applications du filtre et une taille de sortie plus petite, ce qui réduit également les coûts de calcul, mais il peut aussi perdre des caractéristiques des données d'entrée. La taille de sortie T_s d'une couche de convolution est calculée à l'aide de l'équation suivante : $T_s = \frac{(I-R)}{S} + 1$ où, I est la taille de matrice d'entrée, R est la taille de filtre et S est la taille de stride [48].

L'application de filtre avec Stride est similaire que l'application des techniques de réduction des dimensions. Pour chaque couche de convolution, la dimension de la carte des caractéristiques en sortie sera toujours inférieure à celle en entrée. Cela veut dire que la taille de l'entrée diminue trop rapidement avec des pertes de données dans les coins d'image. Cependant, si nous voulons construire un réseau de convolution profond, nous ne voulons pas que la taille de l'entrée diminue trop rapidement. Un petit filtre peut partiellement résoudre ce problème. Mais pour conserver une certaine dimension, nous avons besoin de zéro-remplissage, "zero-padding". Fondamentalement, il ajoute des zéros à la frontière de la matrice d'entrée, comme il est illustré dans la figure III.14. La taille du remplissage de zéro est un hyper-paramètre de la couche de convolution. Cela nous permet de contrôler la taille spatiale de l'image de sortie. Donc, si I est la taille de matrice d'entrée, R est la taille de filtre, S est la taille de stride et P est la quantité de padding nécessaire, la taille de sortie de l'image est alors la suivante [41] : $T_s = \frac{(I-R+2*P)}{S} + 1$.

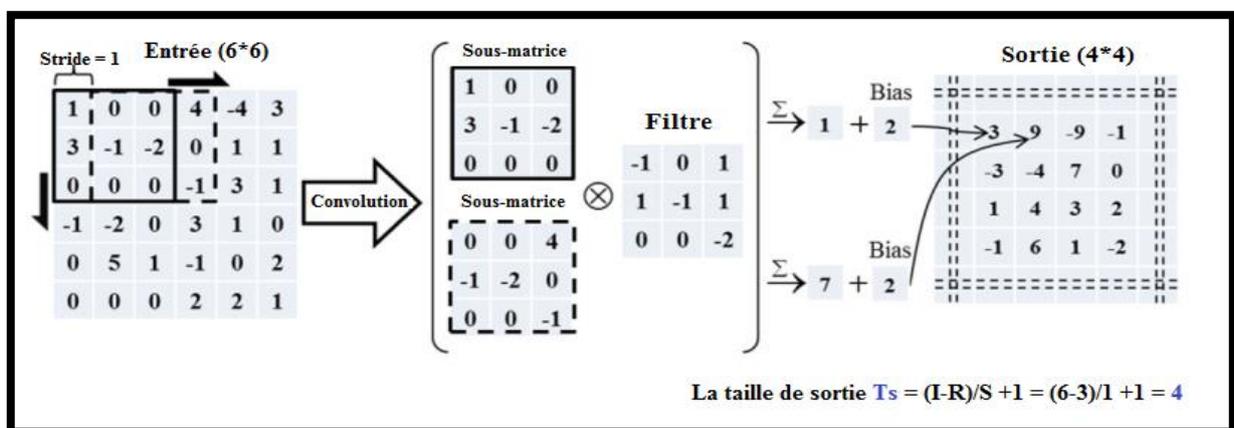


Figure III.13 : Exemple illustrant l'opération des couches convolutionnelles.

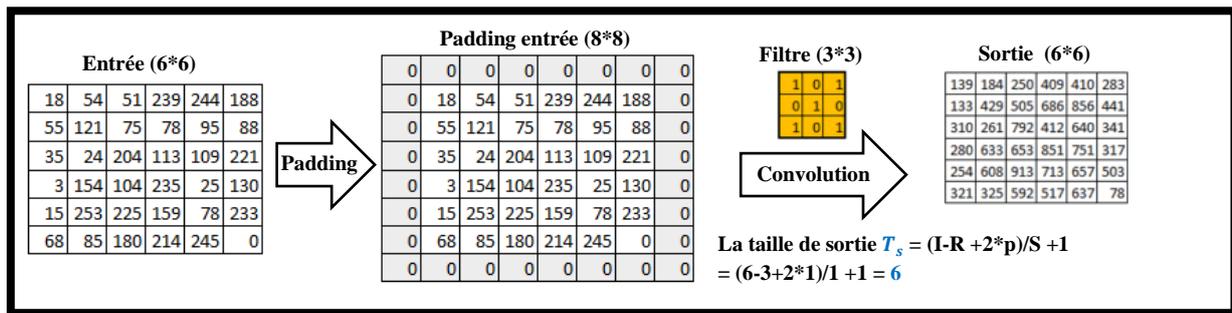


Figure III.14 : Exemple illustre l'opération de zero-padding dans la couche convolutionnelle.

5.3 Couches de correction "ReLU"

Pour effectuer la transformation non linéaire de la carte des caractéristiques dans le but d'introduire la non-linéarité dans le réseau, les couches de convolution devraient être suivies par fonction d'activation ReLU. La correction ReLU est préférable, car il améliore l'efficacité du traitement, il en résulte la formation de réseau plusieurs fois plus rapide, sans faire une différence significative à la généralisation de précision. ReLU effectue la transformation non linéaire de la carte des caractéristiques produite par la couche de convolution et peut générer des représentations des caractéristiques non négatives [48].

5.4 Couche de regroupement

Les couches de regroupement ou les couches de mise en commun, "en anglais (*Pooling layer*)", ou Pool, sont généralement placées entre deux couches convolutives successives. Ce processus effectue des opérations de sous-échantillonnage pour réduire la taille des cartes des caractéristiques et le nombre de paramètres réseau via la pooling maximum ou moyen. Le but de la couche de regroupement est de réduire progressivement la taille spatiale de la représentation pour réduire le nombre de paramètres et de calculs dans le réseau, et donc de contrôler également le sur-apprentissage. La mise en pool maximale prend les valeurs maximales dans les sous-matrices de la matrice d'entrée, tandis que la mise en commun moyen prend les valeurs moyennes. Ces stratégies de mise en commun, qui sont les plus largement utilisées, créent une invariance aux petits déplacements et aux déformations. Plusieurs travaux ont montré que les performances de la mise en pool maximales dans les ensembles de données d'image sont meilleures que celles de la mise en pool moyenne. Le regroupement est souvent appliqué avec des filtres de taille 2×2 appliqués avec une Stride de 2 à chaque tranche de profondeur. Une couche de regroupement de taille 2×2 avec un pas de 2 réduit l'image d'entrée à 1/4 de sa taille d'origine. La figure III.15 illustre la méthode de regroupement avec un Stride de deux, la taille de sortie de la couche de regroupement étant calculée par : $T_s = \frac{(I-P)}{S} + 1$, où I est la taille de matrice d'entrée, P est la taille de filtre et S est la taille de stride [48].

5.5 Couche entièrement connectée et couche de sortie de classification

La couche entièrement connectée, "en anglais (*Fully Connected Layer*) ou FC", constitue les dernières couches dans le réseau CNN qui suivent généralement les couches convolutives et les couches de regroupements.

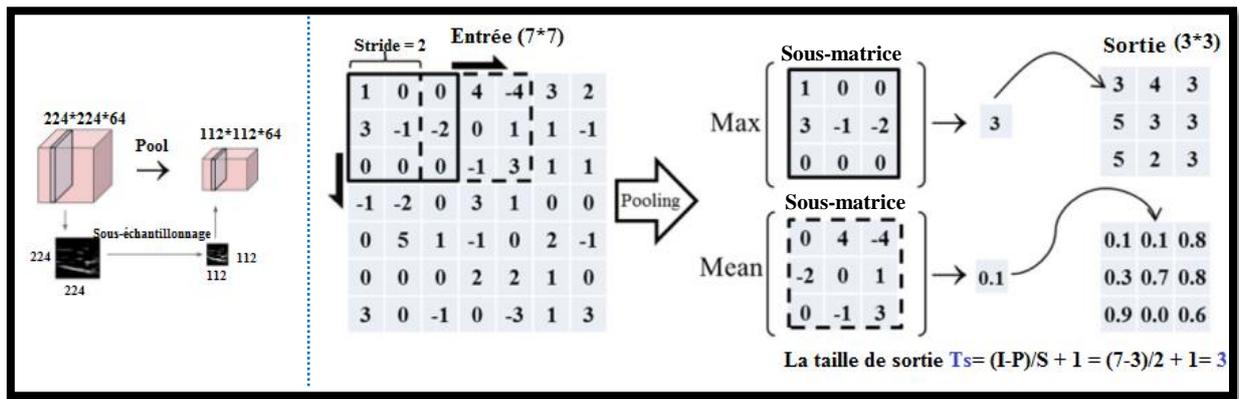


Figure III.15 : Exemple illustrant la méthode de regroupement.

Comme son nom l'indique, tous les neurones de cette couche sont connectés à la carte des caractéristiques de couche précédente, comme dans l'ANN. Si la carte des caractéristiques n'est pas unidimensionnelle, c'est-à-dire $N \times 1 \times 1$, elle est aplatie dans un vecteur unidimensionnel. Plus précisément, dans la dernière couche de regroupement, c'est-à-dire avant les couches entièrement connectées, la carte des caractéristiques regroupée est convertie en un long vecteur linéaire continu à partir de processus, qui s'appelle aplatissement, en anglais "Flattening". Le vecteur unidimensionnel, qui est créé lors de l'étape d'aplatissement, est ensuite passé par une couche entièrement connectée pour classifier les images, où il représente l'entrée de l'ensemble des couches entièrement connectées comme sur la figure III.16. Les informations résumées par les couches entièrement connectées avant la couche finale de CNN à partir des cartes de caractéristiques des couches convolutives précédentes peuvent être utilisées comme un descripteur final, c'est-à-dire une entrée à un classifieur indépendant, comme SVM. Cette couche vise à exploiter les caractéristiques de ces derniers vecteurs avec les étiquettes des classes pour rendre le réseau CNN apprendre les classes d'une image entrée [41].

La couche finale de CNN est généralement associée à une fonction de perte telle que Softmax. La taille de cette couche dépendra de la taille de la sortie, c'est-à-dire le nombre de classe. La probabilité pour chacune des classes est calculée sur cette couche. Softmax calcule les scores de probabilité de chaque classe en générant un vecteur de dimensions C , C étant le nombre de classes et tous les neurones sont connectés à cette couche [48]. La figure III.16 illustre les deux couches.

5.6 Couches auxiliaires

Pour résoudre le problème de sur-apprentissage, des couches d'abandon, dropout Layer, sont utilisées. Généralement, le sur-apprentissage est un problème de long terme dans le domaine de l'apprentissage automatique. Il s'agit d'un phénomène dans lequel un réseau peut classer efficacement une base de données d'apprentissage mais ne fournit pas de résultats de validations et de tests satisfaisants. Un sur-apprentissage survient lorsqu'un modèle apprend les détails et le bruit dans les données d'apprentissage dans la mesure où ils ont un impact négatif sur les performances du modèle avec les nouvelles données. Cela signifie que les fluctuations de bruit ou aléatoires dans les données d'apprentissage sont captées et apprises en tant que concepts par le modèle.

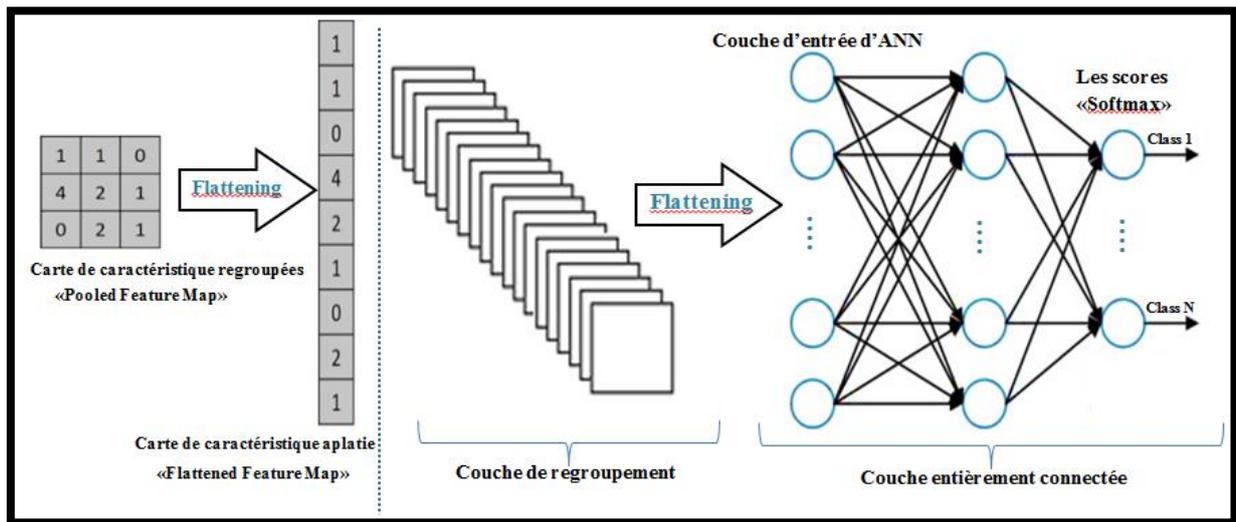


Figure III.16 : Couche entièrement connectée et couche de sortie de classification.

Le problème est que ces concepts ne s'appliquent pas aux nouvelles données et ont un impact négatif sur la capacité des modèles à généraliser. L'idée principale de couche du décrochage est de déconnecter de manière aléatoire les connexions entre neurones de couches connectées avec un certain taux de décrochage. Par conséquent, un réseau peut généraliser les exemples d'apprentissage de manière beaucoup plus efficace en réduisant ces problèmes de sur-apprentissages. La figure III.17 illustre l'application de décrochage [48].

6. Les scénarios de classification basée sur CNN

En général, l'extraction des caractéristiques basées sur CNN s'effectue selon trois scénarios principaux. Plus précisément, CNN peut être utilisé pour la classification dans trois scénarios différents, comme décrit dans [21] et [49] :

- Apprentissage à partir de zéro, en anglais "Training from scratch" ;
- Réglage d'un modèle pré-entraîné, en anglais "Tuning a pre-trained model" ;
- Et le modèle pré-entraîné sans réglage, en anglais "Pre-trained model without tuning".

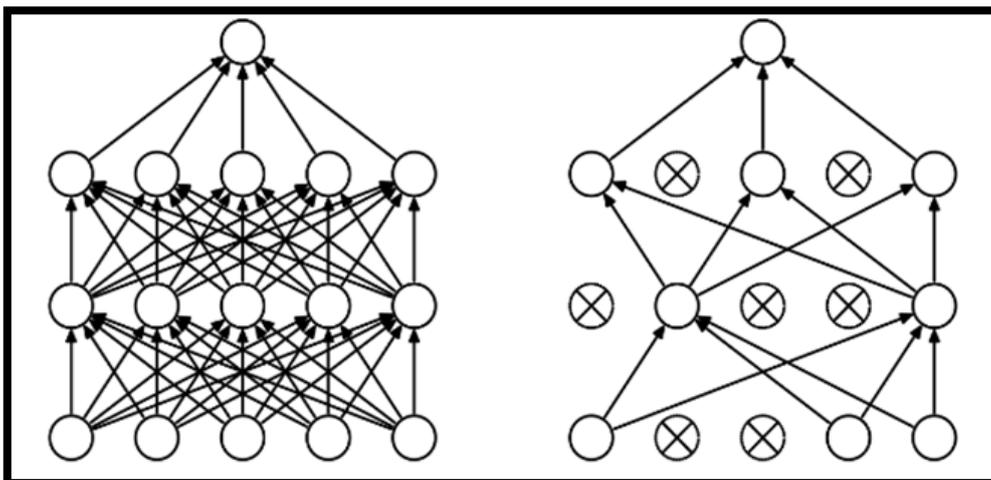


Figure III.17 : Exemple illustrant l'application de décrochage : l'exemple de gauche du réseau entièrement connecté avec deux couches cachées, et l'exemple de droite du réseau après l'application de décrochage.

6.1 Apprentissage de CNN à partir de zéro

Dans ce scénario, la nouvelle architecture CNN a été conçue à partir de zéro en utilisant les couches mentionnées précédemment, avec une initialisation aléatoire des poids de filtre. Cependant, la conception du réseau, c'est-à-dire le choix du nombre de couches, du nombre de neurones et du nombre d'hyper-paramètres dans chaque couche, est une tâche difficile. C'est parce qu'il n'y a pas une définition standard et cela dépend entièrement de nombreux facteurs, tels que les caractéristiques de l'image d'entrée (taille, complexité, etc.), la complexité de l'application, le nombre de classes et le nombre d'échantillons d'apprentissage disponibles [21].

En général, les combinaisons de couches et leurs hyper-paramètres qui conviennent à des applications spécifiques sont identifiés de manière empirique basée sur l'ensemble de données. Ce scénario nécessite une grande quantité de données d'apprentissage pour éviter les problèmes de sur-apprentissage. Par conséquent, cette approche ne peut être adoptée que si une grande quantité d'échantillon d'apprentissage est disponible. Cependant, cela nécessite un temps de calcul très long et des ressources de données énormes. Par conséquent, ce qui explique pourquoi ce scénario est rarement utilisé. Les caractéristiques de cette approche peuvent être qualifiées de caractéristiques spécifiques à l'application et aux données, qui sont apprises en fonction de base d'images d'échantillons provenant de l'application spécifique, par exemple : la détection de la tuberculose, la détection des dommages, etc., et d'images présentant des caractéristiques spécifiques, par exemple : des images médicales, des images aériennes, des images satellitaires, etc. [49].

6.2 Réglage d'un modèle pré-entraîné

Cette approche s'appelle également "Réseau affiné", en anglais "Fine-tuned Network", qui vise à transférer l'apprentissage à l'aide de modèles pré-entraînés ou pré-définis. C'est un autre scénario courant qui adopte le modèle CNN pré-entraîné. L'apprentissage par transfert dans un CNN est réalisé en apprenant des caractéristiques propres à une application en adoptant un modèle pré-entraîné déjà entraîné à l'aide d'un grand nombre d'échantillons (généralement de l'ordre de millions) de domaines pertinents, et en ajustant leurs poids en utilisant les échantillons d'apprentissage de l'application considérée et étudiée [21]. Lors de l'adoption d'un modèle CNN pré-entraîné pour l'apprentissage du transport, il est nécessaire de modifier la couche finale entièrement connectée du modèle CNN pré-entraîné, car sa dimension doit être égale au nombre de classes définies pour l'application étudiée. Également, la structure du réseau peut être modifiée en ajoutant ou en supprimant des couches, si nécessaire.

Cependant, dans la plupart des cas, l'architecture du modèle pré-entraîné est en grande partie conservée, car les caractéristiques qu'ils ont apprises se révèlent souvent plus générales et contiennent des informations importantes pour toute application qui sont très proche. À cette fin, le modèle pré-entraîné est modifié de plusieurs manières et évalué simultanément, c'est-à-dire en supprimant ou en ajoutant des couches et en évaluant simultanément en utilisant d'échantillons de données d'application étudiée. Ces types de modèles CNN pré-entraînés sont de plus en plus disponibles. Cette approche nécessite également un nombre relativement important d'échantillons d'apprentissage [49].

6.3 Les modèle pré-entraîné sans réglage

Dans ce scénario, le modèle CNN pré-entraîné est utilisé comme un outil d'extraction des caractéristiques, telles que la sortie d'une des couches entièrement connectées de l'architecture CNN étant considérée comme la représentation des caractéristiques des images d'application étudiée. Les caractéristiques extraites à l'aide du modèle pré-entraîné sont utilisées pour effectuer une classification basée sur classificateur supervisé quelconque tel que SVM. Cette approche est la plus courante et appropriée lorsqu'une petite quantité de données d'apprentissage est disponible [21].

Plus précisément, si seule une petite quantité d'échantillon d'apprentissage est disponible pour l'application désignée, une approche commune consiste à extraire les caractéristiques à l'aide du modèle pré-entraîné et puis à les utiliser pour effectuer une classification basée sur un classificateur supervisé. De plus, cette approche est la plus appropriée, elle est reconnue efficace pour la classification des images dans divers domaines, notamment les applications de télédétection telles que la classification de la couverture terrestre utilisant des images à très haute résolution et la classification des dommages. Par conséquent, cette approche est adoptée dans cette étude où les caractéristiques pour les images sont obtenues à l'aide du modèle pré-entraîné et un classificateur supervisé indépendant est construit pour effectuer la classification finale [49].

7. Les modèles de CNN pré-entraîné

Dernièrement, plusieurs travaux ont montré que les modèles de CNN pré-entraînés peuvent être appliqués avec succès à un large éventail de tâches, telles que la classification de scènes, détection d'objet et recherche d'image, etc. Les modèles CNN pré-entraînés sont des modèles déjà entraînés à l'aide d'un grand nombre d'échantillons de données, généralement de l'ordre des millions, comme par exemple la base ImageNet qui contient 1,4 millions d'images. Il existe actuellement certain nombre de modèles de CNN pré-entraînés, qui ont démontré une capacité de représentations des caractéristiques robustes et efficaces dans de nombreuses applications, qui sont [50] :

- LeNet-5 ;
- AlexNet ;
- ZF Net ;
- Le réseau VGG-16 ;
- Le réseau VGG-19 ;
- GoogLeNet ;
- Et ResNet.

7.1 LeNet-5

Les premières applications réussies des réseaux CNNs ont été développées par Yann Lecun dans les années 1990. Parmi celles-ci, la plus connue est l'architecture LeNet utilisée pour lire les codes postaux, les chiffres, etc. Le modèle LeNet-5 a été développé en 1998 pour identifier les chiffres manuscrits permettant la reconnaissance du code postal dans le service postal. LeNet-5, un réseau CNN à 7 couches, a été utilisé par plusieurs banques pour

reconnaître les nombres manuscrits sur les chèques numérisés en images d'entrée de niveaux de gris de 32×32 pixels.

La capacité de traiter des images à haute résolution nécessite des couches plus volumineuses et plus convolutives. Cette technique est donc limitée par la disponibilité des ressources informatiques. LeNet-5 reçoit une image d'entrée de $32 \times 32 \times 1$ (image en niveaux de gris) et son objectif était de reconnaître les modèles de chiffres manuscrits. Il utilise un filtre de 5×5 et la foulée est 1. La couche suivante est une couche de regroupement, permettant de calculer la couche de regroupement dans l'architecture LeNet-5. Enfin, il va à la couche entièrement connectée (couche FC) avec 120 nœuds et suivie par une autre couche FC à 84 nœuds. Il utilise les fonctions non linéaires Sigmoid ou Tanh. La variable de sortie avec 10 valeurs possibles des chiffres 0 à 9. Elle est formée sur la base de données de chiffres MNIST avec 60K exemples d'apprentissage. La figure III.18 illustre l'architecture globale de LeNet-5 [50].

7.2 AlexNet

AlexNet, développé par Alex Krizhevsky, Ilya Sutskever et Geoff Hinton, est une architecture CNN profonde et révolutionnaire. C'est un modèle gagnant du défi de reconnaissance visuelle de grande taille ImageNet de 2012 (ILSVRC-2012). AlexNet se compose de cinq couches convolutives, suivies des couches de regroupement ou pooling et des trois couches entièrement connectées, comme l'illustre le schéma de la figure III. 19. Il commence par $227 \times 227 \times 3$ images et la couche de convolution suivante applique 96 filtres sur 11×11 avec une stride de 4. Le volume de sortie réduit sa dimension de 55×55 . La couche suivante est une couche de regroupement qui applique un regroupement maximal de 3×3 filtres avec stride de 2. Il continue et atteint finalement la couche FC avec le paramètre 9216 et les deux couches suivantes avec 4096 nœuds chacun. À la fin, il utilise la fonction Softmax avec 1000 classes de sortie. Il a 60 millions de paramètres [51]. Le succès d'AlexNet repose sur certaines astuces pratiques, telles que la non-linéarité des unités linéaires rectifiées (ReLU), l'augmentation des données et la couche d'abandon. Plus précisément, le ReLU peut considérablement accélérer la phase d'apprentissage. En outre, l'augmentation des données est un moyen efficace d'améliorer la qualité de classification de façon significative et de réduire le sur-apprentissage lors de l'apprentissage d'un grand CNN.

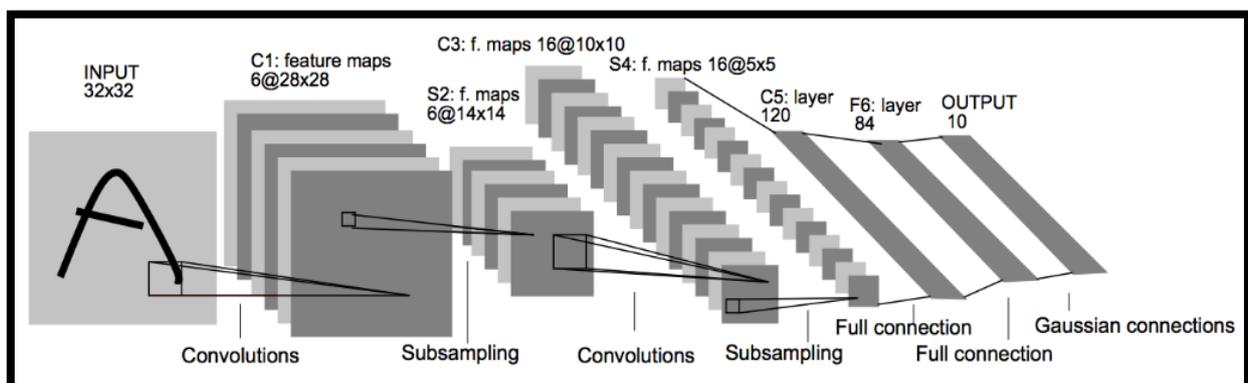


Figure III.18 : L'architecture globale de LeNet-5 [50].

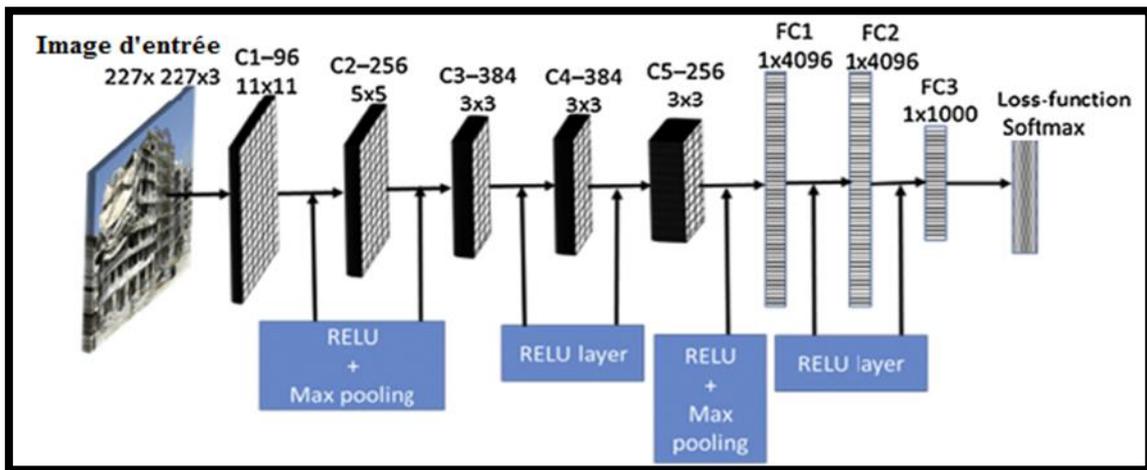


Figure III.19: L'architecture globale d'AlexNet. Elle est composée de cinq couches de convolution (C1-C5) et de trois couches de FC (FC1-FC3) [49].

Cette augmentation est générée plus d'échantillons d'images d'apprentissage en augmentant artificiellement le volume de données à disposition, par exemple en recadrant patches de petite taille et retourner horizontalement ces patches à partir d'images originales. En plus, les couches d'abandon, dropout Layer, sont utilisées au niveau des couches entièrement connectées pour réduire les sur-apprentissages importants. En résumé, le succès d'AlexNet a popularisé l'application des grands CNN dans les tâches de reconnaissance visuelle, et par conséquent, AlexNet est devenu une architecture de base des CNN modernes [51].

7.3 ZF Net

Le réseau ZF Net est un modèle de réseau CNN proposé par Matthew Zeiler et Rob Fergus dans l'article intitulé "Visualizing and Understanding Convolutional Neural Networks". C'est un modèle gagnant du défi de 2013 (ILSVRC-2013). Ce modèle a atteint un taux d'erreur de 11,2%. Il est plus connu sous le nom de ZF Net (abréviation de Zeiler & Fergus Net). Il s'agissait d'une amélioration pour AlexNet en modifiant les hyper-paramètres de l'architecture, notamment en augmentant la taille des couches convolutives moyennes et en réduisant la taille de la foulée et du filtre de la première couche. Cette architecture était plutôt une adaptation de la structure précédente d'AlexNet, mais elle développait toujours quelques idées essentielles pour améliorer les performances. Une autre excellente raison de cet article réside dans le fait que les auteurs ont passé beaucoup de temps à expliquer beaucoup l'intuition de ConvNets et à montrer comment visualiser correctement les filtres et les poids.

Plus précisément, son architecture est très similaire à AlexNet, à l'exception de quelques modifications mineures. Telle que AlexNet a été formé sur 15 millions d'images, tandis que ZF Net a été formé sur seulement 1,3 million d'images. Au lieu d'utiliser des filtres de taille 11×11 dans la première couche (ce qu'AlexNet a implémenté), ZF Net a utilisé des filtres de taille 7×7 et une valeur de foulée réduite. Le raisonnement derrière cette modification est qu'une taille de filtre plus petite dans la première couche de convolution permet de conserver de nombreuses informations sur les pixels d'origine dans le volume d'entrée. Un filtrage de taille 11×11 s'est avéré ignorer de nombreuses informations pertinentes, d'autant plus qu'il s'agit de la première couche de convolution. Les ReLU utilisées pour leurs fonctions d'activation, la perte d'entropie croisée pour la fonction d'erreur

et leur formation à la descente de gradient par lots stochastiques. Ce modèle est formé d'un GPU GTX 580 pendant douze jours [52].

7.4 Le réseau VGG-16

Le réseau VGG-16 est un modèle de réseau CNN très profond proposé par Karen Simonyan et Andrew Zisserman de l'université d'Oxford dans l'article intitulé "Very Deep Convolutional Networks for Large-Scale Image Recognition". C'est un modèle gagnant du défi de reconnaissance visuelle de grande taille ImageNet de 2014 (ILSVRC-2014). Le modèle atteint une précision de 92,7% dans le top 5 des tests dans ImageNet. Il apporte des améliorations par rapport à AlexNet en remplaçant les grands filtres de la taille du noyau (respectivement 11 et 5 dans la première et la deuxième couche de convolution) par plusieurs filtres de la taille du noyau 3×3 , l'un après l'autre. Il a été proposé pour étudier l'effet de la profondeur du réseau. Il a été démontré que la profondeur du réseau joue un rôle important dans l'amélioration de la précision de la classification et donne un résultat impressionnant. Il est actuellement le choix préféré de la communauté pour extraire des caractéristiques d'images [52].

VGG-16 contenant 13 couches de convolution suivie par 3 couches entièrement connectées, comme il est illustré à la figure III. 20. Il commence par des images à une taille fixe de $224 \times 224 \times 3$. L'image est passée à travers une pile de couches convolutives, où les filtres ont été utilisés avec un très petit champ récepteur 3×3 . La foulée de convolution est fixée à 1 pixel. Le pooling spatial est réalisé par cinq couches de max-pooling. La mise en pool maximum est effectuée sur une fenêtre de 2×2 pixels, avec foulée 2. Toutes les couches cachées sont équipées de la non-linéarité de rectification (ReLU) [53]. Trois couches entièrement connectées (FC) suivent une pile de couches convolutives : les deux premières comportent 4096 canaux chacune, la troisième effectue une classification ILSVRC à 1000 classes et contient donc 1000 canaux (un pour chaque classe). La couche finale est la couche soft-max.

La configuration des couches entièrement connectées est la même dans tous les réseaux. VGG-16 comprend 138 millions de paramètres, ce qui peut être un peu difficile à gérer. Il est formé sur 4 GPU pendant 2 à 3 semaines. La figure III.20 montre que dans VGG-16, des blocs avec la même taille de filtre sont appliqués plusieurs fois pour extraire des caractéristiques plus complexes et plus représentatives. Ce concept de blocs est devenu un thème commun dans les réseaux après VGG-16.

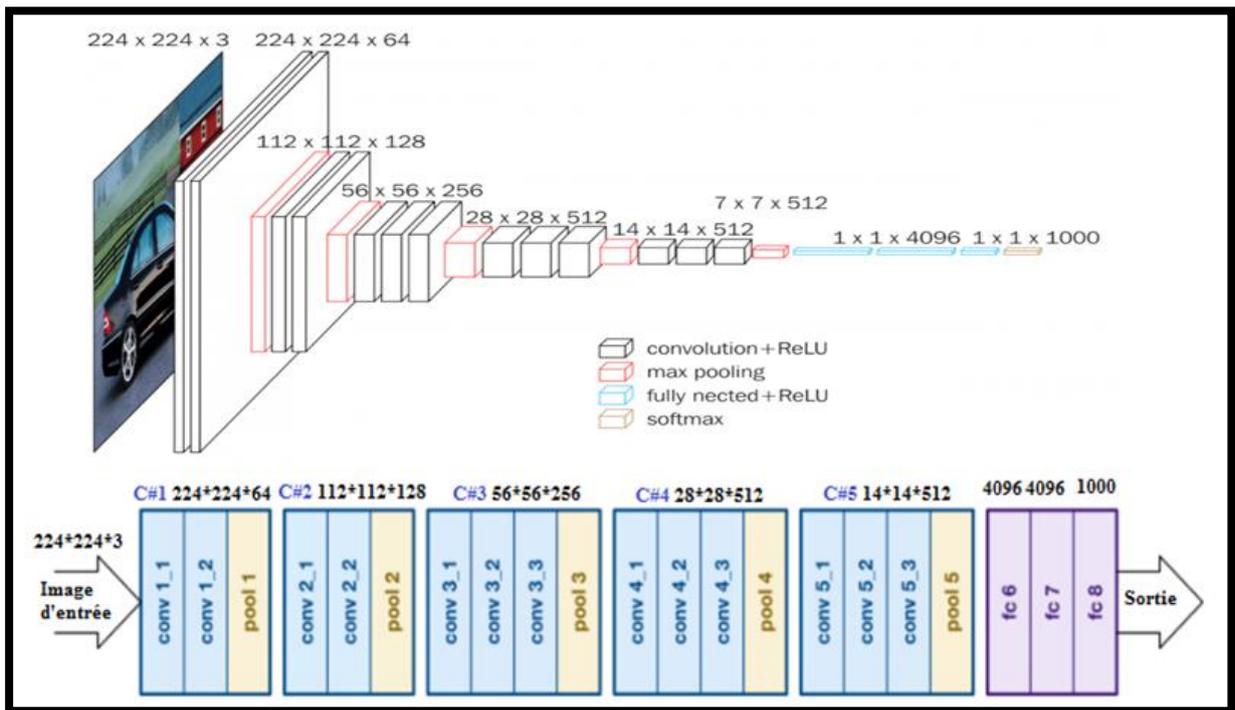


Figure III.20 : L'architecture globale de VGG-16. Elle est composée de cinq blocs de couches de convolution (C#1-C#5) et de trois couches de FC (FC1-FC3).

7.5 Le réseau VGG-19

Le réseau VGG-19 est un autre modèle de réseau CNN très profond, il est proposé par Karen Simonyan et Andrew Zisserman. L'architecture de VGG-19 est similaire de l'architecture VGG-16. Où, il contient cinq blocs de couches de convolution suivi par trois couches entièrement connectées. Cependant, le réseau VGG-19 contient 16 couches de convolution, où il y a 3 couches de convolution supplémentaires représentant la dernière couche de convolution pour les trois derniers blocs par rapport à VGG-16 comme sur la figure III.21 [51].

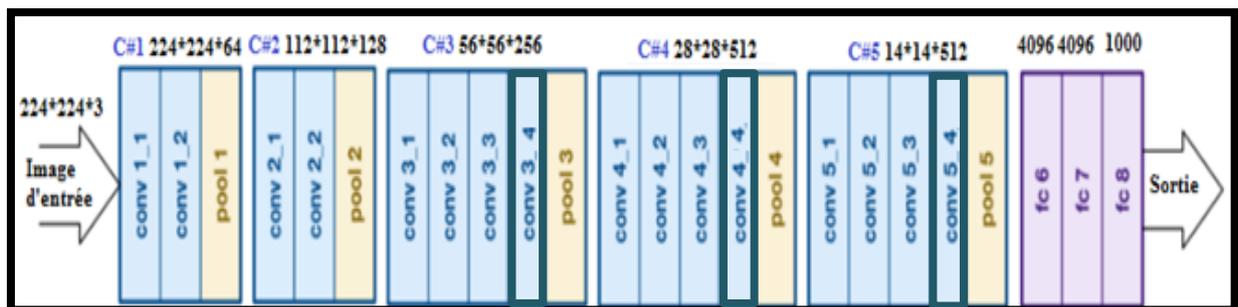


Figure III.21 : L'architecture globale de VGG-19.

7.6 GoogLeNet

En 2014, les chercheurs de Google ont présenté le réseau GoogLeNet qui est un modèle de réseau CNN très profond proposé par Szegedy et al. Il est un modèle gagnant du défi de reconnaissance visuelle de grande taille ImageNet de 2014 (ILSVRC-2014). Il a atteint un taux d'erreur de 6,67% dans le top 5 des tests dans ImageNet. C'était très proche de la performance du niveau humain. Cependant, les réseaux VGG réalisent une précision phénoménale sur ImageNet, son déploiement sur les GPU, même les plus modestes, pose un problème en raison des énormes exigences en matière de calcul, à la fois en termes de mémoire et de temps [52]. Il devient inefficace en raison de la grande largeur des couches convolutives. La principale contribution de GoogLeNet est la mise au point d'un module de démarrage réduisant considérablement le nombre de paramètres dans le réseau. Il est beaucoup plus rapide que VGG. Une autre modification apportée par GoogLeNet a été de remplacer les couches entièrement connectées à la fin par un simple regroupement des moyennes globales qui fait la moyenne des valeurs de canal sur la carte des caractéristiques 2D, après la dernière couche de convolution. Cela réduit considérablement le nombre total de paramètres. L'utilisation d'une largeur et d'une profondeur de réseau importantes permet à GoogLeNet de supprimer les couches FC sans affecter la précision. De plus, ce module utilise le regroupement moyen "Average Pooling" au lieu des couches entièrement connectées, ce qui élimine une grande quantité de paramètres qui ne semblent pas avoir beaucoup d'importance. Leur architecture consistait en un réseau CNN profond de 22 couches mais réduisait le nombre de paramètres de 60 millions (AlexNet) à 4 millions. De plus, un autre point saillant est qu'il possède une couche dite de goulot d'étranglement (convolutions 1×1). Cela contribue à la réduction massive de l'exigence de calcul. Dans une opération de convolution à un emplacement, chaque canal de sortie est connecté à chaque canal d'entrée, ce que nous appelons une architecture de connexion dense. GoogLeNet repose sur l'idée que la plupart des activations dans un réseau profond sont soit inutiles (valeur de zéro), soit redondantes en raison de leurs corrélations. Par conséquent, l'architecture la plus efficace d'un réseau profond aura une connexion fragmentée entre les activations [52]. Donc, GoogLeNet a conçu un module appelé module d'inception, en anglais "*Inception Module*", qui se rapproche d'un CNN clairsemé avec une construction dense normale. Pour chaque cellule, il utilise des convolutions de différentes tailles 1×1 , 3×3 et 5×5 pour capturer des détails à différentes échelles à partir de l'entrée. La figure III.22 montre l'architecture globale de GoogLeNet. Nous pouvons voir qu'il existe de nombreux modules d'inception connectés ensemble pour aller plus profond. En raison de la profondeur de son réseau, il dispose de trois sorties de classification à différentes étapes du réseau pour faciliter la propagation en arrière des erreurs. L'ajout de sorties auxiliaires a principalement profité aux performances finales du modèle, convergeant à une valeur légèrement supérieure à celle de la même architecture réseau sans branche auxiliaire [47].

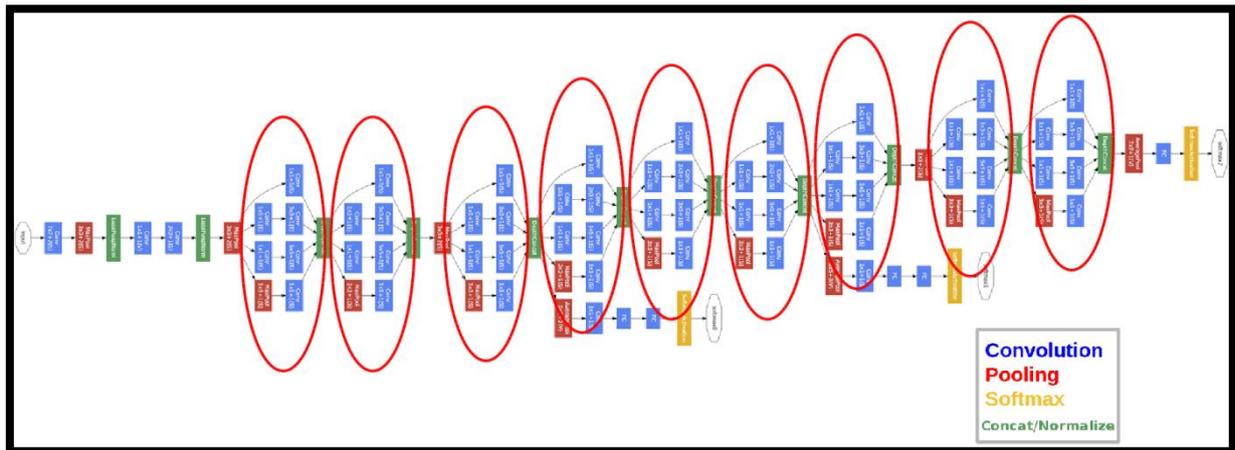


Figure III.22 : L'architecture globale de GoogLeNet.

7.7 ResNet

ResNet est un modèle très profond comparé au modèle précédent AlexNet, VGGNet et GoogLeNet. Le réseau de neurones résiduels développé par Kaiming He et al. a été le gagnant du défi de reconnaissance visuelle de grande taille ImageNet de 2015 (ILSVRC-2015). Il présente des connexions spéciales et une utilisation intensive de la normalisation. Il manque également à l'architecture des couches entièrement connectées à la fin du réseau [52].

Cette architecture a introduit un concept appelé "connexions sautées". Ces connexions de saut sont également appelées unités *gated* ou unités récurrentes *gated* et présentent une forte similitude avec les éléments réussis récents appliqués dans les RNN. Grâce à cette technique, ils ont pu former un NN avec 152 couches tout en ayant une complexité inférieure à celle de VGGNet. Il atteint un taux d'erreur de 3,57% dans le top 5 des tests dans ImageNet, ce qui est supérieur à la performance humaine de cet ensemble de données. Généralement, la matrice d'entrée calcule en deux transformations linéaires avec la fonction d'activation ReLU. Dans un réseau résiduel, il copie directement la matrice d'entrée dans la deuxième sortie de la transformation et additionne la sortie dans la fonction ReLU finale [52].

8. Conclusion

Nous avons présenté, tout au long de ce chapitre, un état de l'art sur l'apprentissage en profondeur. Nous avons expliqué le passage de l'apprentissage automatique à l'apprentissage en profondeur. Par la suite, nous avons présenté les applications et les différentes approches de l'apprentissage en profondeur. En plus, nous avons expliqué l'architecture de réseau CNN et détaillé ses modèles.

Dans le prochain chapitre, nous allons passer revue les travaux existants dans la littérature, qui sont basés sur les méthodes de l'apprentissage automatique, en particulier les méthodes d'apprentissage en profondeur, pour la gestion des catastrophes à partir des données de télédétection.

Chapitre IV

Gestion des catastrophes naturelles par les algorithmes de ML et de DL : état de l’art

À travers ce chapitre, nous passerons en revue les travaux antérieurs portant sur les méthodes de ML, et les méthodes de DL en particulier, pour la gestion des catastrophes naturelles à partir de données de télédétection.

1. Introduction

Au cours des dernières décennies, une augmentation considérable des catastrophes naturelles et de leurs effets sur la population et l’économie a été observée dans le monde entier. Ces catastrophes dévastatrices causent des degrés élevés de dégâts et des milliers de victimes. Pour cela, la rapidité et la précision de l’évaluation des dégâts sont essentiels pour réduire les dommages causés par ces catastrophes et fournissent également des informations importantes pour le sauvetage et l’estimation des pertes économiques. D’autre part, les développements récents de la télédétection ont accéléré la disponibilité des données de télédétection dans le monde entier. Des études antérieures ont montré que l’utilisation des données de télédétection pour l’évaluation des dégâts causés par les catastrophes naturelles est très importante et donnée des informations rapides et essentielles. Plusieurs méthodes automatisées ont été développées pour analyser les images de télédétection pour la détection des dégâts. Récemment, les approches utilisant des algorithmes de ML, et de DL en particulier, se sont révélées performantes pour la détection et la classification des dégâts basées sur les données de télédétection.

Dans le présent chapitre, nous allons passer en revue plusieurs travaux portant sur les méthodes de ML et de DL pour la gestion des catastrophes naturelles à partir des données de télédétection. Dans lequel, nous présenterons les catastrophes naturelles les plus fréquentes et les plus dangereuses, notamment les séismes, les inondations, les incendies, les glissements de terrain, qui causent d’énormes dégâts dans le monde entier et des pertes économiques estimées à des milliards de dollars. Pour chaque catastrophe considérée, nous allons d’abord expliquer son impact sur la vie humaine et l’économie, puis nous allons expliquer sa gestion en exploitant des données de télédétection. Ensuite, nous allons passer en revue les travaux de littérature portant sur les algorithmes de ML et de DL en particulier.

2. Gestion des séismes

2.1 Catastrophes des tremblements de terre

La catastrophe du tremblement de terre ou sismique est considérée comme l’une des catastrophes naturelles les plus destructrices et est devenue plus fréquente autour du monde. Le terme "tremblement de terre" est une bonne description du phénomène naturel qui frappe soudainement une zone, causant des dégâts variables en fonction de l’intensité du séisme et des conditions géologiques locales. Les tremblements de terre frappent soudainement, sans avertissement, et peuvent survenir à tout moment de l’année, jour et nuit. Plus précisément, un tremblement de terre est la secousse de la surface de la terre, résultant de la libération soudaine d’énergie dans la lithosphère terrestre qui crée des ondes sismiques. La taille de tremblement de terre peut aller de ceux qui sont si faibles qu’on ne peut pas les ressentir à ceux qui sont suffisamment violents pour lancer des gens et détruire des villes entières [53] et [69].

Les impacts des tremblements de terre comprennent les pertes humaines avec des morts et des blessures et les dégâts matériels avec la perte d’abris et de moyens de subsistance, la

Chapitre IV – Gestion des catastrophes naturelles par les algorithmes de ML et de DL : état de l'art

perturbation des infrastructures essentielles ou "vital" et la destruction de la communauté. Le tremblement de terre est considéré parmi les catastrophes naturelles les plus meurtrières. Le tremblement de terre peut causer des dégâts majeurs et des décès dans les zones peuplées, mais le tremblement de terre lui-même n'est pas toujours à blâmer. D'autres catastrophes naturelles peuvent être causées par le tremblement de terre et peuvent être également, et parfois plus destructeur. La plupart des décès sont dus à l'effondrement d'un bâtiment ou à des risques secondaires tels que les tsunamis, les incendies, les inondations, les glissements de terrain et la libération de produits chimiques ou de produits toxiques. Les blessures sont généralement dues à des dommages affectant les bâtiments, à la chute ou à la rupture de parties de bâtiments ou de leur contenu, et à l'incapacité de prendre des précautions lors des répliques. Les séismes se produisant près des côtes ou sous les océans peuvent générer des tsunamis avec des vagues de plus de 30 mètres de haut qui peuvent emporter les bâtiments et causer des dégâts considérables dans les régions côtières [54] et [73].

Environ 50 tremblements de terre se produisent chaque jour dans le monde. Les tremblements de terre sont des catastrophes majeures avec une fréquence élevée et constante. Selon l'United States Geological Survey (USGS), il y avait dans le monde plus de dix-sept séismes de magnitude 7 ou plus durant ces dix-huit dernières années. Les grands séismes sont susceptibles de toucher un grand nombre de personnes et une vaste zone géographique. Juste pour énumérer les chiffres de l'un des événements majeurs, le séisme de mai 2008 en Chine a causé la destruction de 420 000 maisons, plus de 70 000 morts et de 5 millions de sans-abri, selon le service d'information sur la recherche du centre d'information multidisciplinaire sur les tremblements de terre. Depuis 2015, plus de 10 000 personnes dans le monde sont mortes des tremblements de terre et des tsunamis. Plus d'un million de séismes se produisent chaque année dans le monde, ce qui équivaut à environ deux séismes par minute. En fait, avec la croissance rapide de l'urbanisation, entre 2001 et 2011, les tremblements de terre ont représenté plus de 60% des décès causés par des catastrophes naturelles, ont provoqué plus de 780 000 décès. Outre ces décès, les séismes ont directement touché 2 milliards de personnes supplémentaires au cours de cette période [54].

La menace de tremblement de terre va probablement augmenter en raison de l'urbanisation mondiale et de la vulnérabilité des grands centres urbains. La plupart des villes les plus peuplées du monde sont sur des lignes de faille et alors des millions de personnes sont exposées aux tremblements de terre. Les tremblements de terre causent non seulement de nombreux décès au départ, mais également de nombreuses blessures graves nécessitant des soins médicaux chirurgicaux et de réanimation complexes. Les tremblements de terre détruisent également les installations médicales, les routes et les ponts, en plus d'interrompre les chaînes d'approvisionnement médicales. En conséquence, les tremblements de terre dévastent la capacité médicale curative locale et créent un besoin important et non satisfait en soins chirurgicaux et médicaux complexes [55].

En outre, les catastrophes sismiques provoquent souvent des dommages environnementaux et structurels importants, ainsi que des pertes économiques de plusieurs milliards de dollars [53]. Le tremblement de terre du Grand Hanshin ou Kobe était un séisme majeur au Japon en 1995. L'épicentre était à Akachi et il était mesuré à 7,2 sur l'échelle de

Chapitre IV – Gestion des catastrophes naturelles par les algorithmes de ML et de DL : état de l’art

Richter. Parmi toutes les grandes villes, Kobe, avec sa population de 1,5 million d’habitants, était la plus proche de l’épicentre et était frappée par les plus fortes secousses. Environ 5 000 personnes ont perdu la vie. Environ 4 600 d’entre eux venaient de Kobe, 350 000 se sont retrouvés sans abri et environ 200 000 bâtiments se sont effondrés. Ce fut l’un des pires séismes au Japon. Des problèmes de centre-ville tels que des zones urbaines densément peuplées avec de vieilles maisons en bois et des rues très étroites ont intensifié les dégâts. Le tremblement de terre a été particulièrement dévastateur car une zone urbaine densément peuplée a été directement touchée par de très fortes secousses au sol [5].

Un puissant séisme a frappé la ville antique de Bam en Iran en 2003. L’épicentre a été signalé par l’USGS à environ 185 km au sud-est de Kerman ville (environ 10 km au sud-ouest de la ville de Bam). Le séisme avait une magnitude locale de 6,6 sur l’échelle de Richter. Le tremblement de terre a été particulièrement dévastateur. La plupart des bâtiments en briques de boue de la ville historique, y compris son ancienne citadelle, se sont effondrés. Le bilan des morts s’élève à 27 000 personnes et à 30 000 blessés. En termes de pertes humaines, le séisme a été le pire de l’histoire iranienne [5]. Le séisme survenu au centre de Java en 2006 s’est produit dans l’état de Yogyakarta, Indonésie. Plus de 5 700 vies humaines ont été perdues, 36 000 personnes ont été blessées en raison du tremblement de terre. Environ 140 000 maisons se sont effondrées et 200 000 ont été endommagés [56] et [57]. Le séisme survenu en Haïti en 2010 de magnitude 7,0, le 12 janvier 2010, a provoqué des dégâts destructeurs à Port-au-Prince et dans les environs, car le séisme s’est produit à faible profondeur et à proximité de la zone urbaine. Plus de 20 000 bâtiments ont été endommagés et environ 10 000 bâtiments ont été complètement effondrés. Le nombre de morts et de blessés était respectivement d’environ 32 000 et 30 000. La perte économique estimée à plus de 8 milliards de dollars américains [58]. Après ce séisme, environ 105 000 maisons ont été complètement détruites, en particulier dans la capitale, Port-au-Prince, où certains bâtiments importants se sont effondrés, tels que la cathédrale de Port-au-Prince et le palais présidentiel.

Les bâtiments sont fondamentaux pour la vie humaine. Cependant, ils sont vulnérables aux catastrophes naturelles. Les bâtiments étaient extrêmement vulnérables aux tremblements de terre, car peu ou pas de plans sismiques ont été échafaudés. Par conséquent, il est essentiel de surveiller l’état des bâtiments et de fournir une évaluation précise des dommages causés aux bâtiments après un séisme, à une échelle détaillée, pour soutenir les activités de secours et d’intervention d’urgence [59].

2.2 Gestion des séismes par télédétection

L’un des problèmes cruciaux après les tremblements de terre est de savoir comment détecter et identifier rapidement et avec précision les zones endommagées ? La superficie et le montant des dégâts, le taux de destruction des bâtiments, le degré des dommages dans la zone touchée, et le type de dommages subis par chaque bâtiment sont des informations essentielles au succès des opérations de sauvetage et de reconstruction dans les zones sinistrées [53].

Traditionnellement, les spécialistes évaluent la zone endommagée, par le biais d’une enquête sur le terrain, à partir d’une inspection manuelle directe sur le terrain. Après le séisme, l’enquête sur le terrain dans les zones touchées avait été menée par de nombreux

Chapitre IV – Gestion des catastrophes naturelles par les algorithmes de ML et de DL : état de l’art

chercheurs pour étudier la distribution des dégâts et leurs causes. Cependant, l’identification de la zone endommagée par l’enquête sur le terrain prend beaucoup de temps. Elle est difficile pour à évaluer les dommages dans des zones inaccessibles et elle est fortement influencée par l’expérience, les compétences et le point de vue des inspecteurs [67]. Pour les activités de secours précoces et la gestion efficace après une catastrophe, il est important de développer une méthodologie permettant d’identifier rapidement la répartition des zones endommagées [57].

Les technologies de télédétection offrent la possibilité d’obtenir des informations sur les dommages plus rapidement et moins chères. La technologie de télédétection est donc un outil très important et plus efficace pour l’évaluation des dégâts sismiques, en particulier pour fournir des informations rapides et fiables aux décideurs, aux intervenants en cas d’urgence [55]. Dans la dernière décennie, divers types de données de télédétection sont utilisés pour le processus d’évaluation des dommages causés par le séisme, tels que l’optique, le SAR et le LiDAR. Afin d’identifier le dégât du tremblement de terre il faut avoir l’expertise pour les reconnaître et ensuite obtenir les données de télédétection appropriées pour avoir des réponses d’une manière efficace [67] et [9].

Les capteurs optiques sont largement utilisés dans l’observation et l’évaluation des dommages après des catastrophes sismiques [5]. De plus, l’utilisation de données optiques est souvent préférée car plus facile à interpréter. Les images optiques sont faciles à interpréter car elles représentent la surface du sol telle qu’elle apparaît à l’œil humain. Des images optiques avec une résolution spatiale inférieure à un mètre peuvent être acquises par de nombreuses plateformes telles que les satellites, les avions pilotés et les drones (UAV) [9]. Les images optiques acquises par les capteurs aériens sont reconnues comme les sources les plus appropriées et facilitent la détection des dommages car ces images capturent à la fois les toits et les façades. Cependant, les capteurs optiques, en particulier, sont limités à la lumière du jour et nécessitent des environnements sans nuages et sans fumée [5].

La capacité des images SAR à être détectées pour la détection des dommages lors de catastrophes naturelles à grande échelle a été démontrée. L’imagerie SAR présente un intérêt particulier lorsque la couverture nuageuse empêche l’acquisition d’imagerie optique. Le grand avantage du SAR est qu’il peut être utilisé quelque soit la lumière du soleil et les conditions météorologiques. La capacité de collecte de données par tous les conditions météorologiques et par tous les temps est offerte par télédétection radar. L’analyse des paramètres des modèles représentant les valeurs d’intensité ou de cohérence rétrodiffusées peut être utilisée pour distinguer les zones endommagées et des zones non endommagées et pour évaluer les dommages. Les principaux problèmes sont les signatures spectrales inhabituelles des capteurs à micro-ondes, qui sont très différentes du système visuel humain et donc difficiles à interpréter, ainsi que la résolution spatiale encore très limitée des capteurs SAR. C’est la raison pour laquelle l’utilisation des données SAR est peu probable dans un premier temps, bien qu’elle soit peut-être la seule donnée disponible en cas de mauvaises conditions météorologiques [49] et [69].

En comparaison avec les capteurs, les capteurs optiques, installés sur une plateforme aériens en particulier, fournissent un fonctionnement plus facile, moins complexe et plus

Chapitre IV – Gestion des catastrophes naturelles par les algorithmes de ML et de DL : état de l’art

fiable, et sont moins coûteux [5]. Les images de la télédétection peuvent être acquises à partir de différentes plateformes avec différentes résolutions spatiales. Les photographies aériennes sont acquises à une résolution spatiale de l’ordre du décimètre, ce qui permet d’identifier et de mesurer la taille des bâtiments, de localiser l’infrastructure de transport et d’évaluer les dommages qui se sont produits. Ces types d’images aériennes peuvent identifier l’état des dommages des bâtiments individuels, mais ils ne peuvent pas couvrir une vaste zone avec un temps d’acquisition. Cependant, des ciels nuageux peuvent empêcher le processus d’acquisition d’images aériennes. Les images satellites ont l’avantage d’être capables d’observer une grande surface en même temps. Si l’on veut obtenir des images sur une grande surface, l’imagerie satellitaire peut être l’outil le plus adapté. Aussi, les images satellitaires n’ont pas des restrictions liées à la météo. Cependant, la plupart des capteurs satellites ont une résolution limitée [5].

La technologie de télédétection est la source la plus prédominante et la plus précoce pour fournir des données permettant d’effectuer de telles évaluations, soit manuellement, soit à l’aide de procédures d’analyse d’images automatisées [55]. L’interprétation visuelle est l’une des méthodes couramment utilisées pour la détection et l’évaluation des dommages après les séismes. Bien que les méthodes d’interprétation visuelle soient actuellement les méthodes les plus précises, car elles peuvent donner des bons résultats pour la détection des dommages et aussi différents interprètes expérimentés peuvent distinguer différents types de sévérités de dommages des images de télédétection, elles nécessitent beaucoup de temps de traitement, ont besoin d’experts expérimentés et conduisent à un travail difficile et laborieux [60].

Pour remédier à ces inconvénients, de nombreuses méthodes automatisées ont été développées dans le but de détecter les zones touchées par des indices de dommages, telles que des tas de gravats, des débris et des effritements dans une région de l’image. Plusieurs études ont montré la détection automatique de régions endommagées à partir des types d’images mentionnés ci-dessus, où les régions lourdement endommagées sont identifiées en reconnaissant les indices de dommages visibles de l’extérieur, tels que l’effritement, les débris, les tas de gravats et les éléments cassés, qui sont des indicateurs puissants de la structure très dommageable [9]. Ces indices de dommages ont une signification spécifique et jouent un rôle majeur dans la classification des dommages. Par exemple, la présence de grandes quantités de tas de débris/gravats autour du bâtiment est un indicateur fort d’un effondrement (partiel) du bâtiment. L’effritement est un indicateur de dommages mineurs ou d’éléments structurels partiellement cassés. Le processus d’identification de ces types de dommages peut être effectué en analysant les caractéristiques extraites par les images de télédétection [53] et [72].

On trouve une quantité considérable d’évaluations des dommages sismiques étudiées dans la littérature utilisant les informations dérivées d’images de télédétection. Ces méthodes sont classées en deux groupes [49] et [53]:

- (1) Des techniques multi-temporelles sont utilisées pour comparer et analyser les changements significatifs entre les images avant et après le séisme. Ces techniques sont largement utilisées pour la détection des régions endommagées, car elles sont rapides et

Chapitre IV – Gestion des catastrophes naturelles par les algorithmes de ML et de DL : état de l’art

automatisées, mais elles présentent certains inconvénients en termes de limitations, comme il est difficile d’obtenir des données proches de la période précédant la catastrophe, c’est-à-dire l’image avant le séisme est difficile à obtenir, il est aussi difficile d’avoir des données avec un court intervalle de temps entre les images avant et après l’événement. De plus, un enregistrement de haute précision entre les données avant et après une catastrophe est délicat et généralement prend beaucoup de temps.

- (2) Les techniques mono-temporelles reposent sur un ensemble de données de télédétection recueillies après un sinistre. Les données post-événement peuvent être rapidement et directement acquises par différents capteurs. Cette méthode est plus rapide et plus pratique pour détecter l’étendue des dommages causés en temps réel, car les données pré-événement ne peuvent pas toujours être disponibles. Même lorsque des données pré-événement sont disponibles, il peut être difficile d’obtenir une paire homogène d’images avant et après le séisme. Les méthodes mono-temporelles reposent principalement sur des méthodes d’apprentissage automatique supervisé conçues pour être efficaces dans la classification des dommages.

2.3 Gestion des séismes par les algorithmes de ML et de DL

2.3.1 Gestion des séismes par les algorithmes de ML

Principalement, les approches utilisant des algorithmes de ML se sont révélées utiles pour la détection et la classification des dommages à partir des images de télédétection, en particulier pour les méthodes mono-temporelles [49]. Comme mentionné précédemment dans le deuxième chapitre, ces approches consistent en deux modules principaux : module d’extraction des caractéristiques et module de classification. Fondamentalement, la performance de ces approches dépend principalement de la puissance des méthodes d’extraction de caractéristiques.

Généralement, le module d’extraction des caractéristiques extrait diverses caractéristiques pour décrire les images d’entrée, telles que les caractéristiques morphologiques, les caractéristiques spectrales, les caractéristiques de texture et les caractéristiques accélérées robustes (SURF), etc. En outre, certaines études ont utilisé des méthodes hybrides pour générer le vecteur de caractéristiques final, telles que les caractéristiques de couleur et de texture, les caractéristiques spectrales et de texture, les caractéristiques morphologiques, textuelles et spectrales [9] et [49]. Basé sur les caractéristiques extraites, le module de classification est conçu pour utiliser un classificateur existant afin de séparer la région en deux classes (endommagées ou non endommagées), par exemple : machine à vecteurs de support (SVM), Random Forest (RF), et Sélection des vecteurs de support et adaptation (SVAS) [53].

Une technique de classification efficace après un séisme a été développée dans [56] pour détecter les régions touchées à partir des données de télédétection. Dans cette étude, les régions endommagées par le séisme survenu au centre de Java en 2006 sont identifiées à partir d’images satellites à haute résolution. Les dommages aux bâtiments ont été identifiés en classant les images satellitaires optiques post-séisme. Dans cette méthode, un indice de dommage a été calculé à partir des images classifiées afin d’évaluer la distribution des

Chapitre IV – Gestion des catastrophes naturelles par les algorithmes de ML et de DL : état de l’art

dommages causés aux bâtiments et les résultats concordaient bien avec ceux observés lors d’une enquête sur le terrain.

Vu et Ban [57] ont développé une approche automatisée basée sur le contenu pour la détection des dégâts sismiques à partir des images satellites à haute résolution. Cette approche repose sur l’identification des régions de débris après le séisme de 2008 au Sichuan en Chine, ce qui a donné de bons résultats pour les régions de test. Les informations contextuelles pertinentes dans cette recherche (y compris la structure, la forme, la taille, la texture des bords, les relations spatiales) décrivant la situation des dommages sont formulées et mises à l’échelle sur une échelle morphologique. Cette approche développée pourrait permettre d’atteindre une précision de plus de 80% pour le calcul du ratio de dommage.

Vu et al. [61] ont développé un algorithme automatisé basé sur les techniques d’apprentissage supervisé pour détecter les régions endommagées par le séisme de Bam à l’aide de données optiques post-événement. Plus de 80% des bâtiments très endommagés ont été classés correctement par rapport aux résultats de l’inspection visuelle.

Ishii et al [62] ont proposé une méthode de détection des régions endommagées à partir des images aériennes à l’aide d’informations sur les couleurs et les bords. A partir des résultats expérimentaux, ils ont confirmé l’efficacité de l’utilisation d’images aériennes en couleurs pour la détection de zones endommagées telles que des bâtiments effondrés.

Yu et al. [63] ont montré que la méthode SVM peut extraire les bâtiments effondrés avec une grande précision à l’aide des images aériennes de la capitale haïtienne, Port-au-Prince, après le séisme du 12 janvier 2010. D’abord, ils ont fait la segmentation d’images en utilisant des paramètres d’échelle et d’homogénéité (couleur, forme et valeur d’altitude). Ensuite, des caractéristiques texturales différentes dérivées de GLCM ont été extraits de l’image. Enfin, les bâtiments effondrés ont été détectés avec un classificateur SVM.

Dans le papier [64], une évaluation des dégâts sismiques a été réalisée avec SVSA de l’image par satellite à haute résolution et deux schémas de classification ont été mis en œuvre : les classifications à base de pixels et les classifications à base de textures et de pixels en combinaison. Avec la classification basée sur les pixels, seules les bandes spectrales, y compris R, G, B et NIR, ont été utilisées lors de la classification. Les caractéristiques de texture ont été obtenues à l’aide de GLCM. Basé sur les deux schémas, les auteurs ont constaté que la performance de la méthode SVSA est meilleure que SVM et KNN en termes de précision de classification.

Sun et al. [65] ont appliqué un classifieur de forêts aléatoires, Random Forest (RF), pour identifier les différents niveaux d’effondrement du bâtiment à partir des images du tremblement de terre de Yushu du 14 avril 2010. Cinq descripteurs de texture ont été utilisés pour l’extraction de texture de région touchée. Les auteurs ont constaté que la plupart des caractéristiques de l’histogramme en niveaux de gris fonctionnent assez bien et que plusieurs autres caractéristiques primitives, telles que les points lumineux isolés provenant de gravats épars, peuvent également aider à discriminer différents niveaux d’effondrement. Les résultats d’interprétation optimaux avec une précision globale de 84,7% ont été obtenus.

Chapitre IV – Gestion des catastrophes naturelles par les algorithmes de ML et de DL : état de l’art

En général, les méthodes basées sur des caractéristiques de bas niveau telles que les caractéristiques de texture, Analyse texture-ondelette, GLCM, Filtre de Gabor, et HOG se sont avérés très performants. Mais l’amélioration des performances de ces caractéristiques est faible et ne se généralise pas bien [9]. En fait, les performances des caractéristiques de bas niveau ont été améliorées lors de l’utilisation d’une stratégie de représentation des caractéristiques telle que le BOW, ce qui est évident dans [9]. Le choix de la stratégie efficace de représentation des caractéristiques joue également un rôle très important dans l’amélioration des performances [9].

Dans le papier [66], une nouvelle approche de la détection des régions endommagées est proposée en utilisant des images de télédétection à haute résolution. La première étape consiste à extraire les descripteurs des caractéristiques dans chaque région de super-pixel. Dans cette approche, la couleur et le HOG sont extraits pour chaque super-pixel en tant qu’un descripteur de caractéristiques. Ensuite, l’image est représentée par BOW. Par conséquent, l’histogramme des régions endommagées et non endommagées peut être obtenu. Basé sur les caractéristiques extraites, un classificateur SVM est utilisé pour séparer les régions endommagées et non endommagées. Les résultats, évalués sur des données réelles, démontrent que cette méthode est réalisable et efficace pour la détection des dommages. Etant donné que l’étape clé pour le modèle BOW est constituée par les descripteurs des caractéristiques, qui déterminent la qualité du résultat de la détection, différents descripteurs de caractéristique locales, notamment Surf et Gabor, ont été comparés à la méthode proposée dans le modèle Visual BOW. Les résultats obtenus ont montré que la méthode proposée est meilleure que les deux autres méthodes en raison de l’effet visuel.

Les auteurs de [9] ont utilisé BOW dans la classification des dommages sismiques et ont démontré qu’elle augmente la précision de la classification par rapport aux représentations globales conventionnelles. Bien que la représentation de BOW améliore la précision de la classification, des problèmes liés à la généralisation restent posés, car elle ne généralise pas assez souvent les localisations géographiques invisibles et aussi lorsque les caractéristiques des images sont fortement variant en termes de la plateforme de capture, de vue du capteur, échelle de l’image et la complexité de la scène. Par conséquent, développer un modèle supervisé robuste et généralisé avec des caractéristiques puissantes pour la classification des dommages après le séisme est fortement souhaitable.

2.3.2 Gestion des séismes par les algorithmes de DL

DL est une méthode récemment bien établie qui a montré des résultats remarquables et qu’elle est capable de dériver de puissantes représentations des caractéristiques dans diverses tâches de vision par ordinateur. Contrairement aux représentations globales conventionnelles, y compris celles qui sont utilisées les méthodes de représentation comme BOW, qui ont besoin d’une quantité considérable de compétences en ingénierie et une expertise de domaine ; la puissance caractéristique d’apprentissage en profondeur est automatiquement apprise à partir des données elles-mêmes [51]. Cela va être à travers les approches d’apprentissage en profondeur telles que les CNN dans de nombreuses applications ; telles que la classification des scènes, la recherche d’images par le contenu, ainsi que la surveillance et l’identification

Chapitre IV – Gestion des catastrophes naturelles par les algorithmes de ML et de DL : état de l'art

des dégâts. De plus, les caractéristiques de DL ont été aussi utilisées pour la détection et la classification d'événements sismiques à partir des images de télédétection [53].

Dans l'étude [21], le potentiel des caractéristiques de DL extraites à partir modèles pré-entraînés de CNN, associé à un classificateur en ligne, a été étudié pour la classification des dommages sismiques à partir d'imagerie satellitaire. En outre, la précision extraite des caractéristiques de CNN a été comparée aux caractéristiques de bas niveaux GLCM et Gabor à l'aide du classificateur en ligne. Les résultats obtenus ont montré que les caractéristiques de CNN se comportent mieux que les deux autres types de caractéristiques.

Dans une étude récente [49], CNN et les caractéristiques tridimensionnelles "caractéristiques du nuage de points 3D", à la fois indépendamment et en combinaison, ont été considérées pour la détection de dommages aux bâtiments causés par des tremblements de terre à l'aide d'images aériennes obliques, et les résultats ont montré que l'intégration des caractéristiques des nuages de points 3D et CNN améliorait considérablement la transférabilité du modèle et que la précision obtenue était améliorée jusqu'à un maximum de 7% comparé au résultat obtenu uniquement par les caractéristiques de CNN. Aussi, cette étude a montré que les caractéristiques de CNN obtenaient de meilleures performances que les caractéristiques de nuage de points 3D.

Les auteurs de [49] ont étudié le transfert des CNN pré-entraîné nommés "Alexnet" pour la classification des dommages. Ils ont utilisé le vecteur d'activation extrait à partir la deuxième couche entièrement connectées (FC2) comme caractéristiques globale de l'image. Ils ont essayé de faire en sorte que les caractéristiques des CNN pré-entraînés se généralisent bien et soient plus expressifs que les caractéristiques de niveau bas et de niveau moyen. De plus, les auteurs ont testé la structure dans trois scénarios différents et ont constaté qu'il n'y avait pas de différence significative dans les performances des caractéristiques apprises par ces différents scénarios. Cependant, les caractéristiques apprises par un modèle CNN pré-entraîné basé sur des images de domaines connexes sont fortement recommandées car elles n'ont besoin que de quelques nombres d'échantillons d'apprentissage par rapport aux deux autres méthodes et sont légères en calcul. Cette approche permet d'atteindre une précision moyenne de détection des dommages d'environ 85% et se généralise bien aux expériences menées pour analyser la transférabilité du modèle basé sur des images de lieux géographiques spécifiques, et a examiné des images de sites invisibles présentant des caractéristiques de scène, une résolution spatiale, des conditions d'éclairage et des caractéristiques radiométriques légèrement différentes.

Le travail [53] a utilisé un modèle de CNN pré-entraîné, nommé VGG-16, comme un outil d'extraction des caractéristiques pour la classification des dommages sismiques. Ce modèle permet d'atteindre à une précision moyenne de détection des dommages d'environ 92.87% et 95.25% pour les bases d'images utilisées aériennes et UAV respectivement.

Dans [58], CNN a été utilisé pour identifier les bâtiments effondrés à partir d'images satellitaires post-séisme. L'objectif de cette étude était d'explorer la performance d'une petite architecture de CNN appelée SqueezeNet en ce qui concerne l'identification de bâtiments

Chapitre IV – Gestion des catastrophes naturelles par les algorithmes de ML et de DL : état de l’art

effondrés à l’aide des images satellitaire après le séisme en Haïti en 2010. Les résultats ont démontré que les informations des bâtiments effondrés peuvent être bien récupérées en utilisant des images post-séisme. SqueezeNet permet de bien classifier les bâtiments effondrés et non effondrés et a atteint une précision moyenne de 78,6% pour toutes les régions testées.

Les auteurs dans le papier [59] ont comparé les performances des caractéristiques de texture et CNN avec le classifieur de forêt aléatoire (Random Forest) pour distinguer les bâtiments effondrés et non effondrés après le séisme en Haïti en 2010, en utilisant des images satellites avant et après l’événement. Un réseau de neurones CNN contenant trois couches de convolution a été mis en œuvre pour apprendre automatiquement les caractéristiques de DL. Les caractéristiques de texture GLCM (notamment le contraste, la dissimilarité, l’entropie, l’homogénéité, la corrélation et le second moment angulaire) ont été extraites à partir de données pré et post-événement. Par la suite, le classifieur de forêt aléatoire a été utilisé comme classifieur pour évaluer les performances de ces deux caractéristiques différentes. Le résultat directement obtenu par un modèle CNN simple a également été considéré. Aussi, le modèle CNN formé peut être utilisé comme un outil d’extraction de caractéristiques et les caractéristiques apprises ont été combinées avec le classifieur de forêt aléatoire. Les résultats ont montré que la caractéristique CNN associée à la méthode de forêt aléatoire obtenait les meilleures performances, avec une précision globale de 87,6%. CNNs ont le potentiel d’extraire des caractéristiques profondes pour identifier les bâtiments effondrés par rapport aux caractéristiques de texture avec la méthode de forêt aléatoire en augmentant la précision globale de 83,4% à 87,6%.

Duarte et al. Dans [67] ont présenté une technique de fusion des caractéristiques basées sur CNN pour identifier une région endommagée après le séisme à partir d’images de télédétection. Pour cela, trois approches de fusion de caractéristiques CNN multi-résolution sont proposées et testées pour effectuer la classification après le séisme. Ils ont constaté que les trois approches hybrides de DL améliorent la performance.

L’article [68] présente une méthodologie visant à augmenter l’efficacité de la détection des dommages sismiques, lorsque la classification des dommages est effectuée à l’aide d’un CNN pré-entraîné et d’un classificateur supervisé SVM sur un ensemble d’images multi-vues d’UAV. Pour réduire le temps de calcul, ils ont dirigé tous les calculs de classification des dommages uniquement vers les régions d’image contenant les façades. Cela augmente la généralisation et la vitesse de détection des dommages sismiques.

Le travail [69] a proposé une méthode de DL, nommée YOLOv3-S-GIoU, pour la détection des bâtiments effondrés à partir d’images post-séisme. Cette étude a appliqué la méthode de détection d’objets (You Only Look Once, YOLOv3) basée sur le réseau de CNN pour détecter les bâtiments effondrés à partir d’images post-séisme. Cette approche permet d’atteindre une précision moyenne de détection des dommages d’environ 90,89%, une meilleure capacité de généralisation, et une vitesse de détection plus rapide.

Les auteurs de [70] ont proposé une méthode basée sur CNN pour l’identification des bâtiments endommagés après le séisme avec des images de différents capteurs et résolutions.

Chapitre IV – Gestion des catastrophes naturelles par les algorithmes de ML et de DL : état de l’art

Ils ont amélioré la précision de leur méthode en combinant des images aériennes et satellitaires de catastrophes.

Bhangale et al. [71] ont conçu un modèle de classification d’apprentissage en profondeur CNN plus performante pour la détection des dommages sismiques basée sur des données de télédétection. Ils ont indiqué que le modèle proposé offre une bonne précision d’environ 85% et une réduction significative du temps de mise en œuvre.

Nex et al. [72] ont créé un réseau CNN pour la détection des dommages structurels visibles à partir d’images satellitaires, aériennes et d’UAV afin de faire la lumière sur ce que les réseaux d’apprentissage en profondeur peuvent actuellement offrir. Ce modèle proposé permet d’atteindre une précision moyenne de détection des dommages d’environ 87.10%, 90.01% et 93.10% pour les bases d’images utilisées : satellites, aériennes et UAV respectivement et une meilleure capacité de généralisation.

Miura et al. [73] ont développé une méthodologie basée sur CNN pour l’identification des dommages sismiques aux bâtiments à partir d’images aériennes post-catastrophe. Ils ont démontré que le modèle CNN proposé serait utile pour l’identification rapide de la distribution des dommages. L’architecture CNN développée dans cette étude a détecté correctement les bâtiments effondrés avec une précision d’environ 95%. Les auteurs ont appliqué cette architecture à des images aériennes d’autre catastrophe (cyclone), avec plus de 90% des dommages sont correctement.

Xiong et al. [74] ont proposé une méthode automatisée d’évaluation des dommages sismiques des bâtiments basée sur CNN à l’aide d’UAV. Un modèle CNN basé sur VGGNet est adopté pour évaluer les dommages sismiques. Le modèle CNN est affiné sur la base d’images de bâtiment étiquetées manuellement obtenues sur Internet. Les auteurs ont indiqué que les dommages peuvent être efficacement obtenus par leur modèle proposé. La distribution des dégâts de la zone est obtenue avec une précision de 89,39%.

Kamilaris et autres dans [110] ont démontré l’efficacité de la méthode de DL, en réalisant un modèle d’apprentissage en profondeur basé sur VGG, pour la surveillance et l’identification des séismes, à partir des images aériennes capturées au moyen d’un UAV.

3. Gestion des inondations

3.1 Catastrophes des inondations

Parmi toutes les catastrophes naturelles du monde, l’inondation est probablement la plus dévastatrice, répandue et fréquente. Les inondations sont plus dangereuses que d’autres catastrophes naturelles, telles que les glissements de terrain, les tremblements de terre et les volcans [75]. Les inondations sont souvent causées par des conditions météorologiques extrêmes. Elles peuvent être causées par de fortes précipitations sur une courte période ou par des précipitations continues sur plusieurs jours. Des inondations peuvent également se produire lorsque la glace ou le vent provoque le débordement d’une rivière ou d’un ruisseau sur les environs [76]. L’occurrence de cette catastrophe devrait augmenter en raison de

Chapitre IV – Gestion des catastrophes naturelles par les algorithmes de ML et de DL : état de l’art

l’augmentation du développement non planifié et de l’urbanisation, de la déforestation accrue et des précipitations continues comme effet du changement climatique dans les régions vulnérables [77].

Il existe plusieurs types d’inondations telles que les inondations fluviales, les crues éclairées, le drainage urbain, les défaillances du sol, les fluctuations du niveau des lacs, les inondations côtières et l’érosion. La dynamique des inondations fluviales varie avec le terrain. Dans les zones plates, des inondations peu profondes, se déplaçant lentement pendant des jours voire des semaines. Dans les régions montagneuses, des inondations peuvent survenir quelques minutes après de fortes pluies ou des orages. Le court préavis, les grandes profondeurs et les vitesses élevées des crues éclairées rendent ces types de crue particulièrement dangereux [78].

Les inondations sont des risques naturels graves, qui posent de grands dommages aux propriétés, aux routes, aux véhicules et même aux personnes, en particulier dans les zones urbaines densément peuplées. Ces dommages apportent une énorme charge utile aux particuliers et aux gouvernements. Les inondations produisent des débris et des déchets, notamment des liquides, des cadavres d’animaux et des matières dangereuses telles que les déchets hospitaliers. Contrairement aux tremblements de terre, dans lesquels les déchets resteront en place, les inondations transporteront les déchets solides avec le débit d’eau. Les débris provoquent des graves menaces sur la santé des personnes et peuvent même bloquer les routes utilisées pour donner une aide d’urgence, aggravant ainsi la situation [79].

Au cours des deux dernières décennies, la fréquence des inondations dans le monde a augmenté de plus de 40%. Dans le répertoire des catastrophes naturelles, les inondations sont souvent les plus dommageables. Les inondations causent des pertes énormes de vies humaines et de dégâts dans les propriétés chaque année dans une partie du monde ou de l’autre. Il est responsable de dizaines de milliers de décès chaque année. En fait, entre 1995 et 2015, environ 109 millions de personnes ont été touchées par les inondations. Avec des dégâts pouvant atteindre 75 milliards de dollars par [79]. Selon le centre de recherche sur l’épidémiologie des catastrophes (Centre for Research of the Epidemiology of Disasters ou CRED), les inondations ont affecté près d’un milliard de personnes au cours de la décennie 2001-2010, entraînant des dommages estimés à environ 142 milliards de dollars américains [80].

Rapporté qu’en 2011 et 2012 seulement, les inondations ont touché environ 200 millions de personnes et entraîné des pertes économiques d’environ 95 milliards de dollars [81]. Les inondations de 2013 ont représenté 31% des pertes économiques résultant de catastrophes naturelles [82]. Selon les États-Unis, les inondations font plus de victimes que les autres catastrophes naturelles [83]. Les inondations causent régulièrement des dégâts catastrophiques et des tragédies humaines, en particulier en milieu urbain, selon aux États-Unis. Le National Weather Service américain de 2016 a signalé environ 126 décès et 10 milliards de dollars de dégâts matériels causés par des inondations. C’est la principale cause de décès liés aux conditions météorologiques aux États-Unis. Par exemple, en octobre 2016, l’ouragan Matthew a provoqué de graves dégâts dans l’est de la Caroline du Nord, en

Chapitre IV – Gestion des catastrophes naturelles par les algorithmes de ML et de DL : état de l’art

dévastant des villes telles que Princeville, Lumberton, Smithfield, Kinston, Fayetteville et Goldsboro. Les dommages qui y sont associés ont été estimés à environ 1,5 milliard de dollars. L’ouragan Florence a provoqué des inondations massives dans le nord et le sud de la Caroline, principalement à la suite des inondations d’eau douce de septembre 2018 [76].

La fréquence des inondations perturbantes augmente et s’accélère dans de nombreux endroits, notamment le long de la côte Est des États-Unis, en raison de l’élévation du niveau de la mer. Sur la base des valeurs moyennes des projections relatives des reflex, d’ici 2050, la plupart des villes côtières américaines connaîtront de fréquentes inondations de trente jours ou plus par an en raison de l’accélération des effets du reflex [84]. En Europe, les inondations sont le type de catastrophes naturelles les plus courantes et représentent 75% de toutes les demandes d’indemnisation résultant des catastrophes naturelles. La Slovénie a été touchée par plusieurs inondations. Les inondations les plus importantes ont eu lieu en septembre 2007 et septembre 2010 [85]. De nombreux pays asiatiques sont fortement touchés par les inondations. La Chine, par exemple, a été le théâtre de la catastrophe naturelle la plus meurtrière du 20ème siècle, l’inondation de 1931 en Chine centrale. Cette inondation a entraîné la mort d’environ 3,7 millions de personnes. Selon les estimations, les dégâts annuels causés par les crues des rivières en Chine s’élevaient à 51 milliards de dollars US en 2010 [79]. Malaisie est l’un des pays qui subissent les effets négatifs des inondations dans certains de ses États, tels que le Kelantan, le Terengganu et de nombreux autres États. Selon les rapports existants, 40% du total des dommages en Malaisie sont causés par des inondations.

Le Département de l’irrigation et du drainage (Department of Irrigation and Drainage ou DID) a déclaré que 9% de la superficie totale de la Malaisie est exposée aux inondations et que les dégâts se chiffrent à environ 0,3 milliard USD par an [77]. En Iran, en particulier dans le nord du pays, des inondations sont observées presque chaque année. Au cours de la dernière décennie, les dommages économiques causés par les inondations ont fortement augmenté, principalement en raison de l’urbanisation croissante entraînant la conversion de forêts en zones résidentielles et la construction de villas [81].

3.2 Gestion des inondations par télédétection

L’extraction rapide et précise des zones inondées est essentielle pour appuyer la planification des interventions d’urgence et fournir une évaluation des dommages dans les mesures spatiales et temporelles [76]. Ainsi, un système capable de détecter les inondations à un stade précoce et d’avertir immédiatement les bureaux concernés sera très utile au public. La détection précoce des inondations peut sauver des vies humaines, du temps et de l’argent pour le gouvernement, ainsi qu’une étape importante pour progresser vers des villes plus intelligentes [86]. Plus précisément, la capacité de cartographier et de surveiller les zones inondées de manière rapide, précise et rentable est essentielle pour les gestionnaires des catastrophes et les autorités nationales, car l’accès à ces informations peut aider à améliorer la gestion des inondations et à atténuer ses effets catastrophiques. Par exemple, les autorités locales ont besoin de ces informations pendant une phase d’urgence pour localiser et identifier les zones touchées, et par conséquent d’organiser des actions de sauvetage et d’atténuation des dommages [87].

Chapitre IV – Gestion des catastrophes naturelles par les algorithmes de ML et de DL : état de l’art

La détection des inondations en temps réel est également essentielle à la préparation aux risques des inondations, car elle permet aux intervenants d’urgence de réagir et de gérer des événements rapides, et de cibler leurs ressources limitées sur les zones les plus prioritaires. Les méthodes de détection des inondations basées sur les enquêtes de terrain sont limitées en termes d’étendue spatiale des zones inondées et peuvent être à la fois laborieuses, des tâches intensives et coûteuses [87].

Dans ce contexte, les données de télédétection ont démontré à maintes reprises leur intérêt et leur utilité au cours des différentes phases du processus de gestion des inondations en fournissant une vue d’ensemble de la situation sur le terrain sans contact direct avec la zone inondée et en permettant aux décideurs de suivre l’étendue de l’eau durant les inondations [88]. Les techniques de télédétection offrent un potentiel de détection efficace des inondations avec les avantages des observations de surface à faible coût, à grande échelle et en temps réel [80]. L’utilisation de données de télédétection pour produire une cartographie des inondations et pour évaluer les risques des inondations presque en temps réel a été populaire au cours des dernières décennies [76]. Par rapport à la mesure sur site, la télédétection peut fournir une couverture complète et continue des inondations, ce qui facilite la surveillance des inondations, d’affiner les plans d’urgence d’intervention et l’évaluation des dommages [83].

En général, les cartes des inondations peuvent être préparées à l’aide de données provenant des satellites, des avions et des UAV. Plusieurs chercheurs ont étudié l’évaluation des risques des inondations à l’aide d’images satellitaire et aériennes pour des projets à grandes échelles [76]. Les images satellitaires offrent une résolution spatiale et temporelle inférieure à celle des avions et des UAV, mais un champ de vision plus grand. Il a été utilisé avec succès dans l’analyse géographique à grandes échelles dans une zone inondable [82].

Dans la littérature, des images radar et optiques ont été utilisées pour identifier les pixels associés aux zones inondées. Il est évident qu’il est plus facile d’identifier les inondations à partir d’images optiques que de les extraire d’images radar en raison de la réponse radiométrique particulière qui caractérise les surfaces d’eau. Des techniques telles que la classification, la segmentation et la fusion ont été appliquées avec succès à l’extraction des inondations à partir d’images optiques [88]. La télédétection optique par la plateforme satellitaire fournit des images en couleur réelles qui permettent une interprétation visuelle et une classification automatique pour fournir des informations de première directes sur les zones inondées. Cependant, l’utilisation de l’imagerie optique pour la cartographie des inondations est limitée par les conditions météorologiques extrêmes, notamment par la présence de nuages. La présence de vastes nuages lors des inondations limite souvent la faisabilité et l’utilité de l’imagerie satellitaire optique [83].

Contrairement aux données optiques, les radars spatiaux permettent l’acquisition de données quelques soient les conditions diurnes et météorologiques et peuvent fournir des informations précieuses pour la surveillance des inondations, car les micro-ondes ont la capacité de pénétrer les nuages, ce qui en fait une alternative efficace aux capteurs optiques pour la différenciation zones inondées et des zones non inondées. Ainsi, les données SAR

Chapitre IV – Gestion des catastrophes naturelles par les algorithmes de ML et de DL : état de l’art

sont particulièrement bien adaptés à la gestion de la cartographie des inondations, grâce à des fonctionnalités telles que la vue synoptique, la capacité de fonctionner dans presque toutes les conditions météorologiques et pendant le jour et la nuit, et la grande sensibilité de la bande hyperfréquence à l’eau qui permet de la distinguer de la terre. De plus, les données SAR peuvent pénétrer dans la végétation et la forêt. Par exemple, les zones inondées peuvent être facilement distinguées des autres types de couverture terrestre si elles ne sont pas perturbées par un vent violent, ce qui entraîne une augmentation de la rugosité de la surface (vagues). Cependant, la répétitivité temporelle des images SAR était un problème critique pour leur utilisation opérationnelle jusqu’à présent pour la surveillance des inondations. Par conséquent, l’un des problèmes majeurs lors de l’analyse d’images satellitaires est la variation des données pour des sites particuliers au fil du temps. Les satellites disponibles au public sont pour la plupart non stationnaires et nécessitent plusieurs jours pour visiter les mêmes lieux [83].

Pour surmonter ce problème, les travaux récents concernent principalement sur l’imagerie aérienne. Comparée aux observations spatiales, la télédétection aérienne est immunisée contre les grands nuages et les revisites-limitations, ce qui se manifeste comme un outil idéal pour la surveillance des inondations. De plus, des images aériennes sont acquises aux niveaux du décimètre et du sous décimètre, représentant des détails beaucoup plus abondants que les données satellitaires. En particulier, les avions pilotés et les UAV sont deux plateformes principales de télédétection aérienne. Le principal inconvénient, des avions pilotés pour la cartographie des inondations urbaines, est qu’il est toujours difficile de trouver des emplacements idéaux pour le décollage et l’atterrissage des avions, en particulier pour les avions à voilure fixe. Cependant, les UAV constituent un moyen beaucoup plus sûr et plus pratique d’acquérir des données dans les zones inondées, car ils peuvent être lancés à la main et atterrir sur un petit terrain plat. De plus, les UAV peuvent être déployés facilement et fréquemment pour répondre aux besoins de surveillance et de cartographie rapides dans les zones inondées à une échelle spatio-temporelle définie par l’utilisateur. La technologie UAV peut générer suffisamment de données plus rapides et plus précises à des coûts bien moindres pour une évaluation rapide des inondations. Plus précisément, les UAV peuvent acquérir efficacement des données haute résolution allant jusqu’au centimètre, pour une détection rapide et précise des zones inondées dans des paysages urbains complexes ainsi que des zones inaccessibles en raison d’environnements dangereux, par rapport à d’autres approches d’acquisition de données. L’extraction précise et immédiate de la région des inondations et son étendue est très importante pour les interventions d’urgence immédiates [76].

Bien que les méthodes d’interprétation visuelle d’images de télédétection soient les méthodes les plus précises et populaires pour détecter les régions d’inondation. Elles peuvent donner de bons résultats pour la détection des dommages, mais elles sont des méthodes longues, imprécises et coûteuses. Elles sont basées sur des connaissances d’experts ; par conséquent, Elles peuvent être erronées [89].

Le seuillage d’une image SAR, Thresholding en anglais, est l’une des méthodes la plus couramment adoptée pour discriminer les zones inondées des zones non inondées. Une telle approche est fondée sur le faible retour radar des corps aérés ouverts qui se comportent comme des réflecteurs spéculaires, mais elle ne donne pas des résultats satisfaisants dans les

Chapitre IV – Gestion des catastrophes naturelles par les algorithmes de ML et de DL : état de l’art

zones de végétation, de forêt et urbaines ou en présence de vent qui rend la surface de l’eau rugueuse. Son avantage est représenté par l’efficacité du calcul qui pourrait le rendre approprié à des fins de cartographie rapide. Cependant, le choix de la valeur de seuil est un aspect critique car il dépend des paramètres environnementaux de la zone d’étude (par exemple, la couverture végétale et l’orographie), ainsi que des paramètres du système (bande de fréquence, polarisation et angle d’observation) [89].

3.3 Gestion des inondations par les algorithmes de ML et de DL

3.3.1 Gestion des inondations par les algorithmes de ML

Les techniques d’apprentissage automatique sont développées pour optimiser l’extraction des régions inondées à partir de données de télédétection [84]. Elles peuvent réduire le temps nécessaire à la détection des inondations [85]. Dans [85], les auteurs ont testé les techniques d’apprentissage automatique pour la détection des inondations. Les résultats ont montré que ces algorithmes pouvaient être utilisés pour détecter les zones inondées avec une grande précision. Des classificateurs connus, SVM, ANN, BN, DT, etc., ont été utilisés pour classer l’image de télédétection en régions inondées et non inondées en fonction des caractéristiques, comme la texture et la couleur [82].

Ichim et Popescu [90] ont présenté une nouvelle méthode basée sur la combinaison des caractéristiques de texture et de couleurs pour la classification des correctifs pour segmenter les zones inondées à partir d’images aériennes UAV. Ils ont présenté une méthode qui combine les informations fournies par divers descripteurs de texture importants, tels que les motifs binaires locaux (LBP), l’histogramme des gradients orientés (HOG) et la dimension fractale en version couleur. Ils ont constaté que la précision de l’approche hybride était meilleure que chaque approche indépendante.

L’étude présentée dans [91] a combiné aussi les caractéristiques de couleur et de texture pour la segmentation d’inondations dans des images UAV. Les auteurs ont présenté une méthode basée sur la combinaison des caractéristiques LBP avec une fractale de couleur pour la classification afin de la segmenter les zones inondées à partir d’images aériennes. La précision de la méthode proposée était meilleure que chaque approche considérée séparément.

Également, Feng et al. [83] ont utilisé les images d’UAV pour la surveillance des inondations en milieu urbain et ont montré que de telles plateformes pouvaient fournir des cartes précises des inondations. Leur objectif général a été de développer une méthode rapide et précise pour identifier des inondations urbaines. La détection d’inondations a été réalisée via l’extraction des caractéristiques à partir d’une GLCM et des méthodes de classifier d’arbres forestiers. Leurs résultats expérimentaux ont indiqué que l’arbre forestier affichait de bonnes performances en cartographie des inondations urbaines avec une précision globale de 87,3%. L’inclusion des caractéristiques de texture peut améliorer considérablement la précision de la classification avec une augmentation de 11,2%. La comparaison avec d’autres classificateurs a révélé que le DT et le ANN, tout en affichant des performances similaires à celles de la SVM. Les résultats obtenus ont démontré qu’un UAV est une plateforme

Chapitre IV – Gestion des catastrophes naturelles par les algorithmes de ML et de DL : état de l’art

exceptionnelle pour la surveillance des inondations urbaines et que la méthode hybride proposée peut fournir des résultats d’extraction précis dans des paysages urbains complexes.

Le classifieur ANN a été utilisé dans la référence [82] pour la détection des zones inondées dans les données UAV. Sa nouveauté réside dans la combinaison des caractéristiques pour différents canaux de couleur avec des caractéristiques sur la distribution spatiale des pixels obtenues à partir de la matrice de cooccurrence chromatique et de la dimension fractale de masse.

La recherche [77] a étudié la potentielle de la SVM à la détection des régions inondées aux inondations à Terengganu, en Malaisie. Les auteurs ont conclu que la technique SVM est un outil efficace et fiable pour l’évaluation de la susceptibilité aux inondations. Les cartes de sensibilité aux inondations qui en résultent peuvent être utiles aux stratégies d’atténuation des inondations.

Le papier [80] a introduit une nouvelle approche pour la surveillance des inondations grâce à l’utilisation combinée d’images satellitaires optiques et radar. La méthode proposée applique SVM et le modèle de contour actif sans arêtes pour la détermination de l’étendue de l’eau. L’approche proposée convient particulièrement à la surveillance des inondations à grande échelle. Elle a été testée lors d’une grave inondation survenue dans le nord-est de la Chine en août 2013, qui a provoqué d’immenses pertes en vies humaines et en biens immobiliers. Des précisions globales élevées de 97,46% pour l’imagerie optique et de 93,70% pour l’imagerie radar sont obtenues grâce à l’utilisation de l’approche présentée dans [80].

3.3.2 Gestion des inondations par les algorithmes de DL

Ces dernières années, la classification des images de télédétection inondée utilisant les CNN a été considérablement améliorée par rapport les méthodes de ML. Les CNNs ont démontré d’excellentes performances dans diverses tâches, notamment la classification des images, l’extraction de caractéristiques et la segmentation [76].

Le papier [76] étudie le potentiel des approches CNN pour extraire les zones inondées à partir d’images UAV. Un réseau CNN basé sur VGG (FCN-16) a été utilisé dans cette recherche. Le modèle FCN-16 a été mis au point et formé pour extraire les zones inondées. Les résultats de la segmentation d’images obtenus à partir de FCN-16 ont été comparés à partir des résultats obtenus à partir de FCN-8, FCN-32 et SVM. Les résultats expérimentaux ont montré que les réseaux CNN pouvaient extraire précisément des zones inondées à partir d’images d’UAV par rapport aux classificateurs classiques tels que SVM. La précision de classification obtenue par les FCN-16, FCN-8, FCN-32 et SVM pour la classe d’eau était respectivement de 97,52%, 97,8%, 94,20% et 89%. Généralement, les résultats expérimentaux ont montré qu’ils avaient réussi à obtenir des résultats de classification prometteurs, même si seulement cent images d’UAV étaient disponibles pour l’apprentissage.

Les auteurs de [86] proposent l’utilisation d’une architecture CNN pour détecter les inondations dans certaines zones sensibles. Ils ont utilisé l’architecture d’apprentissage approfondie d’AlexNet pour former et tester leur base de données. Les images de leur base de

Chapitre IV – Gestion des catastrophes naturelles par les algorithmes de ML et de DL : état de l’art

données sont collectées à partir de deux caméras PTZ (une caméra pan -tilt -zoom capable de contrôler la direction et le zoom à distance) avec des angles de vue différents. Selon les résultats expérimentaux, leur système a obtenu une précision de classification supérieure à 95% sur les deux caméras.

Les auteurs dans [75] ont proposé plusieurs méthodes pour identifier les inondations des images à haute résolution de télédétection en utilisant le DL. Ils ont proposé et analysé la diversité de quatre réseaux profonds distincts pour effectuer la détection des zones des inondations dans les images de télédétection. Les résultats expérimentaux ont montré que leurs méthodes sont efficaces et robustes. Ces résultats montrent que les approches proposées sont efficaces et robustes pour identifier les zones des inondations. Le processus d’identification réalisé à l’aide des algorithmes proposés peut aider les autorités à rester centrées sur les régions les plus vulnérables tout en surveillant les inondations prévues, ce qui peut faciliter la coordination des opérations de sauvetage et aider les victimes.

Bischke et al. [92] ont proposé une approche pour la segmentation des zones inondées sur des images satellites utilisant des modèles CNN. Dans leur approche, ils ont analysé différentes caractéristiques de CNN pour obtenir une représentation d’image puissante. Leurs résultats montrent que les caractéristiques extraites par des réseaux profonds peuvent être utilisées efficacement pour extraire des régions inondées dans l’imagerie satellitaire.

Said et al. [93] ont utilisé des approches de DL pour la classification des inondations et la détection des conséquences des inondations. Ils ont proposé trois techniques de fusion différentes, la fusion précoce, tardive et hybride, en utilisant quatre CNN disponibles dans le commerce comme extracteur de caractéristiques, notamment AlexNet, VGGNet et ResNet. Ils ont combiné les caractéristiques extraites à travers différents modèles profonds à la fusion précoce. Dans la fusion tardive, ils ont considéré la moyenne des résultats de détection des inondations obtenus à travers les modèles individuels. Ils ont combiné les résultats obtenus à partir des deux premières techniques dans leur troisième technique de fusion. Pour la classification, ils ont appliqué un classifieur SVM dans toutes les analyses de fusion soumises. La méthode proposée a obtenu les meilleurs résultats sur MediaEval 2018.

Rahnemoonfar et al. [94] ont proposé un nouveau réseau de neurones récurrent et un CNN pour la segmentation sémantique des zones inondées sur un grand ensemble de données UAV. Le réseau proposé a été appliquée aux images aériennes des zones inondées à Houston, TX. Les résultats expérimentaux démontrent l’efficacité du réseau proposé pour la segmentation sémantique, à la fois qualitativement et quantitativement. Ils ont atteint une précision de 96% dans la détection des zones inondées dans une grande base de données d’UAV.

Dans [95] une méthode efficace de détection de zones inondées à partir d’images UAV a été proposée, dans laquelle l’élément clé est un réseau neuronal de rétro-propagation défini avec une couche de convolution personnalisée basée sur une banque de filtres de Gabor. La nouveauté de cette méthode réside dans le fait que le processus d’acquisition des images d’UAV est définitivement influencé par les conditions météorologiques et que la solution

Chapitre IV – Gestion des catastrophes naturelles par les algorithmes de ML et de DL : état de l’art

proposée par CNN est capable de détecter et de différencier les zones inondables d’autres régions (végétation, maisons, routes ou terrains) dans n’importe quelle orientation des textures présentées dans les images analysées. Malgré le fait que les images testées étaient complexes, les résultats expérimentaux obtenus sur plus de 50 images d’UAV ont montré de bons résultats en termes de précision de la détection d’inondation.

Kamilaris et autres dans [110] ont démontré l’efficacité de la méthode d’apprentissage en profondeur et créé un modèle d’apprentissage en profondeur basé sur VGG, pour la surveillance et l’identification des catastrophes liées à l’inondation, à partir de photos aériennes capturées au moyen d’un UAV.

4. Gestion des incendies

4.1 Catastrophes des incendies

Dans la Grèce antique, le feu était considéré comme l’un des éléments majeurs de l’eau, de la terre et de l’air. Le feu est constitué d’une collection de millions d’atomes. Cela ne peut pas être appelé en tant que matière, mais c’est un effet secondaire visible et tangible du changement de forme de la matière. Il est généré par la réaction chimique entre une sorte de carburant et l’oxygène de l’atmosphère. Cela peut être dû à divers facteurs tels que les types de matériaux, la dimension ou la disposition des bâtiments, etc. [96]. Pendant longtemps, les incendies ont été une source de problèmes. Les incendies, y compris les incendies de forêts, sont des phénomènes naturels qui causent des dommages économiques importants et ont des effets dévastateurs sur l’environnement dans le monde entier et qui entraîne parfois des pertes en vies humaines et laisse beaucoup de dégâts matériels à l’environnement et aux propriétés qui prennent des années et beaucoup d’argent pour restaurer l’environnement [97].

Les forêts peuvent purifier l’eau, stabiliser le sol, recycler les nutriments, tempérer le climat et stocker le carbone. Ils peuvent créer un habitat pour la faune et nourrir des environnements riches en diversité biologique. Ils peuvent aussi contribuer à des milliards de dollars à la richesse économique d’un pays. Cependant, des centaines de millions d’hectares de forêts sont malheureusement dévastés chaque année par des incendies de forêts. Un feu de forêt ou un incendie de forêt désigne tout brûlage incontrôlé de végétation naturelle (herbe, arbuste, bois de la forêt, litière et déchets) dans la zone de nature sauvage. Lorsque la couche de végétation est plus précisément connue, les noms plus spécifiques pourraient être utilisés comme feu de forêts, feu d’herbe ou feu de brousse. Les feux de forêts représentent une menace constante pour les systèmes écologiques, les infrastructures et les vies humaines [98].

En raison des incendies de forêt, plusieurs centaines de millions d’hectares de forêts et d’autres végétaux sont détruits chaque année. Parfois, les incendies de forêts ont forcé l’évacuation des communautés sensibles, en plus de lourds dégâts s’élevant à des millions de dollars [98].

D’après l’enquête forestière réalisée en Inde, 19,27% ou 63,3 millions d’hectares de terres indiennes ont été classés en tant que zones forestières, dont 38 millions d’hectares sont accumulés avec une grande quantité de ressources (densité supérieure à 40%), une menace

Chapitre IV – Gestion des catastrophes naturelles par les algorithmes de ML et de DL : état de l’art

énorme. La dégradation provoquée par les incendies de forêt met en péril les forêts indiennes. Les incendies ont provoqué d’énormes dégâts en 2007, touchant de vastes territoires en plus du nombre important de victimes humaines [99].

Les statistiques du centre national de lutte contre les incendies (NIFC) aux États-Unis montrent que la superficie brûlée a doublé de 1990 à 2015 aux États-Unis. Des incendies de forêts récents dans le nord de la Californie ont déjà entraîné plus de 40 morts et 50 disparus. Plus de 200 000 habitants ont été évacués en urgence. Les incendies se produisent 220 000 fois par an dans le monde, la superficie brûlée annuelle dépassant 6 millions d’hectares [100].

Selon le rapport 2015 de la National Fire Protection Association (NFPA), un total de 1345500 incendies a eu lieu aux États-Unis seulement, entraînant une perte de 14,3 milliards de dollars, 15 500 blessés par des civils et 3 280 morts parmi les civils incendiés. En outre, des civils ont été blessés par balles et ils sont morts toutes les 33,5 minutes et 160 minutes, respectivement. Parmi les décès dus à un incendie, 78% sont dus uniquement à des incendies domestiques [101].

Selon les pronostics, les incendies de forêt, y compris leur destruction dans les forêts tropicales humides, réduiront de moitié le peuplement forestier mondial d’ici à 2030. Chaque année en Europe, plus de 10 000 km² de terrains forestiers sont brûlés et en Russie et aux États-Unis plus de 100 000 km². Le Global incendie au cours de l’année 2008 pour les mois d’août et février. Le fait que plus de 20% des émissions mondiales totales de CO₂ proviennent d’incendies de forêts indique qu’il s’agit d’un phénomène qui doit être traité avec une grande attention [102].

4.2 Gestion des incendies par télédétection

L’option pratique et efficace pour minimiser les dégâts causés par un incendie consiste à détecter les incendies à leurs débuts et à réagir rapidement pour éviter leur propagation [99]. Des efforts importants sont donc déployés pour parvenir à une détection précoce des incendies, qui est traditionnellement basée sur la surveillance des incendies par des observateurs humains placés pendant 24 heures sur des sites de surveillance sélectionnés [103]. Les incendies de forêt et de nombreux autres phénomènes, en particulier ceux qui ont une grande étendue spatiale, font partie des phénomènes spatiaux qui contiennent des modèles spatiaux prévisibles qui sont évidents grâce aux images/produits de télédétection [99].

L’application de la télédétection est actuellement une méthode importante pour la surveillance des incendies de forêts, en particulier dans les zones vastes et isolées. Les chercheurs ont présenté différentes méthodes de détections des incendies de forêts. Où ils obtiennent des informations utiles à partir d’images dans les données de télédétection de la forêt et les utilisent pour déterminer les régions à risque d’incendies en utilisant des algorithmes d’apprentissage automatique [99].

L’approche actuelle en matière de détection des feux de forêt consiste à utiliser un seul véhicule aérien habité, un équipement au sol ou des satellites. Toutefois, les véhicules aériens avec équipage sont généralement gros et coûteux, et nécessitent donc des pilotes humains

Chapitre IV – Gestion des catastrophes naturelles par les algorithmes de ML et de DL : état de l’art

expérimentés. De plus, les environnements dangereux et la fatigue de l’opérateur peuvent potentiellement menacer la vie du pilote. Les mesures au sol sont généralement limitées dans la portée et la manœuvrabilité de la surveillance [98]. Bien que de nos jours le système basé sur les satellites peut surveiller une grande surface, mais la résolution de l’imagerie satellite est faible. Un incendie est détecté lorsqu’il a beaucoup augmenté, il est donc impossible de le détecter en temps réel. De plus, ces systèmes sont très coûteux. Les conditions météorologiques (nuages, par exemple) réduiront considérablement la précision de la détection des feux de forêt par satellite, du fait des limitations dues à la longue période de balayage et à la faible résolution des satellites [102], et les trajectoires de vol et les mises à jour technologiques des satellites fixes orbitaux sont également moins souples.

Par conséquent, le déploiement de drones à petite échelle pour la détection des incendies de forêt est une option naturelle et de plus en plus réaliste. Les UAVs peuvent offrir les avantages suivants :

- (1) Couvrir de vastes zones, en particulier par temps nuageux ;
- (2) Travail de jour, de nuit et de longue durée ;
- (3) Facilement récupérable et relativement peu coûteux ;
- (4) Dans le cas des UAV électriques, c’est aussi un avantage pour l’environnement ;
- (5) Transporter différentes charges utiles pour différentes missions ;
- (6) Couvrir efficacement la zone cible ;
- (7) Et surtout, les missions peuvent être réalisées de manière autonome sans ou avec une implication minimale du pilote/exploitant [98].

Pour éviter les dégâts causés par un incendie à grande échelle, une détection incendie rapide et précise est cruciale. Plus tôt un incendie est détecté, plus de chances d’une meilleure survie. En outre, il est également essentiel de bien comprendre le développement de l’incendie et son emplacement. Le lieu de l’incendie initial, la taille de l’incendie, la direction de propagation de la fumée, le taux de croissance de l’incendie sont des paramètres importants qui jouent un rôle important dans l’analyse de la sécurité et la lutte contre les incendies/atténuation, et sont essentiels pour évaluer le risque d’escalade [104].

4.3 Gestion des séismes par les algorithmes de ML et de DL

4.3.1 Gestion des séismes par les algorithmes de ML

Les recherches antérieures sur la détection d’incendie basées sur le traitement d’images reposent principalement sur des modèles de couleur et de texture [100]. Puisque la couleur est la principale caractéristique visuelle du feu ; les caractéristiques de couleur sont généralement utilisées comme une étape de prétraitement pour la détection d’un incendie potentiel. La plupart de ces méthodes tirent parti des propriétés discriminantes de l’espace colorimétrique pour obtenir des zones d’incendie dans les images [98].

Chen et al. [105] ont utilisé les caractéristiques de couleur et de mouvement basées sur le modèle RGB pour extraire le feu et la fumée réels dans les séquences vidéo. Ils ont utilisé une analyse dynamique des flammes correspondant aux caractéristiques désordonnées des

Chapitre IV – Gestion des catastrophes naturelles par les algorithmes de ML et de DL : état de l’art

flammes avec la croissance des pixels pour vérifier l’existence d’un incendie. Cependant, comme ils mesurent la différence entre les images consécutives, la règle de décision ne permet pas de distinguer les régions d’incendie réelles des régions en mouvement ou du bruit.

Vipin a proposé une méthode de classification des pixels d’incendie utilisant des modèles à base de règles dans deux espaces colorimétriques [102].

Angayarkkani et Radhakrishnan ont présenté un système intelligent de détection efficace des incendies de forêt utilisant des données spatiales, où le réseau neuronal à fonction de polarisation radiale a été testé sur des images numériques des incendies de forêt. Régie par le processus de segmentation, l’image est d’abord convertie de l’espace colorimétrique RGB en espace colorimétrique XYZ, puis une diffusion anisotrope est appliquée pour identifier la région d’incendie [99].

Yuan [98] a proposé une méthode de détection et de suivi des feux de forêt basée sur un UAV. Il a développé un ensemble d’algorithmes de suivi des feux de forêt comprenant le filtrage médian, la conversion d’espace colorimétrique et le seuil d’Otsu. L’idée de base de cette méthode est d’adopter le canal "a" dans le modèle colorimétrique Lab pour extraire les pixels de feu en utilisant les caractéristiques chromatiques du feu. Les résultats des tests expérimentaux démontrent que des performances supérieures peuvent être obtenues en utilisant le canal "a" dans le modèle colorimétrique Lab dans la détection des incendies de forêt sur des images. Il a été constaté que la segmentation dans un canal "a" dans le modèle colorimétrique Lab peut fournir les performances de segmentation incendie les plus satisfaisantes par rapport aux autres modèles colorimétriques [98].

En effet, les recherches antérieures sur la détection des incendies basées sur la vision artificielle sont divisées en deux catégories : la détection de flamme et la détection de fumée [104]. Sous réserve des méthodes utilisées, ces travaux sont fiables sur des scènes spécifiques. Beaucoup sont testés sur des vidéos ou des images à courte portée avec un arrière-plan peu infesté. Le volume de référence de test existant est très petit [100]. Il existe des techniques de détection de fumée et de flammes basées sur les méthodes de vision par ordinateur conventionnelles. Dans des travaux antérieurs, des ondelettes, une machine à vecteurs de support et des automates finis flous étaient utilisés pour la détection de fumée et de flammes [106]. Des méthodes statistiques et d’apprentissage automatique ont été rapportés. Le modèle de mélange gaussien est utilisé dans la détection de flamme dans [107].

Töreyn et les autres associent un modèle de couleur générique basé sur un espace colorimétrique RGB, des informations de mouvement et une analyse améliorée du scintillement du feu selon le processus de Markov pour créer un système de détection d’incendie global. Cependant, l’utilisation de la valeur empirique du nombre de mélanges peut ne pas donner les meilleurs résultats [107].

Également, le classifieur SVM est appliqué dans [108]. Ko et les autres ont proposé une méthode de classification non linéaire utilisant SVM et des cartes de luminescence, montrant que la méthode est robuste dans plusieurs scénarios par rapport aux caractéristiques utilisées précédemment pour la détection de flamme. Il est à noter que lorsque des descripteurs de

Chapitre IV – Gestion des catastrophes naturelles par les algorithmes de ML et de DL : état de l’art

caractéristiques couramment utilisés, tels que les histogrammes de gradients orientés (HOG) de transformée de caractéristique invariante à l’échelle (SIFT), sont utilisés avec ces classificateurs, le taux de fausse alarme n’est pas suffisamment faible.

Cependant, ces travaux sont basés sur des techniques de traitement d’images avec des extracteurs de caractéristiques générées par un expert de domaines, "handcrafted feature extractors". Les résultats dépendent fortement de la précision des paramètres sélectionnés manuellement.

4.3.2 Gestion des séismes par les algorithmes de DL

Jusqu’à présent, la difficulté majeure de la classification des images d’incendies de forêt est l’absence de marques d’identification unifiées, les caractéristiques de la couleur, de la forme, de la texture (fumée, flamme ou les deux à la fois) et de l’arrière-plan pouvant varier considérablement d’une scène à l’autre. Plus précisément, la difficulté fondamentale est comment construire un classificateur adaptatif pour identifier les caractéristiques d’incendie complexes de différentes couleurs, formes et textures à partir d’un arrière-plan varié et encombré. Il a été rapporté que l’architecture CNN atteignait des performances de pointe dans un large éventail de tâches de reconnaissance d’images. Son architecture et son schéma d’apprentissage conduisent à un extracteur efficace de fonctions complexes de haut niveau hautement robuste aux transformations d’entrée [100].

Cependant, la technique d’apprentissage en profondeur a rarement été utilisée pour la détection d’incendies de forêt. Jusqu’à récemment, l’apprentissage en profondeur était signalé dans la reconnaissance du feu.

Les travaux de [100] et [106] a également bénéficié du succès de DL. Ces travaux ont amélioré la performance du système de surveillance des incendies de forêt en développant une nouvelle architecture CNN basée sur l’imagerie UAV. Kim a proposé un modèle CNN à huit couches dans la classification des images au feu [106]. Le base de données d’image d’apprentissage est recadrée et redimensionné manuellement. En outre, l’effet des paramètres clés et des coefficients tels que le taux d’abandon (dropout ratio), le taux d’apprentissage n’a pas été examiné. Zhao et les autres a présenté un réseau CNN à 15 couches nommé Fire_Net en tant qu’extracteur et classificateur de fonctions d’incendie à auto-apprentissage. Pour réduire la taille du réseau, il a été proposé de détecter la saillance afin de localiser et de segmenter la zone centrale d’incendie. Fire_Net garantissait une vitesse de traitement moyenne de 41,5 ms par image pour une inspection en temps réel des feux de forêts avec une précision globale de 98%. Le travail est inspirant, mais il est bien connu que le feu est plus facile à détecter que la fumée. Une comparaison supplémentaire a été effectuée à l’aide de quatre modèles : architecture DCNN à 13 couches, Fire_Net, architecture à 15 couches DCNN et architecture à 20 couches DCNN, dans lesquels Fire_Net et l’architecture à 20 couches DCNN ont obtenu la plus grande précision de validation de 0,98 sur un jeu de données augmenté basé sur la valeur [100].

Le travail [101] a amélioré la précision de la détection d’incendie. Les auteurs proposent une architecture CNN de détection d’incendie rentable pour les vidéos de surveillance. Le

Chapitre IV – Gestion des catastrophes naturelles par les algorithmes de ML et de DL : état de l’art

modèle s’inspire de l’architecture GoogleNet et a été mis au point avec un accent particulier sur la complexité informatique et la précision de la détection. Des expériences ont prouvé que l’architecture proposée par [101] dominait les méthodes de détection d’incendie existantes basées sur des caractéristiques de bas niveaux ainsi que la méthode de détection d’incendie basée sur l’architecture AlexNet.

Le travail [109] propose un réseau de neurones à convolution (CNN) pour identifier un incendie dans des vidéos. Testée sur des séquences vidéo réelles, leur approche permet d’obtenir de meilleures performances de classification que certaines des méthodes de détection d’incendie vidéo conventionnelles pertinentes et indique que l’utilisation de CNN pour détecter les incendies dans des vidéo est très prometteur.

Kamilaris et les autres dans [110] ont démontré l’efficacité de la méthode d’apprentissage en profondeur, créer un modèle d’apprentissage en profondeur basé sur VGG, pour la surveillance et l’identification des catastrophes liées aux incendies de forêts, à partir de photos aériennes capturées au moyen d’un UAV.

Les auteurs de [111] ont proposé une méthode de détection précoce des incendies basée sur des CNN ajustés, fine-tuned CNNs. En intégrant des caractéristiques profondes dans son infrastructure, ils ont montré que les incendies peuvent être détectés plus rapide avec une précision accrue dans divers environnements intérieurs et extérieurs tout en minimisant les fausses alarmes incendie. Par des expériences sur des vidéos contenant des objets en mouvement semblables à des incendies et des incendies réels dans des environnements intérieurs et extérieurs, ils ont confirmé que son infrastructure pouvait détecter les incendies à un stade précoce avec une bonne précision et un minimum de fausses alarmes incendie, tout en assurant une réponse autonome et un contenu représentatif sous surveillance, ce qui peut grandement faciliter les systèmes de gestion des catastrophes.

5. Gestion des glissements de terrain

5.1 Catastrophes des glissements de terrain

Les glissements de terrain sont des phénomènes naturels très complexes et dangereux ayant des effets catastrophiques sur la vie et les propriétés des humains dans le monde entier. Un glissement de terrain est décrit comme un mouvement large dans l’étendue du sol, qui inclut des chutes de pierres, un écoulement de débris peu profonds et des ruptures de talus profonds [112]. Les glissements de terrain, qui constituent l’un des risques géologiques les plus courants au monde, font chaque année des milliers de victimes et de morts, des centaines de milliards de dollars de dégâts et des pertes environnementales [113]. Les glissements de terrain se produisent fréquemment dans les zones montagneuses et les collines [114]. De plus, les glissements de terrain provoquent des situations d’urgence lorsque des zones urbaines ou des constructions, telles que des bâtiments, des ponts, des voies ferrées et des routes, sont affectés [115]. Bien que les glissements de terrain se produisent principalement localement, ils peuvent causer des dommages considérables aux infrastructures naturelles et humaines à différentes échelles dans les zones montagneuses et les collines. En plus des dommages

Chapitre IV – Gestion des catastrophes naturelles par les algorithmes de ML et de DL : état de l’art

physiques, les glissements de terrain ont un impact économique et social direct à long terme sur une zone d’habitation humaine [116].

L’incidence des glissements de terrain est liée à un ensemble de facteurs déclencheurs, tels que les précipitations intenses, les éruptions volcaniques, la fonte rapide des neiges, les niveaux d’eau élevés et les tremblements de terre [114]. Au fil des ans, les glissements de terrain ont gravement menacé de vastes zones situées entre ville et montagne [117]. Un rapport récent sur une catastrophe dans le monde (Fédération internationale des Sociétés de la Croix-Rouge et du Croissant-Rouge, 2001) indique que les inondations, les avalanches et les glissements de terrain représentent 42% l’incidence mondiale des catastrophes naturelles, la perte économique moyenne due aux glissements de terrain se monte à des milliards de dollars (comme au Japon, en Inde, en Italie et aux États-Unis) pour des millions de pays comme le Canada, le Népal et la Suède. En outre, des enquêtes menées entre 1964 et 1999 ont montré qu’il y avait une augmentation constante du nombre de catastrophes causées par des glissements de terrain dans le monde [118].

Chaque année, les glissements de terrain causent 1 000 décès et environ 4 milliards USD de pertes de biens dans le monde [112]. Pour la Chine, de nombreuses régions ont été gravement touchées par des glissements de terrain et ceux-ci ont gravement menacé l’environnement, les colonies de peuplement et les installations industrielles au cours des dernières années [113]. Les inondations et les glissements de terrain en Chine de mai à août 2010 se sont classés au deuxième rang en termes de dommages économiques causés par les catastrophes naturelles, avec des dommages d’un montant de 18 milliards de dollars américains. Au cours de cette catastrophe naturelle, les mouvements de masse ont tué 1 765 personnes et se sont classés dans le top 10 des catastrophes les plus importantes selon le nombre de personnes tuées. Les conditions déclenchant des glissements de terrain, telles que les fortes pluies, les typhons ou les tremblements de terre, peuvent affecter de très vastes zones et parfois provoquer plusieurs milliers de glissements de terrain par événement, par exemple, plus de 9 000 glissements de terrain détectés après le typhon Morakot à Taiwan en août 2009 [119].

Les glissements de terrain de type séisme ont longtemps été reconnus comme des risques secondaires importants des tremblements de terre dans les régions montagneuses et des contributeurs potentiels aux budgets d’érosion à l’échelle de la montagne. Les glissements de terrain associés aux tremblements de terre peuvent contribuer de manière significative à la perte de vies humaines. Par exemple, environ 20 000 des 100 000 décès causés par le tremblement de terre de Mw7,9 en 2008 à Wenchuan en Chine ont été directement attribués à des glissements de terrain. Les débris générés par les dizaines, voire les centaines de milliers de glissements de terrain associés aux grands séismes constituent également une source de sédiments potentiellement importante [120]. Le séisme dévastateur de MW 7,8 de Gorkha en avril 2015 a eu son épicentre à Barpak, à environ 77 km au nord-ouest de Katmandou. Le tremblement de terre et les répliques qui ont suivi ont causé environ 9 000 morts, 23 000 blessés et des dégâts considérables dans les propriétés publiques et privées. C’est devenu la pire catastrophe naturelle qui ait frappé le Népal après le tremblement de terre Népal-Bihar de 1934. Les glissements de terrain ont tué plus de 2 000 personnes dans différentes régions du

Chapitre IV – Gestion des catastrophes naturelles par les algorithmes de ML et de DL : état de l’art

pays. Dans le village de Ghodatabela et Langtang, dans la vallée de Langtang, une avalanche de rocher-neige a emporté environ 329 personnes, alors que plus de 25 personnes ont perdu la vie au camp de base de l’Everest par une autre avalanche liée au même séisme. Les zones les plus durement touchées par les glissements de terrain sont les colonies de peuplement situées autour des districts de Sindhupalchok, Gorkha, Rasuwa, Dhading, Nuwakot et Dolakha [121].

5.2 Gestion des glissements de terrain par des données de télédétection

Les caractéristiques des glissements de terrain rendent la détection et l’identification des glissements de terrain difficiles [118]. La détermination rapide et précise de l’emplacement du glissement de terrain est très importante. La cartographie rapide des glissements de terrain est cruciale pour détecter l’étendue de la zone touchée, y compris le niveau et le type de dommage, pour mesurer l’ampleur et analyser la susceptibilité, les dangers et le risque de séismes, ainsi que pour examiner les schémas de répartition et prévoir les paysages touchés par un glissement de terrain [115]. La cartographie rapide de l’emplacement et de l’ampleur de la déformation superficielle provoquée par de glissements de terrain peut fournir des indications importantes sur la réaction rapide des autorités de protection civile, pour les opérations de secours, humanitaires et de reconstruction dans la zone de crise, ainsi que pour la conception et le déploiement de systèmes de surveillance efficaces [122].

Les enquêtes sur le terrain traditionnelles pour la détection de glissements de terrain sont coûteuses, laborieuses et prennent beaucoup de temps, ce qui a rendu difficile l’obtention d’informations précises et opportunes sur les glissements de terrain, en particulier dans les zones montagneuses. Une enquête sur le terrain peut être utilisée pour vérifier les résultats de l’interprétation de photographies aériennes et de l’analyse d’images satellitaires [118]. De plus, lors d’événements rapides à très rapides, c’est-à-dire des taux de déformation de l’ordre de plusieurs mètres par heure, l’accès à la zone de glissement de terrain peut être trop difficile, rendant les enquêtes sur le terrain trop dangereuses [115].

Par conséquent, la détection de glissement de terrain basée sur la télédétection joue un rôle essentiel en raison de sa réponse rapide, de son large champ de vision et de son coût relativement faible. En raison de la technologie de télédétection, l’efficacité de la surveillance des catastrophes a été améliorée. La technologie de télédétection change le statu quo et présente des avantages évidents pour la surveillance des catastrophes naturelles en raison de sa rapidité, de sa portée et de son économie [117].

Les données de télédétection les plus fréquemment utilisées pour la cartographie rapide des glissements de terrain sont les images satellitaires optiques à très haute résolution. Cependant, étant donné que les fortes pluies sont l’un des déclencheurs les plus fréquents des glissements de terrain, c’est-à-dire qu’il existe une forte probabilité de couverture nuageuse immédiatement après l’événement, les données optiques ne sont pas toujours utiles pour les applications cartographiques rapides [115]. Parmi les données de télédétection, le SAR a aussi démontré son efficacité pour surveiller les changements à la surface de la Terre produits par des processus naturels et induits par l’être humain [122]. Dans la plupart des cas, les données de télédétections SAR pour une zone de crise donnée sont disponibles plus tôt que l’imagerie optique sans nuage. Par conséquent, les procédures de cartographie rapide basées sur les SAR

Chapitre IV – Gestion des catastrophes naturelles par les algorithmes de ML et de DL : état de l’art

permettent une réaction plus rapide aux catastrophes. Cependant, tant pour les capteurs optiques que pour les capteurs SAR, le temps de visite du satellite doit être pris en compte [115].

En plus des données satellitaires de cartographie des glissements de terrain, il existe également d’autres données utilisant par exemple LiDAR, qui utilisent des émetteurs et récepteurs lasers actifs pour acquérir des données d’élévation. De plus, LiDAR a la capacité unique de pénétrer dans des zones à végétation dense et de fournir des informations détaillées sur les terrains à forte densité ponctuelle. En outre, il décrit les caractéristiques de la surface du sol et fournit des informations utiles sur les caractéristiques topographiques dans les zones où les zones de glissements de terrain sont obscurcies par la couverture végétale [114]. Malgré leurs résultats prometteurs et contrairement aux données satellitaires, ces données ne sont pas adaptées à une utilisation dans le monde entier en raison de leur disponibilité limitée [115].

Il existe de nombreuses études sur la détection de glissements de terrain à l’aide de données de télédétection. Certaines études utilisent la méthode de la différence d’image pour la détection des glissements de terrain. Bien que des résultats encourageants aient été obtenus, la plupart des analyses reposent toujours sur une interprétation manuelle ou un traitement d’image assisté par l’homme. Bien que l’interprétation manuelle soit fiable, l’inconvénient majeur est qu’elle prend beaucoup de temps, ce qui entraîne un gaspillage de main-d’œuvre et de ressources matérielles. L’automatisation peut constituer une solution pour remplacer les procédures fastidieuses généralement effectuées par des opérateurs humains. Le défi consiste à substituer une interprétation visuelle des images de télédétection, généralement réalisée par des interprètes photo-experts, à des techniques de classification automatique ou semi-automatique [118].

Il existe deux types de méthodes de détection des glissements de terrain : l’une est la méthode qui n’a besoin que des données post-événement ; l’autre nécessite des données pré-événement et post-événement pour obtenir des modifications temporaires. Ce dernier permet de mieux réduire les zones de recherche et de définir plus clairement la période. Les deux effectuent une classification en utilisant des modèles formés à partir d’échantillons de glissements de terrain, qui sont extraits par des experts [117].

5.3 Gestion des glissements de terrain par les algorithmes de ML et de DL

5.3.1 Gestion des glissements de terrain par les algorithmes de ML

Pour identifier les régions de glissements de terrain, les professionnels ont utilisé la détection visuelle pour reconnaître les glissements de terrain dans les images de télédétection et extraire des éléments tels que la texture, les ombres, la couleur et les tons [117]. Les caractéristiques spatiales, spectrales, de forme et de texture ont été extraites sur la base des images satellitaires optiques pour la classification des glissements de terrain dans l’article [118].

Chapitre IV – Gestion des catastrophes naturelles par les algorithmes de ML et de DL : état de l’art

Récemment, trois algorithmes d’apprentissage automatique sont utilisés afin de produire des cartes de susceptibilité aux glissements de terrain pour le pays de Jumunjin en Corée du Sud dans [112], et leurs résultats de modèle finaux sont comparés les uns aux autres, tel que le meilleur modèle donne une précision de 85,2%.

L’étude [114] propose l’utilisation de réseaux de neurones récurrents (RNN) et de réseaux de neurones de perceptrons multicouches (MLP-NN) pour la détection de glissements de terrain à partir de données LiDAR à haute résolution. Les méthodes proposées ont été testées dans les Cameron Highlands, en Malaisie. Le modèle RNN a montré une meilleure précision dans les zones d’analyse et de test que le modèle MLP-NN.

Les auteurs dans [115] présentent une procédure rapide et transférable pour la cartographie des glissements de terrain, basée sur la détection de changement entre l’imagerie optique pré-événement et l’imagerie radar post-événement. La méthodologie a été appliquée avec succès à deux glissements de terrain de caractéristiques différentes.

Litao Wang [124] a proposé une méthode de détection du glissement de terrain basée sur les caractéristiques de texture. Les résultats montrent que l’algorithme basé sur les caractéristiques de texture peut extraire efficacement les caractéristiques du terrain avec une meilleure précision. Cependant, le risque d’erreur est accru quant à gérer l’image entière directement lors de la détection [117].

Wan Bao-feng et les autres ont proposé algorithme d’identification par l’image de télédétection basé sur l’analyse de texture [125].

Wei Xing [126] a proposé une méthode de reconnaissance d’image basée sur le SVM, qui associe les caractéristiques de texture d’image, les caractéristiques de couleur de différents espaces colorimétriques, puis la combinaison de caractéristiques dans le classificateur SVM afin d’identifier l’apparition du glissement de terrain.

Évidemment, cette méthode nécessite des caractéristiques de conception manuelles telles que la couleur, la texture, la forme, la position, etc. Les caractéristiques de conception manuelle requièrent beaucoup d’expérience, elles doivent comprendre très bien leurs applications et leurs données, elles doivent également faire l’objet d’un débogage important des caractéristiques de conception, et ne s’appuyer que sur la caractéristique pour distinguer seulement dans un certain environnement, la capacité de généralisation et les perspectives d’application sont faibles [123].

Une sélection efficace des caractéristiques, telles que la texture, les caractéristiques sur la bande d’images et les caractéristiques géométriques, est nécessaire pour améliorer la qualité de la cartographie de glissements de terrain. Cependant, la gestion d’une quantité énorme de caractéristiques non pertinentes peut conduire à un sur-apprentissage. L’identification des glissements de terrain dans une zone particulière peut être améliorée en sélectionnant les caractéristiques les plus significatives.

Chapitre IV – Gestion des catastrophes naturelles par les algorithmes de ML et de DL : état de l’art

D’après Van Westen et les autres [127], la sélection des caractéristiques les plus significatives peut aider à distinguer les glissements de terrain des non glissements de terrain. Des chercheurs tels que Stumpf et Kerle [128] ont tenté d’améliorer l’efficacité de la sélection des caractéristiques dans la détection de la région des glissements de terrain. Onze caractéristiques optimisées pour la classification des glissements de terrain ont été sélectionnées à l’aide de l’algorithme génétique (GA), qui donne la meilleure valeur de condition physique pour la classification des glissements de terrain dans [118]. Certains travaux basés sur des objets se sont penchés sur la sélection des caractéristiques pour détecter les glissements de terrain à l’aide de données LiDAR. Récemment, Pradhan et Mezcal [129] ont utilisé efficacement une méthode de sélection pour optimiser la sélection des caractéristiques permettant de détecter les glissements de terrain dans les zones tropicales.

5.3.2 Gestion des glissements de terrain par les algorithmes de DL

Les modèles CNNs sont plus appropriées pour la détection des risques de glissement de terrain. Cette méthode n’a pas besoin de concevoir manuellement les caractéristiques, ni de choisir le classificateur, mais en fonction des données originales avec l’étiquette [123]. Alors que les CNN ont atteint de bonnes précisions pour la reconnaissance d’objet dans les images aériennes, seules quelques études utilisent des méthodes d’apprentissage en profondeur et les CNN pour la détection des glissements de terrain [116].

Ding et al. [117] ont évalué leur méthode CNN pour la détection des glissements de terrain à partir d’images de Shenzhen, Chine. Ce papier présente une nouvelle approche pour la détection de glissements de terrain basée sur CNN et de changement de texture. Cette approche utilise des réseaux de neurones de convolution pour extraire les caractéristiques, puis trouver des zones prometteuses où un glissement de terrain peut survenir. Leur méthode automatisée de détection des glissements de terrain a atteint un taux de détection des glissements de terrain de 72,5%.

Yu et al. [123] ont utilisé CNN pour leurs recherches conjointement avec un algorithme amélioré de croissance de région (RSG_R) pour la détection des glissements de terrain. Ils ont formé leur méthode CNN sur un ensemble d’images de glissements de terrain, extrait les informations discriminantes telles que la surface et les limites des glissements de terrain avec l’algorithme RSG_R, et ont conclu à une grande précision de détection pour l’identification des caractéristiques des glissements de terrain.

L’étude [116] utilise des données optiques pour analyser le potentiel des méthodes d’apprentissage automatique, c’est-à-dire les réseaux de neurones artificiels (ANN), les machines à vecteurs de support (SVM) et les forêts aléatoires (RF), ainsi que différents réseaux de neurones de convolution à apprentissage profond (CNN) pour la détection de glissements de terrain. Ils concluent que la méthode CNN améliore la détection des glissements de terrain.

6. Conclusion

Les études antérieures ont montré d’un part, que l’utilisation des images de télédétection pour l’évaluation des dégâts causés par les catastrophes naturelle est très importante et données des informations rapides et essentielles pour le sauvetage et les secours. D’autre part, l’utilisation des images de télédétection pour la détection des dégâts est considérée comme l’un des problèmes les plus difficiles.

Principalement, les images de télédétection sont très complexes et très diverses dans leurs caractéristiques telles que l’échelle de l’image, vue de la caméra, la plateforme de capture et la complexité de la scène. En plus, le contenu de ces images est lié à une catastrophe naturelle. Par conséquent, le choix d’une approche performant pour détecter les dégâts causés par les catastrophes naturelle en utilisant d’images de télédétection est un grand défi et aussi un problème difficiles.

Dans la littérature, plusieurs approches ont été développées pour analyser les images de télédétection pour la détection des dégâts. En plus, les approches utilisant des algorithmes de ML, et de DL en particulier, se sont révélées performantes pour la détection des dégâts.

Récemment, les travaux antérieurs ont montré que les approches utilisant des algorithmes de DL comme CNN sont capables de généraliser bien et d’optimiser les résultats de la détection des dégâts par rapport à celle qui utilisent des algorithmes de ML. Nous avons aussi confirmé ce résultat dans notre travail.

Nous avons présenté, au long de ce chapitre, les catastrophes naturelles qui sont considérés les plus destructives et plus fréquente autour du monde, notamment : les tremblements de terre, les inondations, les incendies, les glissements de terrain. Pour chaque catastrophe considérée, nous avons d’abord expliqué sa gestion en exploitant des données de télédétection, puis nous avons passé en revue les travaux de littérature portant sur les algorithmes de ML et de DL.

Dans le prochain chapitre, nous allons décrire notre proposition pour la classification des dommages sismiques basés sur les données de télédétection aériennes, puis présenter que les résultats de validation obtenus.

Chapitre V

Optimisation de la classification de dommages via des représentations hybrides de caractéristiques de DL dérivées d'images aériennes post-séisme

À travers ce chapitre, nous présenterons notre proposition pour la gestion des catastrophes naturelles, qui sont sismiques dans notre cas, basée sur des données de télédétection. Nous allons d'abord présenter la description de notre méthode proposée qui est basée sur les méthodes de DL pour la classification de dommages sismiques. Par la suite, nous allons présenter les expériences de validation et les résultats de comparaison de notre méthode proposée.

1. Introduction

Dans les chapitres précédents, nous avons présenté les différentes méthodes de gestion des catastrophes naturelles à partir des données de télédétection, Nous avons expliqué dans le premier chapitre tout ce qui concerne les catastrophes naturelles et les méthodes pour les gérer, en particulier les méthodes qui sont basées sur les algorithmes d'apprentissage automatique. Dans le deuxième chapitre, nous avons abordé les deux différentes phases d'apprentissage automatique. Où nous avons vu les différentes méthodes et les techniques connues dans la boîte à outils de la littérature pour chaque phase, lors de la phase d'extraction des caractéristiques et ensuite la phase de classification. Ensuite, dans le troisième chapitre, nous avons détaillé tout ce qui concerne les méthodes de DL, où nous avons présenté les différentes approches de DL, en particulier le réseau CNN et les modèles que nous allons utiliser dans notre proposition. Dans le quatrième chapitre, nous avons présenté quelques travaux antérieurs qui sont basés sur les méthodes de ML et les méthodes de DL en particulier, à partir des données de télédétection, pour la gestion de quelques catastrophes dévastatrices comme les séismes, les inondations, les incendies et les glissements de terrain.

Dans ce chapitre, nous présenterons notre contribution pour la classification des dommages sismiques basées sur les données de télédétection aériennes, ainsi que les résultats de validation obtenus. Premièrement, nous donnerons l'architecture globale du système proposé, puis nous détaillerons en étudiant séparément chacun de ses deux phases principales. Enfin, nous présenterons notre conception expérimentale, puis analyserons et interpréterons les résultats correspondants.

Il est à noter que les résultats de ce travail décrit dans ce chapitre ont fait l'objet d'une publication dans une revue internationale "International Journal of Image and Data Fusion" [130].

2. Approche proposée

Notre travail vise à construire un modèle d'apprentissage supervisé qui est basé sur les méthodes de DL en utilisant des caractéristiques hybrides pour la classification des dommages sismiques. Notre architecture proposée pour ce but est composé de deux phases principales :

- (1) La phase d'extraction des caractéristiques qui est considérée comme cruciale dans le procédé de notre architecture et s'exécute basé sur plusieurs modèles de CNN ;
- (2) Et la phase de classification des dommages qui est implémenté par un classificateur SVM afin de séparer les images d'entrée en deux classes différentes endommagées et non endommagées.

La figure V.1 illustre le schéma général de notre architecture proposée. Plusieurs vecteurs de caractéristiques sont d'abord extraits de plusieurs modèles de CNN différents. Ensuite, toutes ces caractéristiques CNN sont combinées pour construire un vecteur de caractéristique globale de l'image d'entrée.

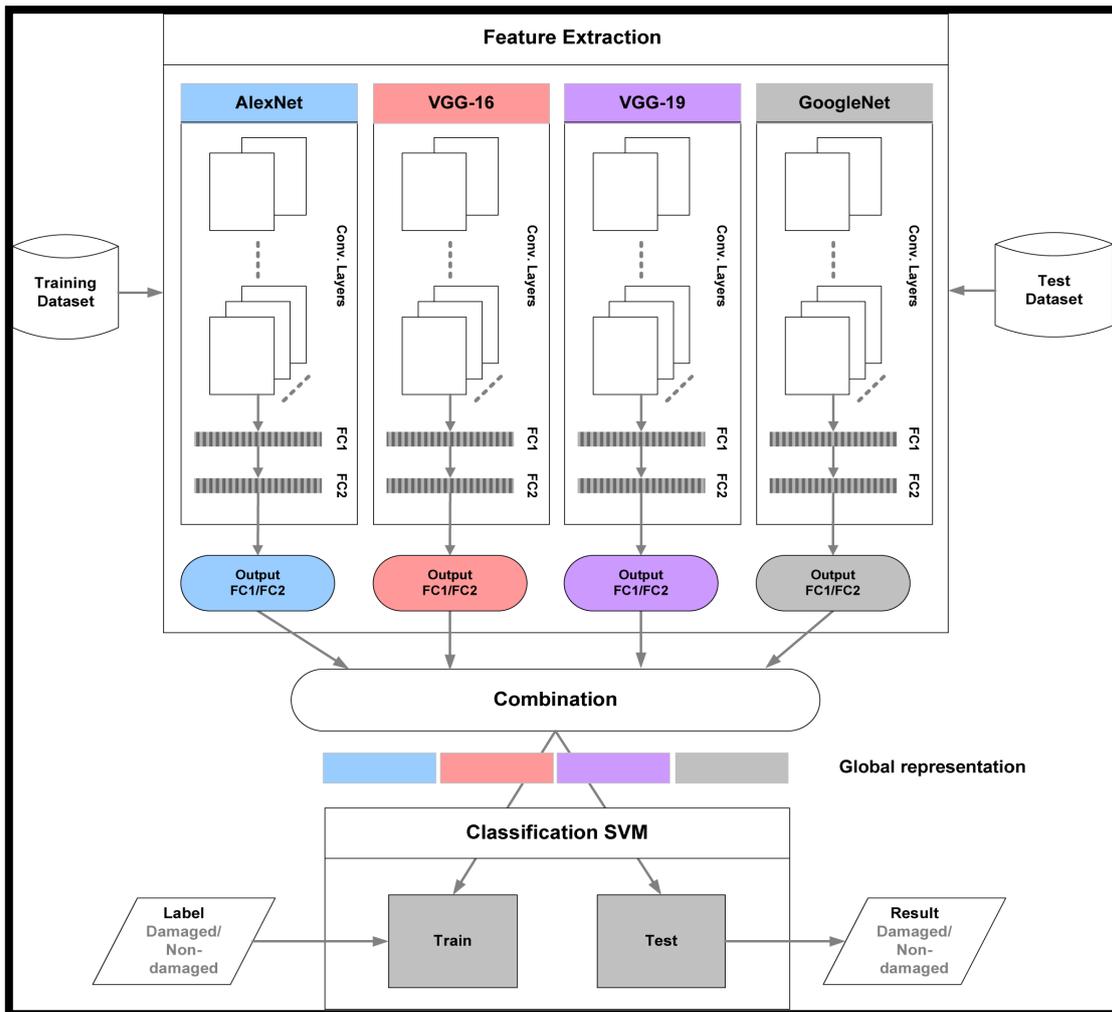


Figure V.1 : Schéma général de notre architecture proposée pour la classification des dommages.

En dernière étape, un classificateur supervisé SVM prend ce vecteur de caractéristiques combiné comme entrée pour donner le résultat de classification.

2.1 La phase d'extraction des caractéristiques

La phase d'extraction des caractéristiques est le processus de transformation des pixels d'image d'entrée à des vecteurs de caractéristiques appropriés. Cette phase est une étape essentielle et cruciale dans notre architecture, car la performance de classification des dommages dépend en grande partie du choix de vecteurs de caractéristiques employés et aussi relative à la qualité des techniques d'extraction des caractéristiques utilisées. En raison que les images de télédétection sont très complexes et très diverses dans leurs caractéristiques, la méthode d'extraction des caractéristiques joue un rôle important pour améliorer la performance de la classification des dommages sismiques. Plus précisément, la performance de classification dépend principalement de la puissance des méthodes d'extraction de caractéristiques. Dans notre travail, on se concentre sur ce point. Notre contribution principale est l'utilisation d'une approche hybride d'extraction des caractéristiques pour classer les dommages post-séismes. Cette approche hybride est basée sur des caractéristiques de DL

extraites à partir de plusieurs modèles pré-entraînés de CNN. Pour ce faire, on passe par deux étapes :

2.1.1 Les caractéristiques extraites à partir d’un modèle CNN particulier

L’extraction des caractéristiques à partir de chaque modèle CNN est la première étape du processus d’extraction de caractéristiques. Pour ce faire, nous avons évalué quatre modèles de CNN pré-entraîné, décrits dans le troisième chapitre : AlexNet, VGG-16, VGG-19 et GoogLeNet. Ces modèles ont été largement utilisés avec succès et ont démontré une capacité de représentation des caractéristiques robustes et efficaces dans nombreux des applications tels que la classification des dommages [49] et [88]. Dans les travaux précédents, ces modèles pré-entraînés sont utilisés comme un outil d’extraction de caractéristiques en supprimant la dernière couche entièrement connectée. Généralement un vecteur d’activation de l’une des deux premières couches entièrement connectées (FC1 ou FC2) et est directement utilisé comme la représentation finale des caractéristiques des images. Contrairement aux études précédentes, dans ce travail, nous avons extraits les vecteurs d’activation à la fois de deux premières couches entièrement connectées (FC1 et FC2) de chaque modèle CNN utilisé. Puis, nous avons combiné ces deux vecteurs de caractéristiques pour obtenir un vecteur hybride de caractéristiques à partir de chaque modèle CNN utilisé. Dans les expériences, le vecteur hybride de chaque modèle CNN est évalué et testé pour la classification des dommages post-séisme.

2.1.1.1 Les caractéristiques extraites à partir d’AlexNet

Le succès d’AlexNet a popularisé l’application des grands CNN dans les tâches de reconnaissance visuelle, et par conséquent, AlexNet est devenu une architecture de base des CNN modernes [51]. AlexNet a été utilisés avec succès dans nombreuses applications, y compris la classification des dommages [93]. Par conséquent, nous avons aussi utilisés AlexNet dans notre travail.

Le tableau V.1 présente l’architecture détaillée d’AlexNet à partir de Matlab. Car les échantillons ont des dimensions différentes, tous les échantillons sont redimensionnés à 227×227 pixels. De plus, AlexNet donne un vecteur de caractéristique de 4096 dimensions à partir de FC1 ou FC2. Par conséquent, le vecteur hybride de caractéristiques obtenu à partir AlexNet peut être considéré comme une collection de 8192-dimensions, comme indiqué dans la figure V.2.

Tableau V.1 : L’architecture détaillée d’AlexNet à partir de Matlab.

Layer Number	Layer Name	Layer Role	Layer Description
1	'data'	Image Input	$227 \times 227 \times 3$ images with 'zerocenter' normalization
2	'conv1'	Convolution	96 $11 \times 11 \times 3$ convolutions with stride [4 4] and padding [0 0 0]
3	'relu1'	ReLU	ReLU
4	'norm1'	Cross Channel Normalization	cross channel normalization with 5 channels per element

Chapitre V – Optimisation de la classification des dommages via des représentations hybrides de caractéristiques de DL dérivées d'images aériennes post-séisme

5	'pool1'	Max Pooling	3 × 3 max pooling with stride [2 2] and padding [0 0 0 0]
6	'conv2'	Convolution	256 5 × 5 × 48 convolutions with stride [1 1] and padding [2 2 2 2]
7	'relu2'	ReLU	ReLU
8	'norm2'	Cross Channel Normalization	cross channel normalization with 5 channels per element
9	'pool2'	Max Pooling	3 × 3 max pooling with stride [2 2] and padding [0 0 0 0]
10	'conv3'	Convolution	384 3 × 3 × 256 convolutions with stride [1 1] and padding [1 1 1 1]
11	'relu3'	ReLU	ReLU
12	'conv4'	Convolution	384 3 × 3 × 192 convolutions with stride [1 1] and padding [1 1 1 1]
13	'relu4'	ReLU	ReLU
14	'conv5'	Convolution	256 3 × 3 × 192 convolutions with stride [1 1] and padding [1 1 1 1]
15	'relu5'	ReLU	ReLU
16	'pool5'	Max Pooling	3 × 3 max pooling with stride [2 2] and padding [0 0 0 0]
17	'fc6'	Fully Connected	4096 fully connected layer
18	'relu6'	ReLU	ReLU
19	'drop6'	Dropout	50% dropout
20	'fc7'	Fully Connected	4096 fully connected layer
21	'relu7'	ReLU	ReLU
22	'drop7'	Dropout	50% dropout
23	'fc8'	Fully Connected	1000 fully connected layer
24	'prob'	Softmax	Softmax
25	'drop7'	Classification Output	crossentropyex with 'tench', 'goldfish', and 998 other classes

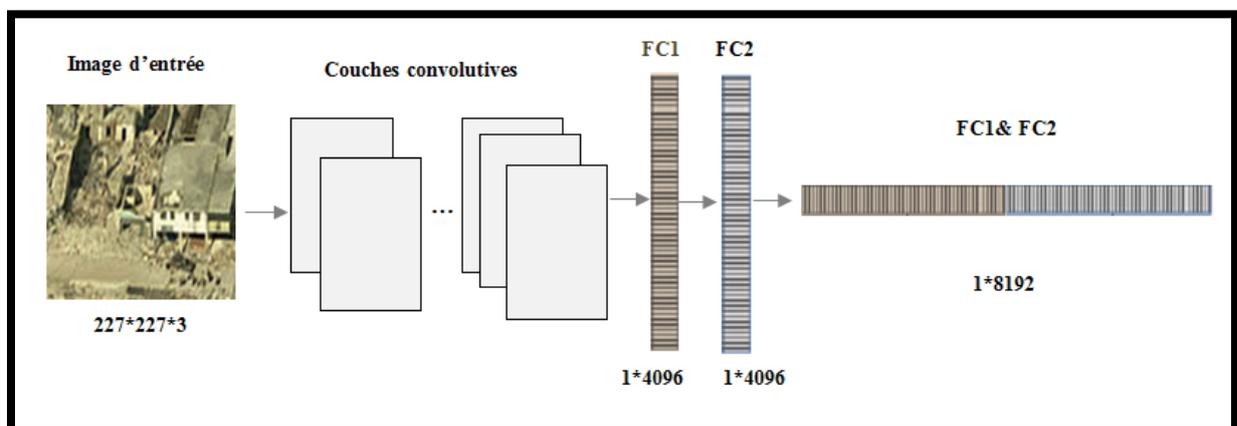


Figure V.2 : Les caractéristiques extraites à partir d'AlexNet.

2.1.1.2 Les caractéristiques extraites à partir de VGG-Net

Les résultats impressionnants des deux CNN très profonds, connus sous le nom de VGG-16 (contenant 13 couches convolutives et 3 couches entièrement connectées) et VGG-19 (contenant 16 couches convolutives et 3 couches entièrement connectées), démontrent que la profondeur du réseau joue un rôle important dans l'amélioration de la performance de la classification. La profondeur de CNN de convolution force le réseau d'apprendre des représentations plus abstraites et plus discriminantes. Plus la couche est profonde, plus les caractéristiques apprises sont abstraites et complexes. C'est la raison pour laquelle nous avons utilisées deux réseaux VGG-16 et VGG-19 dans notre travail.

Les tableaux V.2 et V.3 présentent respectivement l'architecture détaillée de VGG-16 et VGG-19 à partir de Matlab. En outre, car les échantillons ont des dimensions différentes, tous les échantillons sont redimensionnés à 224×224 pixels. De plus, VGG-16 et VGG-19 donnent des vecteurs de caractéristique de 4096 dimensions à partir de FC1 ou FC2. Par conséquent, le vecteur hybride de caractéristiques obtenu à partir de deux VGG-Net peut être considéré comme une collection de 8192-dimensions, comme indiqué dans la figure V.3.

Tableau V.2 : L'architecture détaillée de VGG-16 à partir de Matlab.

Layer Number	Layer Name	Layer Role	Layer Description
1	'input'	Image Input	$224 \times 224 \times 3$ images with 'zerocenter' normalization
2	'conv1-1'	Convolution	$64 \ 3 \times 3 \times 3$ convolutions with stride [1 1] and padding [1 1 1 1]
3	'relu1-1'	ReLU	ReLU
4	'conv1-2'	Convolution	$64 \ 3 \times 3 \times 64$ convolutions with stride [1 1] and padding [1 1 1 1]
5	'relu1-2'	ReLU	ReLU
6	'pool1'	Max Pooling	2×2 max pooling with stride [2 2] and padding [0 0 0 0]
7	'conv2-1'	Convolution	$128 \ 3 \times 3 \times 64$ convolutions with stride [1 1] and padding [1 1 1 1]
8	'relu2-1'	ReLU	ReLU
9	'conv2-2'	Convolution	$128 \ 3 \times 3 \times 128$ convolutions with stride [1 1] and padding [1 1 1 1]
10	'relu2-2'	ReLU	ReLU
11	'pool2'	Max Pooling	2×2 max pooling with stride [2 2] and padding [0 0 0 0]
12	'conv3-1'	Convolution	$256 \ 3 \times 3 \times 128$ convolutions with stride [1 1] and padding [1 1 1 1]
13	'relu3-1'	ReLU	ReLU
14	'conv3-2'	Convolution	$256 \ 3 \times 3 \times 256$ convolutions with stride [1 1] and padding [1 1 1 1]
15	'relu3-2'	ReLU	ReLU

Chapitre V – Optimisation de la classification des dommages via des représentations hybrides de caractéristiques de DL dérivées d'images aériennes post-séisme

16	'conv3-3'	Convolution	256 3 × 3 × 256 convolutions with stride [1 1] and padding [1 1 1 1]
17	'relu3-3'	ReLU	ReLU
18	'pool3'	Max Pooling	2 × 2 max pooling with stride [2 2] and padding [0 0 0 0]
19	'conv4-1'	Convolution	512 3 × 3 × 256 convolutions with stride [1 1] and padding [1 1 1 1]
20	'relu4-1'	ReLU	ReLU
21	'conv4-2'	Convolution	512 3 × 3 × 512 convolutions with stride [1 1] and padding [1 1 1 1]
22	'relu4-2'	ReLU	ReLU
23	'conv4-3'	Convolution	512 3 × 3 × 512 convolutions with stride [1 1] and padding [1 1 1 1]
24	'relu4-3'	ReLU	ReLU
25	'pool4'	Max Pooling	2 × 2 max pooling with stride [2 2] and padding [0 0 0 0]
26	'conv5-1'	Convolution	512 3 × 3 × 512 convolutions with stride [1 1] and padding [1 1 1 1]
27	'relu5-1'	ReLU	ReLU
28	'conv5-2'	Convolution	512 3 × 3 × 512 convolutions with stride [1 1] and padding [1 1 1 1]
29	'relu5-2'	ReLU	ReLU
30	'conv5-3'	Convolution	512 3 × 3 × 512 convolutions with stride [1 1] and padding [1 1 1 1]
31	'relu5-3'	ReLU	ReLU
32	'pool5'	Max Pooling	2 × 2 max pooling with stride [2 2] and padding [0 0 0 0]
33	'fc6'	Fully Connected	4096 fully connected layer
34	'relu6'	ReLU	ReLU
35	'drop6'	Dropout	50% dropout
36	'fc7'	Fully Connected	4096 fully connected layer
37	'relu7'	ReLU	ReLU
38	'drop7'	Dropout	50% dropout
39	'fc8'	Fully Connected	1000 fully connected layer
40	'prob'	Softmax	Softmax
41	'output'	Classification Output	crossentropyex with 'tench', 'goldfish', and 998 other classes

Tableau V.3 : L'architecture détaillée de VGG-19 à partir de Matlab.

Layer Number	Layer Name	Layer Role	Layer Description
1	'input'	Image Input	224 × 224 × 3 images with 'zerocenter' normalization
2	'conv1-1'	Convolution	64 3 × 3 × 3 convolutions with stride [1 1] and padding

Chapitre V – Optimisation de la classification des dommages via des représentations hybrides de caractéristiques de DL dérivées d'images aériennes post-séisme

			[1 1 1 1]
3	'relu1-1'	ReLU	ReLU
4	'conv1-2'	Convolution	64 $3 \times 3 \times 64$ convolutions with stride [1 1] and padding [1 1 1 1]
5	'relu1-2'	ReLU	ReLU
6	'pool1'	Max Pooling	2×2 max pooling with stride [2 2] and padding [0 0 0 0]
7	'conv2-1'	Convolution	128 $3 \times 3 \times 64$ convolutions with stride [1 1] and padding [1 1 1 1]
8	'relu2-1'	ReLU	ReLU
9	'conv2-2'	Convolution	128 $3 \times 3 \times 128$ convolutions with stride [1 1] and padding [1 1 1 1]
10	'relu2-2'	ReLU	ReLU
11	'pool2'	Max Pooling	2×2 max pooling with stride [2 2] and padding [0 0 0 0]
12	'conv3-1'	Convolution	256 $3 \times 3 \times 128$ convolutions with stride [1 1] and padding [1 1 1 1]
13	'relu3-1'	ReLU	ReLU
14	'conv3-2'	Convolution	256 $3 \times 3 \times 256$ convolutions with stride [1 1] and padding [1 1 1 1]
15	'relu3-2'	ReLU	ReLU
16	'conv3-3'	Convolution	256 $3 \times 3 \times 256$ convolutions with stride [1 1] and padding [1 1 1 1]
17	'relu3-3'	ReLU	ReLU
18	'conv3-4'	Convolution	256 $3 \times 3 \times 256$ convolutions with stride [1 1] and padding [1 1 1 1]
19	'relu3-4'	ReLU	ReLU
20	'pool3'	Max Pooling	2×2 max pooling with stride [2 2] and padding [0 0 0 0]
21	'conv4-1'	Convolution	512 $3 \times 3 \times 256$ convolutions with stride [1 1] and padding [1 1 1 1]
22	'relu4-1'	ReLU	ReLU
23	'conv4-2'	Convolution	512 $3 \times 3 \times 512$ convolutions with stride [1 1] and padding [1 1 1 1]
24	'relu4-2'	ReLU	ReLU
25	'conv4-3'	Convolution	512 $3 \times 3 \times 512$ convolutions with stride [1 1] and padding [1 1 1 1]
26	'relu4-3'	ReLU	ReLU
27	'conv4-4'	Convolution	512 $3 \times 3 \times 512$ convolutions with stride [1 1] and padding [1 1 1 1]
28	'relu4-4'	ReLU	ReLU
29	'pool4'	Max Pooling	2×2 max pooling with stride [2 2] and padding [0 0 0 0]
30	'conv5-1'	Convolution	512 $3 \times 3 \times 512$ convolutions with stride [1 1] and

Chapitre V – Optimisation de la classification des dommages via des représentations hybrides de caractéristiques de DL dérivées d'images aériennes post-séisme

			padding [1 1 1 1]
31	'relu5-1'	ReLU	ReLU
32	'conv5-2'	Convolution	512 3 × 3 × 512 convolutions with stride [1 1] and padding [1 1 1 1]
33	'relu5-2'	ReLU	ReLU
34	'conv5-3'	Convolution	512 3 × 3 × 512 convolutions with stride [1 1] and padding [1 1 1 1]
35	'relu5-3'	ReLU	ReLU
36	'conv5-4'	Convolution	512 3 × 3 × 512 convolutions with stride [1 1] and padding [1 1 1 1]
37	'relu5-4'	ReLU	ReLU
38	'pool5'	Max Pooling	2 × 2 max pooling with stride [2 2] and padding [0 0 0 0]
39	'fc6'	Fully Connected	4096 fully connected layer
40	'relu6'	ReLU	ReLU
41	'drop6'	Dropout	50% dropout
42	'fc7'	Fully Connected	4096 fully connected layer
43	'relu7'	ReLU	ReLU
44	'drop7'	Dropout	50% dropout
45	'fc8'	Fully Connected	1000 fully connected layer
46	'prob'	Softmax	Softmax
47	'output'	Classification Output	crossentropyex with 'tench', 'goldfish', and 998 other classes

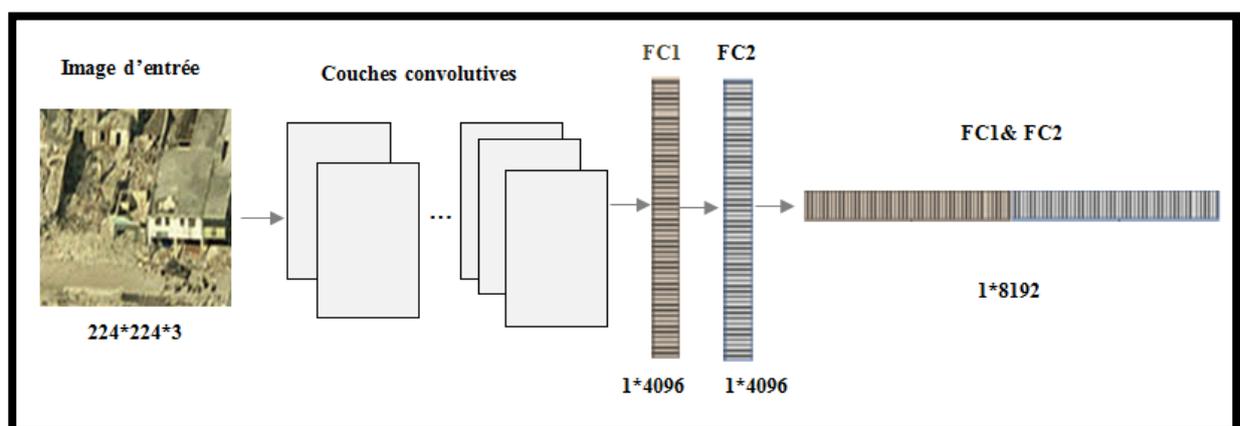


Figure V.3 : Les caractéristiques extraites à partir de VGG-16 et VGG-19.

2.1.1.3 Les caractéristiques extraites à partir de GoogLeNet

GoogLeNet a été l'un des premiers modèles à introduire l'idée que les couches CNN ne devaient pas toujours être empilées de manière séquentielle. En proposant le module Inception, en anglais "*Inception Module*" décrit dans le troisième chapitre, une structuration créative des couches peut conduire à une amélioration des performances de classification et de l'efficacité des coûts de calculs. Grâce à leur architecture et leur profondeur, nous avons évalué leur performance pour la classification des dommages dans notre travail.

Le tableau V.4 présente l'architecture détaillée de GoogLeNet à partir de Matlab. De plus, tous les échantillons sont redimensionnés à 224×224 pixels. En outre, GoogLeNet donne des vecteurs de caractéristiques de 1024 dimensions à partir de la dernière couche de pooling ou couche de drop pooling. Par conséquent, le vecteur hybride de caractéristiques obtenu à partir de GoogLeNet peut être considéré comme une collection de 2048-dimensions, comme indiqué dans la figure V.4.

Tableau V.4 : L'architecture détaillée de GoogLeNet à partir de Matlab.

Layer Number	Layer Name	Layer Role	Layer Description
1	' data'	Image Input	$224 \times 224 \times 3$ images with 'zerocenter' normalization
2	' conv1-7x7_s2'	Convolution	$64 \ 7 \times 7 \times 3$ convolutions with stride [2 2] and padding [3 3 3 3]
3	' conv1-relu_7x7'	ReLU	ReLU
4	' pool1-3x3_s2'	Max Pooling	3×3 max pooling with stride [2 2] and padding [0 1 0 1]
5	' pool1-norm1'	Cross Channel Normalization	cross channel normalization with 5 channels per element
6	' conv2-3x3_reduce'	Convolution	$64 \ 1 \times 1 \times 64$ convolutions with stride [1 1] and padding [0 0 0 0]
7	' conv2-relu_3x3_reduce'	ReLU	ReLU
8	' conv2-3x3'	Convolution	$192 \ 3 \times 3 \times 64$ convolutions with stride [1 1] and padding [1 1 1 1]
9	' conv2-relu_3x3'	ReLU	ReLU
10	' conv2-norm2'	Cross Channel Normalization	cross channel normalization with 5 channels per element
11	' pool2-3x3_s2'	Max Pooling	3×3 max pooling with stride [2 2] and padding [0 1 0 1]
12	' inception_3a-1x1'	Convolution	$64 \ 1 \times 1 \times 192$ convolutions with stride [1 1] and padding [0 0 0 0]
13	' inception_3a-relu_1x1'	ReLU	ReLU
14	' inception_3a-3x3_reduce'	Convolution	$96 \ 1 \times 1 \times 192$ convolutions with stride [1 1] and padding [0 0 0 0]

Chapitre V – Optimisation de la classification des dommages via des représentations hybrides de caractéristiques de DL dérivées d'images aériennes post-séisme

15	' inception_3a-relu_3x3_reduce'	ReLU	ReLU
16	' inception_3a-3x3'	Convolution	128 $3 \times 3 \times 96$ convolutions with stride [1 1] and padding [1 1 1 1]
17	' inception_3a-relu_3x3'	ReLU	ReLU
18	' inception_3a-5x5_reduce'	Convolution	16 $1 \times 1 \times 192$ convolutions with stride [1 1] and padding [0 0 0 0]
19	' inception_3a-relu_5x5_reduce'	ReLU	ReLU
20	' inception_3a-5x5'	Convolution	32 $5 \times 5 \times 16$ convolutions with stride [1 1] and padding [2 2 2 2]
21	' inception_3a-relu_5x5'	ReLU	ReLU
22	' inception_3a-pool'	Max Pooling	3×3 max pooling with stride [1 1] and padding [1 1 1 1]
23	' inception_3a-pool_proj'	Convolution	32 $1 \times 1 \times 192$ convolutions with stride [1 1] and padding [0 0 0 0]
24	' inception_3a-relu_pool_proj'	ReLU	ReLU
25	' inception_3a-output'	Depth concatenation	Depth concatenation of 4 inputs
26	' inception_3b-1x1'	Convolution	128 $1 \times 1 \times 256$ convolutions with stride [1 1] and padding [0 0 0 0]
27	' inception_3b-relu_1x1'	ReLU	ReLU
28	' inception_3b-3x3_reduce'	Convolution	128 $1 \times 1 \times 256$ convolutions with stride [1 1] and padding [0 0 0 0]
29	' inception_3b-relu_3x3_reduce'	ReLU	ReLU
30	' inception_3b-3x3'	Convolution	192 $3 \times 3 \times 128$ convolutions with stride [1 1] and padding [1 1 1 1]
31	' inception_3b-relu_3x3 '	ReLU	ReLU
32	' inception_3b-5x5_reduce'	Convolution	32 $1 \times 1 \times 256$ convolutions with stride [1 1] and padding [0 0 0 0]
33	' inception_3b-relu_5x5_reduce'	ReLU	ReLU
34	' inception_3b-5x5'	Convolution	96 $5 \times 5 \times 32$ convolutions with stride [1 1] and padding [2 2 2 2]
35	' inception_3b-relu_5x5'	ReLU	ReLU
36	' inception_3b-pool'	Max Pooling	3×3 max pooling with stride [1 1] and padding [1 1 1 1]
37	' inception_3b-pool_proj'	Convolution	64 $1 \times 1 \times 256$ convolutions with stride [1 1] and padding [0 0 0 0]
38	' inception_3b-	ReLU	ReLU

Chapitre V – Optimisation de la classification des dommages via des représentations hybrides de caractéristiques de DL dérivées d'images aériennes post-séisme

	relu_pool_proj'		
39	'inception_3b-output'	Depth concatenation	Depth concatenation of 4 inputs
40	'pool3-3x3_s2'	Max Pooling	3 × 3 max pooling with stride [2 2] and padding [0 1 0 1]
41	'inception_4a-1x1'	Convolution	192 1 × 1 × 480 convolutions with stride [1 1] and padding [0 0 0 0]
42	'inception_4a-relu_1x1'	ReLU	ReLU
43	'inception_4a-3x3_reduce'	Convolution	96 1 × 1 × 480 convolutions with stride [1 1] and padding [0 0 0 0]
44	'inception_4a-relu_3x3_reduce'	ReLU	ReLU
45	'inception_4a-3x3'	Convolution	208 3 × 3 × 96 convolutions with stride [1 1] and padding [1 1 1 1]
46	'inception_4a-relu_3x3'	ReLU	ReLU
47	'inception_4a-5x5_reduce'	Convolution	16 1 × 1 × 480 convolutions with stride [1 1] and padding [0 0 0 0]
48	'inception_4a-relu_5x5_reduce'	ReLU	ReLU
49	'inception_4a-5x5'	Convolution	48 5 × 5 × 16 convolutions with stride [1 1] and padding [2 2 2 2]
50	'inception_4a-relu_5x5'	ReLU	ReLU
51	'inception_4a-pool'	Max Pooling	3 × 3 max pooling with stride [1 1] and padding [1 1 1 1]
52	'inception_4a-pool_proj'	Convolution	64 1 × 1 × 480 convolutions with stride [1 1] and padding [0 0 0 0]
53	'inception_4a-relu_pool_proj'	ReLU	ReLU
54	'inception_4a-output'	Depth concatenation	Depth concatenation of 4 inputs
55	'inception_4b-1x1'	Convolution	160 1 × 1 × 512 convolutions with stride [1 1] and padding [0 0 0 0]
56	'inception_4b-relu_1x1'	ReLU	ReLU
57	'inception_4b-3x3_reduce'	Convolution	112 1 × 1 × 512 convolutions with stride [1 1] and padding [0 0 0 0]
58	'inception_4b-relu_3x3_reduce'	ReLU	ReLU
59	'inception_4b-3x3'	Convolution	224 3 × 3 × 112 convolutions with stride [1 1] and padding [1 1 1 1]
60	'inception_4b-relu_3x3'	ReLU	ReLU
61	'inception_4b-5x5_reduce'	Convolution	24 1 × 1 × 512 convolutions with stride [1 1] and padding [0 0 0 0]

Chapitre V – Optimisation de la classification des dommages via des représentations hybrides de caractéristiques de DL dérivées d'images aériennes post-séisme

62	' inception_4b-relu_5x5_reduce'	ReLU	ReLU
63	' inception_4b-5x5''	Convolution	64 $5 \times 5 \times 24$ convolutions with stride [1 1] and padding [2 2 2 2]
64	' inception_4b-relu_5x5'	ReLU	ReLU
65	' inception_4b-pool'	Max Pooling	3×3 max pooling with stride [1 1] and padding [1 1 1 1]
66	' inception_4b-pool_proj'	Convolution	64 $1 \times 1 \times 512$ convolutions with stride [1 1] and padding [0 0 0 0]
67	' inception_4b-relu_pool_proj'	ReLU	ReLU
68	' inception_4b-output'	Depth concatenation	Depth concatenation of 4 inputs
69	' inception_4c-1x1'	Convolution	128 $1 \times 1 \times 512$ convolutions with stride [1 1] and padding [0 0 0 0]
70	' inception_4c-relu_1x1'	ReLU	ReLU
71	' inception_4c-3x3_reduce'	Convolution	128 $1 \times 1 \times 512$ convolutions with stride [1 1] and padding [0 0 0 0]
72	' inception_4c-relu_3x3_reduce '	ReLU	ReLU
73	' inception_4c-3x3'	Convolution	256 $3 \times 3 \times 128$ convolutions with stride [1 1] and padding [1 1 1 1]
74	' inception_4c-relu_3x3'	ReLU	ReLU
75	' inception_4c-5x5_reduce'	Convolution	24 $1 \times 1 \times 512$ convolutions with stride [1 1] and padding [0 0 0 0]
76	' inception_4c-relu_5x5_reduce'	ReLU	ReLU
77	' inception_4c-5x5'	Convolution	64 $5 \times 5 \times 24$ convolutions with stride [1 1] and padding [2 2 2 2]
78	' inception_4c-relu_5x5'	ReLU	ReLU
79	' inception_4c-pool'	Max Pooling	3×3 max pooling with stride [1 1] and padding [1 1 1 1]
80	' inception_4c-pool_proj'	Convolution	64 $1 \times 1 \times 512$ convolutions with stride [1 1] and padding [0 0 0 0]
81	' inception_4c-relu_pool_proj'	ReLU	ReLU
82	' inception_4c-output'	Depth concatenation	Depth concatenation of 4 inputs
83	' inception_4d-1x1'	Convolution	112 $1 \times 1 \times 512$ convolutions with stride [1 1] and padding [0 0 0 0]
84	' inception_4d-relu_1x1'	ReLU	ReLU
85	' inception_4d-	Convolution	144 $1 \times 1 \times 512$ convolutions with stride [1 1]

Chapitre V – Optimisation de la classification des dommages via des représentations hybrides de caractéristiques de DL dérivées d'images aériennes post-séisme

	3x3_reduce'		and padding [0 0 0 0]
86	'inception_4d-relu_3x3_reduce'	ReLU	ReLU
87	'inception_4d-3x3'	Convolution	288 3 × 3 × 144 convolutions with stride [1 1] and padding [1 1 1 1]
88	'inception_4d-relu_3x3'	ReLU	ReLU
89	'inception_4d-5x5_reduce'	Convolution	32 1 × 1 × 512 convolutions with stride [1 1] and padding [0 0 0 0]
90	'inception_4d-relu_5x5_reduce'	ReLU	ReLU
91	'inception_4d-5x5'	Convolution	64 5 × 5 × 32 convolutions with stride [1 1] and padding [2 2 2 2]
92	'inception_4d-relu_5x5'	ReLU	ReLU
93	'inception_4d-pool'	Max Pooling	3 × 3 max pooling with stride [1 1] and padding [1 1 1 1]
94	'inception_4d-pool_proj'	Convolution	64 1 × 1 × 512 convolutions with stride [1 1] and padding [0 0 0 0]
95	'inception_4d-relu_pool_proj'	ReLU	ReLU
96	'inception_4d-output'	Depth concatenation	Depth concatenation of 4 inputs
97	'inception_4e-1x1'	Convolution	256 1 × 1 × 528 convolutions with stride [1 1] and padding [0 0 0 0]
98	'inception_4e-relu_1x1'	ReLU	ReLU
99	'inception_4e-3x3_reduce'	Convolution	160 1 × 1 × 528 convolutions with stride [1 1] and padding [0 0 0 0]
100	'inception_4e-relu_3x3_reduce'	ReLU	ReLU
101	'inception_4e-3x3'	Convolution	320 3 × 3 × 160 convolutions with stride [1 1] and padding [1 1 1 1]
102	'inception_4e-relu_3x3'	ReLU	ReLU
103	'inception_4e-5x5_reduce'	Convolution	32 1 × 1 × 528 convolutions with stride [1 1] and padding [0 0 0 0]
104	'inception_4e-relu_5x5_reduce'	ReLU	ReLU
105	'inception_4e-5x5'	Convolution	128 5 × 5 × 32 convolutions with stride [1 1] and padding [2 2 2 2]
106	'inception_4e-relu_5x5'	ReLU	ReLU
107	'inception_4e-pool'	Max Pooling	3 × 3 max pooling with stride [1 1] and padding [1 1 1 1]
108	'inception_4e-pool_proj'	Convolution	128 1 × 1 × 528 convolutions with stride [1 1] and padding [0 0 0 0]

Chapitre V – Optimisation de la classification des dommages via des représentations hybrides de caractéristiques de DL dérivées d'images aériennes post-séisme

109	' inception_4e-relu_pool_proj'	ReLU	ReLU
110	' inception_4e-output'	Depth concatenation	Depth concatenation of 4 inputs
111	' pool4-3x3_s2'	Max Pooling	3 × 3 max pooling with stride [2 2] and padding [0 1 0 1]
112	' inception_5a-1x1'	Convolution	256 1 × 1 × 832 convolutions with stride [1 1] and padding [0 0 0 0]
113	' inception_5a-relu_1x1'	ReLU	ReLU
114	' inception_5a-3x3_reduce'	Convolution	160 1 × 1 × 832 convolutions with stride [1 1] and padding [0 0 0 0]
115	' inception_5a-relu_3x3_reduce'	ReLU	ReLU
116	' inception_5a-3x3'	Convolution	320 3 × 3 × 160 convolutions with stride [1 1] and padding [1 1 1 1]
117	' inception_5a-relu_3x3'	ReLU	ReLU
118	' inception_5a-5x5_reduce'	Convolution	32 1 × 1 × 832 convolutions with stride [1 1] and padding [0 0 0 0]
119	' inception_5a-relu_5x5_reduce'	ReLU	ReLU
120	' inception_5a-5x5'	Convolution	128 5 × 5 × 32 convolutions with stride [1 1] and padding [2 2 2 2]
121	' inception_5a-relu_5x5'	ReLU	ReLU
122	' inception_5a-pool'	Max Pooling	3 × 3 max pooling with stride [1 1] and padding [1 1 1 1]
123	' inception_5a-pool_proj'	Convolution	128 1 × 1 × 832 convolutions with stride [1 1] and padding [0 0 0 0]
124	' inception_5a-relu_pool_proj'	ReLU	ReLU
125	' inception_5a-output'	Depth concatenation	Depth concatenation of 4 inputs
126	' inception_5b-1x1'	Convolution	384 1 × 1 × 832 convolutions with stride [1 1] and padding [0 0 0 0]
127	' inception_5b-relu_1x1'	ReLU	ReLU
128	' inception_5b-3x3_reduce'	Convolution	192 1 × 1 × 832 convolutions with stride [1 1] and padding [0 0 0 0]
129	' inception_5b-relu_3x3_reduce'	ReLU	ReLU
130	' inception_5b-3x3'	Convolution	384 3 × 3 × 192 convolutions with stride [1 1] and padding [1 1 1 1]
131	' inception_5b-relu_3x3'	ReLU	ReLU
132	' inception_5b-	Convolution	48 1 × 1 × 832 convolutions with stride [1 1]

Chapitre V – Optimisation de la classification des dommages via des représentations hybrides de caractéristiques de DL dérivées d'images aériennes post-séisme

	'5x5_reduce'		and padding [0 0 0 0]
133	'inception_5b-relu_5x5_reduce'	ReLU	ReLU
134	'inception_5b-5x5'	Convolution	128 $5 \times 5 \times 48$ convolutions with stride [1 1] and padding [2 2 2 2]
135	'inception_5b-relu_5x5'	ReLU	ReLU
136	'inception_5b-pool'	Max Pooling	3×3 max pooling with stride [1 1] and padding [1 1 1 1]
137	'inception_5b-pool_proj'	Convolution	128 $1 \times 1 \times 832$ convolutions with stride [1 1] and padding [0 0 0 0]
138	'inception_5b-relu_pool_proj'	ReLU	ReLU
139	'inception_5b-output'	Depth concatenation	Depth concatenation of 4 inputs
140	'pool5-7x7_s1'	Average Pooling	7×7 average pooling with stride [1 1] and padding [0 0 0 0]
141	'pool5-drop_7x7_s1'	Dropout	40% dropout
142	'loss3-classifier'	Fully Connected	1000 fully connected layer
143	'prob'	Softmax	softmax
144	'output'	Classification Output	crossentropyex with 'tench', 'goldfish', and 998 other classes

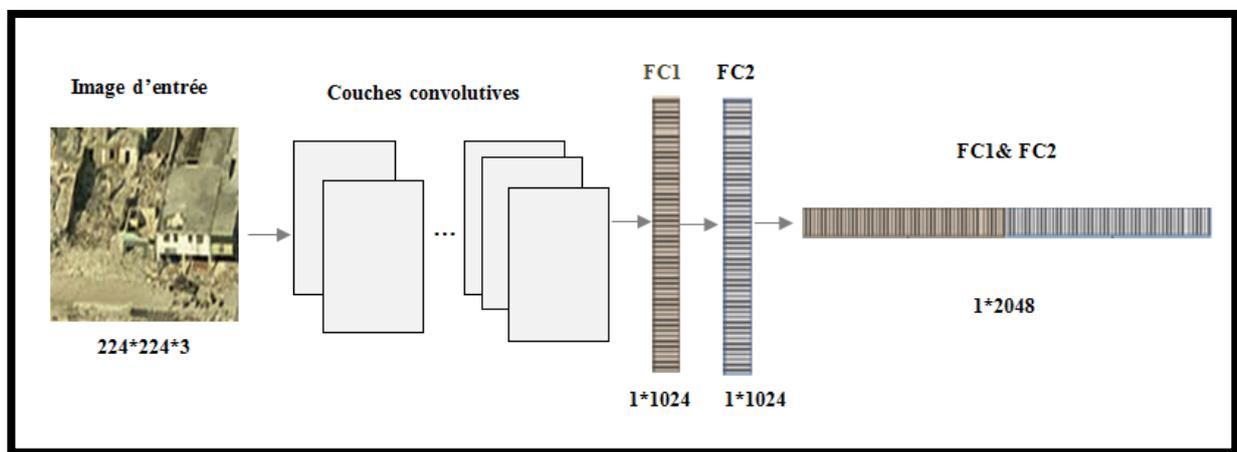


Figure V.4 : Les caractéristiques extraites à partir de GoogLeNet.

2.1.2 La représentation hybride proposée

Notre proposition consiste à combiner les quatre vecteurs des caractéristiques précédemment extraits durant l'étape précédente pour former un vecteur de caractéristiques final, comme le montre la figure V.5. Pour chaque image d'entrée, nous avons obtenu quatre vecteurs des caractéristiques avec les dimensions suivantes, 8192-dimensions, 8192-dimensions, 8192-dimensions, 2048-dimensions, en utilisant, AlexNet, VGG-16, VGG-19 et GoogLeNet, respectivement. Pour cela, chaque image d'entrée est transformée en un vecteur des caractéristiques de 26 624-dimensions. Ce vecteur hybride des caractéristiques est ensuite utilisé dans la phase de classification.

2.2 La phase de classification des dommages

Après l'extraction des caractéristiques, la phase suivante est la classification supervisée. Dans cette phase, un classificateur supervisé prend le vecteur hybride de caractéristiques généré par la phase précédente comme une donnée d'entrée. Par conséquent, le classificateur utilisé doit être capable d'adapter aux données d'entrée hyper-dimensionnelles, car ce vecteur hybride de caractéristiques est de haute dimension "26 624-dimensions". Dans ce travail, le classifieur SVM est appliqué, qui s'est avéré être un modèle de classification supervisée très important en raison de sa capacité à gérer des espaces des caractéristiques de haute dimension (voir le deuxième chapitre pour plus de détails).

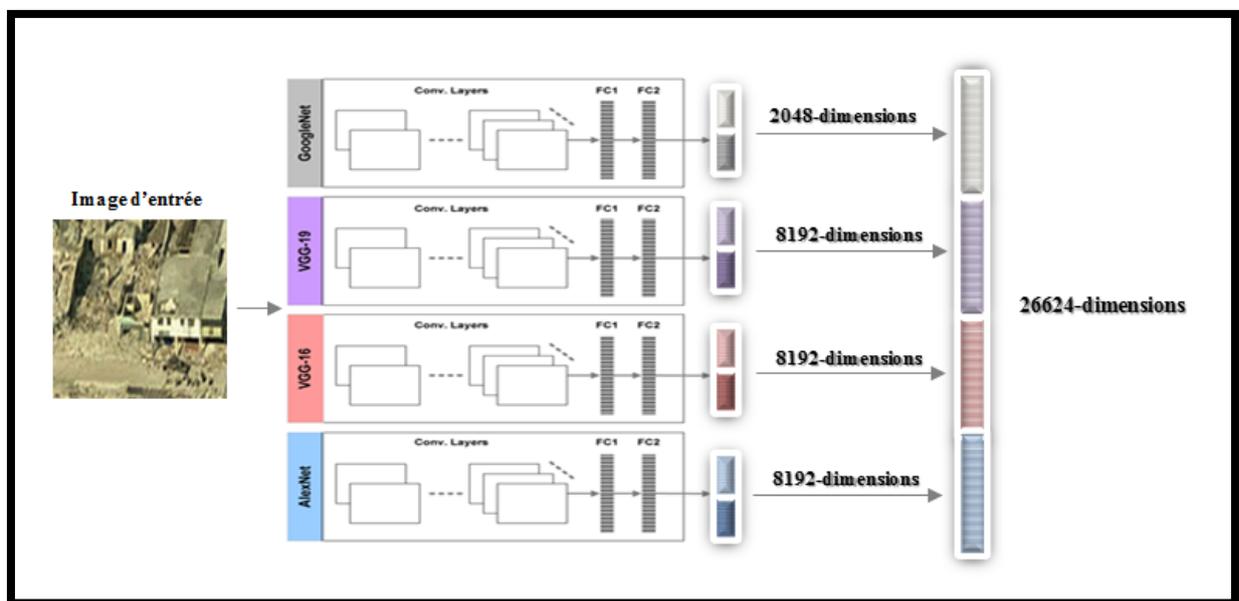


Figure V. 5 : La représentation hybride proposée.

2.2.1 Apprentissage

Dans la phase d'apprentissage, SVM est chargé de prendre des couples qui se compose des vecteurs hybrides de caractéristiques avec les étiquettes de leurs images d'entrée (endommagé ou non endommagé) afin de produire une fonction de décision qui maximise la marge entre deux classes données. Cette phase consiste à initialiser la base des modèles en sauvegardant les caractéristiques de différentes classes avec les étiquettes correspondantes comme illustré sur la figure V.6.

2.2.2 Test

Cette phase consiste à assigner une classe pour chaque nouveau exemple donnée en se basant sur les données de la base des modèles. Plus précisément, dans cette phase, SVM ne prend que le vecteur hybride des caractéristiques sans aucune étiquette et le compare avec les modèles existants dans la base des modèles obtenue à partir de la phase d'apprentissage pour faire une décision de classification comme sur la figure V.6.

3. Conception expérimentale

Dans cette section, nous présenterons la base de données des images utilisées pendant l'évaluation de l'approche proposée, la configuration des expériences, les méthodes étudiées, ainsi que la mesure de performance et les tests statistiques appliqués pour la comparaison des résultats obtenus.

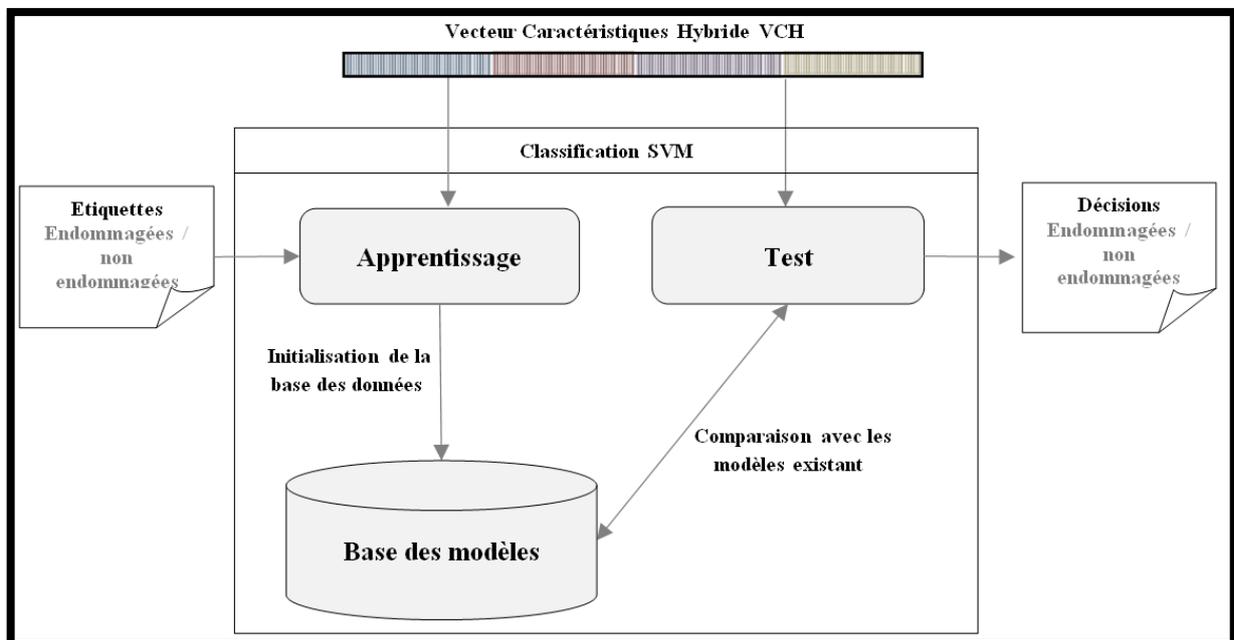


Figure V. 6 : Phase de classification.

3.1 Description de la base d'images utilisée

Pour évaluer l'approche proposée, nous avons besoin d'une base de données d'images différentes par ses caractéristiques telles que l'échelle, la vue de la caméra, la plateforme de capture et la complexité de la scène. Pour cela, nous avons utilisé la base de données récemment utilisés dans [49] avec une grande variabilité des caractéristiques de paysage, des conditions d'éclairage et des caractéristiques d'image. Les images ont été collectées par Vetrivel [49] et ont été capturées par différentes plateformes et capteurs provenant de différents événements sismiques dans plusieurs régions géographiques. Vetrivel a séparé ces images en deux groupes selon la plateforme d'acquisition: les images capturées à partir des avions et les autres à partir des drones, "en anglais (Unmanned Aerial Vehicle ou UAV)".

Le tableau V.5 résume les informations sur les deux types de base d'images.

Tableau V.5 : Description des échantillons d'apprentissage et de test dérivés d'images de deux plateformes pour différentes régions géographiques

Plateforme d'acquisition	Régions géographiques	Année de la catastrophe	Type d'événement
Avion	Port-au-Prince, Haiti	2010	Séisme
	Bidonville, Haiti	2010	Séisme
	L'Aquila, Italie	2009	Séisme
	Onna, Italie	2009	Séisme
	Tempera, Italie	2009	Séisme
	Mirabello, Italie	2012	Séisme
UAV	Equateur, Pérou	2016	Séisme
	Katmandou, Népal	2015	Séisme
	L'Aquila, Italie	2009	Séisme
	Pingtung, Taïwan	2016	Séisme
	Mirabello, Italie	2012	Séisme
	Gronau, Allemagne	2013	Zone industrielle détruite manuellement

La figure V.7 montre des échantillons de chaque base de données d'images. Le premier ensemble d'images capturées par l'avion contient 19133 échantillons et ont été dérivés d'images capturées avec cinq caméras (un nadir et quatre vues obliques), avec une résolution spatiale d'environ 10-16 cm. La classe endommagée de cet ensemble de données contient 9203 échantillons, tandis que la classe non endommagée contient 9930 échantillons.

Le deuxième ensemble d'images capturées à partir UAV contient 11643 échantillons obtenus à partir d'images capturées à différentes hauteurs, vues, caméras et conditions d'éclairage, avec des résolutions spatiales allant de 1 cm à 5 cm. Les classes endommagées et non endommagées de cet ensemble de données contiennent respectivement 5682 et 5961 échantillons. Tous les échantillons de ces deux types ont été obtenus en appliquant la méthode SLIC et en convertissant les super-pixels en des patchs d'image rectangulaires d'une taille d'environ 100 × 100 pixels (voir [49] pour plus de détails).

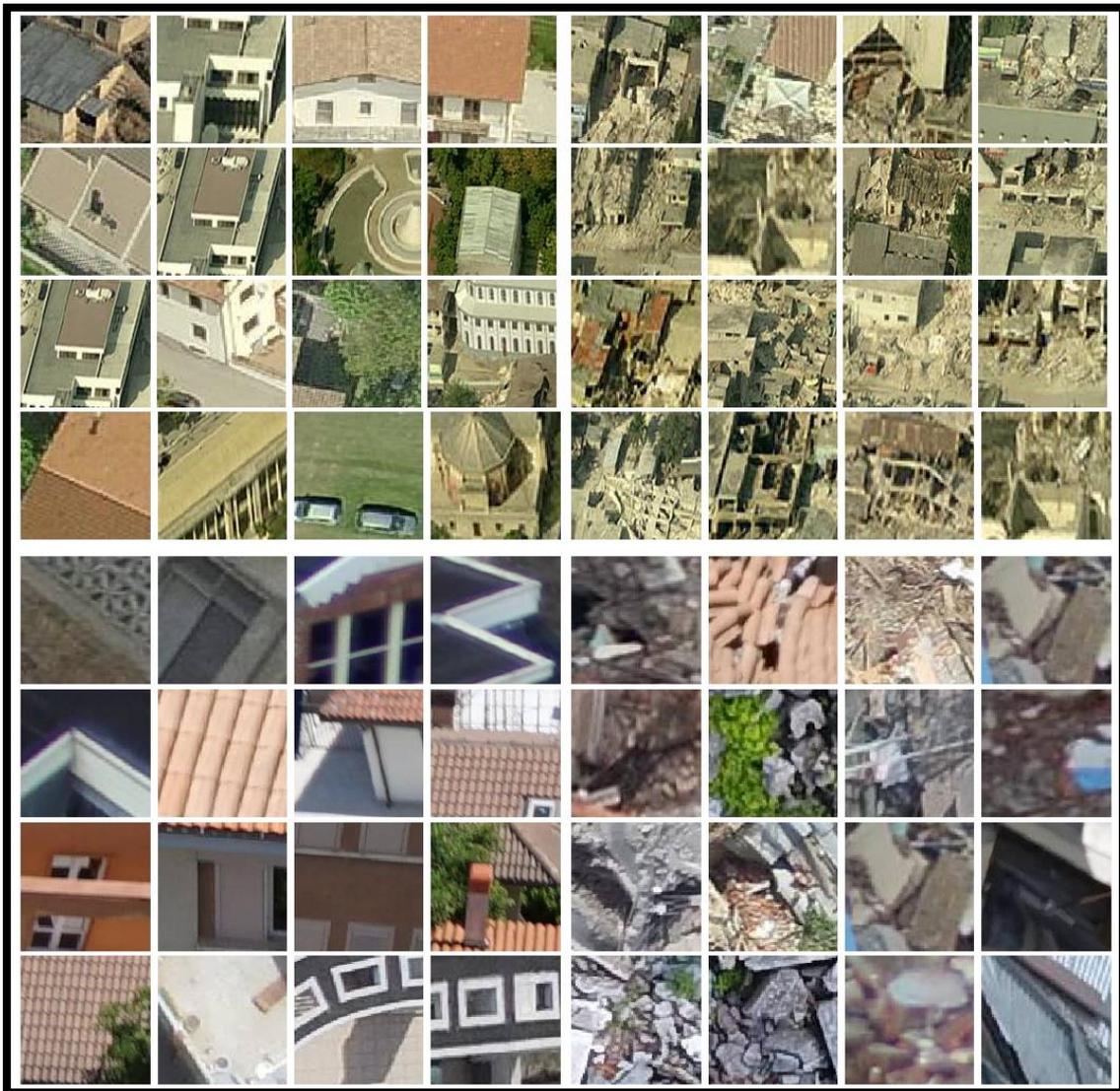


Figure V.7 : Exemples des échantillons de chaque type de base d'images, représentant des échantillons non endommagés (côté gauche) et endommagés (côté droit) des ensembles d'images capturées par l'avion (échantillons du haut) et d'UAV (échantillons du bas).

3.2 Environnement de travail et réglage des paramètres

Toutes les expériences ont été réalisées en utilisant une station de travail avec 64 GB de RAM, un processeur Intel Xeon E5-2620 V3 6 cœurs de 2,4 GHz, système d'exploitation Windows 10 et un programme en Matlab.

Dans ce travail, pour exécuter le programme de notre projet nous avons utilisé le Matlab (signifie Matrix Laboratory) version 2018 qui est un logiciel permet de résoudre des problèmes numériques complexes dans peu de temps et offre une grande simplicité de manipulation d'images. Matlab a été conçu pour rendre le calcul matriciel particulièrement facile. Les fonctionnalités de Matlab peuvent être considérablement étendues par l'ajout de boîtes à outils ou "toolbox". Les boîtes à outils sont des dossiers contenant une multitude de fonctions intégrées et à plusieurs programmes outils testés et regroupés en fonction de leur utilisation.

Ce sont des ensembles de fonctions spécifiques qui ont fourni des fonctionnalités plus spécialisées. Ces fonctionnalités rendent le langage de programmation très efficace pour la mise en œuvre de DL, plus précisément la mise en œuvre de notre projet [131]. Nous pouvons programmer notre projet plus facilement, et ce grâce à cette "toolbox", notamment Deep Learning Toolbox, Deep Learning Toolbox Model for AlexNet Network, Deep Learning Toolbox Importer for Caffe Models, Deep Learning Toolbox Model for VGG-16 Network, etc. Cela signifie que tous les modèles de CNN utilisés dans notre proposition (AlexNet, VGG-16, VGG-19 et GoogLeNet) sont fournis dans Matlab / Simulink.

Plus précisément, les quatre modèles de CNN utilisés se sont appelés à partir des instructions suivantes :

```
net1 = alexnet;  
net2 = vgg16;  
net3 = vgg19;  
net4 = googlenet;
```

Tous les échantillons sont redimensionnés en utilisant la fonction *imresize* dans Matlab / Simulink.

```
I1 = imresize(I, [227,227]);      pour AlexNet.  
I2 = imresize(I, [224,224]);      pour les autres modèles CNN
```

3.3 Méthodes étudiées

Pour évaluer la performance de notre méthode proposée, nous avons effectué deux études expérimentales:

- Le premier ensemble d'expériences a été conçu pour comparer les performances de la classification des dommages sismiques en utilisant les vecteurs de caractéristiques extraits de la première couche entièrement connectée uniquement (FC1), de la deuxième couche entièrement connectée uniquement (FC2) et de tous les deux couches (FC1 et FC2). Il faut noter que l'utilisation de vecteurs de caractéristiques extraits de FC2 est un cas général dans plusieurs études existantes dans la littérature.
- Le deuxième ensemble d'expériences a été réalisé afin de l'évaluation de la performance de la classification des dommages sismiques en combinant des vecteurs de caractéristiques extraits à partir de différents modèles pré-entraînés de CNN. Nous avons étudié trois combinaisons (hybrides) pour les vecteurs de caractéristiques extraits de FC1, FC2 et FC1 & FC2, voir les caractéristiques de chaque hybridation résumées dans le tableau V.6.

Les résultats et l'analyse de ces deux ensembles d'expériences sont présentés séparément dans les sections 4.1 et 4.2.

Tableau V.6 : Les caractéristiques étudiées de chaque hybridation.

Méthode hybride	Modèles CNN utilisés	Taille de vecteur des caractéristiques FC1 ou FC2	Taille de vecteur des caractéristiques FC1 et FC2 combinées
Hybride -1	AlexNet, VGG16, VGG19, etGoogLeNet	13312	26624
Hybride -2	AlexNet, VGG16 et VGG19	12288	24576
Hybride -3	VGG16 et VGG19	819	16384

3.4 Mesure de la performance et validation statistique

Dans toutes nos expériences, chaque base d'images a été scindée en deux ensembles dans les proportions 20% et 80% respectivement pour les échantillons d'apprentissage et pour les échantillons de test par un tirage aléatoire. Afin de collecter des résultats significatifs, nous avons effectué 50 essais (run) indépendants pour chaque méthode. D'autres échantillons d'apprentissage/test sélectionnés de façon aléatoire ont été prélevés dans chaque essais, et toutes les méthodes comparées ont été appliquées aux mêmes 50 échantillons d'apprentissage/test. La performance de classification de chaque essai sont évaluées par le taux de succès (%) qui est calculé par la division de nombre d'échantillons correctement classés sur le nombre total d'échantillons de test, et exprimé selon l'équation suivante:

$$Taux_succès (TS) = \frac{T_P + T_F}{T_P + T_F + F_P + F_N} \quad (5.1)$$

Où T_P , T_F , F_P et F_N désignent les nombres de vrai positif, vrai négatif, faux positif et faux négatif, respectivement.

Afin de tirer des conclusions objectives à partir des résultats numériques, nous avons effectué des tests statistiques sur les résultats de 50 essais des méthodes étudiées. Pour ce faire, nous avons d'abord utilisé des tests de Kolmogorov–Smirnov afin de vérifier si les ensembles de résultats ont des distributions normales (ou gaussiennes) standard. Tous ces tests de normalité ont rejeté les hypothèses nulles et, par conséquent, les résultats n'étaient pas normalement distribués. Par conséquent, nous avons utilisé des tests de Kruskal-Wallis [132] pour comparer les méthodes en fonction du taux de succès. De plus, nous avons utilisé des tests de Friedman afin de fournir un classement des méthodes étudiées en fonction du taux de succès. Dans tous les tests statistiques, une valeur p inférieure à 0,05 a été considérée pour indiquer des différences significatives entre les résultats numériques.

4. Résultats et discussion

4.1 Expérience 1: Combinaison des caractéristiques de FC1 et FC2

Le tableau V.7 montre les taux de succès (%) de la classification de dommages sismiques en utilisant les vecteurs des caractéristiques extraits de FC1, FC2, FC1 & FC2 pour chaque modèle pré-entraînés CNN sélectionné pour les deux bases de données. Dans les colonnes nommées "KW" (pour Kruskal-Wallis) de ce tableau, les symboles "+" indiquent qu'il existe des différences statistiquement significatives entre les résultats de FC1, FC2 et FC1 & FC2 pour le même modèle CNN, tandis que les symboles "•" indiquent non-

significatif. Pour chaque modèle CNN, les meilleurs résultats sont affichés en **gras**. Nous remarquons deux conclusions à partir de ces résultats. Premièrement, les vecteurs de caractéristiques de FC1 ont donné de meilleurs taux de succès que les vecteurs de caractéristiques de FC2 dans tous les cas. Ces comparaisons étaient statistiquement significatives (en utilisant des tests KW avec une valeur p inférieure à 0,05), sauf pour GoogLeNet où les performances étaient statistiquement similaires.

Deuxièmement, pour AlexNet, VGG16 et VGG19, nous pouvons constater que la performance de classification est considérablement améliorée avec l’utilisation des caractéristiques des deux couches FC1 et FC2 combinées que l’utilisation des caractéristiques de chaque couche indépendante, avec des différences statistiquement significatives pour les deux bases de données. Mais pour GoogLeNet, les trois vecteurs de caractéristiques ont montré des taux de succès similaires pour les deux bases de données.

Tableau V.7 : Performances de la classification des dommages sismiques (TS) en utilisant les vecteurs de caractéristiques de FC1, FC2 et FC1&FC2 de quatre modèles CNN pré-entraînés.

Modèle CNN	Base de données aériennes				Base de données UAV			
	FC1	FC2	FC1&FC2	KW	FC1	FC2	FC1&FC2	KW
AlexNet	94.84	94.45	95.47	+	95.47	95.22	95.75	+
VGG-16	96.22	96.09	96.62	+	96.89	96.67	97.01	+
VGG-19	96.52	96.14	96.67	+	96.99	96.87	97.07	+
GoogLeNet	92.06	92.06	92.59	•	94.17	94.17	94.17	•

La figure V.8 illustre une comparaison directe entre les quatre CNN basé sur le taux de succès. Cette figure présente les graphiques des intervalles de comparaison. Le rang moyen de chaque modèle est représenté par un symbole (cercle) et l’intervalle correspondant est représenté par une ligne. Deux intervalles disjoints indiquent que leurs modèles CNN correspondants sont significativement différents, tandis que deux intervalles se chevauchant indiquent la non-signification. Comme l’illustrent ces graphiques, quelques soient les vecteurs des caractéristiques utilisés, et pour les deux bases de données, VGG-19 et VGG-16 occupaient la première position, suivis par AlexNet en troisième position, tandis que les vecteurs de caractéristiques de GoogLeNet ont donné la plus mauvaise performance.

Dans cet ensemble d’expériences, nous pouvons également constater que les caractéristiques de FC1 et FC2 avec la fonction d’activation ReLU permettaient d’améliorer la performance de classification. Dans le tableau V.8, nous comparons FC1 avec ReLU (noté FC1) et sans ReLU (noté FC1\ReLU) pour les quatre modèles CNN. Comme nous pouvons le voir, FC1 avec ReLU a fourni les meilleurs taux de succès pour tous les cas. Ces résultats sont statistiquement significatifs, sauf pour GoogLeNet.

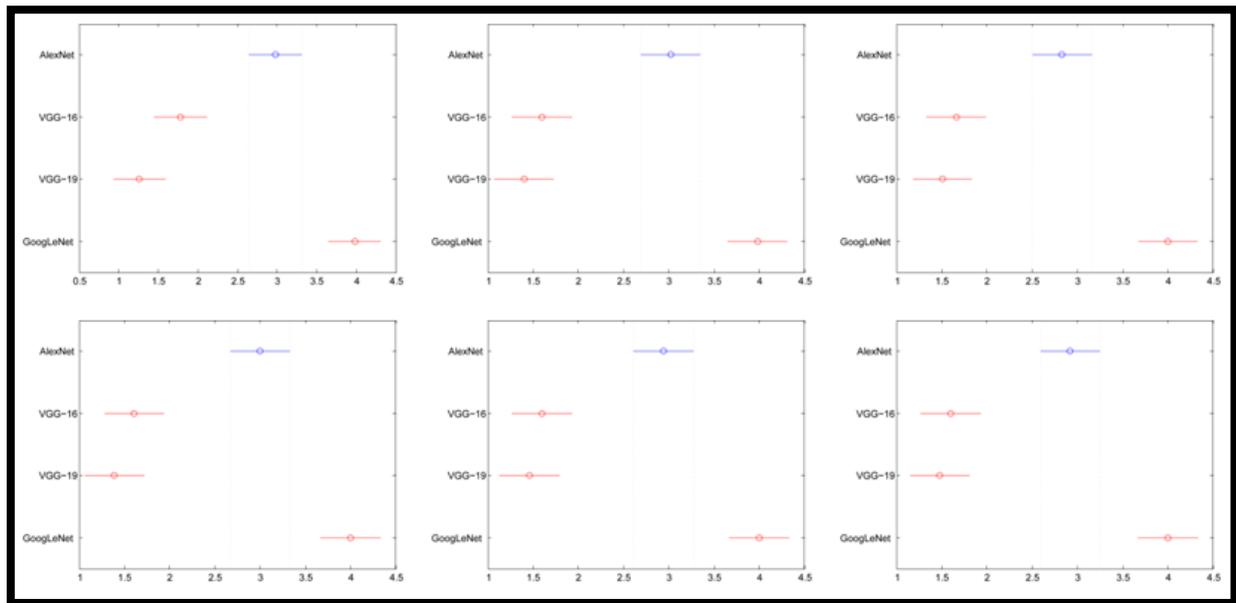


Figure V.8 : Comparaisons multiples de CNN utilisés en fonction du taux de succès. Chaque rectangle contient quatre lignes (partant des cercles de symboles) représentant les rangs moyens et leurs intervalles pour les quatre modèles CNN utilisés. Les trois rectangles du haut concernent la base de données aérienne et les trois rectangles du bas concernent la base de données d’UAV; les colonnes de rectangles de gauche à droite sont respectivement pour représenter les caractéristiques de FC1, FC2, FC1& FC2.

Tableau V.8 : Comparaison (en termes de TS) entre les caractéristiques de CNN avec et sans transformation ReLU.

Modèle CNN	Base de données aériennes		Base de données UAV	
	FC1	FC1\ReLU	FC1	FC1\ReLU
AlexNet	94.84	92.44	95.47	94.08
VGG-16	96.22	94.79	96.89	95.98
VGG-19	96.52	94.84	96.99	96.01
GoogLeNet	92.06	92.06	94.17	94.17

4.2 Expérience 2: Combinaison de caractéristiques de différents modèles CNN

Le tableau V.9 présente la moyenne des taux de succès (%) pour les trois méthodes hybrides indiqués dans le tableau V.6. Le meilleur résultat parmi les trois méthodes hybrides pour chaque vecteur de caractéristique est affiché en **gras**. D’après les résultats statistiques utilisant des tests KW appariés, il existe des différences statistiquement très significatives entre les performances de classification fournies basé sur les représentations hybrides et les performances liées à leurs vecteurs de caractéristiques d’origine. Les figures V.9, V.10 et V.11 montrent les comparaisons des trois méthodes hybrides avec leurs modèles CNN originaux. Comme nous pouvons le remarquer à partir des diagrammes en boîte sur ces figures, quels que soient les vecteurs de caractéristiques utilisé, et pour les deux bases de données, chaque méthode hybride offrait une meilleure classification des dommages que chacun de ses composants de modèle CNN.

Tableau V.9 : Performances de classification de dommage (TS) fournies basé sur les représentations hybrides.

Modèle CNN	Base de données aériennes			Base de données UAV		
	FC1	FC2	FC1&FC2	FC1	FC2	FC1&FC2
Hybird-1	97.51	97.44	97.64	97.62	97.62	97.70
Hybird-2	96.97	97.40	97.61	97.57	96.56	97.67
Hybird-3	97.02	97.12	97.41	97.37	96.26	97.45

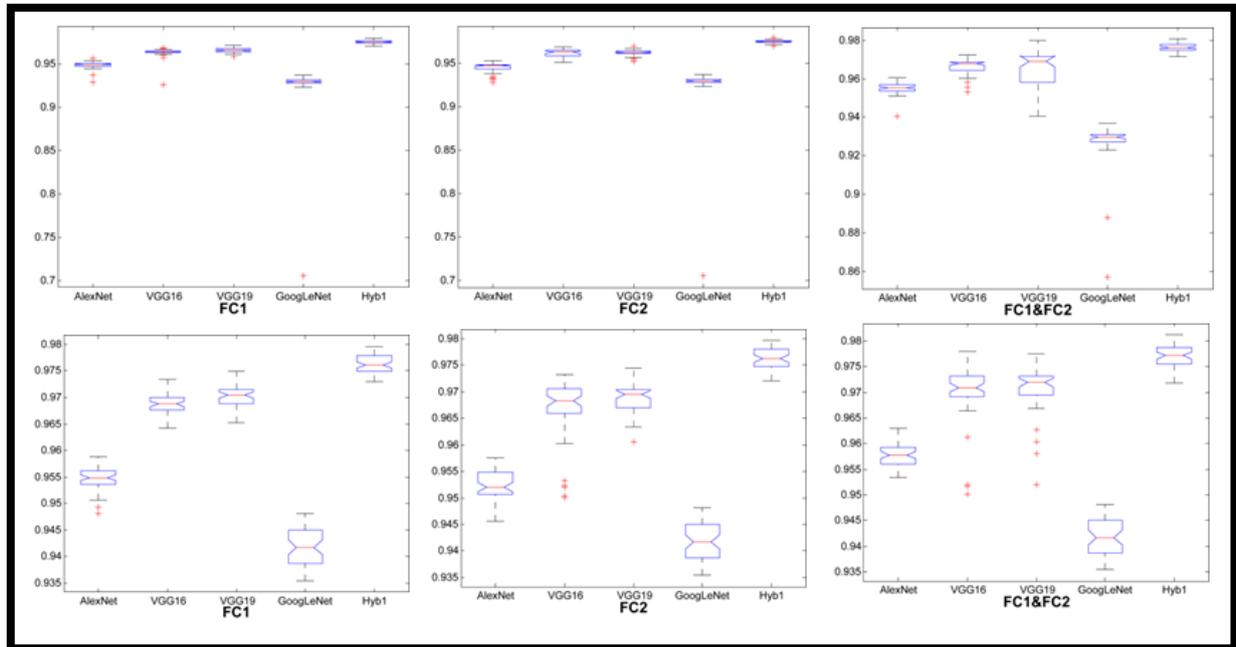


Figure V.9 : Distribution des taux de succès pour la méthode Hybrid-1 et ses composants CNN. Chaque rectangle contient cinq diagrammes en boîte représentant la distribution des taux de succès de la classification des dommages en utilisant les vecteurs de caractéristiques de FC1, FC2 ou FC1 & FC2 pour Hybrid-1 et chacun de ses composants AlexNet, VGG-16, VGG-19 et GoogLeNet. Les trois rectangles du haut concernent la base de données aérienne et les trois rectangles du bas concernent la base de données d’UAV.

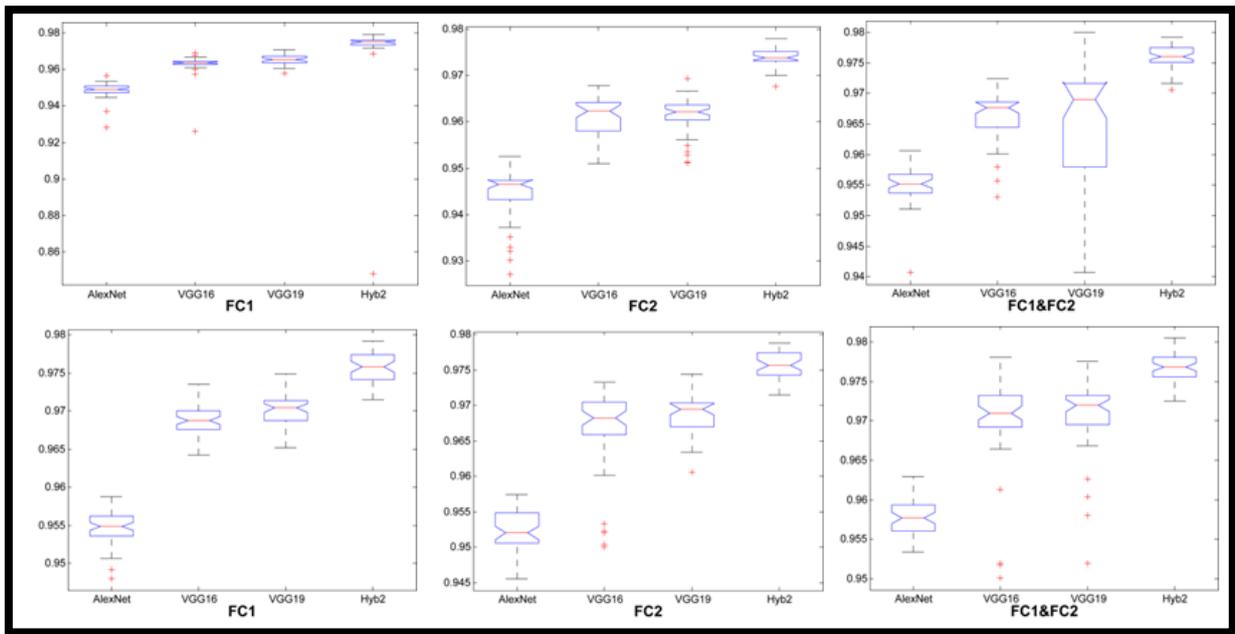


Figure V.10 : Distribution des taux de succès pour la méthode Hybrid-2 et ses composants CNN. Chaque rectangle contient quatre diagrammes en boîte représentant la distribution des taux de succès de la classification des dommages en utilisant les vecteurs de caractéristiques de FC1, FC2 ou FC1 & FC2 pour Hybrid-2 et chacun de ses composants AlexNet, VGG-16 et VGG-19. Les trois rectangles du haut concernent la base de données aérienne et les trois rectangles du bas concernent la base de données d’UAV.

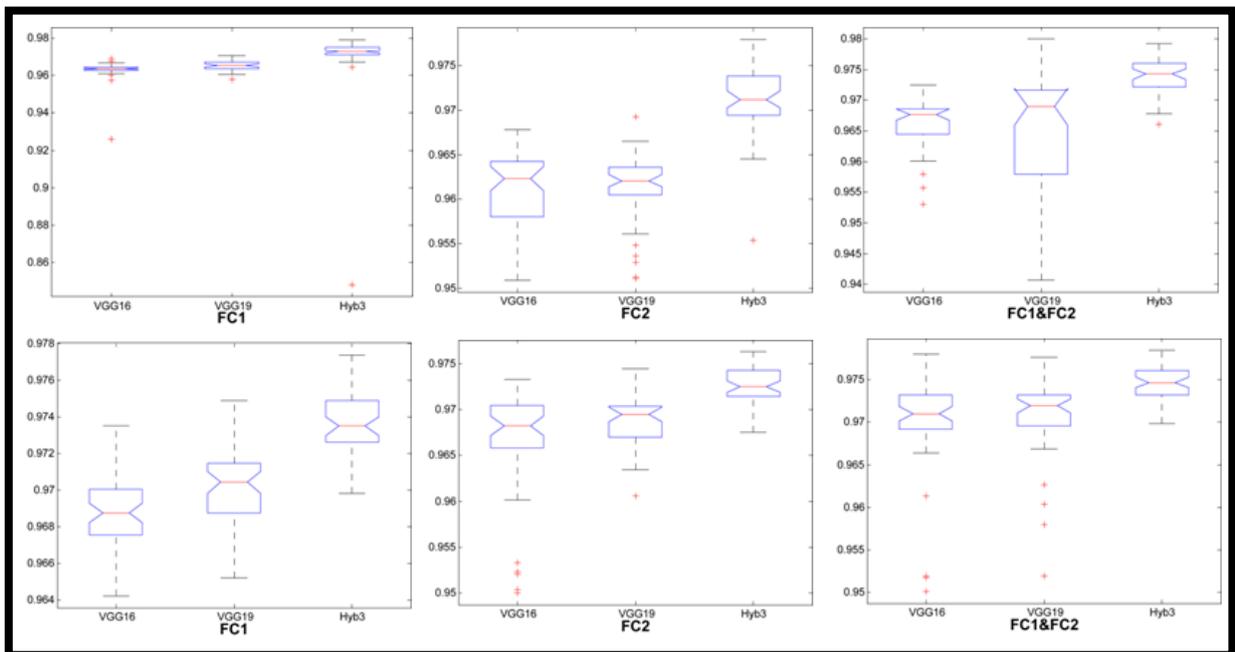


Figure V.11 : Distribution des taux de succès pour la méthode Hybrid-3 et ses composants CNN. Chaque rectangle contient trois diagrammes en boîte représentant la distribution des taux de succès de la classification des dommages en utilisant les vecteurs de caractéristiques de FC1, FC2 ou FC1 & FC2 pour Hybrid-3 et chacun de ses composants VGG-16 et VGG-19. Les trois rectangles du haut concernent la base de données aérienne et les trois rectangles du bas concernent la base de données d’UAV.

Après avoir confirmé que chaque méthode hybride offre des meilleures performances que chaque vecteur de caractéristiques d'origine indépendant, nous allons maintenant comparer ces méthodes hybrides utilisées. La figure V.12 montre les multiples comparaisons des trois méthodes hybrides basées sur les rangs moyens calculés par les tests de Friedman. Comme nous pouvons constater à partir des intervalles de comparaison, Hybrid-1 a toujours occupé la première position avec une signification statistique, tandis que Hybrid-2 est arrivé en deuxième position et Hybrid-3 était le pire, sans différences statistiquement significatives entre eux dans certains cas.

4.3 Comparaison de l'approche proposée avec la littérature

Après avoir confirmé que notre méthode hybride fournit de bonnes performances et optimise la classification des dommages sismiques, nous allons maintenant comparer les résultats obtenus avec les résultats d'autres méthodes de la littérature. Nous pouvons diviser les travaux existants dans la littérature en deux groupes : ceux qui utilisent les méthodes de ML et ceux qui appliquent les méthodes de DL. Bien que notre approche hybride soit basée sur les méthodes de DL, dans cette section, nous allons comparer notre approche avec les deux groupes de méthodes. La phase d'extraction des caractéristiques peut être divisée en trois types, comme décrit dans le deuxième chapitre :

- (1) Les caractéristiques de bas niveaux ;
- (2) Les caractéristiques de niveaux intermédiaires ;
- (3) Et les caractéristiques de plus hauts niveaux.

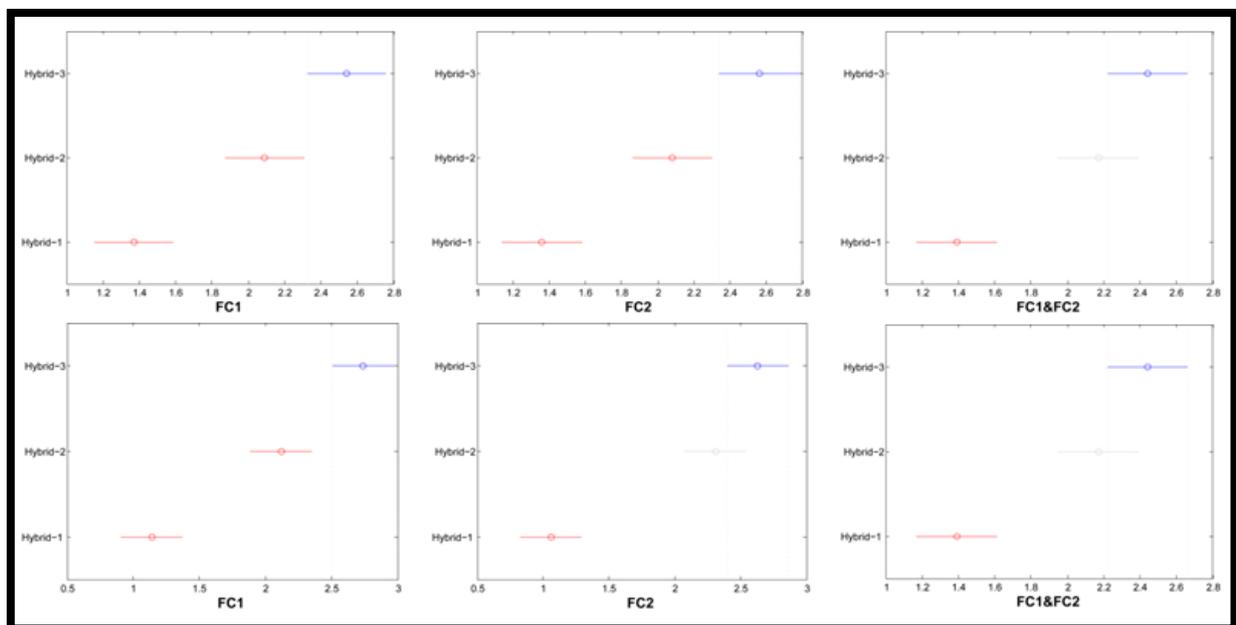


Figure V.12 : Comparaisons multiples des vecteurs des trois méthodes hybrides étudiés. Chaque rectangle contient trois lignes (partant des cercles de symboles) représentant les rangs moyens et leurs intervalles pour les trois méthodes hybrides. Les trois rectangles du haut concernent la base de données aérienne et les trois rectangles du bas concernent la base de données d'UAV.

▪ **Les travaux utilisent les méthodes de ML**

1. Le travail [9] a utilisé BOW avec différents caractéristiques de bas niveaux notamment filtres de Gabor, histogramme de gradient orienté "HOG", accélérées de caractéristiques visuelles robustes "SURF". La méthode proposée a augmenté la précision de la classification des dommages sismiques par rapport aux représentations globales conventionnelles. Une étude comparative a été faite entre les méthodes d'extraction des caractéristiques de bas niveaux et de niveaux intermédiaires, notamment :
 - **Gabor / SVM ;**
 - **HoG / SVM ;**
 - **BoW-Gabor / SVM ;**
 - **BoW- HoG / SVM ;**
 - **Et BoW- SURF / SVM.**
2. Le travail [22] a aussi utilisé BOW pour la classification des dommages sismiques. Un descripteur hybride a été utilisé à partir de combinaison des caractéristiques de couleur et de texture HOG.
 - **BoW- (HoG + Couleur) / SVM.**
3. Dans le travail [49], les caractéristiques tridimensionnelles (caractéristiques du nuage de points 3D) ont été considérées pour la classification des dommages sismiques.
 - **Nuage de points 3D / SVM.**
4. Dans le travail [63], les caractéristiques de texture extrait à partir de GLCM ont été considérées pour la classification des dommages sismiques.
 - **GLCM / SVM.**
5. Le travail [66] a utilisé BOW pour la classification des dommages sismiques. Un descripteur hybride est utilisé à partir de combinaison des caractéristiques de couleur et de texture HOG.
 - **BoW- (HoG + Couleur) / SVM.**

▪ **Les travaux appliquent les méthodes de DL**

1. Dans le travail [49], l'extraction des caractéristiques basées sur CNN s'effectue selon les trois scénarios principaux, comme décrit dans le troisième chapitre :
 - Apprentissage à partir de zéro (Training from scratch, **CNN_S**) ;
 - Réglage d'un modèle pré-entraîné (Tuning a pre-trained model, **CNN_T**) ;
 - Et le modèle pré-entraîné sans réglage (Pre-trained model without tuning, **CNN_F**).

Plus précisément, cette étude a utilisé un modèle de CNN pré-entraîné comme un outil d'extraction de caractéristiques, nommés AlexNet, pour la classification des dommages sismiques. Elle a utilisé le vecteur d'activation extrait à partir de FC2 comme caractéristiques globale d'image. La combinaison des caractéristiques tridimensionnelles et de CNN ont été considérées pour la classification des dommages sismiques.

- **Nuage de points 3D + CNN / SVM.**
2. Le travail [53] a utilisé un modèle de CNN pré-entraîné comme un outil d'extraction de caractéristiques, nommé VGG-16, pour la classification des dommages sismiques.
 - **VGG-16/ SVM.**

3. Le travail [67] a créé une architecture particulière de CNN pour la classification des dommages sismiques.
4. Le travail [69] a proposé une méthode de DL, nommés **YOLOv3-S-GIoU**, pour la détection des bâtiments effondrés à partir d’images post-séisme. Cette étude a appliqué la méthode de détection d’objets (You Only Look Once, YOLOv3) basée sur le réseau de CNN pour localiser les bâtiments effondrés à partir d’images post-séisme.
5. Le travail [73] a créé une architecture particulière de CNN pour la détection des bâtiments effondrés à partir d’images post-séisme.
6. Le travail [110] a utilisé un modèle de CNN pré-entraîné comme un outil d’extraction des caractéristiques, nommé **VGG-16**, pour la classification des dommages sismiques.
7. Nous avons utilisé une méthode hybride basée sur le DL dans la phase d’extraction des caractéristiques pour optimiser la classification des dommages sismiques. Plus précisément, le vecteur de caractéristiques hybride est extrait à partir de quatre CNNs : AlexNet, VGG-16, VGG-19 et GoogLeNet.

- **Hybird-1 /SVM**

Pour faire une comparaison équitable et objective avec les travaux de la littérature, les conditions doivent être équilibrées. Plus précisément, pour comparer notre approche proposée avec la littérature, les approches comparées doivent être à validées et essayées sur le même jeu de données ou sur des jeux de données ayant des caractéristiques similaires. Malheureusement, la plupart de chercheurs ont utilisé leurs propres bases d’images, c’est-à-dire, ils ont considéré un événement sismique spécifique dans une région géographique spécifique pour la validation de leurs approches. Cependant certains travaux de la littérature ont utilisé la même base d’images ou des bases d’images avec des caractéristiques similaires. Nous pouvons donc comparer les résultats de ces travaux avec nos résultats. Le tableau V.10 résume les caractéristiques de chaque base d’images utilisée dans ces travaux.

Tableau V.10 : Caractéristiques de chaque base d’images utilisée dans les travaux de la littérature.

Référence	Base de données aériennes		Base de données UAV	
	Endommagé	Non-endommagé	Endommagé	Non-endommagé
[9]	698	558	482	484
[22]	3625	3367	/	/
[49]	3501	3647	2593	2821
[53]	9203	9930	5682	5961
[66]	8921	4367	/	/
[67]	2754	2701	2505	2692
[69]	8751	2180	/	/
[73]	6057	6057	/	/
[110]	/	/	207	337
Notre travail	9203	9930	5682	5961

Le tableau V.11 présente la performance des méthodes proposées dans de la littérature. Pour les deux approches de ML et de DL, les meilleurs résultats sont affichés en **gras**. Pour la comparaison, nous avons choisi les méthodes qui ont fourni les meilleures performances pour

Chapitre V – Optimisation de la classification des dommages via des représentations hybrides de caractéristiques de DL dérivées d'images aériennes post-séisme

les deux bases de données. La figure V.13 montre les comparaisons entre ces méthodes. Comme nous pouvons constater que notre proposition est la plus généralisée et fournit les meilleures performances dans tous les cas. Plus précisément, notre approche est robuste et la plus généralisée lorsque les caractéristiques des images varient fortement en termes d'échelle, de vue de la caméra, de plateforme de capture et de complexité de la scène par rapport les méthodes de la littérature.

Tableau V. 11 : Comparaison de notre approche avec les méthodes de la littérature (en termes de précision de résultats de classification).

/	Référence	Base de données aériennes	Base de données UAV
ML	[9]_Gabor / SVM	79.00	90.00
	[9]_HoG / SVM	72.00	86.00
	[9]_BoW-Gabor / SVM	88.00	95.00
	[9]_BoW- HoG / SVM	91.00	98.00
	[9]_BoW- SURF / SVM	90.00	94.00
	[22]_BoW- (HoG + Couleur) / SVM	91.00	/
	[49]_Nuage de points 3D / SVM	/	81.64
	[63]_GLCM / SVM	86.1	/
DL	[66]_BoW- (HoG + Couleur) / SVM	91.25	/
	[49]_CNN_S	93.62	89.60
	[49]_CNN_T	94.24	92.33
	[49]_CNN_F / SVM	92.18	91.15
	[49]_Nuage de points 3D + CNN / SVM	/	94.2
	[53]_VGG-16/ SVM	95.25	92.87
	[67]_CNN	91.4	95.04
	[69]_YOLOv3-S-GIoU	90.89	/
	[73]_CNN	95.00	/
	[110]_VGG-16	/	91.0
	[Notre_travail]_Hybird-1 /SVM	97.64	97.70

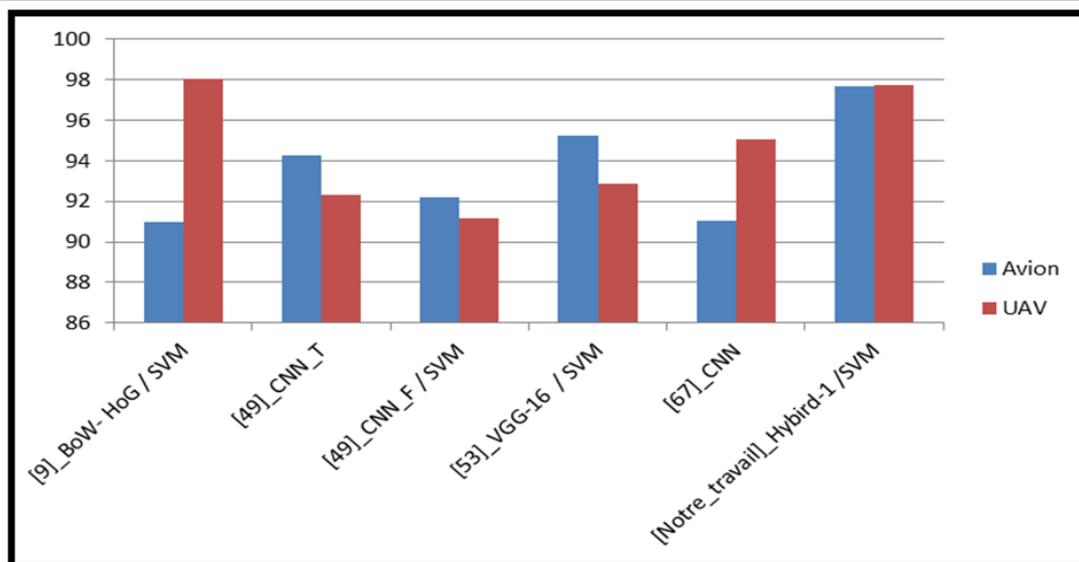


Figure V.13 : Comparaison de notre proposition avec les méthodes de la littérature.

5. Conclusion

Dans ce chapitre, nous avons présenté notre contribution pour optimiser la classification des dommages post-séisme basées sur les données de télédétection aériennes, ainsi que les résultats obtenus et leur validation. Nous avons commencé par illustrer l'architecture globale du système proposé. Ensuite, nous avons présenté séparément chacune de ses deux phases principales. Principalement, nous avons choisi pour la phase d'extraction des caractéristiques une combinaison des vecteurs des caractéristiques extraites à partir de plusieurs modèles pré-entraînés de CNN. Nous avons d'abord exploré la possibilité d'exploiter les vecteurs des caractéristiques des deux couches FC1 et FC2 (FC1 & FC2) des modèles pré-entraînés de CNN. Nous avons ensuite examiné la combinaison des vecteurs des caractéristiques extraites à partir plusieurs modèles pré-entraînés de CNN.

Pour cela, nous avons évalué quatre modèles de CNN pré-entraîné comme un outil d'extraction des caractéristiques : AlexNet, VGG-16, VGG-19 et GoogLeNet. Finalement, nous avons réalisé une étude expérimentale, puis nous avons analysé et interprété les résultats correspondants. Pour la validation, nous avons utilisé deux ensembles d'images capturées par différentes plateformes (avions et UAV), à partir de différents événements sismiques dans plusieurs régions géographiques. Dans toutes nos expériences, nous avons appliqué des tests statistiques sur les résultats numériques pour tirer des conclusions objectives.

Notre premier ensemble d'expériences a été consacré à évaluer le gain de performance lors de l'utilisation de vecteurs de caractéristiques CNN extraits à la fois de FC1 et FC2 (FC1 & FC2). Les résultats de cette session expérimentale ont démontré que :

- (1) Les vecteurs de caractéristiques de FC1 ont conduit à de meilleures performances que les vecteurs des caractéristiques de FC2 dans tous les cas pour AlexNet, VGG-16 et VGG-19.
- (2) Les vecteurs des caractéristiques des deux couches FC1 et FC2 combinées (FC1 & FC2) ont considérablement amélioré les performances des trois modèles CNN (il n'y a pas eu de gain significatif de performances pour GoogleNet).
- (3) Des comparaisons directes entre les quatre CNN étudiés en considérant quelques soient les vecteurs de caractéristiques utilisés FC1, FC2 ou FC1&FC2, et pour les deux bases de données, ont montré que VGG-16 et VGG-19 étaient les gagnants, suivis par AlexNet en deuxième position, tandis que GoogLeNet arrivait en dernière position.

Le deuxième ensemble d'expériences a été réalisé pour évaluer le gain de performance obtenu en combinant des vecteurs de caractéristiques extraits à partir de différents modèles pré-entraînés de CNN. Pour cela, nous avons étudié trois combinaisons pour les vecteurs de caractéristiques extraits de FC1, FC2 et FC1 & FC2, résumées dans le tableau V.6. Cette session expérimentale a montré que :

- (1) Quelques soient les vecteurs de caractéristiques utilisées FC1, FC2 ou (FC1&FC2), chaque méthode hybride étudiée a conduit à de meilleures performances que chaque vecteur de caractéristiques d'origine indépendant.

Chapitre V – Optimisation de la classification des dommages via des représentations hybrides de caractéristiques de DL dérivées d'images aériennes post-séisme

- (2) Des comparaisons entre les méthodes hybrides étudiées ont montré que le vecteur de caractéristiques hybrides qui combine les vecteurs extraits par les quatre CNN utilisés a toujours surpassé les deux autres hybridations.

Conclusion générale

Une augmentation considérable des catastrophes naturelles et de leurs effets sur la population et l'économie a été observée au cours de ces dernières décennies. Chaque année, des catastrophes dévastatrices comme les tremblements de terre causent des degrés élevés de dégâts et des milliers de victimes. La catastrophe naturelle la plus fréquente et la plus dangereuse est la catastrophe sismique qui cause d'énormes dégâts dans le monde entier et des pertes économiques estimées à des milliards de dollars. La détection immédiate de la région endommagée fournit une assistance quasi-instantanée aux blessés et facilite également un bon entretien des structures endommagées. Les développements récents de la télédétection ont accéléré la disponibilité des données d'images satellitaires/aériennes dans le monde entier.

Les approches utilisant des techniques d'apprentissage supervisé se sont révélées utiles pour la détection et la classification des dommages sismiques basées sur les données de télédétection, en particulier pour les méthodes mono-temporelles. Ces approches se composent de deux modules principaux qui sont le module d'extraction des caractéristiques et le module de classification. La performance de ces approches dépend principalement de la puissance des méthodes d'extraction des caractéristiques. Par conséquent, le développement d'une approche d'extraction des caractéristiques qui est capable d'optimiser la classification des dommages sismiques et généraliser bien lorsque les images sont très complexes et très diverses dans leurs caractéristiques telles que l'échelle de l'image, vue de la caméra, la plateforme de capture et la complexité de la scène, c'est fortement désirable. Dans ce travail, nous nous sommes concentrés sur ce problème.

Récemment, des études antérieures ont montré que l'utilisation des méthodes de DL comme CNN sont capables de dériver de puissantes caractéristiques et de donner des résultats remarquables pour la classification des dommages sismiques. Dans ce travail, nous avons utilisé une méthode hybride basée sur le DL dans la phase d'extraction des caractéristiques pour optimiser la classification des dommages sismiques basées sur les données de télédétection aériennes. L'objectif a été atteint en utilisant quatre modèles CNN pré-entraînés : AlexNet, VGG-16, VGG-19 et GoogLeNet et un classificateur SVM. En plus, contrairement aux études précédentes, nous avons exploré dans ce travail la possibilité d'exploiter la combinaison des vecteurs de caractéristiques extraites à la fois de deux premières couches entièrement connectées (FC1 et FC2) de chaque modèle CNN utilisé.

Pour évaluer la performance de l'approche proposée, nous avons utilisé deux ensembles d'images capturées par différentes plateformes (avions et UAV) à partir de différents événements sismiques dans plusieurs régions géographiques. Dans toutes nos expériences, nous avons effectué une validation statistique pour tirer des conclusions objectives et démontrer la performance de notre méthodologie proposée.

Un premier ensemble d'expériences a montré que la combinaison des vecteurs des caractéristiques extraites à la fois de FC1 et FC2 (FC1 & FC2) de chaque modèle CNN utilisé peut améliorer la performance d'un classificateur SVM. Puis, nous avons examiné trois méthodes hybrides en les comparant les unes aux autres puis chacune avec ses vecteurs de

caractéristiques d'origine indépendants. Un autre ensemble d'expériences extensives a montré que la combinaison de vecteurs de caractéristiques tirés de chaque méthode hybride permettait d'obtenir des performances de classification significativement meilleures que tous les vecteurs de caractéristiques CNN indépendant, et le vecteur de caractéristiques hybrides qui combine les vecteurs extraits par les quatre CNN utilisés a toujours surpassé les deux autres hybridations.

D'après les résultats obtenus, nous avons constaté que notre proposition améliore la performance de classification des dommages avec un taux de succès d'environ 97% pour les deux bases d'images (avions et UAV). De plus, nous avons constaté que notre proposition est la plus généralisée et fournit les meilleures performances par rapport les autres méthodes de la littérature, en particulier lorsque les caractéristiques des images varient fortement en termes d'échelle, de vue de la caméra, de plateforme de capture et de complexité de la scène.

Les perspectives de ce travail de thèse sont nombreuses. Parmi les perspectives les plus prometteuses est l'utilisation et la combinaison d'autres modèles CNN pré-entraînés pour améliorer la performance de la classification des dommages sismiques. De plus, comme une autre perspective importante, nous souhaitons étudier la performance de notre proposition pour la classification des dommages d'autres catastrophes telles que les cyclones, les tornades, les inondations, les éruptions volcaniques, etc.

Références

- [1] Sena L. and Woldemichael. K. : "Disaster prevention and preparedness", Lecture Notes - For Health Science Students. Jimma University, Ethiopia Public Heal Train Initiat 2006, 1, 1-80.
- [2] Cours d'introduction à la gestion des catastrophes, Travail collectif, Université virtuelle du common wealth (VUSSC), 2005, pdf sur: https://www.sifec.org/static/uploaded/Files/.../Gestion_Catastrophes_naturelles.pdf, consulté le 21 janvier 2019.
- [3] Wikipedia : "Disaster", sur la page, <https://en.wikipedia.org/wiki/Disaster>, consultée le 20 janvier 2019.
- [4] Site : "The Disaster Management Cycle - GDRC", sur la page, www.gdrc.org/uem/disasters/1-dm_cycle.html, consultée le 22 janvier 2019.
- [5] Rezaeian M. : "Assessment of earthquake damages by image-based techniques", for the degree of Electrical Engineering Doctor of Sciences, University of Tehran, (19178), 149, 2010.
- [6] Site : "Phases of Disaster", sur la page, restoreyoureconomy.org/disaster-overview/phases-of-disaster/, consultée le 21 janvier 2019.
- [7] Ranzato L., Barausse A., Mantovani A., Pittarello A., Benzo M. and Palmeri L. : "PalmeriA comparison of methods for the assessment of odor impacts on air quality: field inspection (VDI 3940) and the air dispersion model CALPUFF", Atmospheric Environment, 61, pp. 570-579, 2012.
- [8] Disaster risk reduction tools and methods for climate change adaptation, Tools and Methods, IATF Working Group on Climate Change and Disaster Risk Reduction, 2005.
- [9] Vetrivel A., Gerke M., Kerle N. and Vosselman G. : "Identification of structurally damaged areas in airborne oblique images using a visual-bag-of-words approach", Remote sensing, 8(3):231, 2016.
- [10] Steves M., Morse E., Gutwin C. and Greenberg S. : "A comparison of usage evaluation and inspection methods for assessing groupware usability", In Proceedings of the 2001 International ACM SIGGROUP Conference on Supporting Group Work. Boulder, Colorado, Sept., ACM Press, pp. 125–134, 2001.
- [11] Yamazaki F., Iwasaki Y., Liu W., Nonaka T. and Sasagawa T. : "Detection of damage to building side-walls in the 2011 Tohoku, Japan earthquake using high-resolution TerraSAR-X images", Proc. SPIE 8892, Image and Signal Processing for Remote Sensing XIX, 889212; doi:10.1117/12.2029465, 2013.
- [12] Ndao M. : "Dynamiques et gestion environnementales de 1970 à 2010 des zones humides au Sénégal : étude de l'occupation du sol par télédétection des Niayes avec Djiddah Thiaroye Kao (à Dakar), Mboro (à Thiès) et Saint-Louis", Thèse de doctorat ès Géographie et Aménagement, Université de Toulouse 2 Le Mirail, cotutelle internationale avec l'Université Gaston Berger de Saint-Louis, Sénégal, 370 p., 2012.
- [13] WU W. : "Application de la géomatique au suivi de la dynamique environnementale en zones Arides - Exemple de la région de Nouakchott en Mauritanie, du Ningxia nord et du Shaanxi nord en Chine du nord-ouest", Thèse de doctorat: Géomatique, L'Université de Paris 1-PanthéonSorbonne et de l'Ecole Pratique des Hautes études,

- 229 p., 2003. Format pdf. Disponible sur: <http://tel.archivesouvertes.fr/docs/00/04/46/46/PDF/Wuthese.pdf>.
- [14] Site : "Télédétection et SIG: Les perspectives d'application", sur la page, <https://www.leconomiste.com/article/teledetection-et-sig-les-perspectives-d-application>, consultée le 23 janvier 2019
- [15] Navalgund R. N., Jayaraman V. and Roy P. S. : "Remote sensing applications: An overview", *current science*, vol. 93, no. 12, pp. 1747-1766, 2007.
- [16] Messoudi W. : "Proposition d'une annotation sémantique floue guidée par ontologie pour l'interprétation des images de télédétection: Application à la gestion des risques naturels", Thèse de Doctorat, Télécom Bretagne, Université de Rennes 1, 2013.
- [17] Module sur les technologies en observation de la terre : "Introduction: les technologies en observation de la terre, Résolution spatiale, taille des pixels et échelle", sur la page, <http://nilerak.hatfieldgroup.com/French/NRAK/EO/html/rsbch6.html>, consultée le 24 janvier 2019.
- [18] A canada centre for Remote Sensing remote sensing tutorial : "Fundamentals of remote sensing", sur la page, <https://www.nrcan.gc.ca/node/9309>, consultée le 24 janvier 2019.
- [19] Ehrlich D., Guo H. D., Molch K., Ma J. W. and Pesaresi M. : "Identifying damage caused by the 2008 Wenchuan earthquake from VHR remote sensing data". *International Journal of Digital Earth* 2 (4), pp. 309-326, 2009.
- [20] Dong L. and Shan J. : "A comprehensive review of earthquake-induced building damage detection with remote sensing techniques". *ISPRS J. Photogrammetry Remote Sens.* 84, pp. 85–99, 2013.
- [21] Vetrivel A., Kerle N., Gerke M., Nex F., Vosselman G. : "Towards automated satellite image segmentation and classification for assessing disaster damage using data-specific features with incremental learning". In *Proceedings of the GEOBIA 2016, Solutions and Synergies*, Enschede, The Netherlands, 14–16 September 2016.
- [22] Tu J., Sui H., Feng W., Sun K., Xu C. and Han Q. : "Detecting building façade damage from oblique aerial images using local symmetry feature and the Gini Index", *Remote Sensing Letters*, 8:7, pp. 676-685, 2017.
- [23] Taskin G., Ersoy O. K. and Kamasak M. E. : "Earthquake-induced damage classification from postearthquake satellite image using spectral and spatial features with support vector selection and adaptation". *J. Appl. Remote Sens.*, 9, 096017, 2015.
- [24] Soumia B. : "Indexation et recherche d'images par contenu dans les grands corpus d'images", Thèse de Doctorat, Informatique : Université Mohamed Boudiaf des sciences et de la technologie, 152 p., 2017.
- [25] Hamadi A. : "Utilisation du contexte pour l'indexation sémantique des images et vidéos", Thèse de Doctorat, du Laboratoire d'informatique de Grenoble et de l'Ecole Doctorale Mathématiques, Sciences et Technologies de l'Information, Informatique (MSTII), 201 p., 2014.
- [26] Cheng G., Han J. and Lu X. : "Remote sensing image scene classification : Benchmark and state of the art", *Proceedings of the IEEE*, vol. PP, no. 99, pp. 1–19, 2017.
- [27] Bessai F. Z., Hamadi A. and Selmoun J. S. : "Indexation et Recherche d'Images par

- le Contenu", Revue d'Information Scientifique et Technique (RIST) 12(02) : pp. 11-25, 2002.
- [28] Houari K. : "recherche d'images par le contenu", Thèse De Doctorat ès Sciences en Informatique, Université mentouri Constantine, 118 p., 2010.
- [29] Desir C. : "Classification automatique d'images, application à l'imagerie du poumon profond", Thèse en vue de l'obtention du diplôme de docteur de l'université de Rouen, 166 p., 2013.
- [30] Lefebvre A. : "Contribution de la texture pour l'analyse d'images à très haute résolution spatiale : application à la détection de changement en milieu périurbain", Thèse de doctorat de géographie, Université Rennes 2 ; Université Européenne de Bretagne, 285 p., 2011.
- [31] Malek S. : "Deep neural network models for image classification and regression", Thèse de doctorat de géographie, Department of Information Engineering and Computer Science University of Trento, 98 p., 2018.
- [32] Doan T. N. and Poulet F. : "Un environnement efficace pour la classification d'images à grande échelle", IRISA, Campus de Beaulieu, 35042 Rennes Cedex, France, 2012.
- [33] Negrel R. : "Représentations optimales pour la recherche dans les bases d'images patrimoniales", Thèse de doctorat de l'université de Cergy-Pontoise, 119 p., 2014.
- [34] Osisanwo F., Akinsola J., Awodele O., Hinmikaiye J. O., Olakanmi O. and Akinjobi J. : "Supervised Machine Learning Algorithms: Classification and Comparison", International Journal of Computer Trends and Technology (IJCTT) 48(3) : pp. 128-138, 2017.
- [35] Soofi A. and Awan A. : "Classification Techniques in Machine Learning: Applications and Issues", Journal of Basic and Applied Sciences 13: 459-465, 2017.
- [36] Kotsiantis S. B. : "Supervised machine learning: A review of classification techniques", Emerging artificial intelligence applications in computer engineering 160 : pp. 3-24, 2007.
- [37] Taskin G., Ersoy O. K. and Kamasak M. E. : "Earthquake-induced damage classification from postearthquake satellite image using spectral and spatial features with support vector selection and adaptation", Journal of Applied Remote Sensing 9(1) : 096017, 2015.
- [38] Pream Sudha V. and Kowsalya R. : "A Survey On Deep Learning Techniques, Applications and Challenges", Department of Computer Science, PSGR Krishnammal College for Women, (India), International Journal of Advanced Technology in Engineering and Science (IJARSE), ISSN-2319-8354(E), Vol. No.4, Issue 03, pp. 311-317, March 2015.
- [39] Abiodun O. I., Jantan A., Omolara A. E., Dada K. V., Mohamed N. A. and Arshad H. : "State-of-the-art in artificial neural network applications : A survey", Heliyon, 4(11), e00938, 2018
- [40] Bhatt N. and Prajapati P. : "Deep Learning : A New Perspective", International Journal of Latest Technology in Engineering, Management & Applied Science (IJLTEMAS), ISSN 2278-2540, pp. 136-140, 2017.
- [41] Site : "CS231n : Convolutional Neural Networks for Visual Recognition", sur la page, <http://cs231n.github.io/neural-networks-1/>, consultée le 4 Mars 2019.
- [42] Zhang C. : "Deep Learning for Land Cover and Land Use Classification", for the

degree of doctor of philosophy, Lancaster Environment Centre, Lancaster University, 2018.

- [43] Site : "Top 15 Deep Learning applications that will rule the world in 2018 and beyond", sur la page, <https://medium.com/@vratulmittal/top-15-deep-learning-applications-that-will-rule-the-world-in-2018-and-beyond-7c6130c43b01>, consultée le 7 Mars 2019.
- [44] Ball J. E., Anderson D. T. and Chan C. S. : "Comprehensive Survey of Deep Learning in Remote Sensing: Theories, Tools, and Challenges for the Community", *Journal of Applied Remote Sensing* 11(4) : 042609, 2017.
- [45] Allili M. S. : "Data analytics and machine learning, Workshop on Deep Learning University of Abderrahmane Mira, Bejaia, 2018.
- [46] Site : "Deep learning—Deep Boltzmann Machine (DBM) ", sur la page, <https://medium.com/datadriveninvestor/deep-learning-deep-boltzmann-machine-dbm-e3253bb95d0f>, consultée le 10 Mars 2019.
- [47] Trivedi S. and Kondor R. : "Lecture 7: Convolutional Neural Networks, CMSC 35246: Deep Learning", University of Chicago, 2017, Cours consultée le 13 Mars 2019. https://ttic.uchicago.edu/~shubhendu/Pages/Files/Lecture7_flat.pdf.
- [48] Cha Y. J., Choi W. and Buyukozturk O. : "Deep learning-based crack damage detection using convolutional neural networks", *Computer-Aided Civil and Infrastructure Engineering* 32(5) : pp. 361-378, 2017.
- [49] Vetrivel A., Gerke M., Kerle N., Nex F. and Vosselman G. : "Disaster damage detection through synergistic use of deep learning and 3d point cloud features derived from very high resolution oblique aerial images, and multiple-kernel-learning". *ISPRS Journal of photogrammetry and remote sensing*, 140 : pp. 45–59, 2018.
- [50] Site : "Common architectures in convolutional neural networks", sur la page, <https://www.jeremyjordan.me/convnet-architectures/>, consultée le 20 avril 2019.
- [51] Hu F., Xia G. S., Hu J. and Zhang L. : "Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery". *Remote Sensing* 7(11): 14680-14707, 2015.
- [52] Site: "The 9 Deep Learning Papers You Need To Know About (Understanding CNNs Part 3)", sur la page, <https://adeshpande3.github.io/The-9-Deep-Learning-Papers-You-Need-To-Know-About.html> , consulté le 21 avril 2019.
- [53] Settou T., Kholadi M. K. and Loid A. : "A Very Deep Learning Feature Representation To Improve Damage Classification Derived From Post-earthquake Aerial Images Using VGG-16 Architecture", the international Conference on Artificial Intelligence and its Applications (AIAP'18), EL OUED, ALGERIA, 04-05 DECEMBER 2018.
- [54] Site : "NASA Disasters Mapping Portal", sur la page, <https://disasters.nasa.gov/earthquakes> , consultée le 01 juin 2019.
- [55] Bartels S. A. and Van Rooyen M. J. : "Medical complications associated with earthquakes". *The Lancet* 379 (9817), pp. 748-757, 2011.
- [56] Miura H., Yamazaki F. and Matsuoka M. : "Identification of damaged areas due to the 2006 Central Java, Indonesia Earthquake using satellite optical images". In: *Proceedings of Urban Remote Sensing Joint Event, Paris, France, 2007.*

- [57] Vu T. T. and Ban Y. : "Context-based mapping of damaged buildings from high-resolution optical satellite images". *International Journal of Remote Sensing* 31 (13), pp. 3411-3425, 2010.
- [58] Ji M., Liu L. and Buchroithner M. : "Identifying Collapsed Buildings Using Post-Earthquake Satellite Imagery and Convolutional Neural Networks: A Case Study of the 2010 Haiti Earthquake", *Remote Sensing*, 10(11) :1689, Nov. 2018.
- [59] Ji. M., Liu L., Du R. and Buchroithner M. : "A Comparative Study of Texture and Convolutional Neural Network Features for Detecting Collapsed Buildings After Earthquakes Using Pre-and Post-Event Satellite Imagery", *Remote Sensing* 11(10) : 1202, 2019.
- [60] Li P., Xu H. and Guo J. : "Urban building damage detection from very high resolution imagery using OCSVM and spatial features", *Int. J. Remote Sens.* 31, pp. 3393–3409. 2010.
- [61] Vu T. T., Matsuoka M. and Yamazaki F. : "Detection and animation of damage using very high-resolution satellite data following the 2003 Bam, Iran earthquake", *Earthquake Spectra* 21 (S1), pp. S319–S327, 2005.
- [62] Ishii M., Goto T., Sugiyama T., Saji H. and Abe K. : "Detection of earthquake damaged areas from aerial photographs by using color and edge information", In: *5th Asian Conference on Computer Vision*, Melbourne, Australia, 2002.
- [63] Yu H., Cheng G. and Ge X. : "Earthquake-collapsed building extraction from LiDAR and aerophotograph based on OBIA", In: *2nd International Conference on Information Science and, Engineering (ICISE)*, pp. 2034-2037, 2010.
- [64] Taskin G., Ersoy O. K. and Kamasak M. E. : "Earthquake-induced damage classification from postearthquake satellite image using spectral and spatial features with support vector selection and adaptation", *J. Appl. Remote Sens.* 9, 096017, 2015.
- [65] Sun W., Shi L., Yang J. and Li P. : "Building collapse assessment in urban areas using texture information from post-event SAR data", *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, Vol. 9, No. 8, pp. 3792-3808, 2016.
- [66] Tu J., Sui H., Feng W. and Song Z. : "Automatic building damage detection method using high-resolution remote sensing images and 3D GIS model", *ISPRS Ann. Photogrammetry, Remote Sens. Spatial Inf. Sci.* III-8, pp. 43-50, 2016.
- [67] Duarte D., Nex F., Kerle N. and Vosselman G. : "Multi-resolution feature fusion for image classification of building damages with convolutional neural networks", *Remote Sensing* 10(10) : 1636, 2018.
- [68] Duarte D., Nex F., Kerle N. and Vosselman G. : "Towards a more efficient detection of earthquake induced façade damages using oblique UAV imagery", *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W6, pp. 93-100. <https://doi.org/10.5194/isprs-archives-XLII-2-W6-93-2017>, 2017.
- [69] Ma H., Liu Y., Ren Y. and Yu J. : "Detection of Collapsed Buildings in Post-Earthquake Remote Sensing Images Based on the Improved YOLOv3", *Remote Sensing*, 12, 44, 2020.
- [70] Duarte D., Nex F., Kerle N. and Vosselman. G. : "Satellite Image Classification Of Building Damages Using Airborne And Satellite Image Samples In A Deep Learning Approach", *ISPRS Annals Of Photogrammetry*, 2018.

- [71] Bhangale U., Durbha S., Potnis A. and Shinde R. : "Rapid Earthquake Damage Detection Using Deep Learning from VHR Remote Sensing Images", IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium, IEEE, pp. 2654-2657, 2019.
- [72] Nex F., Duarte D., Tonolo F. G. and Kerle N. : "Structural building damage detection with deep learning: Assessment of a state-of-the-art cnn in operational conditions", *Remote Sensing*, 11, 2765, 2019.
- [73] Miura H., Aridome T. and Matsuoka M. : "Deep Learning-Based Identification of Collapsed, Non-Collapsed and Blue Tarp-Covered Buildings from Post-Disaster Aerial Images", *Remote Sensing*, 12, 1924, 2020.
- [74] Xiong C., Li Q. and Lu X. : "Automated regional seismic damage assessment of buildings using an unmanned aerial vehicle and a convolutional neural network", *Automation in Construction*, 109, 102994, 2020.
- [75] Nogueira K., Fadel S. G., Dourado I. C., Werneck R. O., Munoz J. A. V., Penatti O. A. B., Calumby R. T., Li L. T., Santos J. A. and Torres R. S. : "Exploiting ConvNet diversity for flooding identification", *IEEE Geoscience and Remote Sensing Letters* 15(9) : pp. 1446-1450, 2018.
- [76] Gebrehiwot A., Hashemi-Beni L., Thompson G., Kordjamshidi P. and Langan T. E. : "Deep Convolutional Neural Network for Flood Extent Mapping Using Unmanned Aerial Vehicles Data", *Sensors* 19(7) : 1486, 2019.
- [77] Tehrany M. S., Pradhan B., Mansor S. and Ahmad N. : "Flood susceptibility assessment using GIS-based support vector machine model with different kernel types", *Catena* 125 : pp. 91-101, 2015.
- [78] Ali K., Bajracharya R. M. and Koirala H. L. : "A review of flood risk assessment", *International Journal of Environment, Agriculture and Biotechnology* 1(4), 2016.
- [79] Pulvirenti L., Pierdicca N., Chini M. and Guerriero L. : "An algorithm for operational flood mapping from synthetic aperture radar (SAR) data based on the fuzzy logic", *Natural Hazard and Earth System Sciences*, 2011.
- [80] Tong X., Luo X., Liu S., Xie H., Chao W., Liu S., Makhinov A. N., Makhinova A. F. and Jiang Y. : "An approach for flood monitoring by the combined use of Landsat 8 optical imagery and COSMO-SkyMed radar imagery", *ISPRS Journal of Photogrammetry and Remote Sensing* 136 : pp. 144-153, 2018.
- [81] Chapi K., Singh V. P., Shirzadi A., Shahabi H., Buid D. T., Phum B. T. and Khosravi K. : "A novel hybrid artificial intelligence approach for flood susceptibility assessment", *Environmental modelling & software* 95 : pp. 229-245, 2017.
- [82] Popescu D., Ichim L. and Stoican F. : "Unmanned aerial vehicle systems for remote estimation of flooded areas based on complex image processing", *Sensors* 17(3) : 446, 2017.
- [83] Feng Q., Liu J. and Gong J. : "Urban flood mapping based on unmanned aerial vehicle remote sensing and random forest classifier-A case of Yuyao, China", *Water* 7(4) : pp. 1437-1455, 2015.
- [84] Cetin M., Iftekharruddin K., Sazara C., Goodall J. and Shen Y. : "Estimating Road Inundation Levels Due to Recurrent Flooding from Image Data", 2019.
- [85] Lamovec P., Velkanovski T., Mikoš M. and Oštir K. : "Detecting flooded areas with machine learning techniques: case study of the Selška Sora river flash flood in

September 2007", *Journal of Applied Remote Sensing* 7(1) : 073564, 2013.

- [86] Son K. Y., Yildirim M. E., Park J. S. and Song J. K. : "Flood detection by using FCN-AlexNet", *Proc. SPIE 11041, Eleventh International Conference on Machine Vision (ICMV 2018)*, 110412P (15 March 2019), DOI: 10.1117/12.2523028; <https://doi.org/10.1117/12.2523028>, 2019.
- [87] Ireland G., Volpi M. and Petropoulos J. P. : "Examining the capability of supervised machine learning classifiers in extracting flooded areas from Landsat TM imagery: A case study from a Mediterranean flood", *Remote Sensing* 7(3) : pp. 3372-3399, 2015.
- [88] Sghaier M. O., Hammami I., Foucher S. and Lepage S. : "Flood extent mapping from time-series SAR images based on texture analysis and data fusion", *Remote Sensing* 10(2) : 237, 2015.
- [89] Pradhan B., Tehrany M. S. and Jebur M. N. : "A new semiautomated detection mapping of flood extent from TerraSAR-X satellite image using rule-based classification and taguchi optimization techniques", *IEEE Transactions on Geoscience and Remote Sensing* 54(7) : pp. 4331-4342, 2016.
- [90] Kaya G. T., Tehrany M. S. and Jebur M. N. : "A comprehensive analysis of earthquake damage patterns using high dimensional model representation feature selection", *Image and Signal Processing for Remote Sensing XIX, International Society for Optics and Photonics*, 2013.
- [91] Ichim L., Popescu D., Jebur M. N. : "Combining Color Fractal with LBP Information for Flood Segmentation in UAV-Based Images", *International Conference on Image Analysis and Processing, Springer*, 2017.
- [92] Bischke B., Bhardwaj P. and Gautam A. : "Detection of Flooding Events in Social Multimedia and Satellite Imagery using Deep Neural Networks", *MediaEval*, 2017.
- [93] Said N., Pogorelov K., Ahmad K. and Riegler M. : "Deep learning approaches for flood classification and flood aftermath detection", *Proceedings of the MediaEval 2018 Workshop, Sophia-Antipolis, France*, 2018.
- [94] Rahnemoonfar M., Murphy R., Miquel M. V., Dobbs D. and Adams A. : "Flooded area detection from uav images based on densely connected recurrent neural networks", *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium, IEEE*, 2018.
- [95] Cirneanu A. L. and Popescu D. : "Flooded Area Detection By Gabor Filtering Based On Convolutional Neural Network", *University Politehnica Of Bucharest Scientific Bulletin Series C-Electrical Engineering And Computer Science* 81(1): 69-80, 2019.
- [96] Kakde A., Arora N. and Sharma D. : "Fire Detection System Using Artificial Intelligence Techniques". *International Journal of Research in Engineering, IT and Social Sciences*, Volume 08, Issue 11, pp. 1-5, 2018.
- [97] Rasouli S., Granmo O. C. and Radianti J. : "A methodology for fire data analysis based on pattern recognition towards the disaster management". *2015 2nd International Conference on Information and Communication Technologies for Disaster Management (ICT-DM)*, IEEE, 2015.
- [98] Yuan C., Liu Z. and Zhang Y. : "UAV-based forest fire detection and tracking using image processing techniques", *2015 International Conference on Unmanned Aircraft Systems (ICUAS)*, IEEE, 2015.

- [99] Angayarkkani K. and Radhakrishnan N. : "An intelligent system for effective forest fire detection using spatial data", *International Journal of Computer Science and Information Security*, Vol. 7, No. 1, arXiv preprint arXiv:1002.2199, 2010.
- [100] Zhao K., Ma J., Li X. and Zhang J. : "Saliency detection and deep learning-based wildfire identification in UAV imagery", *Sensors* 18(3) : 712, 2018.
- [101] Muhammad K., Ahmad J., Mehmood I., Rho S. and Baikl S. W. : "Convolutional neural networks based fire detection in surveillance videos", *IEEE Access* 6 : pp. 18174-18183, 2018.
- [102] Vipin V. : "Image processing based forest fire detection", *International Journal of Emerging Technology and Advanced Engineering* 2(2) : pp. 87-95, 2012.
- [103] Stipaničev D. : "Intelligent Forest Fire Monitoring System—from idea to realization", *Annual of the Croatian Academy of Engineering* 2010 : 58, 2012.
- [104] Çetin A. E., Dimitropoulos K., Gouverneur B., Grammalidis N., Günay O., Habiboglu Y. H., Töreyn B. U. and Verstockt S. : "Video fire detection—review", *Digital Signal Processing* 23(6) : pp. 1827-1843, 2013.
- [105] Chen T., Wu P. and Chiou Y. : "An early fire-detection method based on image processing". In *Proceedings of IEEE International Conference on Image Processing*, Vol. 3, pp. 1707-1710, 2004.
- [106] Kim S., Lee W., Park Y., Lee H. and Lee Y. : "Forest fire monitoring system based on aerial image", *3rd International Conference on Information and Communication Technologies for Disaster Management (ICT-DM)*, pp. 1-6, IEEE, 2016.
- [107] Töreyn B. U., Dedeoğlu Y., Gudukbay U. and Çetin A. E. : "Computer vision based method for real-time fire and flame detection", *Pattern recognition letters* 27(1) : pp. 49-58, 2006.
- [108] Ko B. C., Cheong K. H. and Nam J. Y. : "Fire detection based on vision sensor and support vector machines", *Fire Safety Journal* 44(3) : pp. 322-329, 2009.
- [109] Frizzi S., Kaabi R., Bouchouicha M., Ginoux J. M., Moreau E. and Fnaiech F. : "Convolutional neural network for video fire and smoke detection", *IECON 2016-42nd Annual Conference of the IEEE Industrial Electronics Society*, IEEE, 2016.
- [110] Kamilaris A. and Prenafeta-Boldú F. X. : "Disaster monitoring using unmanned aerial vehicles and deep learning", arXiv preprint arXiv:1807.11805, 2018.
- [111] Muhammad K., Ahmad J. and Baikl S. W. : "Early fire detection using convolutional neural networks during surveillance for effective disaster management", *Neurocomputing* 288 : pp. 30-42, 2018.
- [112] Pourghasemi H., Gayen A., Park S., Lee C. W. and Lee S. : "Assessment of landslide-prone areas and their zonation using logistic regression, logitboost, and naïvebayes machine-learning algorithms", *Sustainability* 10(10) : 3697, 2018.
- [113] Chen W., Xie X., Wang J., Pradhan B., Hong H., Bui D. T., Duan Z. and, Ma J. : "A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility", *Catena* 151 : pp. 147-160, 2017.
- [114] Mezaal M. R., Pradhan B., Sameen M. I., Shafri H. Z. M. and Yusoff Z. M. : "Optimized neural architecture for automatic landslide detection from high-resolution airborne laser scanning data", *Applied Sciences* 7(7) : 730, 2017.
- [115] Plank S., Twele A. and Martinis S. : "Landslide mapping in vegetated areas using

- change detection based on optical and polarimetric SAR data", *Remote Sensing* 8(4) : 307, 2016.
- [116] Ghorbanzadeh O., Blaschke T., Gholamnia K., Meena S.R., Tiede D. and Aryal J. : "Evaluation of different machine learning methods and deep-learning convolutional neural networks for landslide detection", *Remote Sensing* 11(2) : 196, 2019.
- [117] Ding A., Zhang Q., Zhou X. and Dai B. : "Automatic recognition of landslide based on CNN and texture change detection", 2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC), IEEE, 2016.
- [118] Dou J., Qian J., Zhanga H., Chena S., Zhenga X., Zhua J., Xiea Z. and Zoub Y. : "A Landslides detection: a case study in Conghua city of Pearl River delta", Second International Conference on Earth Observation for Global Changes, International Society for Optics and Photonics, 2009.
- [119] Hölbling D., Füreder P., Antolini F., Cigna F., Casagli N. and Lang S. : "A semi-automated object-based approach for landslide detection validated by persistent scatterer interferometry measures and landslide inventories", *Remote Sensing* 4(5) : pp. 1310-1336, 2012.
- [120] Roback K., Clark M. K., West A. J., Zekkos D., Li G., Gallen S. F., Chamlagain D. and Godt J. W. : "The size, distribution, and mobility of landslides caused by the 2015 Mw7. 8 Gorkha earthquake, Nepal", *Geomorphology* 301 : pp. 121-138, 2018.
- [121] Regmi A. D., Dhital M. R., Zhang J. Q., Su L. J. and Chen X. Q. : "Landslide susceptibility assessment of the region affected by the 25 April 2015 Gorkha earthquake of Nepal", *Journal of Mountain Science* 13(11) : pp. 1941-1957, 2016.
- [122] Manconi A., Casu F., Ardizzone F., Bonano M., Cardinali M., Luca C. D., Gueguen E., Marchesini I., Parise M, Vennari C., Lanari R. and Guzzetti F. : "Brief communication : rapid mapping of landslide events : the 3 December 2013 Montescaglioso landslide, Italy", *Natural hazards and earth system sciences* 14(7) : pp. 1835-1841, 2014.
- [123] Yu H., Ma Y., Wang L., Zhai Y. and Wang X. : "A landslide intelligent detection method based on CNN and rsg_r", *IEEE International Conference on Mechatronics and Automation (ICMA)*, IEEE, 2017.
- [124] Wang L., Wang S., Zhou Y., Liu W. and Wang.F. : "Remote sensing change detection method based on texture characteristic in natural disaster monitoring", *Journal of Catastrophology*, Vol. 29, No. 3, pp. 97-101, in Chinese, 2014.
- [125] Wan B. F., Yuan S. H. and Su J. P. : "Landslide Remote Sensing Image Recognition Based on Texture Analysis", *Surveying and Mapping of Geology and Mineral Resources*, Issue 2, pp. 11-14, 2009.
- [126] Wei X. : "Image Recognition of Landslide Hazard Based SVM", *Information Technology and Image Processing*, issue 08, pp. 65-70, 2013.
- [127] Van Westen C. J., Castellanos E. and Kuriakose S. L. : "Spatial data for landslide susceptibility, hazard and vulnerability assessment, an overview", *Eng. Geol.* 2008, 102, pp. 112-131, 2008.
- [128] Stumpf A. and Kerle N. : "Object-oriented mapping of landslides using Random Forests", *Remote Sens. Environ.*, 115 p., 2564–2577, 2011.
- [129] Pradhan B. and Mezaal M. R. : "Optimized Rule Sets for Automatic Landslide Characteristic Detection in a Highly Vegetated Forests", In *Laser Scanning Applications in Landslide Assessment*; Springer: New York, NY, USA, pp. 51-68,

2017.

- [130] Settou T., Kholadi M. K. and Ben Ali A. : "Improving damage classification via hybrid deep learning feature representations derived from post-earthquake aerial images", International Journal of Image and Data Fusion, DOI : 10.1080/19479832.2020.1864787, 2020.
- [131] Site : "Understanding Why MATLAB Is Best Suited For Deep Learning", sur la page, <https://analyticsindiamag.com/understanding-why-matlab-is-best-suited-for-deep-learning/> , consulté le 25 Novembre 2020.
- [132] Corder G. W. and Foreman D. I. : "Nonparametric statistics for non-statisticians", John Wiley & Sons, Inc., 2011.