

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Institut National Polytechnique de Toulouse (Toulouse INP)

Discipline ou spécialité :

Informatique et Télécommunication

Présentée et soutenue par :

M. BOUKAYE BOUBACAR TRAORE

le vendredi 23 novembre 2018

Titre :

Modélisation des informations et extraction des connaissances pour la gestion des crises

---

Ecole doctorale :

Aéronautique-Astronautique (AA)

Unité de recherche :

Laboratoire de Génie de Productions de l'ENIT (E.N.I.T-L.G.P.)

Directeur(s) de Thèse :

M. BERNARD KAMSU-FOGUEM

M. FANA TANGARA

Rapporteurs :

M. ALBERT DIPANDA, UNIVERSITE DE BOURGOGNE

Mme SAMIA BOUKIR, INSTITUT POLYTECHNIQUE DE BORDEAUX

Membre(s) du jury :

M. ALBERT DIPANDA, UNIVERSITE DE BOURGOGNE, Président

M. BERNARD KAMSU-FOGUEM, ECOLE NATIONALE D'INGENIEUR DE TARBES, Membre

M. FANA TANGARA, UNIV. DES SCIENCES TECH & TECHNO BAMAKO, Membre

Mme ANNIG LACAYRELLE, UNIVERSITE DE PAU ET DES PAYS DE L ADOUR, Membre

Mme MAGUELONNE TEISSEIRE, IRSTEA, Membre

Modélisation des informations et extraction des pour la gestion des crises

Boukaye Boubacar connaissances TRAORE

Le 23 novembre 2018

Thèse dirigée par
M. Bernard KAMSU-FOGUEM et M. Fana TANGARA

Jury

Mme Samia BOUKIR, Professeur des Universités, Rapporteur
M. Albert DIPANDA, Professeur des Universités, Rapporteur
Mme Annig LACAYRELLE, Maître de Conférences, Examinateur
Mme Maguelonne TEISSEIRE, Professeur des Universités, Examinateur
M. Bernard KAMSU-FOGUEM, Maître de Conférences, Directeur de thèse
M. Fana TANGARA, Professeur des Universités, Co-directeur de thèse

Résumé

L'essor des technologies émergentes de collecte de données offre des opportunités nouvelles pour diverses disciplines scientifiques. L'informatique est appelé à jouer sa partition par le développement de techniques d'analyse intelligente des données pour apporter un certain éclairage dans la résolution de problèmes complexes. Le contenu de ce mémoire de recherche doctorale s'inscrit dans la problématique générale de l'extraction des connaissances à partir de données par les techniques informatiques. Ce travail de thèse s'intéresse dans un premier temps à la problématique de la modélisation des informations pour la gestion de crise nécessitant des prises en charge médicale, à l'aide d'une collaboration des applications informatiques de la télémédecine. Nous avons proposé une méthodologie de gestion d'une crise à distance en trois étapes. Elle est principalement axée sur la collaboration des actes de télémédecine (Téléconsultation, Téléexpertise, Télésurveillance, Téléassistance, et la Régulation médicale), de la phase de transport des victimes à la phase de traitements médicaux dans et/ou entre les structures de santé. Cette méthodologie permet non seulement de mettre à la disposition des gestionnaires de crise un système d'aide à la décision informatisé, mais aussi de minimiser les coûts financiers et réduire le temps de réponse des secours à travers une gestion organisée de la crise. Dans un deuxième temps, nous avons étudié en détail l'extraction de la connaissance à l'aide des techniques de data mining sur les images satellitaires afin de découvrir des zones à risques d'épidémie, dont l'étude de cas a porté sur l'épidémie de choléra dans la région de Mopti, au Mali. Ainsi, une méthodologie de six phases a été présentée en mettant en relation les données collectées sur le terrain et les données satellitaires pour prévenir et surveiller plus efficacement les crises d'épidémie. Les résultats nous indiquent qu'à 66% le taux de contamination est lié au fleuve Niger, en plus de certains facteurs sociétaux comme le jet des ordures en période hivernale. Par conséquent, nous avons pu établir le lien entre l'épidémie et son environnement d'évolution, ce qui permettra aux décideurs de mieux gérer une éventuelle crise d'épidémie. Et enfin, en dernier lieu, pendant une situation de crise d'épidémie, nous nous sommes focalisés sur l'analyse médicale, plus précisément par l'usage des microscopes portables afin de confirmer ou non la présence des agents pathogènes dans les prélèvements des cas suspects. Pour ce faire, nous avons présenté une méthodologie de six phases, basée sur les techniques du deep learning notamment l'une des techniques des réseaux de neurones convolutifs, l'apprentissage par transfert qui tire parti des systèmes complexes avec des invariants permettant la modélisation et l'analyse efficace de grandes quantités de données. Le principe consiste à entraîner les réseaux de neurones convolutifs à la classification automatique d'images des agents pathogènes. Par exemple dans notre cas d'étude, cette approche a été utilisée pour distinguer une image microscopique contenant le virus de l'épidémie de choléra appelé Vibrio choléra d'une image microscopique contenant le virus de l'épidémie du paludisme appelé Plasmodium. Ceci nous a permis d'obtenir un taux de réussite de classification sur la base de validation et de test de 97%. Par la suite, l'idée est de déployer cette solution de reconnaissance d'images d'agents pathogènes dans les microscopes portables intelligents pour les analyses de routine et applications de diagnostic médical dans la gestion de situations de crise. Ce qui permettra de combler le manque de spécialistes en manipulation microscopique et un gain de temps considérable dans l'analyse des

prélèvements avec des mesures précises favorisant l'accomplissement du travail dans de meilleures conditions.

Mot-clés : Fouille des données, Gestion à distance d'une crise, Images satellitaires, Zones à risque d'épidémie, Réseaux de neurones convolutifs, Classification d'images d'agents pathogènes.

Abstract

The rise of emerging data collection technologies offers new opportunities for various scientific disciplines. IT is expected to play its part by developing intelligent data analysis techniques to provide some insight into solving complex problems. The content of this doctoral research dissertation is part of the general problem of extracting knowledge from data by computer techniques. This PhD work focuses, firstly, to the problem of information modeling for crisis management requiring medical care, using a collaboration of computer applications of telemedicine. We proposed a methodology for managing a remote crisis in three stages. It is mainly focused on the collaboration of telemedicine acts (Teleconsultation, Teleexpertise, Telemonitoring, Remote Assistance, and Medical Regulation), from the transport phase of victims to the phase of medical treatment in and / or between health structures. This methodology not only helps to provide crisis managers with a computerized decision aid system, but also to minimize the financial costs and to reduce the response time of emergency through an organized management of the crisis. Secondly, we studied in detail the extraction of knowledge using data mining techniques on satellite images to discover epidemic r risk areas, including the case study focused on the cholera epidemic in the region of Mopti, Mali. Thus, a methodology of six phases was presented by relating the data collected in the field and satellite data to prevent and more effectively monitor the epidemic crises. The results show that 66% of the contamination rate is related to the Niger River, in addition to certain societal factors such as garbage dumps in winter. As a result, we have been able to establish the link between the epidemic and its development environment, which will enable decision makers to better manage a possible crisis of epidemics. Finally, during an epidemic crisis situation, we focused on medical analysis, more specifically by the use of portable microscopes to confirm or not the presence of pathogens in samples of case suspects. To do this, we have presented a methodology in six phases, based on the techniques of deep learning including one of convolutional neural network techniques, transfer learning that take advantage of complex systems and analysis of large amounts of data. The idea is to train networks convolutional neural automatic image classification pathogens. For example, in our case study, this approach was used to distinguish a microscopic image containing the cholera epidemic virus called Vibrio cholerae from a microscopic image containing the malaria epidemic virus called Plasmodium. This allowed us to obtain good performances with a classification accuracy of 97% on validation and test databases. Subsequently, the idea is to deploy this pathogen image recognition solution in intelligent portable microscopes for routine analysis and medical diagnostic applications in crisis management. This will make it possible to fill the lack of specialists in microscopic manipulation and a considerable time saving in the analysis of the samples with precise measures favoring the accomplishment of the work under better conditions.

Keywords: Data Mining, Remote Crisis Management, Satellite Images, Epidemic Risk Areas, Convolutional Neural Networks, Pathogen Image Classification.

Remerciements

Au terme de mes travaux de recherche, je me vois dans l'obligation nonobstant la pauvreté des mots pour exprimer mes sentiments de remerciement.

Mes remerciements s'adressent en premier lieu à mon directeur de thèse Monsieur **Bernard KAMSU-FOGUEM** (Maître de Conférences à l'École Nationale d'Ingénieurs de Tarbes) et à mon codirecteur de thèse Monsieur le Doyen de la Faculté des Sciences et Techniques (FST) **Fana TANGARA** (Professeur à l'Université des Sciences, des Techniques et des Technologies de Bamako) dont la confiance, les conseils, la disponibilité et la diligence ont permis la réalisation de cette thèse.

Ensuite mes remerciements vont aux membres du jury :

- Je remercie le professeur **Albert DIPANDA** d'avoir accepté de présider le jury de la soutenance de thèse.
- Merci à Madame **Samia BOUKIR**, Professeur à l'Institut Polytechnique de Bordeaux et à Monsieur **Albert DIPANDA**, Professeur à l'Université de Bourgogne pour m'avoir fait l'honneur d'être rapporteurs de cette thèse. Merci pour leurs commentaires et leurs remarques intéressantes pour l'amélioration de mes travaux de recherche.
- Je suis également très reconnaissant envers Mesdames **Annig LACAYRELLE**, Professeur à l'Université de Pau et **Maguelonne TEISSEIRE**, Professeur à l'Institut National de Recherche en Sciences et Technologies pour l'Environnement et l'Agriculture de Montpellier, pour avoir accepté d'examiner ce travail.

Je voudrais également remercier mon père Dr El Hadj Boly TRAORE, Maitre-assistant à Université des Lettres et des Sciences Humaines de Bamako, Mali (ULSHB) pour la relecture de mon manuscrit de thèse ; de même, Monsieur Shester GUEUWOU pour la relecture approfondie de mes articles publiés en anglais.

Je ne saurai laisser passer sous silence la disponibilité et l'amabilité des enseignants chercheurs, des doctorants, du personnel administratif du Laboratoire Génie de Production (LGP) de Tarbes. Je garde un souvenir inoubliable des services rendus par le secrétariat du LGP entre autres la qualité de l'accueil, de l'installation et de la bonne collaboration notamment, Mesdames Eliane CASTA, Marie Cécile De BARROS, Marie TEIXEIRA.

Qu'il me soit permis de remercier aussi la coopération bilatérale entre le Mali et la France à travers la réalisation de cette thèse en cotutelle internationale entre l'Institut National Polytechnique de Toulouse (INPT) et l'Université des Sciences, des Techniques et des Technologies de Bamako (USTTB). Egalement, je voudrais remercier le Programme de Formation de Formateurs (PFF) des Universités de Bamako pour le financement de cette thèse (mention spéciale à son coordinateur Dr Sinaly DEMBELE).

Je remercie le Professeur Ouaténi DIALLO, Vice-recteur de l'USTTB, pour sa confiance et ses multiples conseils pour mener à bien ce travail de recherche. Un grand merci à l'ensemble du

personnel de la FST, spécialement du Département de Mathématiques et Informatique et son chef Dr Yaya KONE, Maître de Conférences.

Je remercie aussi tous mes amis, particulièrement Modibo BOUARE pour les multiples accueils chaleureux à Paris, Tamba CAMARA, Abdoulaye DIAMOUTENE, Lassana COULIBALY, Solemane COULIBALY, Mamadou Bilo DOUMBOUYA, Alsane SENE et la famille KAMSU pour les multiples invitations pendant mes séjours à Tarbes.

Enfin, mes remerciements vont à ma famille : mon père Dr El Hadj Boly TRAORE, ma mère Madame TRAORE Doussou TRAORE, mes sœurs Assitan TRAORE, Fatoumata TRAORE et Salimatou TRAORE, mon épouse Madame TRAORE Makan DEMBELE et à mon fils Sidy Pertin TRAORE pour m'avoir soutenu et encouragé avec patience durant toutes ces années d'études.

Table des figures

Figure 1.1– Gestion de crises à l'aide de la modélisation des informations et l'ECD 7
Figure 1.2 – Principaux actes de télémédecine (Barge et al., 2012) 9
Figure 1.3– Les niveaux d'interopérabilité (Kotze et al., 2013) 11
Figure 1.4 – Principe de la télédétection spatiale (El Merabet et al., 2015) 12
Figure 1.5 – Télédétection passive 13
Figure 1.6 – Télédétection active 13
Figure 1.7 – Principe de la réflectance solaire 14
Figure 1.8 – Quelques exemples d'application de la télédétection spatiale : carte d'occupation du sol, suivi des conditions atmosphériques 15
Figure 1.9 – Le spectre électromagnétique (Khan Academy, 2015) 16
Figure 1.10 – les différents niveaux de résolution d'une image satellitaire 17
Figure 1.11 – La résolution radiométrique 17
Figure 1.12 – Processus d'extraction des connaissances à partir des données (Fayyad et al., 1996) 19
Figure 1.13 – Neurone biologique à gauche et neurone formel à droite 23
Figure 1.14 – Le perceptron 1957- F. Rosenblatt (Dejasmin, 2018) 24
Figure 1.15 – Fonction d'activation échelon avec un seuil de 0 25
Figure 1.16 – Les perceptrons multicouches (Dejasmin, 2018) 25
Figure 1.17 – Description d'une scène d'image par les techniques du deep learning 26
Figure 1.18 – Principe de la convolution 27
Figure 1.19 – Exemple d'une convolution dont la configuration est : Opération = Argument maximal, pas horizontal = 1 pixel, pas vertical = 1 pixel 28
Figure 1.20 – Architecture d'un réseau de neurones convolutif 28
Figure 1.21 – Réduction de la taille de la carte de caractéristiques 29
Figure 2.1– Les étapes de la gestion de crise 36
Figure 2.2 – Les différents actes de télémédecine 37
Figure 2.3 – Les étapes d'EDM pour la gestion des catastrophes naturelles 39
Figure 2.4 – Organisation de l'unité de gestion de crise (SIRACEDPC et CYPRES, 2014) 40
Figure 2.5 – Le processus EMS typique (Aboueljinane et al., 2013) 41
Figure 2.6 – Framework général de gestion de crise (Bénaben, 2016) 44
Figure 2.7 – Processus de gestion de crise à distance 46
Figure 2.8 – Gestion à distance de l'exercice de simulation de Tsunami à Cannes 54
Figure 3.1 – Processus de découverte des zones à risque d'épidémie 64
Figure 3.2 – Capteur satellite LANDSAT 8 (15 m) 65
Figure 3.3 – Résultat de la discrétisation sur un échantillon (Xie et al., 2016) 68
Figure 3.4 – Répartition des cas de choléra au Mali de 1995 à 2004 par région 73
Figure 3.5 – Zone d'étude : Région de Mopti (département de Mopti en bande rouge) 73
Figure 3.6 – Image satellite de Mopti (Chemin = 197 et Ligne = 50) 74
Figure 3.7 – Fichier NDVI 75
Figure 3.8 – Graphique correspondant à la base de données 78
Figure 4.1 – Ancienne méthode d'extraction des caractéristiques 87

Figure 4.2 – Deep Learning avec plusieurs couches cachées....87
Figure 4.3 – L'architecture générale de CNN....88
Figure 4.4 – Profondeur de la couche 89
Figure 4.5 – Le pas de déplacement du filtre 89
Figure 4.6 – La marge 89
Figure 4.7 – Processus de Max pooling 90
Figure 4.8 – Couche entièrement connectée 91
Figure 4.9 – Vibrio choléra (Wang et al., 2017) 94
Figure 4.10 – Plasmodium falciparum (Walker et al., 2018)....95
Figure 4.11 – Apprentissage par transfert sur les images d'agents pathogènes épidémiques....95
Figure 4.12 – Quelques images de notre base d'apprentissage....96
Figure 4.13 – Architecture VGG-16 (Monasse and Nadjahi, 2018) 97
Figure 4.14 – Processus d'extraction de caractéristiques (Chollet, 2018) 98
Figure 4.15 – Quelques images de la base de test....101
Figure 4.16 – Evaluation du modèle sur la base d'évaluation et la base de test 101
Figure 4.17 – La courbe représentative de la précision....103
Figure 4.18 – La courbe représentative de la perte 103
Figure 4.19 – Résultat de classification de quelques images de test....104
Figure 5.1 - Cadre général de gestion des crises 111

Liste des tableaux

Tableau 1.1 – Exemple de table d'une base de données 21
Tableau 2.1 – Classification des risques majeurs (Gupta et al., 2002) 35
Tableau 2.2 – Les caractéristiques du tsunami 52
Tableau 3.1 – Bandes spectrales OLI et TIRS (LANDSAT 8) 65
Tableau 3.2 – Définition des indices spectraux 67
Tableau 3.3 – Répartition des cas de choléra au Mali de 1995 à 2004 par région 72
Tableau 3.4 – Images satellitaires traitées 76
Tableau 3.5 – Précisions de la classification des différentes images satellitaires 77
Tableau 3.6 – Résultats après classification 77
Tableau 3.7 – Base de données des cas confirmés et des zones à risque 78
Tableau 3.8 – Système de codification de la base de données 78
Tableau 3.9 – Base de données codifiée 79
Tableau 3.10 – Interprétation des résultats obtenus 79
Tableau 4.1 – Les paramètres des différentes couches d'une architecture CNN 91
Tableau 4.2 – Représentation de l'architecture VGG-16 sans les dernières couches entièrement connectées 98
Tableau 4.3 – Représentation de la couche entièrement connectée 100
Tableau 4.4 – La dernière étape de l'entraînement du modèle 102

Table des matières

Résumé i
Abstract iii
Remerciements v
Table des figures viii
Liste des tableaux x
Table des matières xi
Introduction générale 2
Contexte 2
Objectifs 2
Organisation du mémoire 3
1. État de l'art 7
1.1 Introduction 7
1.2 Télémédecine 7
1.2.1 Définition 8
1.2.2 Les domaines d'application de la télémédecine 9
1.2.3 Principe d'interopérabilité des applications de télémédecine 10
1.3 Télédétection 11
1.3.1 Définition 11
1.3.2 Le principe de la télédétection spatiale 12
1.3.3 Les domaines d'applications de la télédétection 14
1.3.4 Les images satellitaires 15
1.3.4.1 Définition 15
1.3.4.2 Les différents types de résolution des images satellitaires 16
1.3.5 Logiciels de traitement d'images satellitaires 18
1.4 Data Mining 19
1.4.1 Définition 19
1.4.2 Les domaines d'application du data mining 20
1.4.3 Les différents types de données 21
1.4.4 Les différentes tâches du Data Mining 21
1.4.5 Quelques logiciels du Data Mining 22
1.5 Deep learning 23

1.5.1 Définition 23
1.5.2 Les algorithmes du Deep Learning 26
1.5.3 Les réseaux de neurones convolutifs 27
1.5.3.1 Le but de la partie convolutive des modèles CNN 28
1.5.3.2 Entraînement des réseaux de neurones convolutifs 30
1.5.4 Les logiciels du deep learning 31
1.6 Conclusion 32
2. Modélisation des informations pour la Gestion de crise avec la télémédecine 35
2.1 Introduction 35
2.1.1 Définition d'une crise 35
2.1.2 Les différentes étapes de gestion de crise 35
2.1.3 Les différents actes de télémédecine 36
2.2 Travaux connexes 38
2.2.1 Approches basées sur les systèmes informatiques d'aide à la décision d'urgence 38
2.2.2 Approches orientées vers la gestion de crise en différentes étapes 39
2.2.3 Approches axées sur les modèles mathématiques 41
2.2.3.1 La description fonctionnelle de la gestion de crise 42
2.2.3.2 Les flux d'informations liés aux fonctions de gestion de crise 42
2.2.4 Synthèse des différentes approches 44
2.3 Méthodologie adoptée pour la gestion de crise à distance 45
2.3.1 Processus de gestion de crise à distance 45
2.3.2 Modélisation mathématique 49
2.3.3 Algorithme proposé 51
2.4 Étude de cas 51
2.5 Résultat et discussion 55
2.6 Conclusion 56
3. Les techniques de Data Mining sur les images satellitaires pour découvrir les zones à risque d'épidémie 59
3.1 Introduction 59
3.2 Contexte et travaux connexes 59
3.2.1 Contexte 59
3.2.2 Travaux connexes 61
3.3 Méthodologie de découverte des zones à risque d'épidémie 63

3.3.1 Collecte des données 64
3.3.1.1 Collecte des informations sur le terrain à propos de l'épidémie 64
3.3.1.2 Collecte des images satellitaires 64
3.3.2 Sélection et transformation des données (Données dérivées d'images) 66
3.3.3 Mesures de télédétection 68
3.3.4 Discrétisation des données 68
3.3.5 Traitement des données 69
3.3.6 Interprétation des résultats 71
3.4 Étude de cas : Épidémie de choléra à Mopti (Mali) 71
3.4.1 Collecte des données 71
3.4.1.1 Collecte des informations sur le terrain à propos de l'épidémie de choléra 71
3.4.1.2 Collecte des images satellitaires de la zone d'étude 73
3.4.2 Sélection et transformation des données 74
3.4.3 Mesures de télédétection 74
3.4.4 Discrétisation des données 75
3.4.5 Traitement des images : classification 76
3.4.6 Interprétation des résultats 78
3.5 Discussions 80
3.6 Conclusion 81
4. Classification des images d'agents pathogènes épidémiques à l'aide des réseaux de neurones convolutifs 85
4.1 Introduction 85
4.2 Travaux connexes 86
4.2.1 Définition du Deep learning et travaux connexes 86
4.2.2 Les réseaux de neurones convolutifs (CNN) 87
4.2.3 Les différentes couches constitutives d'un réseau de neurones convolutifs 88
4.2.3.1 La couche de convolution 88
4.2.3.2 Couche ReLU 90
4.2.3.3 Couche de Pooling 90
4.2.3.4 Couche entièrement connectée 90
4.2.4 Les paramètres des différentes couches 91
4.2.5 Apprentissage du réseau de neurones convolutifs 92
4.3 Méthodologie et logiciel utilisés 93

4.3.1 Méthodologie pour la classification des images des agents pathogènes 94
4.3.1.1 Constitution de la base d'apprentissage 95
4.3.1.2 Constitution de la base de validation 96
4.3.1.3 Application de l'apprentissage par transfert : Extraction de caractéristiques ... 96
4.3.1.4 Modèle généré par apprentissage par transfert 100
4.3.1.5 Soumission de la base de test 101
4.3.1.6 Evaluation d'images classées 101
4.3.2 Logiciel utilisé : framework Keras 104
4.4 Discussions 104
4.5 Conclusion 105
Conclusion générale 109
Bilan 109
Contributions 109
Perspectives 111
Annexe A : Algorithme de gestion de crise avec les actes de télémédecine 113
Annexe B : Codes de l'apprentissage par transfert sur les images d'agents pathogènes épidémiques 117
Bibliographie 120
Bibliographie de l'auteur 129

Introduction générale

Contexte.....2
Objectifs.....2
Organisation du mémoire 3

Introduction générale

Contexte

Le monde d'aujourd'hui est confronté à une explosion de données entraînant un fort besoin de modélisation d'une part, et d'extraction de la connaissance à partir de ces grands volumes de données d'autre part. La modélisation des données fait référence à l'analyse, à la formalisation et à la conception de l'ensemble des données nécessaires dans le développement d'un système informatique. L'extraction de connaissances à partir des données (ECD) quant à elle se définit comme « *l'acquisition de connaissances nouvelles, intelligibles et potentiellement utiles à partir de faits cachés au sein de grandes quantités de données* » (Fayyad et al., 1996). En d'autres termes, on cherche surtout à extraire des connaissances intéressantes (règles, régularités, patterns …) via la généralisation d'une information non triviale, implicite, non connue précédemment et potentiellement utile.

Certes beaucoup de travaux ont été menés dans ce domaine, mais rares sont ceux qui ont mis l'accent sur la gestion des crises notamment celles à caractère épidémique à l'aide de la modélisation et de l'extraction des connaissances à partir des données. C'est pourquoi ce travail de recherche doctorale s'inscrit dans la problématique générale de l'extraction des connaissances à partir des données par les techniques informatiques dans le cadre de la gestion de crise. Selon Girard et ses collègues, une crise est définie comme l'apparition d'un état de désordre, d'anomalie, de dysfonctionnement, de déséquilibre profond, ou de trouble grave (Girard et al., 2006). Entraînant ainsi, la nécessité de gérer les crises qui surviennent de manière multiforme dans le monde. La gestion de crise est « *une série de fonctions ou de processus permettant d'identifier, d'étudier et de prévoir les problèmes de crise et de définir des moyens spécifiques permettant à une organisation de prévenir ou de faire face à une crise* » (Kash and Darling, 1998).

Objectifs

L'objectif principal de cette thèse consiste donc à mettre au profit de la gestion de crise notamment épidémique les nouvelles techniques de modélisation et d'extraction des connaissances à partir des données. Ainsi, pour atteindre cet objectif général, nous nous sommes intéressés à trois problématiques, à savoir : l'interopérabilité des systèmes d'information, le Data mining et le Deep learning à différents niveaux de gestion de crise. Ci-dessous, les définitions de ces trois problématiques considérées :

- Interopérabilité des systèmes d'information

Selon le Comité Européen de Normalisation, l'interopérabilité est « *un état entre deux applications lorsque, pour une tâche spécifique, une application peut accepter des données de l'autre pour effectuer cette demande de manière appropriée et satisfaisante sans une intervention de l'opérateur externe* » (Guiddir et al., 2010).

- Le Data mining

Le data mining est une partie intégrante du processus ECD et il se situe à la croisée des statistiques, de l'intelligence artificielle, et des bases de données. Le data mining est défini comme un processus inductif, itératif et interactif de découverte dans les bases de données larges de modèles de données valides, nouveaux, utiles et compréhensibles (Talbi, 2018).

- Le Deep learning

Le Deep learning quant à lui est à la croisée des mathématiques et de l'informatique. Le Deep learning (ou les réseaux de neurones en plusieurs couches) est un ensemble de techniques d'apprentissage automatique utilisant une architecture comprenant une cascade de transformations linéaires et de non-linéarités. En d'autres termes, il permet de faire apprendre à un système une tâche précise en observant un grand nombre d'exemples (LeCun et al., 2015).

Organisation du mémoire

Le présent mémoire de thèse est structuré comme suit : une introduction générale, quatre chapitres et une conclusion générale.

Dans le premier chapitre, nous nous intéresserons à l'état de l'art relativement aux différentes problématiques liées à la modélisation des informations et l'extraction des connaissances dans le cadre de la gestion des crises. En effet dans ce chapitre, nous décrivons en détail l'ensemble des thématiques abordées dans la thèse à savoir la Télémédecine, la Télédétection, le Data Mining et le Deep learning. A cet effet, pour chaque problématique abordée, nous donnons sa définition, son principe de fonctionnement, ses domaines d'application et quelques logiciels utilisés. Ceci dans le but de maîtriser le contour du sujet, car il s'agit des notions qui seront utilisées dans nos différentes contributions.

Les chapitres suivants (2 à 4) regroupent l'ensemble de nos contributions pour cette thèse.

Le deuxième chapitre intitulé « Modélisation des informations pour la Gestion de crise avec la télémédecine » constitue notre première contribution relative à la problématique de la modélisation des informations pour la gestion de crise à l'aide d'une collaboration des applications de télémédecine. Cette contribution a pour objectif d'améliorer la gestion de crises à l'aide des outils informatiques d'aide à la décision. Pour ce faire, nous commençons par présenter les notions complémentaires sur la gestion de crise et les différents actes de télémédecine. Ensuite, nous présentons les travaux existants dans le domaine de la gestion de crise à l'aide des outils d'aide à la décision et nous terminons avec une synthèse des différentes approches. Puis, nous présentons notre méthodologie en trois grandes phases (1- Définition de la crise, 2- Analyse de la crise et 3- Gestion de la crise) et principalement axée sur la collaboration des actes de télémédecine (Téléconsultation, Téléexpertise, Télésurveillance, Téléassistance, et la Régulation médicale). Ensuite, nous avons validé notre méthodologie à travers une étude de cas sur la simulation d'une crise de tsunami à Cannes (France). Enfin, les résultats obtenus sont présentés, suivis d'une discussion.

Le troisième chapitre ayant pour titre « Les techniques de Data Mining sur les images satellitaires pour découvrir les zones à risque d'épidémie » expose notre deuxième contribution relative à la prévention d'une crise d'épidémie par les techniques d'extraction de connaissances à

partir de données (ECD). Cette contribution a pour objectif de prévenir et de surveiller des crises. Nous nous sommes intéressés aux techniques de data mining sur les images satellitaires pour la découverte des zones à risques d'épidémie. A cet effet, premièrement nous avons mené une étude sur les travaux connexes pour dégager notre positionnement. Deuxièmement, nous avons proposé notre méthodologie en six phases (1.A- Acquisition d'informations sur le terrain concernant l'épidémie, 1.B-Acquisition de données satellitaires, 2- Sélection et transformation de données (Données dérivées d'images), 3- Mesures de télédétection, 4- Discrétisation de données, 5- Traitement des données, et enfin 6- Interprétation des résultats). Cette méthodologie met un accent particulier sur les données collectées sur le terrain et les données satellitaires afin de prévenir et surveiller plus efficacement les crises d'épidémie. Troisièmement, la méthodologie a été appliquée à un cas d'étude qui a porté sur l'épidémie de choléra dans la région de Mopti, au Mali. Et enfin, ce chapitre sera clôturé par une discussion et une conclusion.

Le quatrième chapitre intitulé « Classification des images d'agents pathogènes épidémiques à l'aide des réseaux de neurones convolutifs » est dédié à la troisième et dernière contribution de cette thèse qui s'inscrit également dans le cadre des techniques d'ECD. Une fois dans une situation de crise, cette contribution a pour objectif d'améliorer la gestion de la crise. Ce chapitre est consacré à la reconnaissance d'images à l'aide des techniques de Deep learning pour classer automatiquement les images microscopiques des agents pathogènes responsables des épidémies. Pour ce faire, dans un premier temps, il était nécessaire de faire une étude complémentaire et approfondie sur les réseaux de neurones convolutifs. Cette étude nous a permis de nous intéresser aux techniques de l'apprentissage par transfert ou le transfert learning, plus adaptées aux petites bases de données pour la reconnaissance de nos images. Deuxièmement, nous présenterons notre approche de six phases (1- Constitution de la base d'apprentissage, 2- Constitution de la base de validation 3- Application de l'apprentissage par transfert 4- Génération du modèle final 5- Soumission de la base de test au modèle généré et enfin 6- Evaluation du processus de classification) pour classer les images des agents pathogènes épidémiques tout en présentant à chaque étape de l'approche les résultats obtenus. Pour l'étude de cas, il s'agira de distinguer une image microscopique contenant les virus de l'épidémie de choléra appelés Vibrio choléra d'une image microscopique contenant les virus de l'épidémie du paludisme appelés Plasmodium. Et en dernier lieu, nous mènerons une discussion et nous clôturons avec une conclusion.

En enfin, la dernière partie est réservée à la conclusion générale portant à la fois sur le bilan de nos recherches, sur l'ensemble de nos contributions et quelques perspectives pour la suite de ce travail.

Chapitre 1

État de l'art

1.1 Introduction 7
1.2 Télémédecine 7
1.2.1 Définition 8
1.2.2 Les domaines d'application de la télémédecine 9
1.2.3 Principe d'interopérabilité des applications de télémédecine 10
1.3 Télédétection 11
1.3.1 Définition 11
1.3.2 Le principe de la télédétection spatiale 12
1.3.3 Les domaines d'applications de la télédétection 14
1.3.4 Les images satellitaires 15
1.3.4.1 Définition 15
1.3.4.2 Les différents types de résolution des images satellitaires 16
1.3.5 Logiciels de traitement d'images satellitaires 18
1.4 Data Mining 19
1.4.1 Définition 19
1.4.2 Les domaines d'application du data mining 20
1.4.3 Les différents types de données 21
1.4.4 Les différentes tâches du Data Mining 21
1.4.5 Quelques logiciels du Data Mining 22
1.5 Deep learning 23
1.5.1 Définition 23
1.5.2 Les algorithmes du Deep Learning 26
1.5.3 Les réseaux de neurones convolutifs 27
1.5.3.1 Le but de la partie convolutive des modèles CNN 28
1.5.3.2 Entraînement des réseaux de neurones convolutifs 30
1.5.4 Les logiciels du deep learning 31
1.6 Conclusion 32

1. État de l'art

1.1 Introduction

Dans ce premier chapitre, nous allons présenter l'état de l'art des différentes problématiques liées à la modélisation des informations et extraction des connaissances pour la gestion des crises notamment la télémédecine, la télédétection, le data mining et le deep learning (Figure 1.1). La première partie traitera les notions de la télémédecine avec ses différents actes et le principe d'interopérabilité des applications de télémédecine, ceci dans le but de mettre la télémédecine au service de la gestion des crises particulièrement celles qui nécessitent des prises en charge médicale. Dans la deuxième partie, consacrée à la télédétection, nous donnerons une définition générale de la télédétection en expliquant son principe de fonctionnement, pour par la suite répertorier ses domaines d'applications, définir les images satellitaires et présenter quelques logiciels de traitement des images satellitaires. En troisième lieu, une présentation détaillée sera faite sur data mining le processus d'extraction des connaissances à partir des données, avec un focus sur les trois principales tâches du Data mining à savoir la classification, le partitionnement et les règles d'association. Des indications sont fournies sur quelques outils indispensables du data mining. Les techniques du data mining ont été mises à profit pour mettre en évidence les zones à risque d'épidémie dans le troisième chapitre dans le cadre l'extraction de connaissances à partir des données. En dernier lieu, une introduction générale du deep learning sera présentée avec les algorithmes du Deep Learning tout en mettant un accent particulier sur les réseaux de neurones convolutifs qui ont été particulièrement utilisés dans le quatrième chapitre de la thèse. Egalement, ce chapitre s'inscrit dans le cadre l'extraction de connaissances à partir des données (Figure 1.1).

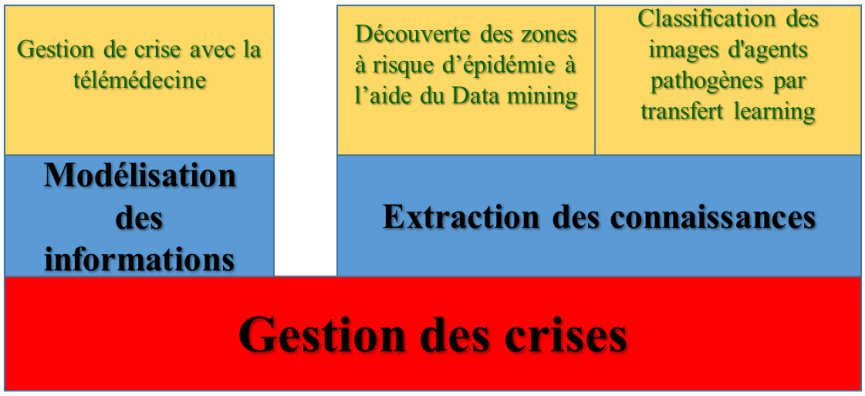


Figure 1.1– Gestion de crises à l'aide de la modélisation des informations et l'ECD

1.2 Télémédecine

Les crises majeures nécessitant des prises en charge médicales, les situations d'épidémies, le vieillissement de la population et la pénurie de médecins spécialistes dans certains territoires constituent des arguments en faveur du suivi de ces patients par les techniques de la télémédecine. La télémédecine peut être une solution à ces problématiques, et apporter un bénéfice tant en matière de qualité et de sécurité des soins, qu'en matière de réduction des dépenses de santé grâce aux déploiements de différents actes à distance. Il s'agit alors de faire voyager les

données médicales issues de plusieurs capteurs (comme le poids, la tension ou la glycémie obtenus d'une balance, d'un tensiomètre ou d'un glucomètre, etc.) plutôt que les patients.

1.2.1 *Définition*

La télémédecine est définie par l'Organisation Mondiale de la Santé (OMS) en 1997 comme « *la partie de la médecine qui utilise la transmission par télécommunication d'informations médicales (images, comptes rendus, enregistrements, etc.), en vue d'obtenir à distance un diagnostic, un avis spécialisé, une surveillance continue d'un malade, une décision thérapeutique* ». Selon le Décret n◦ 2010-1229 du 19 octobre 2010, les principaux actes de la télémédecine sont la téléconsultation, la télé expertise, la télésurveillance, la téléassistance et la réponse médicale (Barge et al., 2012; Simon and Williatte Pellitteri, 2012)(Figure 1.2):

- La **téléconsultation**, qui a pour objet de permettre à un professionnel médical de donner une consultation à distance à un patient. Un professionnel de santé peut être présent auprès du patient et, le cas échéant, assister le professionnel médical au cours de la téléconsultation. L'acte de la téléconsultation nécessite la présence active du patient (le patient est acteur et peut témoigner de la façon dont l'acte a été réalisé).
- La **téléexpertise**, qui a pour objet de permettre à un professionnel médical de solliciter à distance l'avis d'un ou de plusieurs professionnels du domaine médical en raison de leurs formations ou de leurs compétences particulières, sur la base des informations médicales liées à la prise en charge d'un patient.
- La **télésurveillance** médicale, qui a pour objet de permettre à un professionnel médical d'interpréter à distance les données nécessaires au suivi médical d'un patient et, le cas échéant, de prendre des décisions relatives à la prise en charge de ce patient. L'enregistrement et la transmission des données peuvent être automatisés ou réalisés par le patient lui-même ou par un professionnel de santé.
- La **téléassistance** médicale, qui a pour objet de permettre à un professionnel médical d'assister à distance un autre professionnel de santé au cours de la réalisation d'un acte.
- La **réponse médicale** qui est apportée dans le cadre de la régulation médicale mentionnée à l'article L. 6311-2 et au troisième alinéa de l'article L. 6314-1.

Dans tous les cas, les actes de télémédecine sont des actes médicaux sous la responsabilité d'un médecin (professionnel médical).

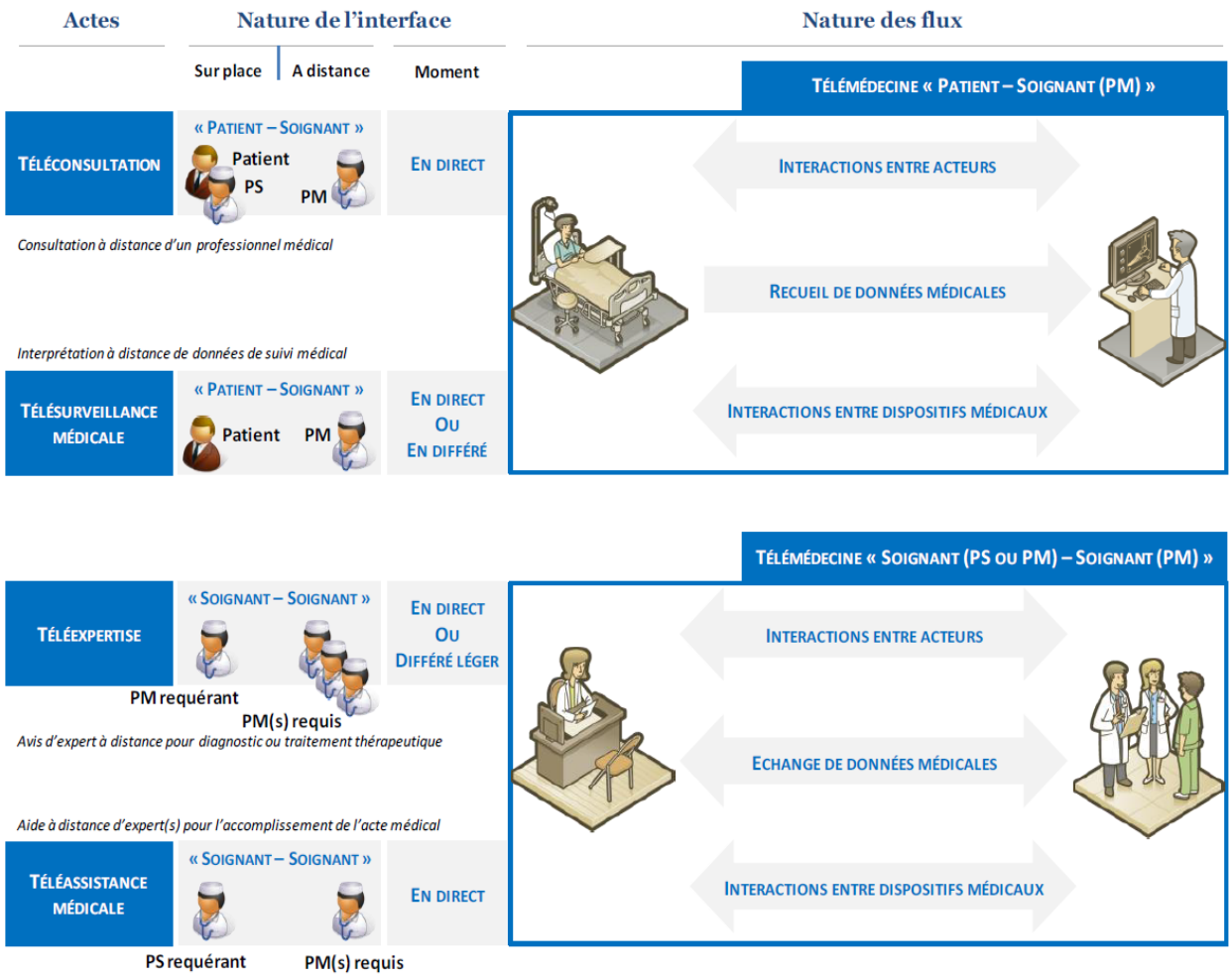


Figure 1.2 – Principaux actes de télémédecine (Barge et al., 2012)

Ces différents actes sont classés en deux sous-groupes : (i) l'un mettant en relation patient et personnel médical (ii) l'autre mettant en relation Personnel Soignant (PS : professions médicales, professions pharmaceutiques, auxiliaires médicaux, aides-soignants, auxiliaires de puériculture et ambulanciers) ou Personnel Médical (PM : médecin, chirurgien-dentiste, sage-femme) et Personnel Médical.

1.2.2 Les domaines d'application de la télémédecine

La télémédecine est en pleine expansion, et offre à ce jour de nombreux produits et services de par le monde. Le nombre de services et de produits de télémédecine parait illimité. En effet, ces applications comprennent de multiples logiciels incluant souvent, une multiplicité de modules pour la gestion d'une variété d'informations. Ces informations se rapportent tant aux patients qu'au personnel soignant et qu'à leurs collaborateurs. Ainsi, ci-dessous, nous listons quelques progrès réalisés à l'aide des technologies de la télémédecine (HERVEG, 2007):

- Informatisation du dossier patient et son accès possible par le patient par voie télématique ;
- Numérisation de l'imagerie médicale pour accélérer leur circulation ;

- Envoi sécurisé des courriers électroniques entre les praticiens et souvent les patients ;
- Apparition des formats électroniques des prescriptions de médicaments ;
- Accès des patients aux multiples sites d'information sur la santé dont certains sont labellisés, garantissant de cette façon une sorte de fiabilité quant à leur contenu informatif, grâce à l'Internet ;
- Disponibilité de nouveaux dispositifs médicaux implantables ou ambulatoires comme les systèmes de télésurveillance ;
- Mises en réseau des bases de données médicales à l'échelle locale, régionale, européenne, internationale et mondiale à des fins très variées, mais principalement aujourd'hui à des fins de prise en charge optimale du patient, de contrôle de la qualité et des dépenses des soins de santé, ou de recherche scientifique ;
- Apparition des vêtements intelligents : ils sont munis de capteurs pour collecter différents paramètres ou ils permettent la délivrance à distance de substances directement au patient ;
- Renforcement du travail coopératif entre les professionnels de la santé à l'aide des applications supportées par des réseaux télématiques qui autorisent aussi les prestations de santé à distance comme la téléconsultation ou la téléchirurgie ;
- Gestion de la prise en charge individuelle et collective du patient de manière informatique notamment pour la prise de rendez-vous et la gestion de l'occupation des lits d'hôpitaux ;
- Développement de l'assistance automatisée aux prestations de soins de santé grâce à des outils d'aide au diagnostic, au traitement et à la médication, transférant la connaissance médicale à la machine et en transformant le praticien de la santé en une sorte d'interface en contact avec le patient ;
- Création des activités de contrôle et de financement des soins de santé par leurs propres systèmes d'information afin de réaliser plus efficacement leurs missions.

1.2.3 Principe d'interopérabilité des applications de télémédecine

L'interopérabilité est « *un état entre deux applications lorsque, pour une tâche spécifique, une application peut accepter des données de l'autre pour effectuer cette demande de manière appropriée et satisfaisante sans une intervention de l'opérateur externe* » (Guiddir et al., 2010). Ainsi, pour favoriser la coopération entre les établissements de santé et les professionnels de santé dans le cadre des parcours de santé des patients, et aussi pour aider la décision médicale, il est nécessaire que les différents composants des applications ou plateformes de télémédecine mises en place puissent communiquer entre elles, par l'intermédiaire de standards et de protocoles de communication internationaux du domaine médical. Kotze et ses collègues définissent quatre niveaux d'interopérabilité pour faire face au problème d'hétérogénéité des applications informatiques de la télémédecine à savoir (Kotze et al., 2013):

- **Niveau technique** : il permet l'échange de données entre systèmes hétérogènes, mais ne garantit pas que le système de réception qui saura utiliser les données échangées de manière significative.
- **Niveau syntaxique** : il assure la conservation de l'objectif clinique durant la transmission des données vers les systèmes de santé.
- **Niveau sémantique** : il permet à plusieurs systèmes d'interpréter de la même manière les informations qui sont en train d'être échangées à travers des concepts prédéfinis partagés.

- **Niveau organisationnel** : il facilite l'intégration des processus métiers et les flux au-delà des limites d'une seule organisation.

Ces différents niveaux d'interopérabilité sont illustrés dans la Figure 1.3

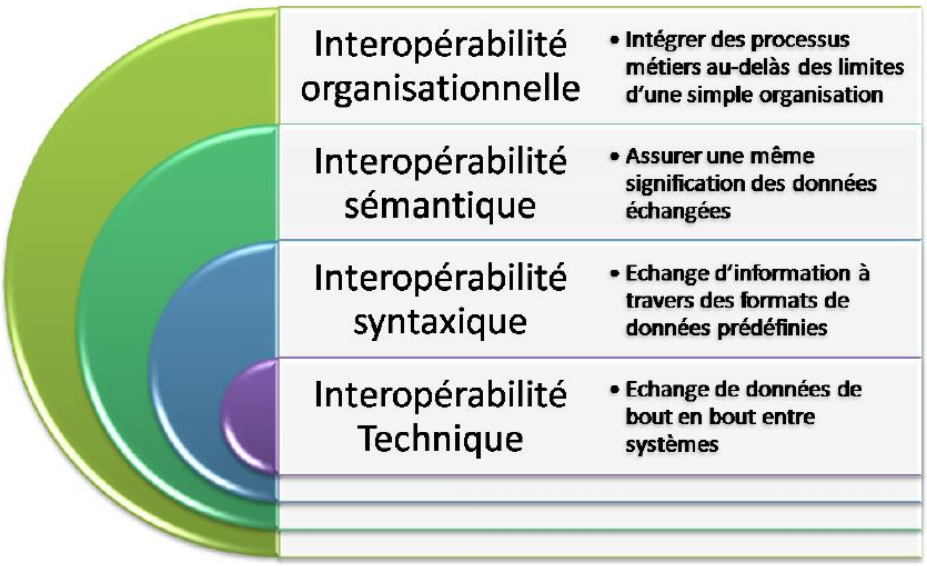


Figure 1.3– Les niveaux d'interopérabilité (Kotze et al., 2013)

Pour échanger, communiquer et se comprendre, les applications de télémédecine doivent être communicantes afin de favoriser la coopération des professionnels de santé dans le cadre des parcours de santé centrés sur le patient, et améliorer la qualité de la décision médicale. Pour faire face à ce problème d'interopérabilité des systèmes d'information de santé, nous avons proposé dans le deuxième chapitre une méthodologie de travail collaboratif pour résoudre d'une situation de crise nécessitant des prises en charge médicale telle que : les tremblements de terre, les inondations, les feux de forêt, la libération de gaz ou de liquide toxique, etc.

1.3 Télédétection

1.3.1 *Définition*

La définition officielle de la télédétection est « *l'ensemble des connaissances et techniques utilisées pour déterminer des caractéristiques physiques et biologiques d'objets par des mesures effectuées à distance, sans contact matériel avec ceux-ci* » (COMITAAS, 1988). Ainsi, la télédétection permet d'élargir la connaissance de notre environnement et faciliter l'interprétation des multiples processus qui affectent notre planète. Ses principales sources sont les mesures et les images obtenues à l'aide de plateformes aériennes et spatiales. Il s'agit par exemple d'un avion, d'un engin spatial, d'un satellite ou encore d'un bateau de n'importe quel type d'instrument permettant l'acquisition d'informations sur l'environnement. Ce qui implique l'usage des instruments tels que les appareils photographiques, les lasers, les radars, les sonars, les sismographes ou les gravimètres.

1.3.2 Le principe de la télédétection spatiale

Le principe de la télédétection spatiale se présente comme suit (*Figure 1.4*).

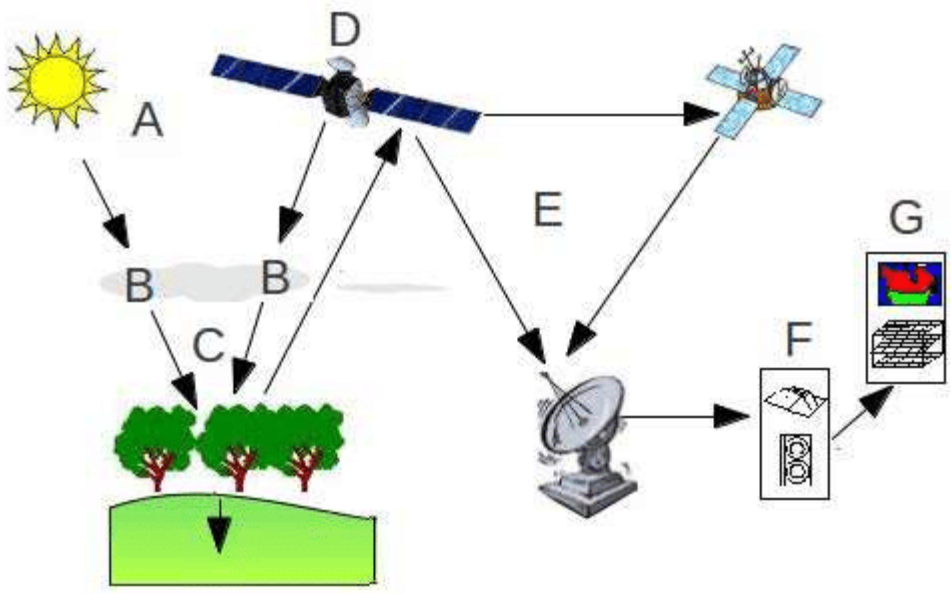


Figure 1.4 – Principe de la télédétection spatiale (El Merabet et al., 2015)

Le rayonnement émis par une source d'énergie ou d'illumination (A) parcourt une certaine distance et interagit avec l'atmosphère (B) avant d'atteindre la cible (C). L'énergie interagit avec la surface de la cible, en fonction des caractéristiques du rayonnement et des propriétés de la surface. Le rayonnement est réfléchi ou diffusé vers le capteur (D), qui l'enregistre et peut ensuite transmettre l'énergie par des moyens électroniques à une station de réception (E) où l'information est transformée en images (numériques ou photographiques). Une interprétation visuelle et/ou numérique de l'image (F) est ensuite nécessaire pour extraire l'information que l'on désire obtenir sur la cible. La dernière étape du processus consiste à utiliser l'information extraite de l'image pour mieux comprendre la cible, pour nous en faire découvrir de nouveaux aspects ou pour aider à résoudre un problème particulier (G).

L'on distingue en général les moyens de télédétection active et passive (González et al., 2013):

- **Télédétection passive**: enregistrement du rayonnement naturel, fourni par la lumière ou la chaleur, qu'il soit émis, réfléchi ou réfracté (exemple : photographies aériennes du paysage éclairé par la lumière du soleil ainsi que certaines images satellitaires comme (SPOT, LANDSAT, IKONOS,…) (*Figure 1.5*).

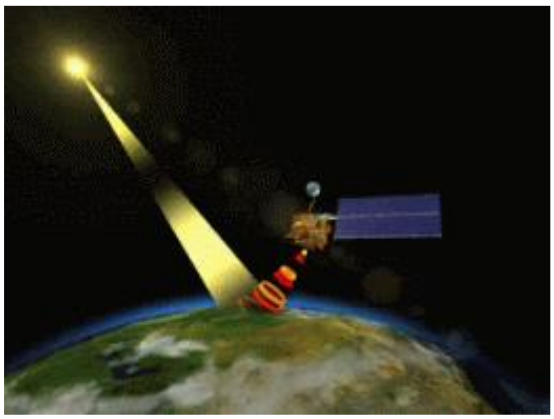


Figure 1.5 – Télédétection passive

- **Télédétection active** : enregistrement du rayonnement que réfléchit l'objet ou le paysage « illuminé » par l'opérateur (exemple : images radar) (*Figure 1.6*).

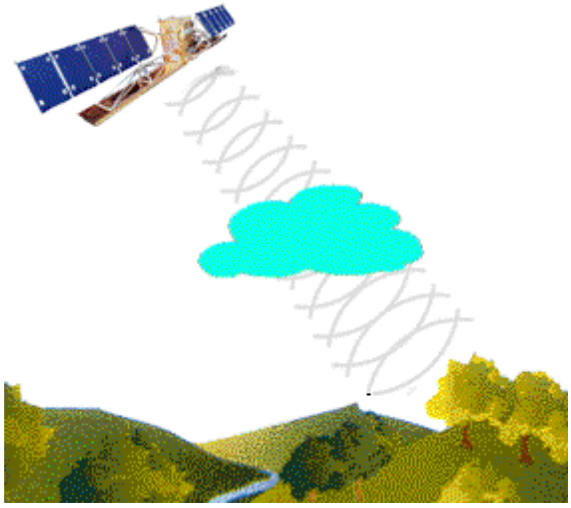


Figure 1.6 – Télédétection active

L'acquisition des données par la télédétection est la mesure des rayonnements électromagnétiques émis ou réfléchis des objets étudiés dans un certain domaine de fréquence (infrarouge, visible, et micro-ondes). Ceci est rendu possible par le fait que les objets étudiés (plantes, maisons, surfaces d'eau ou masses d'air) émettent ou réfléchissent du rayonnement à différentes longueurs d'onde et intensités selon leur état. Certains instruments de télédétection utilisent des ondes sonores de façon similaire, et d'autres mesurent des variations dans des champs magnétiques ou gravitaires.

Réflectance solaire

La réflectance solaire se définit comme le rapport entre l'énergie solaire réfléchie et l'énergie solaire incidente sur une surface (Energie Solaire Réfléchie / Energie Solaire Incidente) (*Figure 1.7*). Par exemple, une réflectance de 100% signifie que la surface en question réfléchit toute l'énergie solaire dans l'atmosphère et n'en absorbe aucune fraction.

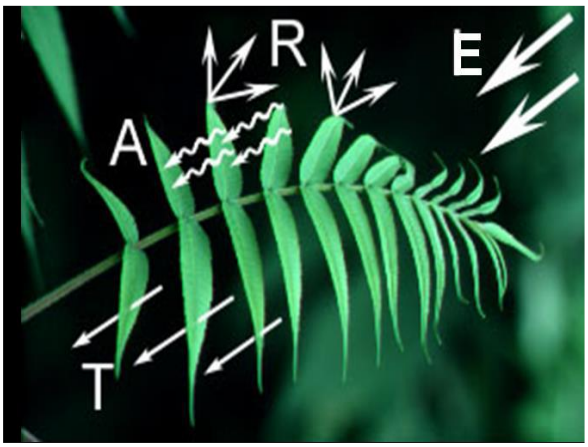


Figure 1.7 – Principe de la réflectance solaire

Le rayonnement solaire atteint les objets (E). Les objets peuvent en observer une partie (A), en transmettre une partie (T), et en réfléchir une autre (R). Ainsi :

- ✓ La réflectance est : $\rho = \frac{R}{E}$
- ✓ L'absorptance est : $\alpha = \frac{A}{E}$
- ✓ La transmittance est : $\tau = \frac{T}{E}$

$$\rho + \alpha + \tau = 1$$

1.3.3 Les domaines d'applications de la télédétection

Les domaines d'applications de la télédétection spatiale sont très nombreux, on peut citer entre autres :

- ➢ l'évaluation et la surveillance de l'environnement (croissance urbaine, et déchets dangereux) ;
- ➢ la détection et la surveillance des changements globaux (appauvrissement de l'ozone atmosphérique, déforestation, et réchauffement climatique) ;
- ➢ l'exploration des ressources non renouvelables (minéraux, pétrole, et gaz naturel) et les ressources naturelles renouvelables (océans, forêts et terres) ;
- ➢ la météorologie (prévision météorologique et les processus dynamiques atmosphériques) ;
- ➢ les systèmes d'informations géographiques, la cartographie (topographie, utilisation des terres, génie civil,), etc. ;
- ➢ les modèles numériques de terrain peuvent être produits par interférométrie, méthode consistant à enregistrer une série de mesures de la cible à partir d'un avion, d'un satellite ou d'une navette spatiale. La combinaison des données issues de ces mesures offre une carte détaillée contenant de l'information sur la couverture du sol (*Figure 1.8*), le relief ou encore mouvement à une échelle centimétrique. Les données couvrent généralement des bandes de plusieurs kilomètres de largeur.

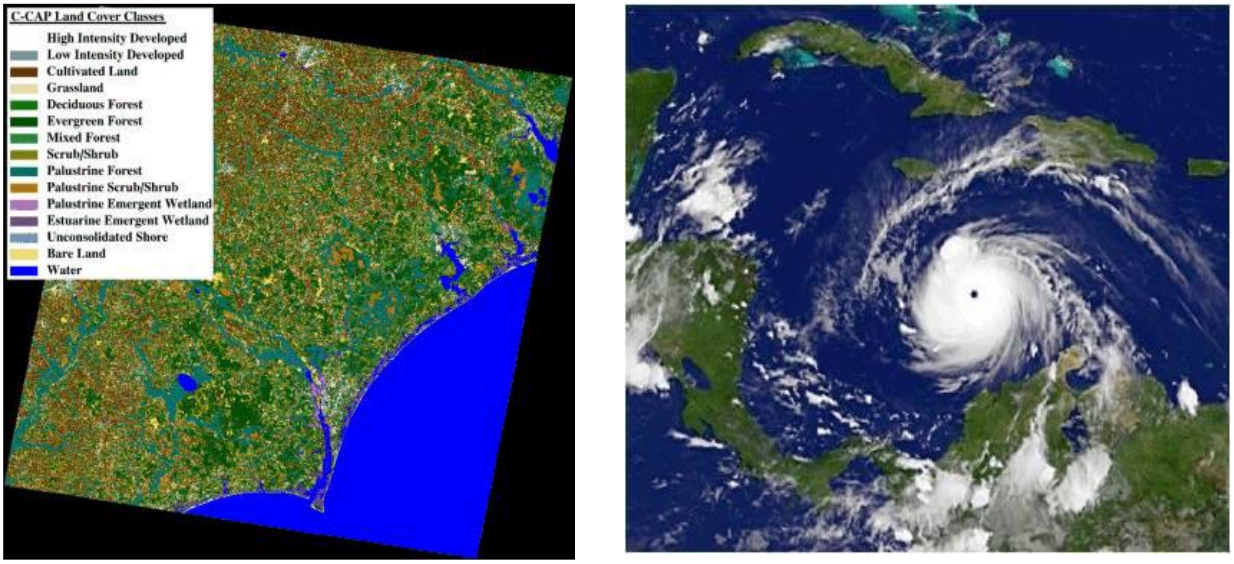


Figure 1.8 – Quelques exemples d'application de la télédétection spatiale : carte d'occupation du sol, suivi des conditions atmosphériques

Dans le cadre de nos travaux, nous nous intéressons particulièrement aux satellites comme capteurs de télédétection et donc aux images satellitaires comme données de télédétections. Les images satellitaires feront l'objet d'une présentation ci-dessous.

1.3.4 *Les images satellitaires*

1.3.4.1 *Définition*

Depuis l'espace, nous pouvons observer la terre et utiliser ces observations pour prévoir le temps, étudier l'évolution des climats, mesurer si le trou dans la couche d'ozone s'agrandit, détecter et suivre des feux de forêt ou des nappes de pétrole, créer des cartes... Actuellement, les images satellitaires sont partout : dans les informations météo à la TV, dans les journaux et même parfois dans les publicités. Une image satellitaire est comme une photo d'une partie de la Terre, prise de l'espace. Cette image est prise par un satellite placé en orbite autour de la planète. A la différence des images obtenues d'un appareil photo, ou dessinées sur du papier, qui voit la terre en effet de la même manière que les yeux humains, sensible à la portion de lumière qu'on appelle la lumière visible, un satellite observe au-delà de la lumière visible, s'étend à d'autres parties du spectre électromagnétique comme l'infrarouge, l'ultraviolet ou les micro-ondes (*Figure 1.9*).

Le spectre électromagnétique est la description de l'ensemble des rayonnements électromagnétiques classés par fréquence, longueur d'onde ou énergie. Il s'agit de la propagation du rayonnement électromagnétique de l'énergie dans l'espace par exemple la chaleur d'un feu, la lumière du soleil, les rayons X utilisés en médecine, les micro-ondes et les ondes radio (Khan Academy, 2015). Ces instruments (scanners) balaient la surface de la Terre et enregistrent des mesures de lumière qui sont ensuite utilisées par des programmes informatiques pour créer des images.

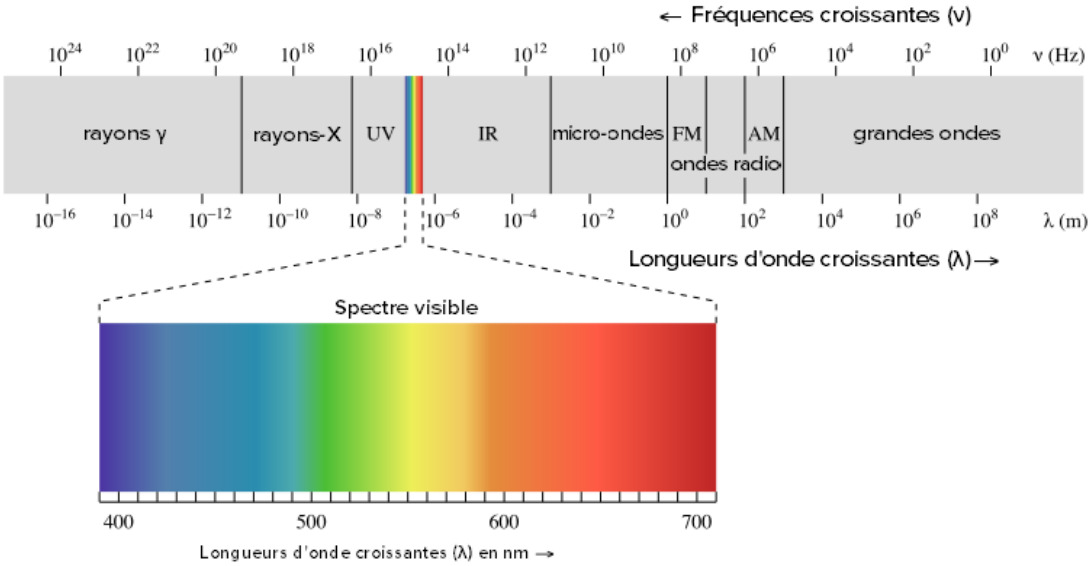


Figure 1.9 – Le spectre électromagnétique (Khan Academy, 2015)

À bord des satellites, le capteur enregistre les images sous forme de bandes spectrales. Chacune des bandes spectrales donne une image en noir et blanc, où les tons de gris varient. En superposant trois bandes spectrales avec des filtres de couleur, il est possible d'obtenir une image en couleurs ou un composé coloré. Chaque composé coloré donne des renseignements différents sur l'occupation du territoire. L'image numérique est fabriquée par un assemble de pixels, représentés par une série de bits, ayant chacun comme caractéristique un niveau de couleur ou de gris prélevé à l'emplacement correspondant dans l'image réelle, cet emplacement étant défini par ses coordonnées cartésiennes dans un plan d'ordonné y, d'abscisse x et de centre O.
Grâce à leur altitude, les satellites fournissent des images de grande taille. Dans le jargon, l'image prise est appelée « scène ». Effectivement, une image satellite représente en moyenne une « scène » de 60km de côté, mais cela peut varier selon le satellite. La résolution des images joue un rôle important : la précision d'un pixel peut aller de 20 à 30 mètres en couleur et jusqu'à 1 ou 2 mètres en noir et blanc. Ceci dépend des capacités et de l'altitude du satellite. Les résolutions les plus hautes sont malheureusement encore réservées aux services militaires. Grâce à la technologie qui s'améliore de jour en jour, les satellites les plus récents fournissent des images comparables à des photos aériennes. La qualité des images satellitaires est étroitement liée à la plateforme d'obtention des images en question. Ainsi, les facteurs pour la sélection d'une plateforme se résument comme suit (González et al., 2013):

- l'étendue de la zone à couvrir,
- la vitesse de développement des phénomènes observés,
- la disponibilité et le coût des données,
- les caractéristiques fonctionnelles des instruments et des capteurs (résolutions).

1.3.4.2 Les différents types de résolution des images satellitaires

González et ses collègues ont répertorié quatre types de résolutions: la résolution spatiale, la résolution spectrale, la résolution temporelle et la résolution radiométrique (González et al., 2013).

La résolution spatiale apporte une indication sur la taille du pixel de l'image de télédétection. Les trois niveaux de résolution spatiale sont : haute résolution spatiale : 0.5 - 4 m, moyenne résolution spatiale 4 - 30 m, faible résolution spatiale 30 - 1000 m (*Figure 1.10*).

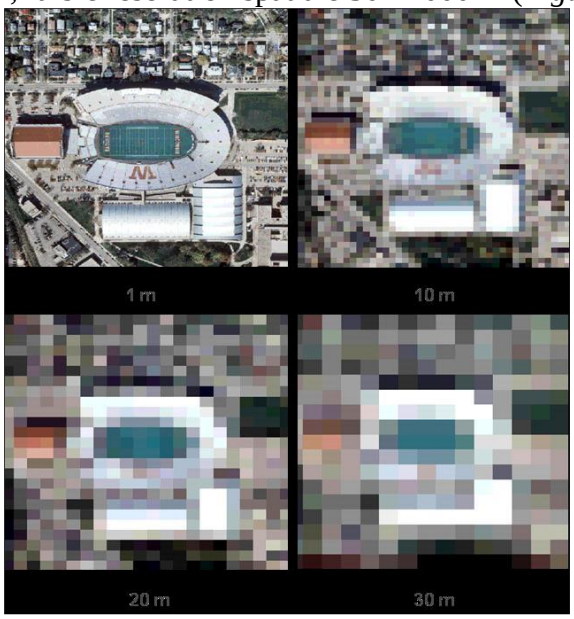


Figure 1.10 – les différents niveaux de résolution d'une image satellitaire

La résolution spectrale est l'aptitude du capteur à distinguer des rayonnements électromagnétiques de fréquences différentes. Plus le capteur est sensible à des différences spectrales fines (intervalles de longueur d'onde étroits), plus la résolution spectrale du capteur est élevée. Elle indique le nombre de bandes spectrales dans lesquelles le capteur peut capturer les radiations ainsi que la position des bandes dans le spectre électromagnétique.

La résolution radiométrique est la sensibilité d'un détecteur aux variations de l'intensité de l'énergie électromagnétique émise, réfléchie ou diffusée qui est détectée (*Figure 1.11*).

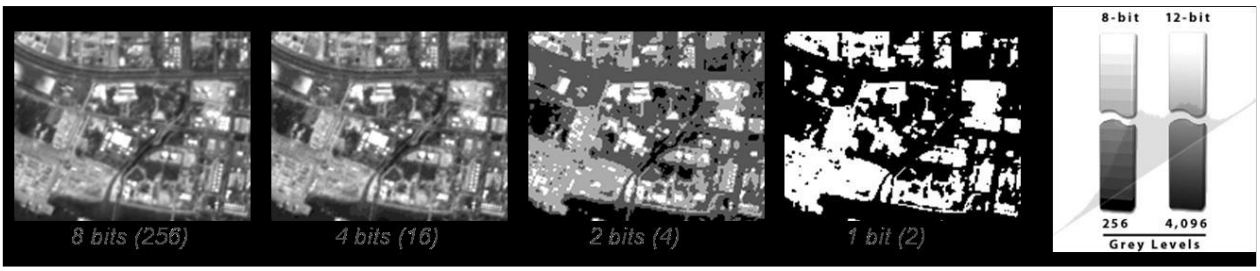


Figure 1.11 – La résolution radiométrique

La résolution temporelle précise la fréquence de révision du satellite pour un emplacement spécifique. Elle est plus importante lorsque : les nuages persistants offrent une vue directe limitée de la surface de la Terre, les phénomènes à court terme (inondations, marées noires, etc.), les phénomènes qui changent rapidement (atmosphérique). Les trois niveaux de résolution temporelle sont : haute résolution temporelle soit de 1 à 3 jours, moyenne résolution temporelle allant de 4 à 16 jours, faible résolution temporelle supérieure à 16 jours.

1.3.5 *Logiciels de traitement d'images satellitaires*

Les informations recueillies par les capteurs de télédétection (satellites) sont traitées et analysées à l'aide des logiciels connus comme des outils de traitement de télédétection. Il y a un certain nombre de logiciels libres et beaucoup d'autres de paiement à traiter ce genre d'information. Ensuite on présente les outils de chaque type les plus utilisés pour le traitement d'images de télédétection. En outre, pour chacun d'eux, on inclut une information résumée de ses principales caractéristiques.

Outils commerciaux

ENVI : Programme axé sur le traitement et l'analyse des images de toutes sortes. Il est assez flexible avec les formats de lecture et semble très amiable à l'utilisateur. Il utilise le langage IDL et il permet d'intégrer des modules supplémentaires.

ERDAS : Probablement le plus grand programme d'implantation commerciale, en étant un standard pour les formats d'échange, notamment avec les SIG. Il est disponible pour une grande variété de plateformes. Il peut être programmé à l'aide du Model Maker.

ESRI : ArcGIS est un ensemble de produits SIG. Ils sont regroupés en plusieurs applications pour la capture, l'édition, l'analyse, le traitement, la conception, l'impression et la publication de l'information géographique. ArcGIS Desktop, gamme d'applications SIG, est l'un des plus utilisés, compris ArcReader, ArcMap, ArcCatalog, ArcToolbox, ArcScene et ArcGlobe, en plus de diverses extensions.

Idrisi : Outil qui intègre de nombreux utilitaires d'analyse numérique d'image, et pour le bas prix, a été largement utilisé dans l'enseignement de la télédétection. Il a été développé par l'Université Clark (USA).

E-Cognition : Programme qui intègre de multiples aspects de l'analyse du contexte et de la morphologie mathématique. Il est idéal pour la reconnaissance des modèles spatiaux.

ER-Mapper : Programme australien qui reprend la plupart des caractéristiques communes en matière de télédétection, en étant très efficace dans la gestion des images volumineuses en intégrant des algorithmes virtuels et des systèmes innovants de compression.

PCI–Geomatica : Nouvelle version du programme canadien PCI, qui incorpore dans un schéma plus intégré ses paquets d'analyse numérique d'images, l'ortho rectification, la production des cartes et SIG. Il fournit des outils d'analyse très puissants grâce à sa connexion avec le Centre Canadien de Télédétection. On peut utiliser une version de démonstration limitée, appelée Freeview.

DRAGON : Programme à faible coût en environnement Windows pour l'analyse numérique d'images. Il est très approprié pour un cadre éducatif.

Miramon : Programme de SIG et télédétection développé par CREAF de l'Université de Barcelone. Faible coût et des nombreuses fonctionnalités.

Outils gratuits

Grass : Programme axé sur SIG et traitement d'image. Développé pour UNIX, sur différentes plateformes, mais il existe des versions pour Linux et Windows (winGRASS). Il a été initialement développé comme un outil militaire de l'armée américaine. Parmi les organisations qui l'utilisent on souligne NASA, NOAA, USGS, etc.

Orfeo-Monteverdi : Orfeo Toolbox (OTB) est une bibliothèque de traitement d'images de télédétection développé par le CNES en 2006 pour faciliter l'utilisation des images de différents capteurs. En particulier la constellation Orpheus (Pléiades et Cosmo-SkyMed). Monteverdi est la plus robuste et opérationnelle.

SeaDAS : Logiciel de la NASA pour la visualisation, le traitement et l'analyse des images de capteurs de Couleur de l'Océan. Il prend en charge les principales missions de la NASA et de l'ESA récemment.
Multispec : Programme développé par l'Université de Purdue, un pionnier dans les années 70. Il est très approprié pour l'environnement d'apprentissage, parce qu'il a peu de requis matériels. Il y a également une nouvelle version de 32 bits (32-bit version).
SPRING : Programme développé par l'INPE (Institut national de recherche spatiale du Brésil). Il a une longue liste de fonctionnalités, mais est essentiellement orientée vers l'environnement SIG. Des versions existent en portugais, anglais et espagnol.

1.4 Data Mining

1.4.1 *Définition*

Le Data mining a vu le jour suite à l'explosion des données souvent provenant de supports hétérogènes dont l'analyse et l'exploitation par les méthodes d'analyse classiques s'avèrent peu efficaces. Il se situe à la croisée des statistiques, de l'intelligence artificielle et des bases de données. Le data mining est la recherche et l'extraction de l'information utile et inconnue dans de gros volumes de données, stockées dans des bases de données ou des entrepôts de données (Han et al., 2011). Plus formellement, le data mining est défini comme un processus inductif, itératif et interactif de découverte dans les bases de données larges de modèles de données valides, nouveaux, utiles et compréhensibles (Talbi, 2018). En d'autres termes, ce processus concerne l'extraction des connaissances intéressantes (règles, régularités, patterns …) via la génération d'une information non triviale, implicite, non connue précédemment et potentiellement utile.

Les motivations sont multiples notamment l'amélioration de la productivité des entreprises, le renforcement des systèmes de prise de décisions par l'exploitation des données historiques pour prédire l'avenir et anticiper le marché, la disponibilité des supports de stockage et d'exécution des données souvent en temps réel. Le data mining est une partie intégrante du processus ECD (extraction de connaissances à partir de données) (*Figure 1.12*).

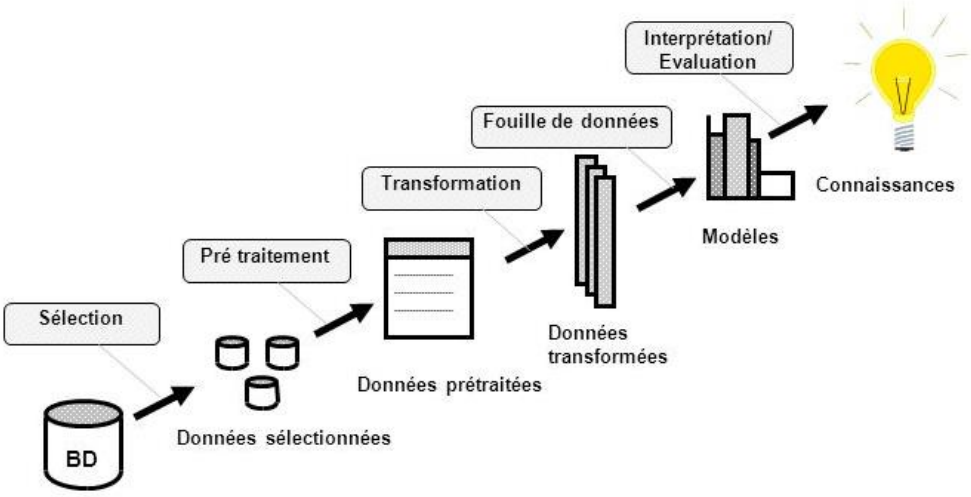


Figure 1.12 – Processus d'extraction des connaissances à partir des données (Fayyad et al., 1996)

L'extraction de connaissances se déroule en sept étapes :

1. Nettoyage des données (pour supprimer le bruit et les données incohérentes)
2. Intégration des données (où plusieurs sources de données peuvent être combinées)
3. Sélection des données (lorsque les données pertinentes pour la tâche d'analyse sont extraites de la base de données)
4. Transformation de données (lorsque les données sont transformées ou consolidées dans des formulaires appropriés pour l'extraction d'informations en effectuant des opérations de résumé ou d'agrégation, par exemple)
5. Data mining (un processus essentiel où des méthodes intelligentes sont appliquées pour extraire les modèles de données afin de décrire le comportement actuel des données et ou prédire le comportement futur des données)
6. Évaluation de modèle (pour identifier les modèles vraiment intéressants représentant la connaissance basée sur des mesures d'intérêt)
7. Présentation des connaissances (où les techniques de visualisation et de représentation des connaissances sont utilisées pour présenter les connaissances générées à l'utilisateur)

Le processus de l'ECD peut se résumer généralement en trois grandes phases à savoir : une phase de prétraitement qui se focalise sur la préparation des données, la phase d'application des techniques de Data mining (par exemple : classification, clustering, association …) qui consiste à extraire de la connaissance et enfin la dernière phase le post-traitement pour l'interprétation des résultats.

1.4.2 Les domaines d'application du data mining

Dans un monde concurrentiel, les entreprises sont soucieuses de fournir de meilleurs services, s'adapter aux clients, pratiquement tous les domaines sont concernés par le data mining :

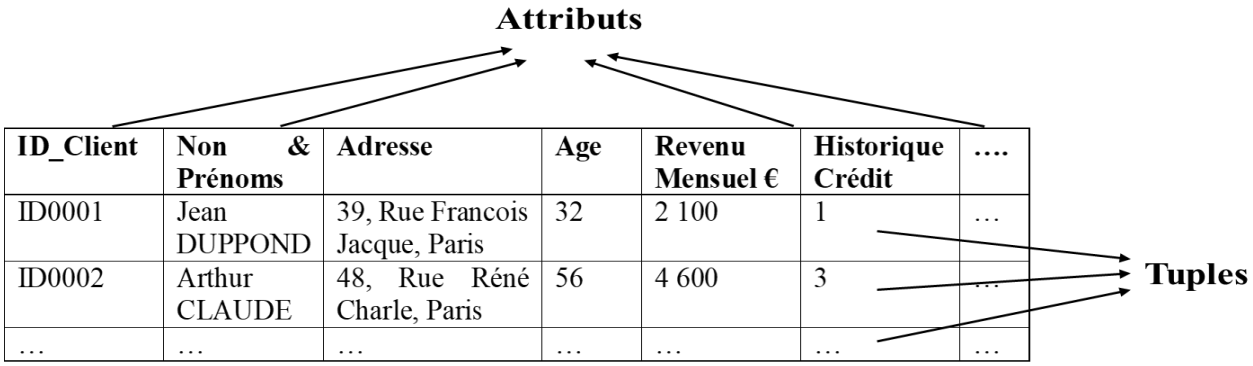
- La gestion de stocks : moment de la commande des produits, détermination de la quantité requise, etc. ;
- La gestion et analyse de risque : Assurances, Banques (crédit accordé ou non en fonction du profil du client, de sa demande et des expériences passées de prêts) ;
- La détection de fraudes : par exemple dans le domaine des télécommunications ;
- L'organisation des rayonnages dans les supermarchés en regroupant les produits qui sont généralement achetés ensemble (pour que les clients n'oublient pas d'acheter un produit parce qu'il est situé à l'autre bout du magasin). Par exemple, on extraira une règle du genre : les clients qui achètent le produit X en fin de semaine, pendant l'été, achètent généralement le produit Y ;
- Le marketing direct : ciblage des offres selon l'âge, le sexe, la profession, l'adresse… ;
- Le diagnostic médical : les patients ayant certains symptômes et demeurant dans des agglomérations de plus de 1000 habitants développent couramment telle pathologie ;
- L'analyse du génome et bio-informatique plus généralement ;
- La classification d'objets (astronomie ...) ;
- Les moteurs de recherche sur internet : fouille du web ;
- L'extraction d'information à partir des textes : fouille de textes ;
- L'évolution des données dans le temps : fouille de séquences, etc.

1.4.3 *Les différents types de données*

Les données se situent au cœur du processus d'extraction des connaissances. Elles sont collectées et stockées de façon exponentielle à longueur de journée non seulement par les géants du web tels que Google, Facebook, Tweater, gmail, yahoo …, mais aussi par les activités de commerce électronique, les transactions bancaires, les capteurs tels que les satellites, drones, les systèmes à temps réel, etc. En principe, les techniques de data mining peuvent être appliquées à tous types de données en l'occurrence les bases de données relationnelles, transactionnelles, multimédias, spatiales, textes, web, etc (Han et al., 2011). Les techniques d'extraction peuvent différer d'un type de base de données à un autre. Ainsi, il est nécessaire de faire un prétraitement des données afin de les adapter aux techniques de data mining souhaitées.

Dans ce qui suit, nous allons plus détailler la base de données relationnelle, car elle fait partie des bases de données plus utilisées et les plus disponibles. Une base de données relationnelle est une collection de tables, chacune ayant un nom unique. Chaque table se compose d'un ensemble d'attributs (colonnes ou champs) et stocke généralement un grand nombre de tuples (enregistrements ou *instance*) (Tableau 1.1). Chaque tuple dans une table relationnelle représente un objet identifié par une clé unique et est décrit par un ensemble de valeurs des attributs de cette table.

Tableau 1.1 – Exemple de table d'une base de données



ID_Client	Non & Prénoms	Adresse	Age	Revenu Mensuel €	Historique Crédit	….
ID0001	Jean DUPPOND	39, Rue Francois Jacque, Paris	32	2 100	1	…
ID0002	Arthur CLAUDE	48, Rue Réné Charle, Paris	56	4 600	3	…
…	…	…	…	…	…	…

Un attribut peut être de nature qualitative ou quantitative en fonction de l'ensemble des valeurs qu'il peut prendre. Les valeurs des attributs d'une base de données représentent une certaine mesure d'une quantité dans le monde. Ainsi, quand on fait de l'extraction de la connaissance particulièrement avec le data mining, on effectue de nombreuses opérations arithmétiques qui selon la valeur de l'attribut ne sont pas souvent valables. Dans ce cas, il faut un prétraitement des données par exemple la discrétisation afin d'adapter les données aux techniques de data mining envisagées. Ces notions seront plus détaillées dans le troisième chapitre.

1.4.4 Les différentes tâches du Data Mining

Le data mining est l'ensemble des techniques et méthodes destinées à l'exploration et l'analyse de grandes bases de données en vue de détecter dans ces données des règles, des associations, des tendances inconnues (non fixées a priori), des structures particulières restituant de façon concise

l'essentiel de l'information utile pour l'aide à la décision. A cet égard, le data mining est constitué de plusieurs tâches notamment la **classification** qui consiste à prédire si une nouvelle instance de donnée est membre d'un groupe ou d'une classe prédéfinie. Une autre tâche consiste, dans une base de données, à repérer des groupes d'instances qui se ressemblent ; en ce moment il s'agit d'un problème de **partitionnement** (ou **clustering**). Toutefois, on peut également disposer d'un ensemble de données pour lequel on veut détecter des relations entre la valeur de leurs attributs. Il s'agit alors de chercher des **règles d'association.** On peut aussi disposer d'un ensemble de données qui décrivent l'évolution au cours du temps d'une certaine entité. On veut alors prédire le comportement futur à partir du comportement observé : c'est un problème de **prédiction**. Il y a des cas où on doit déterminer les valeurs des attributs dont les valeurs sont manquantes. Ce qui est un problème **d'estimation** ; pour plus de détails sur les concepts du data mining voir (Fayyad et al., 1996; Han et al., 2011; Hand, 2007; Tan et al., 2006).

1.4.5 *Quelques logiciels du Data Mining*

Le choix de l'outil de data mining dépend de plusieurs critères notamment : le type de données à exploiter, les sources de données, les méthodes de data mining appropriées, la visualisation (des données, des résultats, etc), interactivité avec l'utilisateur et aussi l'extensibilité des données, etc. Ci-dessous, nous allons citer quelques outils de data mining.

Logiciels libres et/ou gratuits

Tanagra : une grande partie de la panoplie des méthodes de DATA MINING intégrées dans une seule structure. Le mode d'utilisation du logiciel est au standard logiciel du domaine, avec notamment la définition des opérations à réaliser sur les données à l'aide d'une représentation visuelle. La section didacticiels de ce site mérite une attention particulière.

Sipina : du même auteur que TANAGRA, il se distingue surtout par sa large palette de méthodes d'induction par arbres de décision, avec la possibilité d'interagir directement avec les modèles construits. Distribué depuis une dizaine d'années, ce logiciel est très connu dans le monde de la recherche.

R-project : Un logiciel gratuit que l'on associe souvent aux statisticiens, mais qui en réalité convient très bien pour la fouille de données. La bibliothèque des fonctions est impressionnante et s'enrichit chaque jour grâce au système des packages. Toutefois, il fonctionne à l'aide d'un interpréteur de commandes qui peut restreindre son accessibilité. Il faut un peu de pratique pour en tirer véritablement parti.

Weka : est un logiciel de data mining complet du groupe apprentissage automatique de l'université de Wakato, Nouvelle-Zélande. Les algorithmes peuvent être appliqués directement à un ensemble de données ou appelés à partir de votre propre code Java. Weka contient des outils pour le prétraitement des données, la classification, la régression, le partitionnement, les règles d'association et la visualisation.

Spmf : est une bibliothèque libre de data mining écrite en Java, plus spécialisée dans les règles d'association, l'extraction de règles séquentielles, la prédiction de séquence, etc. Il implémentation 150 algorithmes de data mining.

Logiciels commerciaux

SPAD de COHERIS : un logiciel français du DATA MINING, il comporte plusieurs méthodes de data mining.

IBM SPSS Modeler : c'est également un bon logiciel qui est largement utilisé.

STATISTICA Data Miner : un autre pro des statistiques qui se lance dans le DATA MINING à travers une représentation graphique des traitements et de leur enchaînement. Les méthodes sont très complètes avec une documentation particulièrement riche (cf. Electronic Statistics Textbook).

Entreprise Miner de SAS : outil complet de data mining et orienté statistique.

1.5 Deep learning

Un autre domaine de l'Intelligence artificielle en pleine croissance est le Deep learning (apprentissage profond) qui est à la croisée des mathématiques et de l'informatique. Ce vaste domaine est relativement nouveau et s'incorpore dans le Machine Learning (apprentissage automatique). Le but du Machine Learning est d'apprendre à un système à effectuer des tâches de façon autonome à partir des données.

1.5.1 *Définition*

Le Deep learning quant à lui est un ensemble de techniques d'apprentissage dans les réseaux de neurones. Le Deep learning ou les réseaux de neurones en plusieurs couches sont actuellement la meilleure solution à de nombreux problèmes de reconnaissance d'image, de texte, de voix, de visage ; la surveillance routière, des piétons, le développement des systèmes autonomes (voiture, robot…). En gros, le deep learning permet de faire apprendre à un système une tâche précise en observant un grand nombre d'exemples (LeCun et al., 2015).

Un réseau neuronal est l'association, en un graphe plus ou moins complexe, d'objets élémentaires : des neurones formels. Les réseaux se distinguent par leur architecture, c'est-à-dire le nombre de neurones et le nombre de couches.

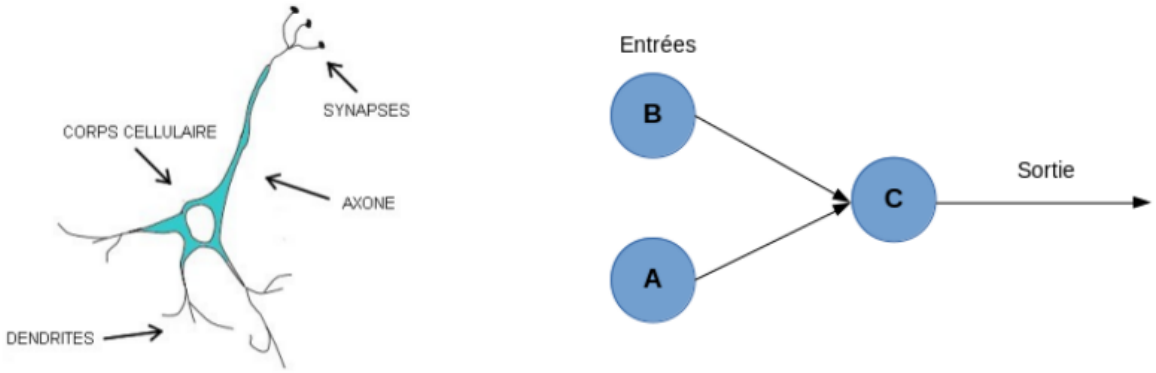


Figure 1.13 – Neurone biologique à gauche et neurone formel à droite

Un neurone biologique est une cellule caractérisée par (*Figure 1.13*) :

- des synapses, connexions avec les autres neurones ;
- des dendrites, entrées du neurone;
- les axones, les sorties;

le noyau qui active les sorties en fonction des stimulations en entrée.
Le neurone formel est inspiré du neurone biologique qui est une représentation mathématique et informatique d'un neurone biologique. Le neurone formel possède plusieurs entrées et une sortie qui correspondent respectivement aux dendrites et à l'axone. Un neurone reçoit des signaux électriques par le biais des dendrites et lorsque le neurone reçoit suffisamment de signaux en un temps donné alors il déclenche ses propres signaux. Le fonctionnement d'un neurone est relativement simple, mais lorsqu'un neurone est connecté à des milliers d'autres et qu'il y a des milliards de neurones, cela crée des réseaux capables de résoudre des problèmes extrêmement complexes.

Le neurone active sa sortie (sortie active = 1) selon si ses entrées dépassent un certain seuil. À partir de ce neurone, il est possible de construire n'importe quel réseau de neurones artificiels capable de résoudre des opérations logiques.

En ce qui concerne l'historique du deep learning (Dejasmin, 2018; Gagné, 2016), le réseau de neurones le plus simple est le perceptron et il a été inventé en 1957 par F. Rosenblatt. Le perceptron est composé de neurones qui fonctionnent différemment du neurone formel. Ces neurones sont appelés des unités linéaires à seuil (LTU, Linear Treshold Unit). Les LTU ont en entrées des nombres quelconques (contrairement aux neurones formels qui ont des entrées binaires) et chaque entrée est pondérée par un poids (w) (*Figure 1.14*) :

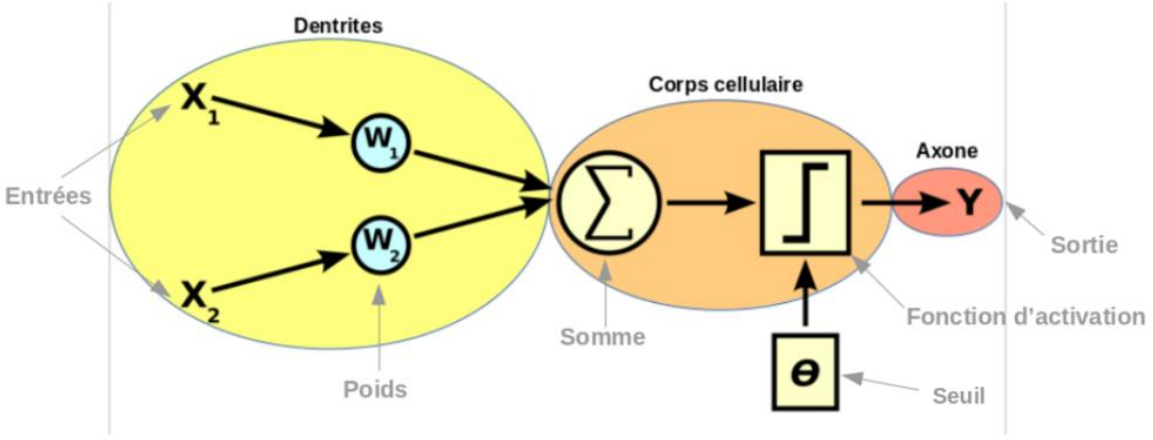


Figure 1.14 – Le perceptron 1957- F. Rosenblatt (Dejasmin, 2018)

Le LTU sur le schéma ci-dessus fonctionne de la façon suivante : X1 et X2 sont deux entrées (cela correspond aux signaux qu'un neurone biologique reçoit par le biais des synapses d'un autre neurone). W1 et W2 sont des poids qui vont respectivement pondérer X1 et X2 (poids synaptiques). Ensuite, le symbole Σ (grand sigma) indique une somme, on va donc faire la somme des deux entrées X1 et X2 pondérées par W1 et W2. Le résultat de cette somme est la valeur d'entrée de la fonction d'activation (dernier cadre). Cette fonction détermine la sortie Y. Le symbole σ (petit sigma) est un seuil. Lorsque la valeur en entrée de la fonction d'activation est supérieure à ce seuil

le neurone est actif (la sortie est égale à 1), lorsque cette valeur est inférieure à ce seuil il est non-actif (la sortie vaut alors 0 ou -1). Quand la valeur en entrée de la fonction d'activation est aux alentours du seuil, on est dans ce que l'on appelle la phase de transition.

Par exemple avec une fonction d'activation échelon, la phase de transition est en théorie inexistante, la sortie vaut 1 si l'entrée de la fonction échelon vaut 0, autrement la sortie vaut 0 (*Figure 1.15*) :

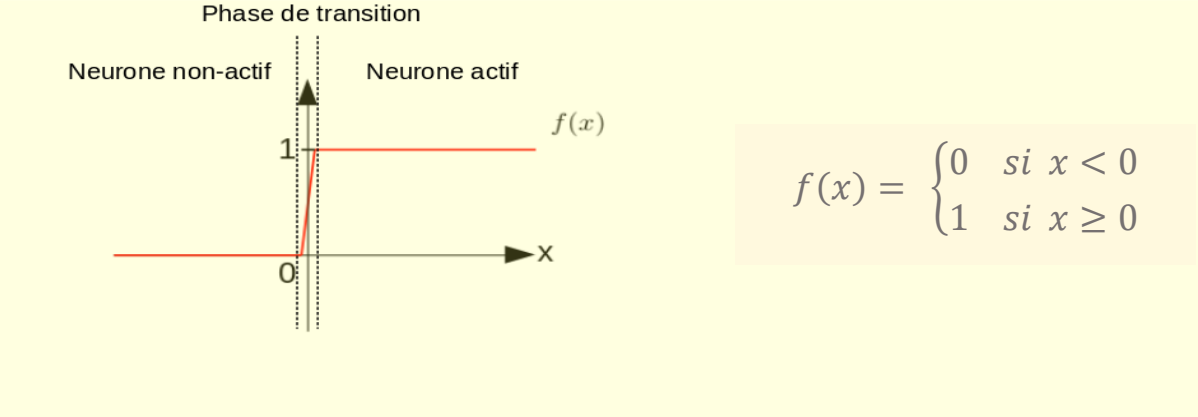


Figure 1.15 – Fonction d'activation échelon avec un seuil de 0

Un perceptron peut être composé d'un ou plusieurs LTU en entrées et d'une ou plusieurs sorties. Les perceptrons permettent de résoudre des problèmes de classification linéaires.

Par la suite, les perceptrons multicouches (MLP) ont vu le jour en 1986 et se représentent généralement sous la forme suivante (*Figure 1.16*) :

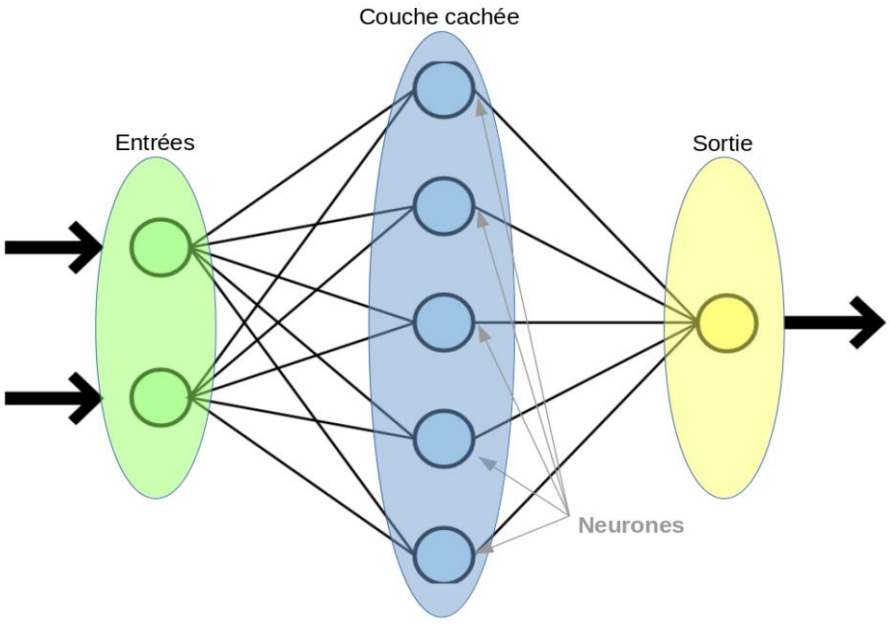


Figure 1.16 – Les perceptrons multicouches (Dejasmin, 2018)

On distingue une couche d'entrée en vert sachant qu'il peut y avoir autant d'entrées que l'on souhaite, une couche cachée en bleu (ici avec 5 neurones) et une couche de sortie en jaune (il peut y avoir plus d'une sortie). Les perceptrons multicouches ont pour objectif de classer différentes données selon leur étiquette. Pour cela le perceptron observe chacune des données qu'il possède et met à jour chaque poids de chaque neurone de son réseau afin de classifier au mieux cette base de données. L'algorithme que les perceptrons utilisent pour mettre à jour leurs poids s'appelle la rétropropagation du gradient de l'erreur (LeCun et al., 2015).

Les années 2010 ont vu la naissance du deep learning autrement appelé un réseau de neurones profond qui est un perceptron avec au minimum deux couches cachées (il peut y en avoir autant qu'on le souhaite). La révolution du deep learning est liée à la puissance des ordinateurs qui ne cesse d'augmenter et aussi de la quantité de données qui ne cesse de s'accumuler par exemple l'explosion des images a révolutionné la performance de la reconnaissance d'images.

Un modèle de réseaux de neurone profond est capable aujourd'hui de reconnaître chaque élément d'une scène pourvu qu'il ait été entraîné pour cela. À partir de la sémantique de cette reconnaissance, un réseau de neurones profond peut même générer une légende à cette scène.

Par exemple, aujourd'hui il est assez courant de pouvoir obtenir le résultat suivant automatiquement (*Figure 1.17*) :

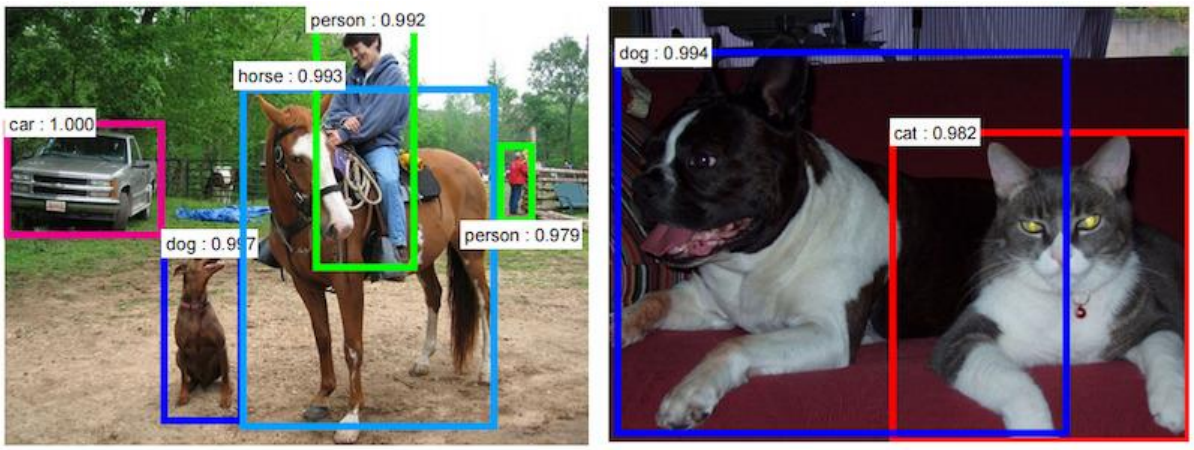


Figure 1.17 – Description d'une scène d'image par les techniques du deep learning

Le fonctionnement de la reconnaissance d'image par intégration d'une méthode basée sur le deep learning est basé sur les pixels (Richards and Richards, 1999).

1.5.2 *Les algorithmes du Deep Learning*

Il existe différents algorithmes de Deep Learning. Nous pouvons ainsi citer :

- Les **réseaux de neurones profonds** : Ces réseaux sont similaires aux réseaux MLP, mais avec plus de couches cachées. L'augmentation du nombre de couches permet à un réseau de neurones de détecter de légères variations du modèle d'apprentissage, favorisant le sur-apprentissage ou sur-ajustement.

- Les **réseaux de neurones convolutionnels** : Le problème est divisé en sous-parties, et pour chaque partie, un « cluster » de neurones sera créé afin d'étudier cette portion spécifique. Par exemple, pour une image en couleur, il est possible de diviser l'image sur la largeur, la hauteur et la profondeur (les couleurs).
- La **machine de Boltzmann profonde** : Ces algorithmes fonctionnent suivant une première phase non supervisée, suivie de l'entraînement classique supervisé. Cette étape d'apprentissage non supervisée permet, en outre, de faciliter l'apprentissage supervisé.

Dans ce travail de thèse, nous nous sommes plus intéressés aux réseaux de neurones convolutifs pour la reconnaissance d'image par apprentissage profond. Les réseaux de neurones convolutifs sont plus efficaces et plus utilisés dans ce domaine, la section suivante y est consacrée.

1.5.3 *Les réseaux de neurones convolutifs*

La convolution est un outil mathématique simple qui est très largement utilisé pour le traitement d'image, ce qui explique que les réseaux de neurones à convolution soient particulièrement bien adaptés à la reconnaissance d'image. La convolution agit comme un filtrage. On définit une taille de fenêtre qui va se balader sur toute l'image (sachant qu'une image peut être vue comme étant un tableau). Au tout début de la convolution, la fenêtre sera positionnée tout en haut à gauche de l'image puis elle va se décaler d'un certain nombre de cases (appelé le pas) vers la droite et lorsqu'elle arrivera au bout de l'image, elle se décalera d'un pas vers le bas, ainsi de suite jusqu'à ce que le filtre ait parcouru la totalité de l'image (*Figure 1.18*).

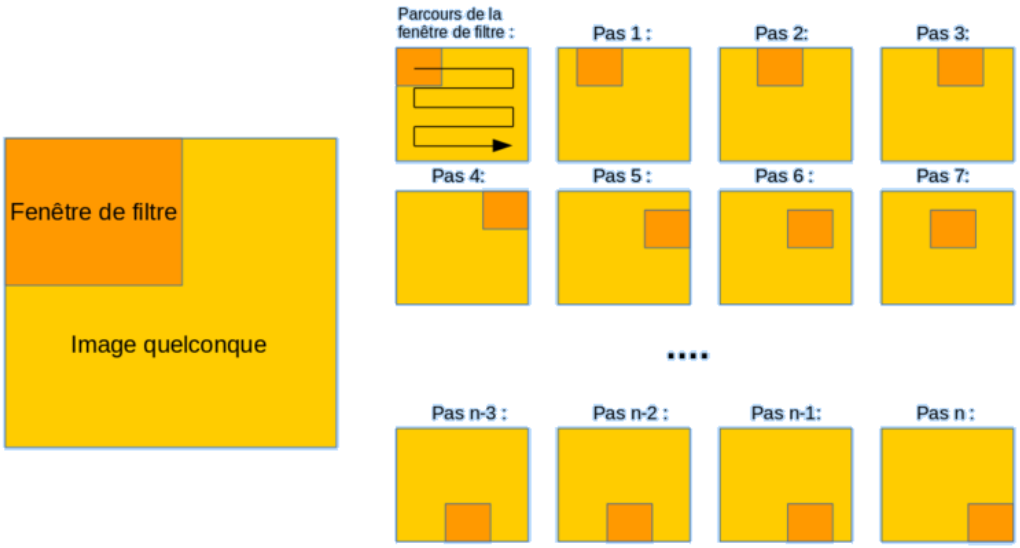


Figure 1.18 – Principe de la convolution

Le but est de se servir des valeurs présentes dans le filtre à chaque pas. Par exemple si l'on définit une fenêtre 3 par 3, cela représentera 9 cases du tableau (c'est-à-dire 9 pixels). La convolution va effectuer une opération avec ces 9 pixels. Il peut s'agir de n'importe quelle opération, par exemple on extrait la valeur la plus grande (soit le pixel avec la plus grande valeur) (*Figure 1.19*).

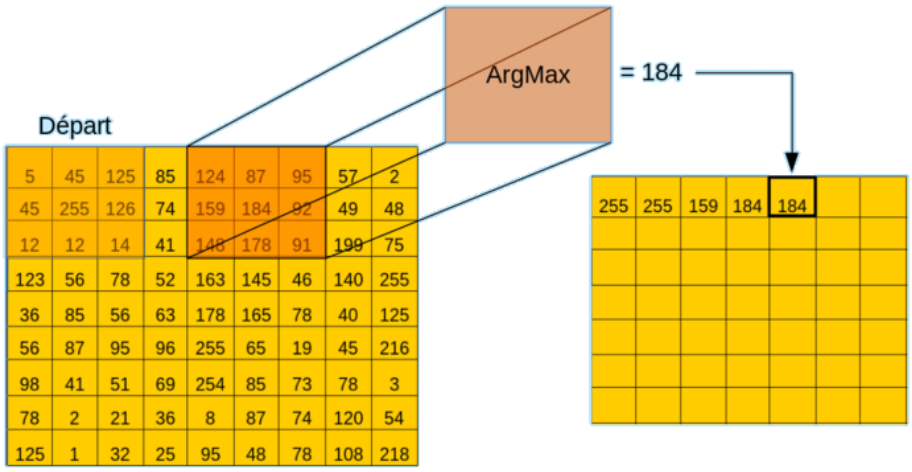


Figure 1.19 – Exemple d'une convolution dont la configuration est : Opération = Argument maximal, pas horizontal = 1 pixel, pas vertical = 1 pixel.

L'image obtenue à la suite de la convolution est appelée « carte de caractéristiques », avec une dimension plus petite que celle de l'image en entrée. Les réseaux de neurones convolutifs sont directement inspirés du cortex visuel des vertébrés. Un réseau de neurones à convolution, appelé aussi convnet (pour « Convolutional Network »), ou encore CNN (pour « Convolutional Neural Network ») se représente en général sous la forme suivante (*Figure 1.20*) :

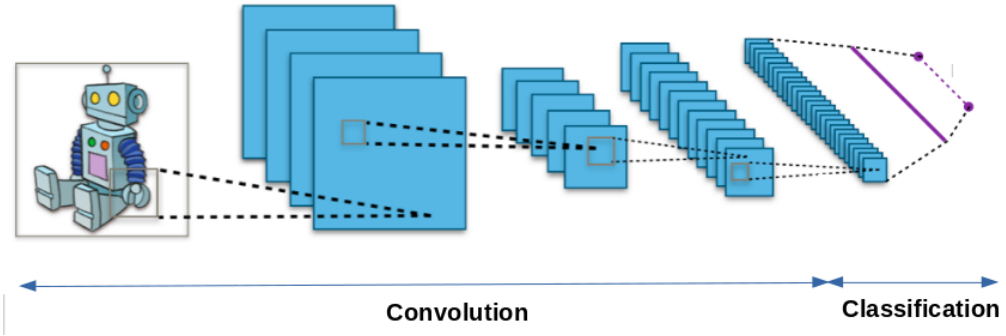


Figure 1.20 – Architecture d'un réseau de neurones convolutif

On distingue deux parties, une première partie appelée la partie convolutive du modèle et la seconde partie appelée la partie classification du modèle qui correspond à un modèle MLP (Multi Layers Perceptron).

1.5.3.1 Le but de la partie convolutive des modèles CNN

Considérons par exemple comme entrée de notre réseau de neurones, une image quelconque de 512 pixels de côté. Cette image comporte donc 262.144 pixels (512*512). Imaginons maintenant que nous n'avions jamais entendu parler des CNN et que nous utilisons donc un réseau de neurones profonds classique (MLP). Ce réseau, par exemple, présente sur sa première couche cachée 512 neurones. Cela a pour conséquences que sur la première couche cachée uniquement nous avons plus de 13 millions de poids à calculer (262.144 * 512). Ce n'est pas envisageable.

Le CNN et en particulier sa partie convolutive permet de pallier ce problème, car elle va grandement diminuer le nombre de poids à calculer dans le modèle. En effet, comme nous l'avons vu ci-dessus, la convolution va avoir pour effet de réduire la dimension de « la carte de caractéristiques » que l'on obtient après convolution (en comparaison avec la taille de l'image en entrée). Si l'on répète ce processus plusieurs fois, en prenant comme nouvelle entrée (sur laquelle nous allons effectuer la convolution) la sortie de la convolution précédente, nous allons diminuer de plus en plus la taille de la carte de caractéristiques, et donc nous diminuons également le nombre de calculs (*Figure 1.21*) :

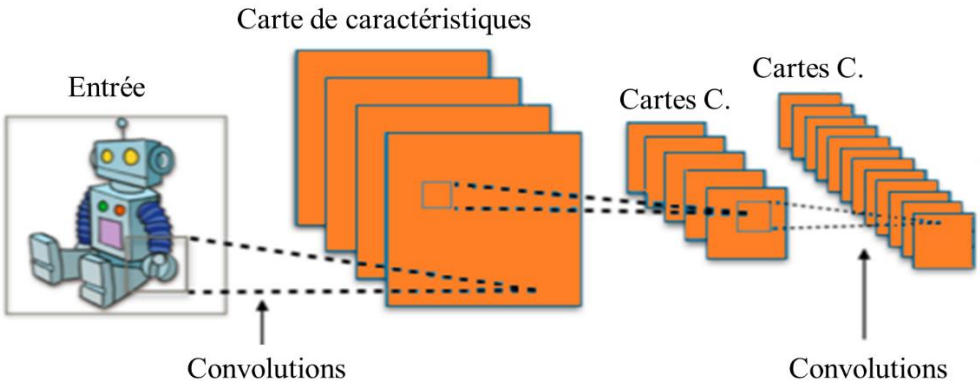


Figure 1.21 – Réduction de la taille de la carte de caractéristiques

Nous remarquons que la convolution a pour avantage de n'avoir que quelques poids à calculer (ceux du filtre) et qu'elle les réutilise pour l'ensemble de l'image, tandis que les MLP vont avoir un poids unique à calculer pour chaque neurone. Cela réduit encore considérablement le nombre de calculs. Dans le but de simplifier, il n'a été mentionné que la carte de caractéristiques en 2D. Mais il est notable qu'une image se représente en 3D, il y a 2 dimensions qui correspondent à la largeur et à la hauteur de l'image et une troisième dimension qui correspond à la composante couleur (un pixel est composé d'une composante rouge, d'une verte et d'une bleue). Ainsi, une image sera un tableau 3D de taille : (largeur, hauteur, 3).

Lorsque l'on effectuera une convolution sur une image, nous aurons donc en sortie 3 cartes de caractéristiques (une pour chaque composante couleur). Il est également notable que nous pouvons effectuer plusieurs convolutions sur une même image (ou sur une même carte de caractéristiques), l'idée est que chaque convolution effectuée sur une image correspondra à une caractéristique particulière de cette image. Finalement, en sortie d'une première convolution d'une image nous aurons 3 *n cartes de caractéristiques différentes (3 = composante rouge, verte et bleu de l'image et n =nombre de caractéristiques choisies).

À la sortie de la partie convolutive nous avons un long vecteur qui comprend les caractéristiques les plus pertinentes de l'image. Nous connectons chaque valeur de ce vecteur à un neurone du réseau de neurones de la partie classification, que l'on appellera réseau de neurones entièrement connecté. Cette partie agit comme un MLP et permet de classer les images que l'on souhaite.

1.5.3.2 *Entraînement des réseaux de neurones convolutifs*

Lorsque l'on fait de la reconnaissance d'image par apprentissage profond, le plus important c'est d'avoir une grande base de données, c'est à dire avoir le plus d'images possible. Comme la écrit Jean-Claude Heudin (auteur de « Comprendre le deep learning ») : « Ce n'est pas forcément celui qui a le meilleur algorithme qui gagne, c'est celui qui a le plus de données. » Pour les réseaux de neurones convolutifs performants, il faut être en mesure de fournir un maximum d'images au modèle qu'on souhaite former tout en faisant attention à lui fournir une base de données cohérente et utile. Le processus d'entraînement d'un réseau de neurones convolutifs est le suivant :

1. Préparation de la base de données

Lorsque que l'on crée un modèle de réseaux de neurones convolutifs ou tout simplement quand on utilise un modèle déjà existant, nous devons entraîner ce modèle avec des données afin qu'il puisse apprendre et être capable de reconnaître un objet ou quelque chose en particulier. Pour cela on utilise une base de données contenant toutes nos images.

Une partie de cette base de données servira à entraîner le modèle CNN, c'est ce qu'on appelle « la base de données d'entraînement ». La partie de la base de données restante sert pour évaluer le modèle, on l'appelle la « base de données de validation ». Ce qui est important de savoir c'est que toutes les images que nous utilisons doivent être préalablement étiquetées, c'est-à-dire que chacune des images que nous utilisons doit appartenir à une classe en particulier. Par exemple si nous souhaitons reconnaître des voitures, des vélos et des motos, cela nous donne 3 classes. Chaque image de notre base de données doit appartenir à une de ces trois classes. C'est ce que l'on appelle l'apprentissage supervisé.

2. Processus d'apprentissage

L'entraînement d'un CNN consiste à déterminer et à calculer empiriquement la valeur de chacun de ses poids. Le principe est le suivant : le CNN traite une image (de la base de données d'entraînement) et en sortie il fait une prédiction, c'est-à-dire qu'il dit à quelle classe il pense que cette image appartient. Sachant qu'on connaît préalablement la classe de chacune des images d'entraînement, on peut vérifier si ce résultat est correct.

En fonction de la véracité de ce résultat, on met à jour tous les poids du CNN selon un algorithme qui s'appelle la rétropropagation du gradient de l'erreur.

Lors de la phase d'entraînement du modèle, le processus expliqué ci-dessus est répété plusieurs fois et avec la totalité des images de la base de données d'entraînement. Le but étant que le modèle classifie au mieux ces données.

Lorsque le modèle a fini de mettre à jour ses poids, on évalue le modèle en lui présentant la base de données de validation. Il classe toutes ces images (qui sont des images que le modèle n'a jamais vues) et on calcule son taux de bonne classification, c'est ce qu'on appelle la précision du modèle.

A la fin du processus d'apprentissage, trois cas de figure peuvent se présenter :

- Le modèle est aussi performant sur les données d'entraînement que sur les données de validation. Cela est le cas de figure idéale, ça signifie que le modèle a très bien fait son travail et qu'il reconnaît aussi bien les images qu'il connaît que celles qu'il n'a jamais vues.

- Le modèle reconnaît très bien les images d'entraînement et moins bien celles de validation. Le modèle aura une faible capacité prédictive, il n'arrive pas à généraliser. On parle alors de **sur-entraînement**. Dans ce cas-là on peut ajouter davantage d'images pour pallier ce problème.
- Le modèle ne reconnaît pas très bien les images d'entraînement et pas très bien non plus les images de validation. On parle alors de **sous-apprentissage**. Dans ce cas-ci, ajouter plus d'images ne servira à rien, c'est généralement le modèle choisi qui ne convient pas, il faudrait utiliser un modèle plus complexe.

1.5.4 *Les logiciels du deep learning*

Il existe de nombreux outils, librairies et framework de deep learning. Ci-dessous, nous allons présenter quelques outils les plus utilisés.

TensorFlow : à ce jour, reste le framework le plus utilisé notamment par les grandes entreprises telles que Google, Ebay, Cocacola, Twitter, Intel, Nvidia, etc. Il est développé en C++ et python à l'origine par des chercheurs et des ingénieurs de l'équipe Google Brain au sein de l'organisation AI de Google, il est plus orienté pour les technologies de l'apprentissage automatique, du deep learning en particulier et du calcul scientifique et d'autres domaines scientifiques. C'est TensorFlow qui a été utilisé dans le cadre de nos travaux.

Theano : il est créé en 2007 par Yoshua Bengio et l'équipe de recherche de l'Université de Montréal et a été le premier outil de DL (Deep Learning) largement utilisé. Theano est une bibliothèque Python, extrêmement rapide et puissante, mais critiquée pour être un framework d'apprentissage en profondeur de bas niveau. L'équipe derrière Theano a annoncé en 2017 qu'après la publication de la dernière version il n'y aura pas d'autres développements.

Torch : Torch est un framework open source de calcul scientifique et de deep learning prenant en charge les algorithmes basés principalement sur les GPU. Il est écrit en langage de script LuaJIT, très souple et facile de construire des modèles. Il a été utilisé par Google, Twitter et Facebook et d'autres entreprises. **Pytorch**, une nouvelle version de Torch basée sur Python, réalisé par Facebook en 2017, ce qui a augmenté sa popularité. Pytorch est un framework assez rapide et flexible.

Keras : a été développé par François Chollet comme une interface facile à utiliser pour simplifier la construction de réseaux de neurones pour plus de rapidité. Il est écrit en Python et peut être utilisé comme une couche supérieure sur TenserFlow et Theano. Keras sera inclus par Google dans les prochaines versions de TenserFlow à partir de 2018.

Caffe2 : C'est une librairie open source capable de fonction avec TensorFlow, CNTK ou Theano, développée par le centre de recherche en Intelligence Artificielle (IA) de Berkeley et écrit en C++ et Python. Il se caractérise par son interface facile à utiliser, sa modularité et son extensibilité. Il supporte également les réseaux de neurones récurrents et les réseaux de neurones convolutifs et fonctionne sur toutes les plateformes. Il existe d'autres comme MXNet, Deeplearning4j, CNTK de Microsoft…

Actuellement, il existe des centaines d'applications de l'intelligence artificielle basées sur le deep learning. Par exemple, **Amazon** utilise des outils de deep learning pour recommander aux clients ce qu'il faut acheter et ces recommandations couvrent près d'un tiers de ce qu'ils achètent réellement. La société gère également ses centres physiques en employant près de 80 000 robots. **Microsoft** utilise la reconnaissance vocale dans Bing pour les traductions de langages en temps réel et apprend plus d'algorithmes de deep learning avec CNTK de Microsoft. **YouTube** exploite une reconnaissance vidéo artificielle sur les classifications des contenus afin de protéger les utilisateurs des mauvaises surprises. Le système de reconnaissance vocale de **Google Home Assistant** connaît l'utilisateur parfaitement et ses préférences. **Gmail** comprend le texte de votre boîte de réception et vous suggère des réponses personnalisées. **Uber** fournit des systèmes de prévision en temps réel pour offrir un transport plus fiable en adaptant les méthodes du deep learning. Dans cette thèse, nous avons utilisé le deep learning pour la classification automatique d'images des agents pathogènes responsables des maladies épidémiologiques.

1.6 Conclusion

Ce chapitre nous a permis de définir et comprendre dans les détails des techniques existantes relativement à ces quatre thématiques notamment : la télémédecine, la télédétection, le data mining et le deep learning pour faciliter la compréhension du sujet de recherche : ceci à commencer par les notions de télémédecine et d'interopérabilité des applications de télémédecine, ensuite, les techniques de data mining et du deep learning pour l'extraction de la connaissance à partir des données et le choix motivé des outils utilisés.

Nous allons donc voir par la suite les améliorations et contributions que nous avons proposées pour traiter les différentes problématiques énumérées ci-dessus liées à la modélisation des informations et extraction des connaissances pour la gestion des crises, notamment :

- Dans le cadre de la gestion de crises, la modélisation des informations pour la gestion de crise avec la télémédecine ;
- Puis dans le cadre de la prévention et de la surveillance des crises, les techniques de Data Mining sur les images satellitaires pour découvrir les zones à risque d'épidémie ;
- Et enfin, une fois dans une situation de crise, on s'intéresse particulièrement à l'aspect d'analyse médicale par l'usage des microscopes pour la classification des images d'agents pathogènes épidémiques à l'aide des réseaux de neurones convolutifs.

A cet effet, dans le chapitre suivant, nous allons traiter et présenter notre contribution sur la problématique de la modélisation des informations pour la gestion de crise nécessitant des prises en charge médicale, à l'aide d'une collaboration des applications informatiques de la télémédecine.

Chapitre 2

Modélisation des informations pour la Gestion de crise avec la télémédecine

2.1 Introduction 35
2.1.1 Définition d'une crise 35
2.1.2 Les différentes étapes de gestion de crise 35
2.1.3 Les différents actes de télémédecine 36
2.2 Travaux connexes 38
2.2.1 Approches basées sur les systèmes informatiques d'aide à la décision d'urgence 38
2.2.2 Approches orientées vers la gestion de crise en différentes étapes 39
2.2.3 Approches axées sur les modèles mathématiques 41
2.2.3.1 La description fonctionnelle de la gestion de crise 42
2.2.3.2 Les flux d'informations liés aux fonctions de gestion de crise 42
2.2.4 Synthèse des différentes approches 44
2.3 Méthodologie adoptée pour la gestion de crise à distance 45
2.3.1 Processus de gestion de crise à distance 45
2.3.2 Modélisation mathématique 49
2.3.3 Algorithme proposé 51
2.4 Étude de cas 51
2.5 Résultat et discussion 55
2.6 Conclusion 56

2. Modélisation des informations pour la Gestion de crise avec la télémédecine

2.1 Introduction

Ce chapitre présente notre première contribution relative à la problématique de la modélisation des informations pour la gestion de crise à l'aide d'une collaboration des applications de télémédecine. Pour mener à bien cette étude, ce chapitre est organisé en cinq parties. La première partie présente les travaux existants dans le domaine de la gestion de crise à l'aide des outils d'aide à la décision. Dans la deuxième partie, nous présentons notre méthodologie pour la gestion des crises à distance. Puis dans la troisième partie, nous proposons une application de notre méthodologie à travers une étude de cas sur la simulation d'une crise de tsunami à Cannes (France). Dans la quatrième partie, nous présentons les résultats obtenus et nous présentons une discussion objective. Et enfin, la dernière partie est réservée à la conclusion. Mais avant d'explorer ces différentes parties, quelques notions seront abordées notamment sur la crise, les différentes étapes générales de gestion de crise et la définition des actes de télémédecine.

2.1.1 Définition d'une crise

À travers le monde, en différents points du globe, l'humanité est confrontée à une multitude de crises qui peuvent affecter la santé humaine, le climat et l'environnement (Son and Thanh, 2018). Une crise est définie comme l'apparition d'un état de désordre, d'anomalie, de dysfonctionnement, de déséquilibre profond, ou de trouble grave (Girard et al., 2006). La gestion de crise est « *une série de fonctions ou de processus permettant d'identifier, d'étudier et de prévoir les problèmes de crise et de définir des moyens spécifiques permettant à une organisation de prévenir ou de faire face à une crise* » (Kash et Darling, 1998).

2.1.2 Les différentes étapes de gestion de crise

La gestion de crise est un sujet complexe, car elle nécessite une bonne compréhension des phénomènes impliqués et leurs conséquences (notamment les scénarii, les événements et les résultats), la connaissance des processus sous-jacents et l'analyse des causes profondes (Jensen et Aven., 2018). Une crise peut être causée par deux types de risques majeurs (Tableau 2.1).

Tableau 2.1 – Classification des risques majeurs (Gupta et al., 2002)

Risque naturel	**Risque technologique**
Inondation et forte pluie	Accident industriel
Feu de forêt	Accident électrique
Chute de neige abondante	Accident chimique
Avalanche	Accident nucléaire
Cyclone	Échec du barrage
Tremblement de terre	Transport de marchandises dangereuses
Mouvement de terre	Libération de gaz toxiques
Éruption volcanique	Libération de liquide toxique

En gros, la résolution d'une crise consiste non seulement à revenir à une situation normale, mais aussi à minimiser ses conséquences. Cela nécessite une planification préalable en termes de planification, préparation et exécution. Ce processus se divise en trois étapes principales (Figure 2.1) : (1) diffusion des messages d'alerte (2) évacuation des victimes et (3) protection des zones évacuées (Devlin, 2006; Zhou et al., 2017).



Figure 2.1– Les étapes de la gestion de crise

En situation d'urgence, les activités médicales à distance délivrées à l'aide de la télémédecine sont importantes pour la protection de la vie humaine dans la gestion de crise. Les gestionnaires de crise et les acteurs seront ainsi mieux équipés pour faire face à des risques et des situations sensibles, ainsi ceci permet de réduire l'impact des problèmes médicaux complexes causés par les risques naturels ou industriels. En outre, le développement des logiciels de télémédecine et le déploiement des plateformes ont réalisé d'importants progrès au cours de ces dernières années, impliquant une importante augmentation du nombre d'outils de communications et de réseaux de collaboration. Donc, il y a un fort besoin de systèmes d'information interopérables pour faciliter leur communication efficace, ce qui implique le partage et l'échange de données entre ces systèmes (dossier médical patient, images radiologiques, données d'analyses médicales, etc.).

L'objectif principal de cette étude est de fournir un nouveau cadre méthodologique pour la gestion de crise à distance à l'aide de la collaboration des actes de télémédecine. Elle est basée sur une spécification formelle des flux d'informations et une caractérisation des moyens d'exploitation pour fournir aux gestionnaires de crise des moyens pertinents pour faciliter la collaboration en vue de prendre des décisions critiques. L'application de cette méthodologie a deux principaux avantages : (1) le renforcement de la collaboration à distance entre professionnels de santé axée sur les logiciels de télémédecine existants (Saidi et al., 2018) et (2) l'amélioration des soins médicaux afin de réduire le nombre de décès dû au manque d'assistance médicale (ASIP, 2012). De plus, le traitement médical commence pendant la phase de transport des victimes de la crise, ce qui permet de soulager à distance les patients avant leur arrivée dans les structures de santé.

En général, la gestion de crise à distance est soutenue par un environnement de travail collaboratif utilisant des systèmes de workflow, aidant ainsi les décideurs à gérer et à résoudre les problèmes de crise (Mak et al., 1999). Dans ce travail, la modélisation conceptuelle peut être appliquée à tous les types de crises avec la prise en charge médicale des victimes. Par conséquent, les actes de télémédecine sont pertinents dans le contexte du système d'information collaboratif pour aider les victimes de la crise.

2.1.3 Les différents actes de télémédecine

La télémédecine est l'utilisation des technologies de télécommunication et de l'information afin de fournir des soins de santé cliniques à distance (Najeeb, 2012). La loi française portant le numéro de décret 2010-1229 du 19 octobre 2010 (articles R.6316-1 à R.6316-9 du code de la santé publique) précise les cinq actes réalisables de la télémédecine (Figure 2.2) :

- **Téléconsultation :** un médecin donne une consultation à distance à un patient, qui peut être assisté par un professionnel de la santé (professions médicales, professions de la pharmacie et de la physique médicale, et des auxiliaires médicaux aux ambulanciers). Le patient et/ou le professionnel de la santé à ses côtés fournissent l'information, le médecin peut également faire le diagnostic à distance;
- **Téléexpertise** : un médecin sollicite à distance l'avis d'un ou de plusieurs collègues sur la base d'informations médicales relatives à la prise en charge d'un patient ;
- **Télésurveillance** : un médecin surveille et interprète à distance les paramètres médicaux d'un patient. L'enregistrement et la transmission des données peuvent être automatisés ou effectués par le patient lui-même ou par un professionnel de la santé ;
- **Téléassistance** : un médecin aide à distance un autre professionnel de la santé lors de l'exécution d'un acte ;
- **Réponse médicale** : les médecins des services médicaux d'urgence utilisent le téléphone pour établir un diagnostic initial afin de déterminer et de déclencher la réponse la plus adaptée à la nature de l'appel.

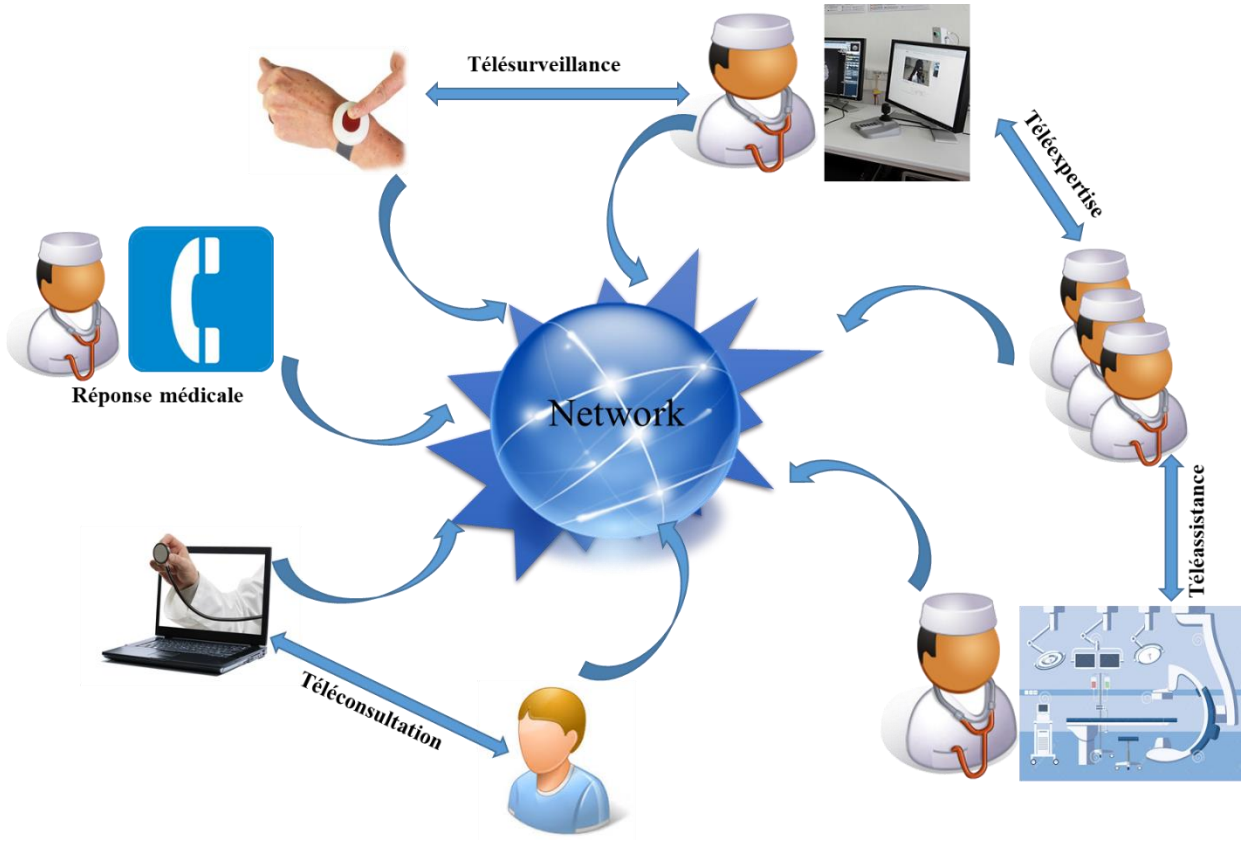


Figure 2.2 – Les différents actes de télémédecine

L'intégration de la télémédecine fait partie d'une initiative qui vise à améliorer la gestion des crises. En effet, la télémédecine consiste à promouvoir l'offre des services de santé à distance et l'échange d'informations médicales. L'utilisation de la télémédecine est bénéfique pour les soins médicaux aux victimes de crises et aider les communautés à améliorer la mise en œuvre d'une politique de

gestion de crise. La télémédecine et ses actes apportent une aide précieuse dans la gestion des crises par le partage d'informations, la concertation entre professionnels de santé et la communication entre tous les acteurs impliqués (par exemple entre le poste de commandement et l'équipe logistique) en situation de crise. Par exemple, la téléconsultation peut aider à établir les niveaux de priorité qui déterminent l'ordre dans lequel les patients seront traités et évacués. Le but de ce tri est de sauver le maximum de victimes.

2.2 Travaux connexes

Nous nous intéressons aux travaux relatifs à la gestion de crises en général. L'objectif principal de la gestion de crise consiste à réduire les dommages matériels et humains, notamment à travers une prise de décision d'urgence et de façon organisée (Emergency Decision-Making – EDM). C'est ainsi que nous avons recensé trois principales tendances générales dans la littérature pour traiter l'EDM (Zhou et al., 2017) (Figure 2.3) :

- la première tendance est basée sur les systèmes informatiques d'aide à la décision d'urgence (Système d'Information Géographique - SIG, Systèmes Multi Agent - SMA, Raisonnement par cas et combinant apprentissage automatique et Traitement automatique du langage),
- la deuxième tendance concerne le découpage du problème en différentes étapes basées sur l'évolution de la situation
- la dernière tendance est basée sur les modèles mathématiques.

2.2.1 Approches basées sur les systèmes informatiques d'aide à la décision d'urgence

Pour les systèmes informatiques d'aide à la décision d'urgence, Vescoukis et ses collègues (Vescoukis et al., 2012) ont proposé des technologies d'architecture orientée service (SOA) flexibles pour la planification et l'aide à la décision dans la gestion des crises environnementales utilisant des outils SIG. Keramitsoglou et ses collègues (Keramitsoglou et al., 2004) ont proposé une approche également basée sur le SIG pour la gestion des crises de feux de forêt en Europe du Sud. Les SIG constituent un support important ; par exemple la cartographie (cartes géographiques et géologiques) des inondations est un moyen potentiel de surmonter les risques d'inondations en milieu urbain avec une sensibilisation préalable des populations aux catastrophes (Padawangi et al., 2016). D'autres chercheurs (Traore et al., 2017; Voigt et al., 2007) ont utilisé la télédétection comme meilleure option des technologies informatiques pour améliorer les décisions concernant la planification de la répartition des services publics, de l'équipement et du personnel sur le terrain de façon appropriée. Slam et ses collègues (Slam et al., 2015) ont proposé une plateforme pour les systèmes d'aide à la décision de crise intégrant les incertitudes, le raisonnement et l'apprentissage avec des réponses en temps réel pour renforcer la prise de décision. Ben Othman et ses collègues ont adopté une approche similaire, mais plutôt basée sur les systèmes multiagents (SMA) pour la gestion des chaînes d'approvisionnement d'urgence et la coopération multiagents garantissant la livraison des ressources des zones d'approvisionnement aux zones touchées par la crise (Ben Othman et al., 2017).

Cependant, en gestion de crise d'urgence, il est difficile de connaître exactement à l'avance les différentes zones touchées. Selon une étude empirique antérieure des systèmes d'aide à la décision (Arnott and Pervan, 2005), la question de la pertinence est le point faible de nombreuses recherches, car les acteurs, les clients et les utilisateurs des différentes applications sont mal identifiés. Il est donc essentiel d'adapter les systèmes d'aide à la décision de manière à prendre en compte les degrés

de désirabilité pour mesurer les conséquences des décisions prises (Nokhbeh Foghahaayee et al., 2014).

2.2.2 Approches orientées vers la gestion de crise en différentes étapes

Zhou et ses collègues (Zhou et al., 2017) ont défini l'EDM en quatre étapes: atténuation des catastrophes, préparation aux catastrophes, intervention en cas de catastrophe et reprise après catastrophe (voir Figure 2.3).

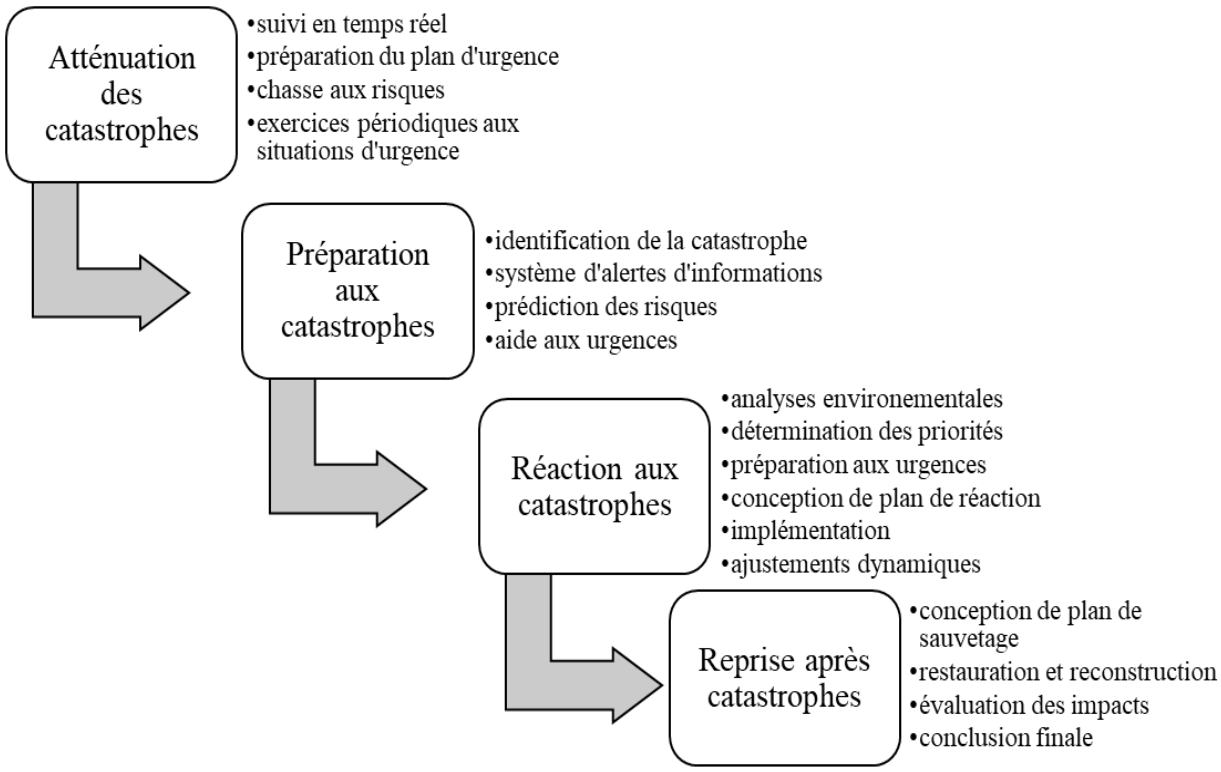


Figure 2.3 – Les étapes d'EDM pour la gestion des catastrophes naturelles

Ce type d'organisation de gestion de crise peut être déployé pendant une crise pour renforcer la surveillance, la sécurité des zones dangereuses, l'information des populations, l'évacuation des zones touchées et l'hébergement des évacués. Pour montrer l'importance de la phase de réaction aux crises, Scheer et ses collègues (Scheer et al., 2012) ont proposé un cadre générique d'évacuation des victimes d'une crise en trois étapes comme suit: (i) disposer d'une première version valable d'un plan d'évacuation (ii) installer et diffuser le plan d'évacuation et le déployer (iii) intégrer et maintenir le plan à long terme. L'exécution des plans de gestion de crise nécessite une coordination des acteurs impliqués en unités de gestion de crise. Par exemple, en France, le plan de sauvegarde communale définit clairement le rôle et la composition d'une unité de gestion de crise (Figure 2.4).

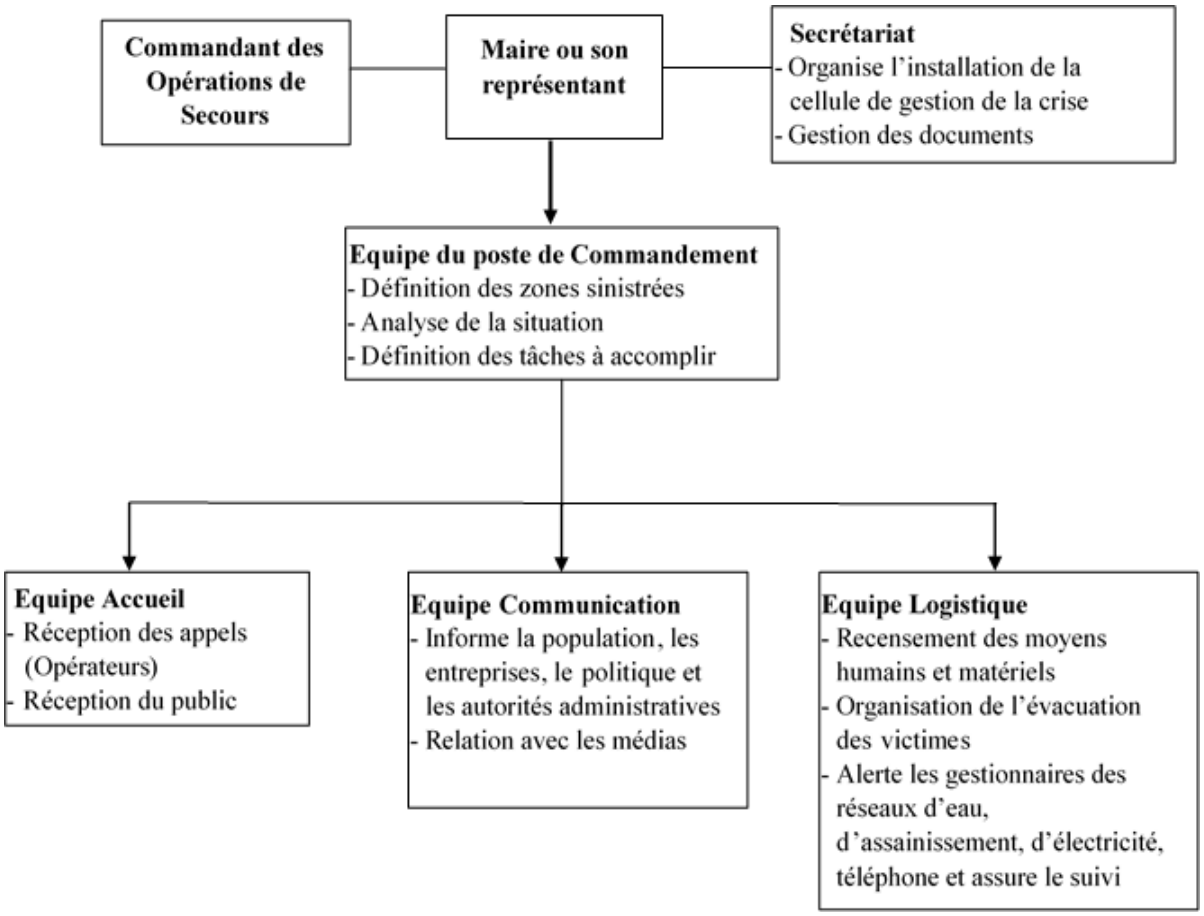


Figure 2.4 – Organisation de l'unité de gestion de crise (SIRACEDPC et CYPRES, 2014)

L'unité de gestion de crise est un groupe de réflexion interdisciplinaire qui peut réagir immédiatement à des événements graves ou à des risques majeurs afin de permettre au maire d'une commune administrative concernée de prendre les mesures les plus appropriées. Cette cellule de crise constitue alors un poste de commandement (PC) et elle doit conseiller et proposer au maire des actions concrètes pour limiter les effets des catastrophes, sécuriser et protéger la population. Pour la composition d'une unité de gestion de crise, on peut distinguer deux catégories d'acteurs : (i) personnel permanent qui peut être convoqué quel que soit le type de catastrophe dont il est question au sein de l'unité ; et (ii) les services ou les conseillers techniques qui sont des professionnels ou des spécialistes d'un risque particulier. Les unités de gestion de crise peuvent être structurées en plusieurs équipes : un poste de commandement, une équipe de communication, une équipe logistique et une équipe de réception pour recevoir les appels publics et téléphoniques, comme le montre la Figure 2.4. Ce type d'unité de gestion de crise peut être déployé pendant une crise de tsunami.

Abouelјinane et ses collègues (Abouelјinane et al., 2013) ont fourni un aperçu des modèles de simulation appliqués aux services médicaux d'urgence (Emergency Medical Service – EMS). Ils décrivent l'EMS comme étant la régulation médicale des urgences et les systèmes de sécurité publique responsables de la stabilisation durant la préhospitalisation et du transport des victimes d'accidents ou la prise en charge des cas critiques (malaise, maladie ou grossesse). La réalisation

de l'EMS nécessite un centre d'opérations pour gérer les différents appels et des opérations externes pour apporter des premiers secours aux victimes. Une description détaillée des étapes du processus des premiers secours est présentée dans la Figure 2.5. Les opérations EMS gèrent trois types de décisions :

- Les décisions à long terme sont des décisions politiques en termes d'économie, de considération sociale, de ressources humaines, d'équipement et de localisation des services.
- Les décisions à moyen terme déterminent le nombre d'équipes de sauvetage dans chaque base et consistent également à établir une programmation des véhicules de service disponibles afin de satisfaire la demande des équipes de secours.
- Les décisions à court terme sont : (i) l'organisation de l'envoi des équipes de secours (ii) le choix du bon hôpital pour réduire le temps de transport (iii) le redéploiement stratégique pour assurer une meilleure couverture des futurs appels entrants (iv) la gestion du temps de la préparation du véhicule (nettoyage et remplacement de l'équipement) pour une nouvelle intervention.

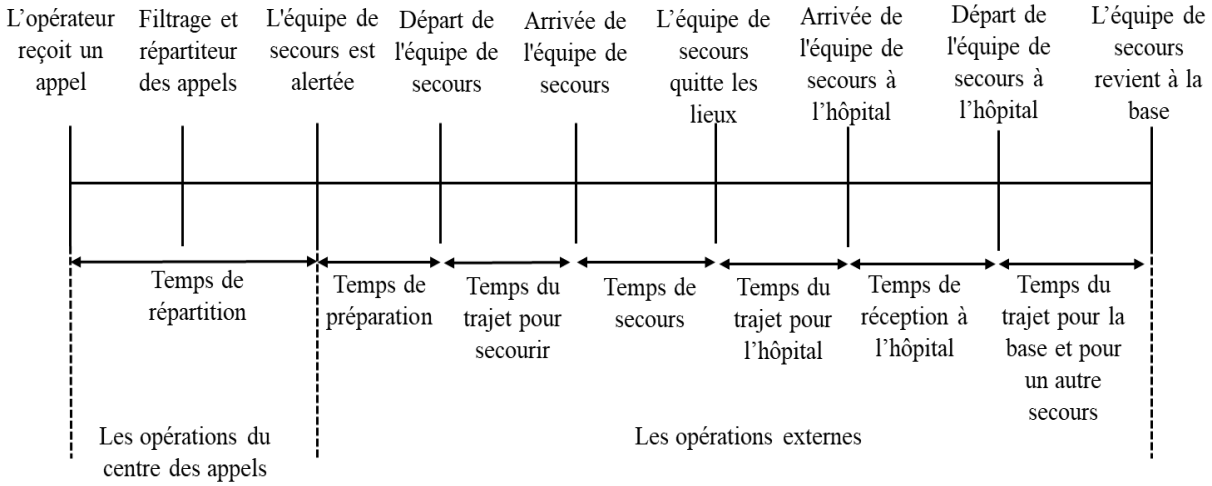


Figure 2.5 – Le processus EMS typique (Aboueljinane et al., 2013)

Il existe trois principaux types de mesures de performance utilisées dans les modèles de simulation EMS : la rapidité d'exécution, les taux de survie et les coûts déployés.

2.2.3 Approches axées sur les modèles mathématiques

Cette section est principalement basée sur les travaux de Bénaben concernant la modélisation mathématique de la gestion de crise. En fait, Frédérick Bénaben (Bénaben, 2016) a proposé un modèle mathématique général basé sur deux principaux éléments: (1) la nature des informations disponibles ou requises et (2) les fonctions qui devraient être en mesure de les utiliser pour fournir au domaine de gestion de crise un système d'aide à la décision pertinente. Selon Frédérick Bénaben (Bénaben, 2016), les concepts de crise peuvent être caractérisés par la localisation, le type et la gravité. Pour la gestion de crise, il faut également considérer des concepts tels que le contexte, les partenaires et les objectifs. La description formelle des concepts et de la gestion des crises est définie dans la sous-section suivante (Figure 2.6).

2.2.3.1 La description fonctionnelle de la gestion de crise

Pour réaliser une gestion de crise efficace, les trois objectifs principaux sont : définir la stratégie, réaliser la stratégie (considérant que ce n'est pas parce que le plan de gestion de crise est correctement décrit qu'elle sera réalisable) et maintenir la stratégie (en considérant également que la situation de crise peut évoluer ou que la stratégie adoptée pour la crise peut ne pas avoir les conséquences attendues). Ces trois objectifs sont décrits comme des fonctions : considérons d'abord O comme un ensemble d'objectifs de gestion de crise O_i (principalement "prévenir un risque" ou "traiter un effet"), F comme un ensemble de fonctions partenaires F_i et P comme un ensemble de processus métier de gestion de crise Pi. Donc nous avons :

$$\textbf{\textit{Définir: } } O^n \times F^n \rightarrow P$$

($[O_1...O_n]$, $[F_1...F_n]$) → P= le processus atteignant les objectifs avec les fonctions disponibles.

La fonction *Définir* est chargée de prendre en compte les objectifs réels de la gestion de la situation de crise (le vecteur objectif) et les fonctions disponibles des partenaires (le vecteur fonction) pour construire le processus métier dédié à atteindre ces objectifs avec ces fonctions. Évidemment, il peut y avoir plusieurs processus construits grâce à plusieurs invocations de la fonction *Définir*.

$$\textbf{\textit{Réaliser: } } P \times F^n \times O^n \rightarrow O^n$$

(Pi, $[F_1...F_n]$, $[O_1...O_n]$) → $[O'_1 \ ... \ O'_n]$= le statut des objectifs mis à jour "à la volée", selon les progrès du processus.

La fonction *Réaliser* est responsable de l'exécution du processus créé par la fonction *Définir*. Par conséquent, il est dédié à invoquer les fonctions partenaires pertinentes selon le schéma proposé par le processus. De plus, la fonction *Réaliser* fournit continuellement un ensemble d'objectifs mis à jour. Ces objectifs sont le statut attendu de la situation de crise en fonction de l'avancement du processus (c'est-à-dire la situation attendue à ce stade). Il peut y avoir plusieurs exécutions de plusieurs processus grâce à de nombreuses invocations de la fonction *Réaliser*.

$$\textbf{\textit{Maintenir: } } O^n \times O^n \times F^p \rightarrow P$$

($[O_1...O_n]$, $[O'_1 \ ... \ O'_n]$, $[F_1...F_n]$) → Pi = le processus qui correspond le mieux aux objectifs de mise à jour.

La fonction *Maintenir* est responsable de l'agilité dans la réponse. En fait, cette fonction vise à comparer la situation attendue (objectifs fournis par la fonction *Réaliser*) et la situation réelle (objectifs issus de l'image réaliste de la situation). Ainsi, cette fonction compare ces deux ensembles d'objectifs puis, en fonction des différences potentielles entre ces ensembles d'objectifs, elle utilise la fonction *Définir*, l'ensemble des objectifs actuels et l'ensemble des fonctions pour déduire un nouveau processus plus adapté aux situations actuelles.

2.2.3.2 Les flux d'informations liés aux fonctions de gestion de crise

Les concepts de crise peuvent être caractérisés par les premiers paramètres d'identification (localisation, type et gravité). Considérons G comme l'ensemble des zones géographiques / sociales l_i (utilisé pour décrire l'emplacement et le périmètre d'une situation de crise), comme l'ensemble des types de situations de crise t_i (telles que les catastrophes naturelles, les accidents industriels, les attaques, etc.) et la gravité g est définie comme une valeur comprise entre 0 et 1 (peu importe,

comment elle est calculée). Considérons aussi S comme l'ensemble des enjeux S_i, IR / ER comme l'ensemble des risques intrinsèques / émergents IRi / ERi, et E comme l'ensemble des effets Ei. En conséquence, nous pouvons décrire :

Fonction: G x T x]0,1[→ F^n

(l_i,t_i,g_i)→ [F_1...F_n]= l'ensemble des fonctions disponibles et réparables

La fonction est chargée de délimiter la sous-partie affectée du monde pour déduire, à partir du type et de la gravité, les intervenants disponibles et leurs compétences.

Enjeu: G x T x]0,1[→ S^n

(l_i,t_i,g_i)→ [S_1...S_n]= l'ensemble des enjeux potentiellement impactés

La fonction *Enjeu* est chargée de délimiter la sous-partie affectée du monde pour en déduire du type et de la gravité, les enjeux éventuellement menacés dans la zone.

IntrinsicRisk : G x T x]0,1[→ IR^n

(l_i,t_i,g_i)→ [IR_1...IR_n]= l'ensemble des risques intrinsèques à prendre en compte

La fonction *IntrinsicRisk* est chargée de délimiter la sous-partie impactée du monde pour déduire du type et de la gravité, les risques intrinsèques à considérer dans la zone.

Objectif: S^n x IR^n x ER^n x E^n → O^n

([S_1...S_n], [IR_1...IR_n], [ER_1...ER_n], [E_1...E_n]) → [O_1...O_n]= l'ensemble des objectifs de gestion de crise, en fonction de tous les risques, effets et enjeux concernés.

La fonction *Objective* est chargée d'identifier la liste complète des objectifs de la gestion de crise en prenant en compte tous les risques (intrinsèques et émergents), tous les effets et les enjeux.

Les trois premières fonctions de gestion de crise (*Fonction, Enjeu,* et *Risque Intrinsèque*) sont très similaires : délimitation d'un espace géographique / social impacté par une crise identifiée (en termes de types et de gravité) afin de déduire les fonctions disponibles, les risques intrinsèques concernés et les enjeux menacés. Plus précisément, le troisième (*Objectif*) concerne l'agrégation des risques et des effets (en ce qui concerne les enjeux) pour construire un ensemble d'objectifs (la prévention des risques ou l'atténuation des effets) pour piloter le plan de gestion de crise.

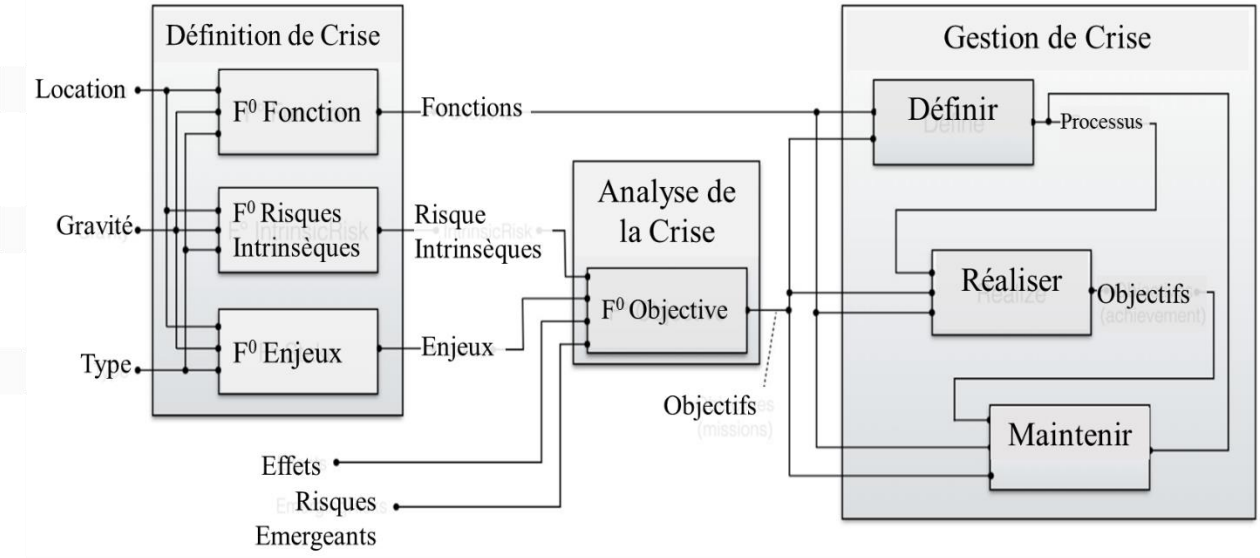


Figure 2.6 – Framework général de gestion de crise (Bénaben, 2016)

Le Framework général de gestion de crise (Figure 2.6) est décrit comme un diagramme fonctionnel où: (i) l'information est utilisée comme entrée pour toute la structure (ii) la description fonctionnelle est regroupée dans la case de gestion de crise et (iii) la description fonctionnelle est distribuée dans les cases de définition de crise et d'analyse de crise.

2.2.4 Synthèse des différentes approches

Finalement, il ressort que les trois tendances ont chacune leurs avantages et leurs inconvénients. Ainsi, l'avantage principal dans l'approche axée sur le découpage de la gestion de crise en étape est d'une importance capitale tant sur le plan organisationnel que sur le plan de la rapidité dans la maitrise de la crise. Cependant, très souvent il n'est pas facile d'exécuter à cent pour cent un plan de gestion de crise, car certaines crises causent des dégâts nécessitant de revoir complètement le plan de gestion. Par exemple, lors d'une crise d'inondation les voies d'accès des véhicules de secours peuvent être endommagées donc cela nécessite de revoir le plan de secours aux victimes. De même pour l'approche basée sur les modèles mathématiques, il faudra s'assurer de l'implémentation du modèle choisi aussi bien en termes de ressources humaines que matériel. Pareillement pour l'approche axée sur les technologies de l'information et de la communication (TIC) qui sont assez pertinentes tels que les SIG et l'imagerie satellitaire qui permettent une meilleure appréciation de certaines crises comme la crise d'inondation, suivi de l'évolution des épidémies, des feux de forêt, etc. Toutefois, leur déploiement peut s'avérer impossible dans certaines situations telles que pendant les crises entraînant des coupures d'électricité et autres aléas.

C'est pourquoi nous avons opté pour une approche fédératrice des approches précédentes, car le succès de la gestion de crise est lié non seulement à l'exécution d'un processus bien élaboré, mais à une analyse approfondie soutenue par des modèles mathématiques et aussi bien à la disponibilité d'outils TIC pour faciliter la collaboration des acteurs impliqués. Pendant que de nombreuses recherches ont été menées dans le domaine des systèmes d'aide à la décision d'urgence, à notre connaissance, aucune d'entre elles ne s'est focalisée sur l'utilisation organisée des actes de

télémédecine pour les systèmes d'aide à la décision d'urgence, en particulier pendant la phase d'évacuation des victimes. Ainsi, dans la partie suivante, nous présentons notre méthodologie qui mettra un accent particulier sur l'usage coordonné des technologies de la télémédecine pour faire face à une situation de crise.

2.3 Méthodologie adoptée pour la gestion de crise à distance

Cette section présente notre approche de gestion de crise à distance à l'aide des actes de télémédecine. Les logiciels et les plateformes de télémédecine ont fait des progrès significatifs ces dernières années, impliquant un fort besoin de standardisation pour faciliter leur interopérabilité. Cette interopérabilité concerne le partage et l'échange de données entre ces systèmes, notamment en cas de crise nécessitant une évacuation (par exemple le service d'aide médicale urgente - SAMU en France) et nécessitant des soins médicaux immédiats (en cas de crise de tsunami, accident industriel, épidémie, etc.). Les données collectées et échangées dans le cadre d'un acte de télémédecine sont les données des dispositifs médicaux (mesures : pression, pouls, température, glycémie ...), le dossier médical du patient, l'imagerie médicale, les rapports et prescriptions médicales issues de la télémédecine, etc. Dans la médecine d'urgence, on peut réaliser certains actes à distance : la télé radiologie, la télé psychiatrie, la télé cardiologie, la télé ophtalmologie, la télé audiologie, la télédétection, la télé dermatologie, etc. Cependant, cette approche peut être déployée dans tout type de crise nécessitant une prise en charge médicale des victimes.

L'objectif de cette méthodologie consiste à décrire un système collaboratif informatisé pertinent à l'aide des applications de télémédecine, spécifiquement dédiée à la gestion de crise et à fournir aux décideurs des informations pertinentes. Les applications de la télémédecine doivent collaborer pour gérer les aspects médicaux des situations de crise. Cette collaboration est basée sur le concept de l'architecture orientée des services (SOA) pour permettre l'interopérabilité des applications médicales (Benaben et al., 2015; Traore et al., 2016; Vescoukis et al., 2012). Chaque acte de télémédecine est considéré comme une application médicale distante et rend des services aux autres et vice versa. Par conséquent, les prestations de services peuvent être améliorées par l'interopérabilité des applications pour offrir des services complexes. En déployant des systèmes d'applications collaboratives assurant le partage de l'information et permettant l'interprétation des services, des applications très spécialisées et complexes pourraient être réalisées pour améliorer les temps de réponse des services d'urgence.

2.3.1 Processus de gestion de crise à distance

Pour la gestion de la crise, nous avons proposé une méthodologie en trois phases à l'aide des applications de télémédecine comme suit : (1) *Définition de la crise* (2) *Analyse de la crise* et (3) *Gestion de la crise* (Figure 2.7).

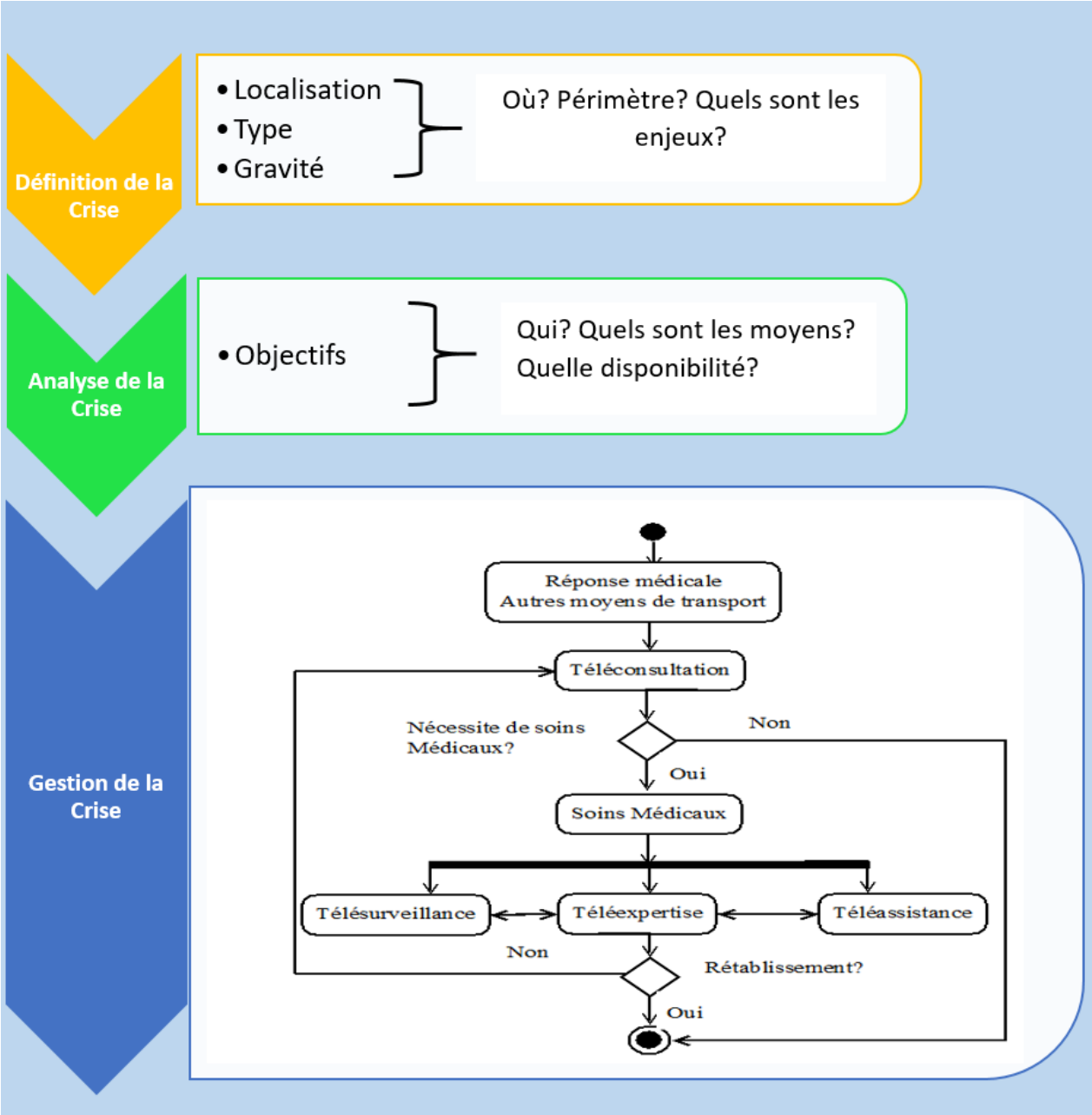


Figure 2.7 – Processus de gestion de crise à distance

Dans la gestion de crise, des décisions appropriées doivent être prises pour minimiser les dommages causés par les risques naturels ou industriels, et aussi pour faciliter le retour à une situation normale avec une meilleure protection des victimes. Pour cela, dans un premier temps, il est nécessaire de collecter les bonnes informations dans le système informatique, en deuxième lieu il est judicieux d'utiliser ces informations pour analyser la crise en question et en dernier lieu il est pertinent de passer à la planification proprement dite de la gestion de la crise.

1. ***Définition de la crise***
   - Localisation : Où ? Quel périmètre ? Ces paramètres définissent et délimitent la zone géographique de la crise afin de préparer un plan d'urgence et réduire significativement les surprises. Par exemple, dans une crise d'inondation, nous pouvons établir la liste des rivières, la liste des zones habitées sous les menaces dans l'application.
   - Type : Les paramètres de catégorisation de la crise aident les décideurs à identifier la nature de la crise, à la qualifier, et à l'insérer dans une catégorie de gestion de crise, et ce à travers les questions pertinentes : De quoi s'agit-il ? Et quel type de crise est-ce ? Par exemple, dans la crise de tsunami, des centres spécialisés de surveillance en temps réel peuvent être sollicités pour plus de détails.
   - Gravité : Quels sont les enjeux ? Ce paramètre permet de déterminer les priorités d'urgence en termes de sauvetage de vies, d'infrastructures et d'économies. Par conséquent, il est nécessaire d'identifier les ressources humaines et matérielles et de planifier diverses activités pour parvenir à une gestion de crise pertinente.
2. ***Analyse de la crise pour l'atteinte des objectifs***
   - Quelles sont les erreurs majeures à éviter ? Il faut renforcer la capacité à identifier les maillons faibles de l'appareil, la flexibilité de la réponse, éviter la négligence ;
   - Quels réseaux d'acteurs ? Ceci s'applique à la fois à la sensibilité des réseaux de supervision et de surveillance, à la réactivité vis-à-vis du processus de génération d'alerte et la capacité de gestion de crise ;
   - Quelles sont les initiatives à mettre en œuvre pour mobiliser les sources d'énergie ? La recherche de solutions alternatives pour la gestion et la rationalisation des ressources pouvant présenter des défaillances techniques et fonctionnelles (problèmes affectant les différents réseaux d'eau, d'électricité, de gaz et de communication) provoquées par la crise. Nous devons nous préparer à la diversification des ressources fondamentales (par exemple l'utilisation des énergies alternatives et renouvelables).

3. ***Gestion de la crise :*** Cette partie est principalement basée sur le processus de prise en charge médicale de toutes les victimes d'une crise en commençant par la phase de transport des victimes vers les structures de santé jusqu'au rétablissement définitif des patients à l'aide de la réalisation des actes de télémédecine:

- Premièrement, la réponse médicale organise la régulation médico-sanitaire des urgences avec le transport des victimes vers une structure de santé dans des conditions appropriées. Les victimes peuvent également prendre d'autres moyens de transport pour se rendre dans les hôpitaux pour des soins critiques ou d'urgence ;
- Deuxièmement, les premiers examens médicaux sont effectués par téléconsultation pendant la phase de transport (par exemple, mesure de la température, questionnaires, observations cliniques, etc.). Après cela, si nécessaire, les soins médicaux peuvent commencer pendant le trajet ;
- En troisième lieu, la victime est placée en télésurveillance ; ceci est particulièrement important surtout en cas de crise d'épidémie ;

- Enfin, en même temps, le personnel médical peut solliciter deux actes de télémédecine à savoir la téléassistance et la téléexpertise. Après les phases de téléassistance ou de téléexpertise, si la victime n'est pas rétablie, une nouvelle téléconsultation peut être sollicitée et ainsi de suite.

Chaque acte a une fonctionnalité particulière et spécifique dans le sauvetage des victimes. Les actes sont utilisés en fonction des besoins contextuels. Par conséquent, pour une situation de crise donnée en fonction des besoins médicaux, les cinq actes seront partiellement ou entièrement déployés. Les actes de télémédecine collaborent pour offrir des services de qualité. Cette collaboration doit être bien organisée pour tirer pleinement profit des avantages des différents actes de télémédecine.

Il doit y avoir un plan à suivre du début à la fin non seulement pour faire face à une situation de crise dès l'apparition des premiers signes, mais aussi pour apporter des soins médicaux nécessaires jusqu'au rétablissement définitif des victimes. C'est dans cette perspective que nous avons proposé une méthodologie (Figure 2.7) pour améliorer la collaboration des actes de télémédecine.

Les avantages de cette méthodologie sont :

- Le sauvetage de la vie des premières victimes y compris les professionnels de santé (en cas de crise d'épidémie) grâce à l'utilisation coordonnée des actes de télémédecine,
- Le gain de temps considérable pendant les prises en charge médicale,
- Le contrôle dans le temps et dans l'espace de la propagation de la crise.

Pour de meilleurs résultats, quatre conditions sont considérées pour évaluer l'exactitude du modèle :

- la nomination d'un coordinateur général pour assurer la cohérence des activités, car la gestion d'une crise implique un grand nombre de partenaires, notamment des structures sanitaires, des services d'urgence, des décideurs politiques, etc. ;
- l'équipement des véhicules de secours et aussi la structure de santé avec les dispositifs télé médicaux appropriés ;
- la formation des manipulateurs des équipements de télémédecine afin de réduire considérablement leur temps de réponse ;
- la disposition d'une connaissance approfondie de la crise en fournissant les bons paramètres pour se doter d'un système d'aide à la décision performant.

L'efficacité d'un système d'alerte est strictement liée à une connaissance approfondie de la crise. Les paramètres à considérer changent d'une crise à une autre, par exemple une crise d'épidémie n'a pas les mêmes paramètres qu'une crise de tsunami. Particulièrement, les recherches visant à préciser les zones à risque de tsunami devraient intégrer les études d'hydrogéologie et de vulnérabilité. De même, les enquêtes sur le site et les études post-tsunami sont essentielles pour une meilleure compréhension des traces historiques et de la dynamique d'évolution de la crise. Du point de vue d'un système d'alerte précoce, la modélisation est également essentielle pour connaître

l'impact d'un tsunami sur la côte. Cependant, ils ne sont efficaces que si la bathymétrie approchant la côte est suffisamment précise pour tenir compte des effets du site.
Pour réagir de manière appropriée en cas de crise, un centre d'alerte efficace doit pouvoir émettre un message 24 sur 24 indiquant l'occurrence d'un séisme susceptible de générer un tsunami (exemple : CENtre d'Alerte aux Tsunamis - CENALT, en France). Le message doit également parvenir aux autorités compétentes responsables de l'organisation de l'aide afin que celles-ci sachent parfaitement ce qu'il y a lieu de faire à tous les niveaux. Compte tenu des temps de réaction très courts et d'un grand nombre d'acteurs impliqués, l'improvisation doit être évitée. La conception des cartes d'inondation et d'évacuation devrait être encouragée, car elles permettent notamment d'évaluer à l'avance l'étendue des dommages et d'identifier les itinéraires pouvant être utilisés pour la livraison des secours. Les principaux ports et les zones côtières très peuplées devraient être couverts en premier. En outre, des exercices de formation et de simulation sont essentiels pour identifier les dysfonctionnements et améliorer la gestion des crises (exemple : simulation de tsunami à Cannes, France).

2.3.2 Modélisation mathématique

La modélisation mathématique concerne la phase de gestion de crise proposée dans la méthodologie. Considérons un ensemble d'actes de télémédecine $(TAct_1, ..., TAct_n)$ qui doivent collaborer pour la prise en charge médicale des victimes. Dans ce processus de modélisation, la victime de la crise correspond au patient médical. Le processus de gestion de crise a été présenté ci-dessus dans la Figure 2.7.

Définissons un domaine $D : (C, P)$ composé d'un ensemble d'actes de télémédecine $C = (TAct_1, ... , TAct_n)$ et un ensemble de prédicat $P= \{p_1, p_2, ... , p_n)$ pour spécifier les propriétés éventuelles des objets (Patients, Professionnels de santé, Documents générés par les logiciels informatiques) et les relations entre eux. Concrètement, nous avons adopté cette codification : TéléConsultation (*TC*), TéléSurveillance (*TS*), TéléAssistance (*TA*), TeleExpertise (*TE*), Réponse médicale *(RM)* et ainsi, en remplaçant $TAct_{1...n}$ par les différents actes de télémédecine, nous obtenons $C = (TC, TM, TA, TE, RM)$ et un ensemble de prédicats définis par $P = (casVulnerable\ P, casConfirme\ P, retabli\ P, decede\ P)$ où *P* est une personne victime. Le prédicat *casVulnerable P* est une propriété d'une personne qui signifie que *P* est plus vulnérable aux maladies (par exemple une épidémie ou infection parasitaire). *P* est un objet au point de vue structure des données.

Un acte de télémédecine est défini par $TAct = (Pentree, Psortie, Pentreesortie, Prec , Effet)$ où:

- $Pentree = \{pentree_1, ..., pentree_n\}$ est l'ensemble des paramètres d'entrée de *TAct*,
- $Psortie = \{psortie_1, ... , psortie_n \}$ est l'ensemble des paramètres de sortie de *TAct*,
- $Pentreesortie = \{ pentreesortie_1, ... , pentreesortie_n \}$ est l'ensemble des paramètres d'entrée-sortie de *TAct*,
- *Prec* est un ensemble de conditions à remplir par les objets de *Pentree*,
- *Effet* est l'ensemble des effets d'exécution de la réalisation de l'acte de télémédecine.

Les actes de *TC* sont formalisés comme suit :

- *Pentree = {MedecinTraitant, Patient, DossierMedicalPatient }*
- *Psortie = {PrescriptionMedicamenteuse , RapportTC , autreTActDemande }*
- *Pentreesortie = {PlateformeTM }*
- *Prec = {casVulnerable Patient, PlateformeTM operationnellle , DossierMedicalPatient disponible }*
- *Effect = {PrescriptionMedicamenteuse genere , indiquerSi autreTActDemande, RapportTC genere }*

Pour réaliser la *TC*, nous avons des *entrées* qui sont : le médecin traitant, le patient et son dossier médical. *Les sorties* sont : la prescription médicale, le rapport complet de *TC* qui peut demander un autre *TAct*. Ensuite, la plateforme TIC est nécessaire pour réaliser tous les actes de télémédecine à l'*entrée* et à la *sortie*. Il y a également quelques conditions préalables à remplir : le patient doit être un cas vulnérable, la plateforme TIC doit être opérationnelle, le dossier médical patient disponible. Enfin, les résultats de *TC* sont : la génération d'ordonnances médicales, des rapports de *TC* et une suggestion si un autre *TAct* est nécessaire.

Les actes de *TS* sont formalisés comme suit :

- *Pentree = {MedecinTraitant, Patient, DonneesSurveillanceMedicale }*
- *Psortie = {InterpretationResultat , RapportTS, autreTActDemande }*
- *Pentreesortie = {PlateformeTM }*
- *Prec = {casConfirme Patient, PlateformeTM operationnelle }*
- *Effect = {InterpretationResultat genere, indiquerSi autreTActDemande, RapportTS genere }*

Pour réaliser *TS*, les *entrées* sont : le médecin traitant, le patient et ses données de surveillance médicale (un indicateur clinique, radiologique et biologique). Aussi, *les sorties* sont : les résultats interprétés, le rapport complet de *TS* et l'indication de la nécessité d'un autre acte de télémédecine *TAct*. Ensuite, la plateforme TIC est nécessaire pour réaliser tous les actes de télémédecine à l'*entrée* et à la *sortie*. Et aussi, nous avons quelques conditions préalables à remplir : le patient doit être un cas confirmé et la plateforme TIC doit être opérationnelle. Enfin, les résultats de *TS* sont : la génération des résultats interprétés des différents examens et analyses médicaux, un rapport complet de *TS* et une suggestion si un autre *TAct* est nécessaire.

Les actes de *TE* sont formalisés comme suit :

- *Pentree = {MedecinRequerant , MedecinSollicite , DossierMedicalPatient }*
- *Psortie = {AvisExpert , RapportTE , autreTActDemande }*
- *Pentreesortie = {PlateformeTM }*
- *Prec = {PlateformeTM opérationnelle, DossierMedicalPatient disponible}*
- *Effet = {AvisExpert genere, indiquerSi autreTActDemande, RapportTE genere }*

Pour réaliser la *TE*, les *entrées* sont : le médecin requérant, les médecins sollicités et le dossier médical du patient. *Les sorties* sont : l'avis des experts, le rapport complet de *TE* qui peut également demander un autre *TAct*. Ensuite, la plateforme TIC est nécessaire pour réaliser tous les actes de télémédecine à l'*entrée* et à la *sortie*. De plus, nous avons certaines conditions préalables à remplir

: la plateforme TIC doit être opérationnelle et le dossier médical du patient doit être disponible. Enfin, les résultats de *TE* sont : la génération des avis des experts, les rapports *TE* et une suggestion si un autre *TAct* est nécessaire.

Les actes de *TA* sont formalisés comme suit :

- *Pentree = {MedecinRequerant, MedecinSollicite, ActeMedical}*
- *Psortie = {RapportTA , RequestAnotherTA }*
- *Pentreesortie = {PlateformeTM }*
- *Prec = {PlateformeTM operationnelle, DossierMedicalPatient disponible }*
- *Effet = {AssistanceExpert, indiquerIf RequestAnotherTA, RapportTA genere}*

Pour réaliser la *TA*, les *entrées* sont : le médecin requérant, les médecins sollicités et l'acte médical à réaliser. *Les sorties* sont : le rapport complet de *TA* qui peut également demander un autre *TAct* . Ensuite, la plateforme TIC est nécessaire pour réaliser tous les actes de télémédecine à l'*entrée* et à la *sortie*. Et nous avons aussi quelques conditions préalables à remplir : la plateforme TIC doit être opérationnelle et le dossier médical du patient doit être disponible. Enfin, les résultats de la *TA* sont: l'assistance des experts, les rapports de *TA* et une suggestion si un autre *TAct* est nécessaire.

2.3.3 Algorithme proposé

L'algorithme proposé est présenté en annexe A (Page 113) ; il permet de dérouler la méthodologie adoptée pour une éventuelle implémentation (Figure 2.7). En ce qui concerne l'implémentation informatique de l'approche, tout acte de télémédecine est considéré comme une application distincte qui fournit des services à d'autres applications. Cette collaboration de logiciels de télémédecine est gérée d'une part, par les techniques d'interopérabilité pour permettre un échange et un partage efficace des données de santé tout en assurant leur confidentialité, et d'autre part, par l'interopérabilité sémantique afin de traiter et de comprendre la signification des données basées sur des terminologies médicales communes (ASIP, 2012). L'interopérabilité des logiciels de télémédecine sera mise en œuvre par des architectures orientées service (SOA) basées sur le langage XML (Extensible Markup Language) et MDA (Model Driven Architecture ou Architecture dirigée par les modèles) (Benaben et al., 2015 ; Traore et al., 2016) ou d'autres technologies existantes telles que le cloud computing ou informatique en nuage (Church et al., 2015). Dans le cloud computing, tout est considéré comme un service (logiciels, donnés et matériels) et les principaux services proposés en cloud computing sont le SaaS (Software as a Service ou logiciel en tant que service), le PaaS (Plateforme as a Service ou Plateforme en tant que service) et le IaaS (Infrastructure as a Service ou Infrastructure en tant que service) (Liu et al., 2017).

2.4 Étude de cas

En octobre 2017, un exercice de simulation d'un tsunami s'est déroulé à Cannes, ce qui est une première en France. L'étude de cas concerne la gestion de cette simulation qui s'inscrit dans le cadre de modélisation des informations. La commune n'a pas communiqué sur les détails concernant les données manipulées sur la simulation de l'exercice notamment en termes de chiffres des unités déployées. De ce fait, nous nous sommes principalement basés sur les articles de Binacchi et de Lioult, publiés sur la simulation en question, tout en faisant allusion aux données publiées dans ces deux articles (Binacchi, 2017; Lioult, 2017). Certes dans un premier temps, on procède dans l'urgence à l'évacuation des zones littorales, mais il peut y avoir soit des refus ou des

cas isolés. De plus, la crise peut s'étendre au-delà de la zone prévue entraînant ainsi une réelle crise qui nécessite de secourir les victimes. Le déploiement des différents actes de télémédecine est bénéfique pour secourir efficacement les victimes nécessitant des soins médicaux.

Les Tsunamis d'origine géologique sont causés par la pénétration ou de la disparition dans le fond marin d'une quantité significative de matériaux géologiques, ce qui entraîne le déplacement significatif d'une grande masse d'eau déferlante et potentiellement destructrice sur les rivages terrestres. Trois types d'événements sont susceptibles de provoquer un tsunami : glissement de terrain sous-marin, un séisme et les explosions volcaniques (Ando et al., 2018; Röbke and Vött, 2017). L'exercice tsunami à Cannes était strictement réservé aux professionnels qui doivent intervenir dans une cellule de crise qu'est le Centre opérationnel de gestion interministérielle des crises (COGIC). Le Centre d'Alerte aux Tsunamis (CENALT) envoie au COGIC des informations selon lesquelles un séisme de magnitude 7,2 sur la plaque algérienne est susceptible de provoquer un tsunami sur les côtes méditerranéennes, son heure d'arrivée est estimée entre 95 à 100 minutes. Les caractéristiques "tsunamigènes" sont vérifiées, la vague et sa vitesse sont calculées ainsi que le temps de l'arrivée de ces vagues sur les côtes françaises comme indiqué dans le Tableau 2.2.

Tableau 2.2 – Les caractéristiques du tsunami

Caractéristiques du tsunami	Magnitude ou volume	Amplitude maximale des ondes sur la rive	Heure d'arrivée	Zone côtière française concernée
Tremblement de terre de la marge du Nord algérien	M = 7,2	4m à St-Tropez, Cannes 3m à La Ciotat, Nice, Villefranche	95 min à 100 min	De Marseille à Menton (en passant par Cannes)

Il est important de considérer les paramètres relatifs à la hauteur et à la profondeur des inondations, mais aussi la durée et le nombre d'onde tsunamigène qui se propage sur terre ainsi que la direction et la vitesse des courants. En outre, il est également nécessaire d'intégrer d'autres paramètres (bâtiments et infrastructures) le long de la zone côtière, ce dernier peut influencer la direction de la propagation de l'eau. En effet, pour une gestion dynamique de la crise du tsunami, ces paramètres influent sur le transport des victimes, la détermination des zones d'évacuation et l'approvisionnement en eau, nourriture, médicaments, couvertures et produits de première nécessité.

Ces paramètres sont transmis immédiatement par le COGIC aux autorités nationales et départementales (Préfecture). En conséquence, la préfecture des Alpes-Maritimes relaye l'information au maire de Cannes environ une heure avant l'arrivée des premières grandes vagues. Le conseil municipal dirige la gestion de l'événement au sein de la cellule de crise avec le soutien du CENALT et du COGIC pour atteindre les objectifs ci-dessous :

1. l'évacuation des ports, des plages et le littoral avec des mesures de protection idoines ;
2. la gestion de crise pendant la durée du tsunami ;

3. le déploiement des secours pour atténuer les dommages collatéraux après le tsunami.

Pour atteindre ces objectifs en appliquant la méthodologie proposée de gestion de crise, nous obtenons les étapes suivantes :

Étape 1 : Définition de la crise

Lieu ← Cannes ;
Type ← tremblement de terre ;
Gravité ← magnitude 7,2 ;

Étape 2 : Analyse de la crise // consiste à fournir des informations sur l'analyse de la crise

Objectifs ← Sauver des vies ;
Ressources_Disponibles ← COGIC (Centre opérationnel de gestion interministérielle des crises), CENALT (Centre d'alerte aux tsunamis), CYPRES (Centre d'information pour la prévention des risques majeurs) ;
Limites ← ressources matérielles et humaines, temps ;

Étape 3 : Gestion de la crise

1. Coordination des activités des différentes ressources : COGIC, CENALT, CYPRES
   1.1. Organisation d'un poste de commandement communal,
   1.2. Rédaction de messages d'alerte,
   1.3. Mise à jour d'un dossier d'information commun sur les risques majeurs,

2. Calcul de la hauteur de la vague, sa vitesse et son heure d'arrivée sur la côte.

3. Evacuation des ports et du littoral, et l'organisation de secours en utilisant des actes de télémédecine :
   3.1. Pour chaque patient pendant les phases de transport, réaliser les actes suivants :

Téléconsultation (Patient)
Télésurveillance (Patient)
Téléassistance (Patient)
Téléexpertise (Patient)

} Ces actes sont partiellement ou totalement utilisés en fonction des besoins contextuels, ce qui permet de gagner du temps.

Cet exercice de simulation de la gestion de la crise de tsunami à Cannes nous a permis de valider notre méthodologie en trois étapes. L'utilisation coordonnée de différents actes de télémédecine est très importante lors de la phase de transport vers les structures de santé, car elle permet d'améliorer le processus d'évacuation avec des résultats intéressants notamment pour la réduction du temps d'intervention et la diminution des risques de complication médicale. Ces améliorations contribuent fortement à diminuer le temps de déploiement des secours et à sauver plus de victimes. Les différentes interactions entre les différents acteurs de la gestion de crise sont représentées par le diagramme de séquence UML (Figure 2.8).

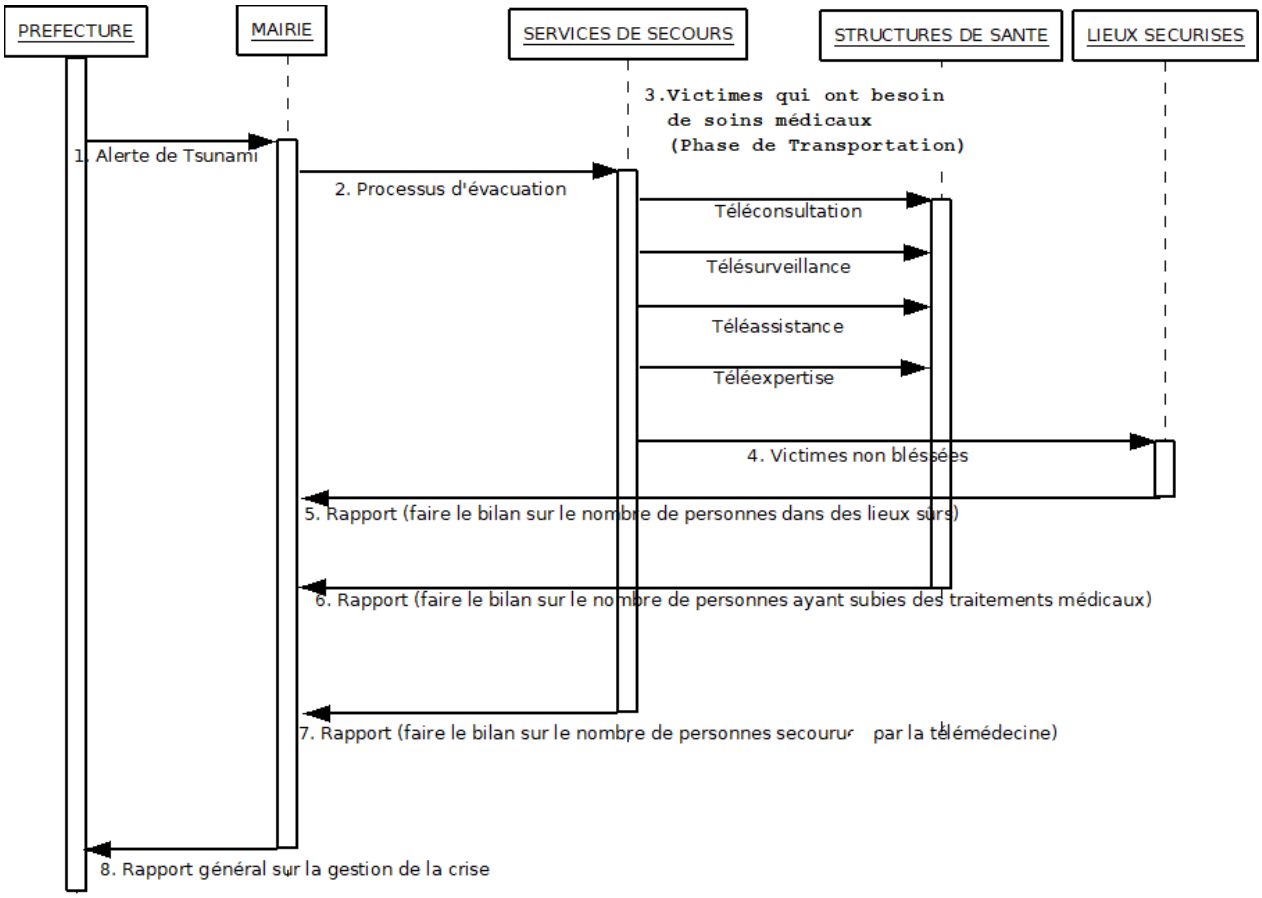


Figure 2.8 – Gestion à distance de l'exercice de simulation de Tsunami à Cannes

A travers ce schéma, nous pouvons observer une gestion collaborative d'une crise de tsunami qui peut se produire à Cannes, en France (une ville située sur la Côte d'Azur). Dans un premier temps, le CENALT envoie une alerte de tsunami au COGIC qui informe la Préfecture avec les paramètres pertinents du Tsunami, à savoir l'amplitude, la hauteur de la vague, sa vitesse et son heure d'arrivée sur la côte. Ensuite, la Préfecture avertit le maire une heure avant la première grande vague du tsunami. Après cela, le service de sauvetage des ports évacue les plages, et l'ensemble du littoral. Au cours de la phase de transport, les services de secours réalisent des activités de télémédecine dans les moyens de transport agréés. Ces actes de télémédecine sont la téléconsultation, la télésurveillance, la téléassistance, la téléexpertise et ils sont partiellement ou totalement réalisés en fonction du contexte afin de fournir des soins médicaux optimaux. Les personnes non blessées sont transportées dans des lieux sûrs (par exemple les établissements d'enseignement, les salles polyvalentes, gymnases, hôtels, chalets ...). Les différents rapports des services de sauvetage, des structures de santé et des endroits sûrs seront retournés à la mairie afin d'évaluer la gestion de la crise du tsunami. Les leçons apprises, les forces et les faiblesses seront identifiées par retour d'expérience pour constituer des pistes d'amélioration pour la prévention, la sensibilisation au risque de tsunami et la gestion des futures crises.

2.5 Résultat et discussion

Au cours d'une situation de crise, la gestion du temps et la connaissance de la situation sont des caractéristiques importantes afin de réduire les effets de la crise. La gestion d'une crise nécessite une gestion organisée des activités. Il existe plusieurs approches existantes ; mais très peu d'entre elles intègrent une collaboration des logiciels de télémédecine. Notre méthodologie considère que les actes de télémédecine dont les services sont accessibles à distance doivent collaborer pour faire face à une situation de crise dans le but de sauver plus de vies à travers des opportunités pour les soins intensifs de la télémédecine.

La méthodologie proposée (Figure 2.7) repose sur une formalisation mathématique de la gestion de crise. Cette formalisation concerne tous les actes de télémédecine en spécifiant les paramètres d'entrée, de sortie, d'entrée-sortie, les conditions préalables à remplir avant d'effectuer un acte de télémédecine et les résultats éventuels après chaque acte de télémédecine. A cet effet, un algorithme générique a été présenté pour permettre une implémentation concrète d'une plateforme informatique pour la gestion de crise. Cette plateforme informatique favorisera l'échange, le partage et le traitement d'informations pour la collaboration à distance efficace des professionnels de la santé pendant la phase de transport et les soins médicaux des patients.

L'objectif principal était de fournir un modèle amélioré de système informatique pour le travail collaboratif notamment pour la gestion de crise à distance à l'aide des actes de télémédecine. Cette méthode donne de meilleurs résultats notamment en termes d'organisation de travail. Plus il y a de vies sauvées par la collaboration des actes de télémédecine, meilleur est le résultat. Par conséquent, la méthode utilisée pour évaluer l'exactitude du modèle pourrait être le comptage du nombre de victimes sauvées par la pratique de la télémédecine. Le succès du modèle est strictement lié à l'interopérabilité des logiciels de télémédecine, à l'équipement adéquat des structures sanitaires, la disponibilité des véhicules de secours et à la formation des manipulateurs du matériel de télémédecine. Cependant, si le déploiement de cette méthodologie est un atout dans la gestion de crise, nous devons être conscients de ses limites, à savoir : l'équipement des structures de santé et l'acquisition des véhicules de secours nécessitent des coûts considérables et aussi bien la formation de manipulateurs du matériel de télémédecine.

Par ailleurs, nous identifions l'utilisation des données météorologiques, climatiques, environnementales pour la réalisation de systèmes de décision sur une échelle de temps d'une minute, une heure ou même un siècle, pour aider les décideurs dans les prises de décision dans l'acquisition d'information et les campagnes de sensibilisation du public (Brunet et al., 2009). Il y a aussi dans la gestion de crises, les systèmes utilisant les satellites pour fournir des informations d'alerte de détresse et de localisation aux services de recherche et de sauvetage du monde entier pour les utilisateurs maritimes, aériens et terrestres en détresse (Balogh et al., 2010), mais de tels systèmes manquent de collaborations réelles entre les différents acteurs. Ceci explique le développement de certains modèles de simulation appliqués aux services médicaux d'urgence (EMS) lors de l'apparition d'une situation de crise jusqu'à l'acheminement du patient à l'hôpital. L'EMS est intéressant pour une bonne coordination des équipes de secours avec la définition des décisions à court, moyen et long terme afin de réduire le temps de réponse (Aboueljinane et al., 2013). L'EMS est une partie intégrante de notre approche principalement dans la phase de la régulation médicale des urgences. Cependant, elle se limite uniquement au bon acheminement des victimes dans les établissements de santé les plus proches.

2.6 Conclusion

Nous avons présenté dans ce chapitre une nouvelle méthodologie de la gestion de crise à l'aide des actes de télémédecine ceci dans le but de fournir aux gestionnaires de crise un système d'aide à la décision informatisé pertinent. Chaque acte de télémédecine (Téléconsultation, Téléexpertise, Télésurveillance et Téléassistance) est représenté par son application informatique correspondante qui fournit des services à d'autres et vice versa. L'application de cette méthodologie à la simulation d'un exercice de tsunami à Cannes (France) a permis non seulement de valider la méthodologie proposée, mais aussi de déceler ses limites. Les principaux résultats de cette étude sont: (i) une approche méthodologique pour une meilleure protection des villes contre les grandes vagues généralement provoquées par des tsunamis (ii) des systèmes d'information avec des composants interopérables pour une meilleure prise en charge médicale des victimes (iii) un cadre formel de gestion des crises pour capitaliser les expériences et exploiter les leçons apprises.

Après avoir fourni le cadre conceptuel de gestion de crise, il serait plus souhaitable de surveiller et de prévenir l'avènement d'une crise qui occasionnera forcément des conséquences négatives sur l'environnement et sur les humains. Cette prévention peut se faire à travers la mise à disposition des gestionnaires de crise des systèmes d'alerte précoce, des cartes de risque, etc. C'est pourquoi il est nécessaire de s'intéresser de plus près à l'étude des paramètres (environnementaux, climatiques, spatiaux, démographiques, etc.) impliqués dans l'avènement d'une crise. Ainsi, le chapitre suivant sera consacré à l'étude de la prévention d'une crise d'épidémie à l'aide des techniques de data mining sur les images satellitaires. L'objectif final de cette étude est de fournir aux acteurs de la santé publique des outils adéquats leur permettant de surveiller et de prévoir les crises ayant pour origine le risque naturel des épidémies.

Chapitre 3

Les techniques de Data Mining sur les images satellitaires pour découvrir les zones à risque d'épidémie

3.1 Introduction 59
3.2 Contexte et travaux connexes 59
3.2.1 Contexte 59
3.2.2 Travaux connexes 61
3.3 Méthodologie de découverte des zones à risque d'épidémie 63
3.3.1 Collecte des données 64
3.3.1.1 Collecte des informations sur le terrain à propos de l'épidémie 64
3.3.1.2 Collecte des images satellitaires 64
3.3.2 Sélection et transformation des données (Données dérivées d'images) 66
3.3.3 Mesures de télédétection 68
3.3.4 Discrétisation des données 68
3.3.5 Traitement des données 69
3.3.6 Interprétation des résultats 71
3.4 Étude de cas : Épidémie de choléra à Mopti (Mali) 71
3.4.1 Collecte des données 71
3.4.1.1 Collecte des informations sur le terrain à propos de l'épidémie de choléra 71
3.4.1.2 Collecte des images satellitaires de la zone d'étude 73
3.4.2 Sélection et transformation des données 74
3.4.3 Mesures de télédétection 74
3.4.4 Discrétisation des données 75
3.4.5 Traitement des images : classification 76
3.4.6 Interprétation des résultats 78
3.5 Discussions 80
3.6 Conclusion 81

3. *Les techniques de Data Mining sur les images satellitaires pour découvrir les zones à risque d'épidémie*

3.1 Introduction

Dans ce troisième chapitre, nous allons présenter notre contribution pour la prévention d'une crise d'épidémie par les techniques d'extraction de connaissances à partir des données (ECD) notamment le data mining sur les images satellitaires pour la découverte des zones à risques d'épidémie. Pour ce faire, ce chapitre se scinde en quatre parties. La première partie sera dédiée à la présentation du contexte de l'étude et des travaux connexes pour dégager notre positionnement. Ensuite, nous présenterons dans la deuxième partie notre méthodologie adoptée pour prévenir une crise d'épidémie à l'aide des techniques de data mining sur les images satellitaires en six phases. Puis, la troisième partie sera consacrée à l'application de la méthodologie à travers un cas d'étude qui a porté sur l'épidémie de choléra dans la région de Mopti, au Mali. Et enfin, dans la dernière partie, nous clôturerons ce chapitre par une discussion et une conclusion.

Ces dernières années, les informations issues des satellites d'observation de la Terre et de la météorologie, associées aux systèmes d'information géographiques et aux systèmes de positionnement par satellites (GNSS pour *Global Navigation Satellite System*), ont été utilisées pour étudier l'épidémiologie des maladies. C'est ainsi qu'a émergé la téléépidémiologie comme un nouvel axe de recherche combinant l'épidémiologie et la télédétection pour une meilleure utilisation des analyses écologiques, environnementales, et autres facteurs contribuant à la propagation des maladies. La téléépidémiologie associe des données d'observation satellitaire (géographique, climatologique, environnementale, foncière, etc.) à des ensembles de données pluridisciplinaires (données de santé, température ou humidité, données socio-économiques, données biologiques, etc.) (Marechal et al., 2008). Cela permet d'analyser les mécanismes d'émergence, de propagation et de transmission des maladies épidémiques. De plus, l'étude des relations tripartites Climat - Environnement - Santé, et l'étude des liens entre les maladies épidémiques et leur environnement d'évolution deviennent nécessaires surtout à l'aide des technologies de télédétection (Lacaux et al., 2007). A cet effet, des produits statistiques et non statistiques ont été développés pour faciliter le travail des épidémiologistes : quantification des cas de maladie, identification des facteurs caractéristiques des zones affectées, identification des zones à risque élevé, développement de modèles de prévision, gestion des programmes de prévention, etc.

3.2 Contexte et travaux connexes

3.2.1 Contexte

Les maladies infectieuses gagnent chaque année du terrain, ce qui augmente le risque d'épidémies. Elles ont en commun des petits organismes, souvent des arthropodes (virus, moustiques, mouches, tiques, poux ou puces), qu'on appelle des vecteurs parce qu'ils transmettent l'agent infectieux d'un hôte à un autre. La zone de distribution de ces vecteurs s'étend progressivement ; également le changement climatique ouvre certaines possibilités de réapparition de la maladie dans des

populations précédemment protégées. Parfois, de nouvelles régions vulnérables sont ainsi gravement touchées. C'est le cas de l'épidémie de Choléra qui reste un problème majeur de santé publique dans les pays en voie de développement. Plus de cinquante pays dans le monde souffrent du choléra soit 200 000 cas avec un taux de létalité de 4% (Piarroux, 2002). A ce jour, il n'y a pas de vaccin efficace disponible dans les programmes de vaccination de l'OMS. Le Mali situé au cœur de l'Afrique de l'Ouest n'échappe pas à la règle. Selon une étude faite par Dao et ses collègues, le facteur géographique joue un rôle majeur dans cette épidémie, car dans 78% des cas les victimes se baignaient dans les lacs, les rivières ou les fleuves, en particulier pendant les périodes d'hivernage (Dao et al., 2009). En effet, au Mali, les localités touchées par l'épidémie de choléra sont situées le long du fleuve Niger qui traverse le pays du Sud au Nord.

Pour surmonter les limites des techniques traditionnelles de lutte contre les épidémies, il est nécessaire de profiter des avantages des technologies spatiales pour lutter et surveiller les maladies infectieuses. Les avantages potentiels de l'usage des technologies spatiales sont nombreux en termes :

- D'économie : les satellites couvrent de grandes surfaces, ce qui évite le déplacement des équipes de terrain ;
- De sécurité : les satellites couvrent des zones dangereuses ou inaccessibles ;
- De dynamique : décryptage de la dynamique spatio-temporelle des maladies ;
- De suivi : mise à disposition de cartes de risques et d'une plateforme de surveillance épidémiologique pour un meilleur suivi et une anticipation des épidémies.

Le but de ce travail est de démontrer le lien entre le facteur géographique, le fleuve Niger et la propagation de l'épidémie de choléra dans les localités affectées par des images satellitaires et des données de terrain à l'aide des techniques de data mining. Ainsi, ce travail contribue à renforcer la politique de surveillance épidémiologique d'une zone particulière du Mali et par la suite peut s'étendre sur toute l'étendue du territoire national.

Les techniques de data mining sur les données satellitaires sont une opportunité pour le secteur de la santé, particulièrement en épidémiologie. Le data mining ou la fouille des données est un processus de découverte automatique d'informations dans de grandes bases de données (Tan et al., 2006). Le data mining permet d'extraire des modèles à partir des données pour générer des connaissances intéressantes (Fayyad et al., 1996), telles que des zones à risque (à travers des cartes de risque) à partir d'énormes bases de données issues du traitement des images satellitaires. Ces modèles peuvent être considérés comme une description des données, et seront utilisés dans une analyse plus approfondie pour prédire de nouvelles données, par exemple étendre la production des cartes de risque d'épidémie sur l'ensemble du territoire national cible.

Les trois principales phases du data mining sont le prétraitement, le traitement et le post-traitement. Le gros du travail de data mining se situe dans le post et le prétraitement des données soit environ 80% et seulement 20% est réservé à l'application des techniques de traitement du data mining (application d'algorithmes) (Qin et al., 2009; Zhang et al., 2003). La plupart des algorithmes de data mining ne sait pas manipuler les données numériques, mais plutôt les données catégoriques. Pour résoudre ce problème, il est nécessaire de discrétiser les données numériques. La discrétisation consiste à convertir des valeurs numériques en valeurs catégoriques et elle est positionnée dans la phase de prétraitement des données. La discrétisation fait partie des problèmes actuels du data mining, car elle a beaucoup d'impacts sur les résultats et peut conduire facilement à des interprétations erronées.

La télédétection est une méthode d'observation de la surface terrestre ou de l'atmosphère depuis l'espace en utilisant des systèmes satellitaires qui enregistrent des images pouvant être interprétées afin de produire des informations utiles. La télédétection est utilisée dans la cartographie des territoires (écosystèmes terrestres, marins et côtiers) et elle a de nombreuses applications notamment dans l'occupation du sol, l'agriculture, l'observation militaire, la déforestation, la croissance urbaine, etc. De nos jours, la télédétection a donné naissance à de nouveaux axes de recherche notamment l'aspect spatio-temporel des épidémies pour identifier, prévoir et étudier les zones à risque d'épidémie. Les détails sur la télédétection sont donnés dans le premier chapitre de la thèse.

De nombreux satellites sont positionnés autour du globe. Certains sont géostationnaires c'est-à-dire ne couvrent que la même zone géographique, d'autres tournent autour de la terre pour couvrir le globe terrestre. Ils envoient tous un flux continu de données à la surface de la Terre, précisément vers les stations de réception. Ce qui fait qu'aujourd'hui, il y a une énorme quantité de données satellitaires stockées dans les bases de données. Ces données renferment de la connaissance et doivent être mises à la disposition des décideurs dans plusieurs domaines d'activité tels que la santé, l'agriculture, la planification, etc. Pour découvrir ces connaissances cachées, nous nous sommes intéressés aux techniques de data mining.

Les données satellitaires d'observation de la terre utilisées dans ce travail proviennent du programme Landsat, car d'une part c'est la plus ancienne entreprise dans l'acquisition d'images satellitaires de la terre et d'autre part les images Landsat sont libres (gratuites) en téléchargement et sont de bonne qualité. Le 23 juillet 1972, le premier satellite du programme Landsat nommé Earth Resources Technology a été lancé et Landsat 8 a été lancé le 11 février 2013. Les satellites Landsat ont acquis des millions d'images et elles peuvent être consultées et téléchargées sur le site Web de l'US Geological Survey (USGS: Earth Explorer, Landsat Look Viewer et GloVis). Par la suite, les techniques de data mining seront appliquées sur les images satellitaires pour identifier les zones à risque épidémique.

3.2.2 Travaux connexes

La télédétection en épidémiologie ou la téléépidémiologie est une perspective intéressante pour la cartographie et la modélisation des facteurs de risque d'épidémies (Emmanuel et al., 2011). La téléépidémiologie est définie par Bicout et ses collègues comme l'usage de l'ensemble des technologies spatiales au service de la lutte contre les épidémies transmises par l'eau, l'air ou des vecteurs en mettant en relation le climat, l'environnement et la santé (Bicout et al., 2015). Les données de télédétection sont utilisées pour comprendre l'épidémie de choléra notamment les facteurs environnementaux favorables à l'apparition et à la propagation de la maladie tels que les eaux stagnées, les fleuves, la forêt, le déplacement des troupeaux, etc. Cependant, dans certaines situations une seule scène satellite (image satellitaire ou sa partie considérée, traitée et analysée comme un élément distinct) n'est pas suffisante pour couvrir la zone d'étude, par exemple dans le cas d'une crise d'inondation ou d'incendie. Dans ce cas, on peut utiliser plusieurs scènes en vue de mieux couvrir la zone d'étude. C'est pourquoi (Liu and Hodgson, 2016) ont proposé un modèle d'optimisation spatiale pour résoudre le problème d'acquisition d'images satellitaires à grande échelle dans le contexte d'intervention d'urgence.

A partir d'une étude de recherche bibliographie menée par (Mabaso and Ndlovu, 2012), il ressort que l'un des défis majeurs concernant la gestion des crises d'épidémie est le développement de systèmes d'alerte précoce opérationnels prenant en compte l'aspect climatique et environnemental. A cet effet, plusieurs applications ont été développées ; (Kansakar and Hossain, 2016) ont présenté une revue générale des applications basées sur l'imagerie satellitaire dans de nombreux domaines tels que la cartographie du sol, l'évaluation de la biomasse carbone, la sécurité alimentaire, la santé, la gestion des catastrophes, la gestion des océans et la qualité de l'air. Dans le domaine de la santé humaine (Liu et al., 2015) ont développé une application épidémiologique basée sur le traitement des données spatiales, appelée EASTWeb. Cette application est relative aux maladies transmises par les moustiques (par exemple le Paludisme, la Fièvre de la Vallée du Rift). Les facteurs environnementaux concernés sont la température, la précipitation, l'évapotranspiration, l'indice spectral. Cependant, elle ne prend pas en compte les images satellitaires, mais elle est plus focalisée sur les données de terrain issues des facteurs environnementaux précédemment cités.

Dans le but de contrôler, prévenir et éradiquer les vecteurs arthropodes, en fonction de facteurs environnementaux (Jamison et al., 2015) utilisent la recherche multidisciplinaire intégrant le climat, les systèmes d'information géographique, la télédétection et l'écologie. Egalement au Kenya, des données de télédétection et des facteurs écologiques de la vallée du Rift ont été associés pour produire la carte des risques de la fièvre de la vallée du Rift (RVF) (Mosomtai et al., 2016). Pour cette étude, les principales variables écologiques utilisées étaient : les mesures des flux de sols humides, l'indice de végétation par différence normalisée (Normalized Difference Vegetation Index – NDVI) et l'évapotranspiration (ET). Ainsi, la carte finale obtenue des zones à risques ressemble davantage à une application de Système d'Information Géographique (SIG), car ils ont superposé les trois couches représentant les trois variables utilisées (c'est à dire les trois mesures précédemment effectuées). Halimi et ses collègues ont développé une application SIG à partir de la forte relation entre les flambées de climat et de fasciolase. Ils produisent une carte des risques d'épidémie de fasciolose en Iran en utilisant l'indice de risque de fasciolose d'Ollerenshaw. Ainsi, ils ont trouvé quatre classes de risque d'épidémie de fasciolase (Halimi et al., 2015). Cette étude est basée sur les relations entre le facteur climatique et l'épidémie de fasciolose, mais ne tient pas compte du processus d'imagerie satellitaire par télédétection. Toujours dans le domaine de la téléépidémiologie (Xu et al., 2014) établissent l'impact de la température sur la pneumonie infantile à Brisbane en Australie en utilisant des données de télédétection combinées aux données des visites aux urgences.

Pour prédire les épidémies de choléra dans la région du delta du Bengale en utilisant la télédétection par satellite (Jutla et al., 2012) ont étudié la variabilité de l'espace-temps de la chlorophylle. La bactérie Vibrio choléra ne peut pas être mesurée directement depuis l'espace. Le facteur choléra considéré est la chlorophylle, qui est estimée à partir d'images satellites en mesurant la partie pigmentaire (chlorophylle) existant dans le plancton. Cette approche est limitée, car l'observation de la partie pigmentaire dans tout le plancton ne signifie pas à 100% la présence de chlorophylle. Il convient de préciser la période et le pourcentage de chlorophylle impliqués dans tous les cas d'épidémie de choléra dans la région. Il serait intéressant d'intégrer les données de terrain pour améliorer la précision et l'exactitude des zones à risque d'épidémie d'arthropodes. Ce qui est certain, c'est que le choléra est une infection diarrhéique aiguë due à la consommation d'eau ou d'aliments contaminés par la bactérie Vibrio choléra. Dans la plupart des pays concernés, le choléra est une épidémie saisonnière généralement pendant la saison des pluies, ce qui est également le cas du Mali.

Par ailleurs, l'accessibilité d'énormes quantités de données de santé (données non structurées et structurées) présente des opportunités et des défis considérables pour le suivi en temps réel des maladies, la prédiction des épidémies, la détection des pathogènes et des causes de leur émergence (Asokan and Asokan, 2015). Les ressources humaines de la santé ont beaucoup à gagner et à comprendre du déploiement de l'analyse géospatiale et de l'exploitation des sources de données spatiales généralement disponibles pour la modélisation épidémiologique ainsi que le suivi sanitaire de manière durable. Il y a un fort besoin de nouveaux moyens pour l'analyse, et la modélisation de données interdisciplinaires. Par exemple, afin d'améliorer la précision de la prédiction de l'identification des facteurs liés aux glissements de terrain à l'aide d'un système d'information géographique, des modèles de data mining peuvent être utilisés conjointement avec d'autres modèles (par exemple les réseaux de neurones artificiels et la logique de floue) (Song et al., 2012). Dans de telles circonstances, les techniques de data mining peuvent être utilisées pour examiner les caractéristiques topographiques et végétatives pour la vérification des glissements de terrain impliqués dans des conditions climatiques spécifiques (par exemple en cas de fortes pluies) (Tsai et al., 2013). Indépendamment de la technique utilisée, il existe des problèmes de recherche ouverts dans les techniques de data mining spatiales telles que les tests de validité, la sélection des caractéristiques pertinentes, l'interprétation des modèles et le traitement des données multidimensionnelles (Arentze, 2009).

Finalement, il ressort de cet état de l'art que d'une part les relations tripartites « Environnement-Climat-Santé » sont au cœur de la découverte des zones à risque des épidémies et d'autre part de la nécessité d'explorer en profondeur les données de terrain et satellitaires disponibles pour en découvrir les connaissances cachées notamment par les techniques de data mining. Ainsi, la section suivante sera consacrée à l'exploration de cette piste pour surveiller et prévenir les crises d'épidémie à l'aide des techniques de data mining sur les images satellitaires.

3.3 Méthodologie de découverte des zones à risque d'épidémie

L'objectif final de cette méthodologie consiste à découvrir des zones à risque des épidémies. C'est dans ce sens que nous avons établi une méthodologie qui combine à la fois les données épidémiologiques collectées sur le terrain et les données issues du traitement des images satellitaires. La méthodologie se scinde en six phases décrites par la figure ci-dessous (Figure 3.1): (1) la première phase commence par la collecte des données du terrain (les données épidémiologiques et les images satellitaires) (2) la seconde phase concerne la sélection et la transformation des données (3) la troisième phase concerne le choix des mesures de télédétection (par exemple NDVI, application des techniques de correction des images satellitaires) et la collecte du nombre de cas confirmés de l'épidémie sur le terrain (4) la quatrième phase concerne la préparation des données par des techniques de discrétisation (5) la cinquième phase consiste à l'application des techniques de data mining et enfin (6) la sixième phase est dédiée à l'interprétation des résultats obtenus pour la production de cartes des zones à risque d'épidémie.

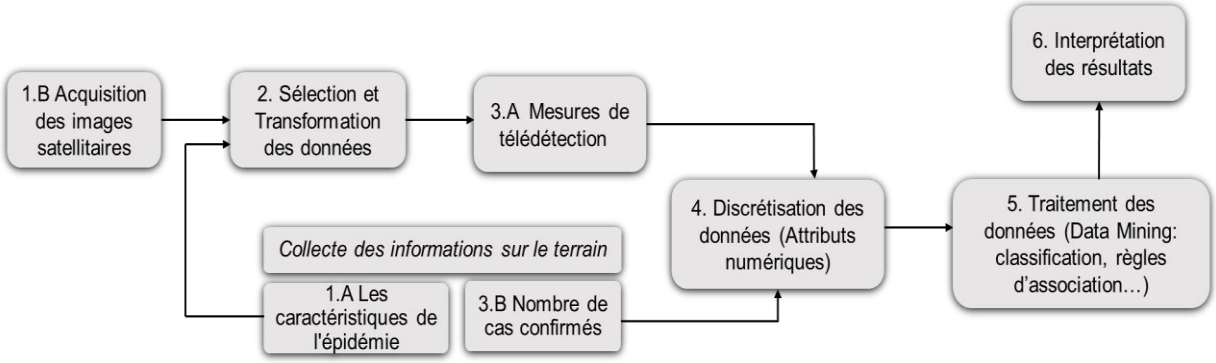


Figure 3.1 – Processus de découverte des zones à risque d'épidémie

3.3.1 Collecte des données

3.3.1.1 Collecte des informations sur le terrain à propos de l'épidémie

L'acquisition d'informations sur le terrain se fait auprès des structures de santé et des spécialistes de terrain.

Il y a deux types d'information à savoir :

- *Nombre de cas confirmés (données épidémiologiques du terrain) :* il s'agit d'obtenir le nombre de cas confirmés dans des bases de données des structures de santé (données de santé) d'un endroit à un autre et d'une période à une autre. Ces données permettent de voir l'évolution de l'épidémie dans différents lieux et à différentes périodes.
- *Les caractéristiques de l'épidémie :* les maladies épidémiologiques telles que l'épidémie de la maladie à virus Ebola, le paludisme, le choléra, la maladie de Lyme ... elles ont toutes leurs caractéristiques particulières, mais elles sont toutes transmises par des vecteurs. Le vecteur (oiseaux, animaux, tiques, moustiques, mouches, vers…) est l'agent porteur des virus et des bactéries provenant d'une source de risque épidémique capable d'infecter la population. Ainsi, nous devons déterminer les liens entre les vecteurs et les environnements favorables à leur émergence ; ces environnements sont en constante expansion en raison du changement climatique, des activités humaines nuisibles (par exemple, des pratiques inappropriées d'utilisation des terres). Les environnements favorables aux épidémies (eau, végétation, etc.) doivent être détectables par les satellites.

Il est important d'analyser l'émergence, la propagation et la transmission des maladies infectieuses et d'établir les relations qui existent entre les maladies infectieuses et l'environnement, tout en utilisant la technologie spatiale. La prochaine étape est l'acquisition de données satellitaires.

3.3.1.2 Collecte des images satellitaires

De nombreux satellites sont positionnés autour du globe : certains sont géostationnaires au-dessus d'une région, d'autres sont en orbite autour de la Terre, mais tous envoient un flux continu de données à la surface: Landsat (Figure 3.2), SPOT, Ikonos, ASTER , AVIRIS, AVHRR, Quickbird, RadarSat, TMS, DTED ou WorldView ...



Figure 3.2 – Capteur satellite LANDSAT 8 (15 m)

Les images de certains satellites sont libres et gratuites, c'est le cas de Landsat (Figure 3.2), qui a l'avantage d'être disponible gratuitement par l'Institut d'études géologiques des États-Unis (United States Geological Survey – USGS). Une fois les images téléchargées sur le site web de l'USGS (http://earthexplorer.usgs.gov/), les images sont fournies sous la forme de plusieurs fichiers tiff, chaque fichier correspond à une bande (ou canal). Une image Landsat 7 est constituée de 8 bandes tandis qu'une image Landsat 8 est composée de 11 bandes. Un exemple d'image Lansat 8 est donné dans le Tableau 3.1.

Tableau 3.1 – Bandes spectrales OLI et TIRS (LANDSAT 8)

Bande	**Longueur d'onde**	**Résolution**	**Description**
Bande 1 - Coastal / Aérosol	0.43 - 0.45	30 m	Côtières et aérosols études
Band 2 - Bleu	0,45 - 0,51	30 m	Cartographie bathymétrique, distinguant les sols de la végétation et des feuillus de la végétation résineuse
Band 3 - Vert	0,53 - 0,59	30 m	Met l'accent sur la végétation de pointe, ce qui est utile pour évaluer la vigueur de la plante
Band 4 - Rouge	0,64 - 0,67	30 m	Discrimine les pentes de végétation
Bande 5 – Proche Infrarouge (NIR)	0,85 - 0,88	30 m	Souligne le contenu de la biomasse et les rives
Bande 6 - SWIR 1 Infrarouge moyen 1	1,57 - 1,65	30 m	Discrimine la teneur en humidité du sol et de la végétation ; pénètre les nuages minces
Bande 7 - SWIR 2 Infrarouge moyen 2	2,11 - 2,29	30 m	Amélioration de la teneur en humidité du sol et de la végétation et pénétration du nuage mince
Band 8 - Panchromatique	0,50 - 0,68	15 m	Résolution de 15 mètres, définition d'image plus nette

Band 9 - Cirrus	1,36 - 1,38	30 m	Détection améliorée de la contamination par les nuages de cirrus
Bande 10 - Infrarouge thermique (TIRS) 1	10,60 - 11,19	100 m	Résolution de 100 mètres, cartographie thermique et humidité estimée du sol
Bande 11 - Infrarouge thermique (TIRS) 2	11h50 - 12h51	100 m	Résolution de 100 mètres, amélioration de la cartographie thermique et estimation de l'humidité du sol

Cette information relative aux bandes nous aide à sélectionner les bandes utiles pour calculer des indices de télédétection. Par la suite, nous décrivons le processus de sélection et la transformation des données.

3.3.2 Sélection et transformation des données (Données dérivées d'images)

Le but de cette étape est d'identifier et de sélectionner des données pertinentes et importantes impliquées dans la recherche de facteurs environnementaux favorables à l'épidémie. Ce sont ces facteurs qui seront utilisés dans l'étude et c'est la raison pour laquelle on doit les mettre en exergue. A cet effet, pour faire la sélection des bandes utiles, il est important de connaître :

- les informations sur l'origine et les types d'environnements favorisant l'émergence et la propagation de l'épidémie. Cette information est fournie par les spécialistes de l'épidémie en question.
- les coordonnées des images satellitaires du champ d'études.

Le but d'une méthode de télédétection consiste à extraire les informations à partir d'une seule bande ou à partir d'une superposition de bandes afin de mettre en exergue l'environnement qui nous intéresse. Pour cela, des techniques ont été proposées pour faciliter l'analyse des images satellitaires telles que les indices de végétation, les techniques de prévision spatiale (Hengl et al., 2004), les produits de télédétection thermique (Takahata et al., 2010), les estimations biophysiques, les techniques de classification (Alrababah and Alhamad, 2006).
Plusieurs indices ont été développés spécifiquement pour l'analyse des images satellitaires, mais dans la littérature, les indices de végétation sont les plus utilisés. C'est pourquoi nous avons également opté pour les indices de végétation.

Ces indices de végétation utilisent le principe de différenciation entre les zones de végétation et leurs sous-zones (par exemple les zones forestières et arides) et les zones d'eau (par exemple, lac, rivière et étang). Dans la littérature, différents indices de végétation ont été développés (voir Tableau 3.2) (Prabhakara et al., 2015) : (Freden et al., 1974) ont proposé l'indice de végétation par différence normalisée (NDVI) pour estimer la vigueur des plantes, Huete a présenté un indice de végétation ajusté (SAVI) (Huete et al., 2002). Un indice de végétation triangulaire a été présenté par (Broge and Leblanc, 2001) l'indice de végétation par différence de gradient à trois bandes (TGDVI) par (Tan et al., 2006), et l'Indice de bassin de différence normalisé (NDPI) ainsi que l'Indice de turbidité différentielle normalisée (NDTI) ont été présentés par (Bicout et al., 2015; Lacaux et al., 2007).

Dans le Tableau 3.2, les bandes sont désignées dans les formules par R (Rouge), B (Bleu), G (Vert), NIR (Near Infrared) et RE (Red-Edge).

Tableau 3.2 – Définition des indices spectraux

Indice	Désignation	Référence	Formule
NDVI	Indice de végétation par différence normalisée	(Tucker, 1979)	(NIR-R) / (NIR + R)
NGRDI	Indice normalisé de différence rouge vert	(Tucker, 1979)	(GR) / (G + R)
GNDVI	Indice de végétation par différence normalisée verte	(Moges et al., 2005)	(NIR-G) / (NIR + G)
SR	Ratio simple	(Tucker and Sellers, 1986)	NIR / R
SAVI	Indice de végétation ajusté du sol (L = 0,5)	(Huete, 1988)	[(NIR-R) /(NIR + R + L)] (1 + L)
EVI	Indice de végétation amélioré	(Huete et al., 2002)	2,5(NIR-R)/(NIR+6*R-7,5*B+1)
TVI	Indice de végétation triangulaire	(Broge and Leblanc, 2001)	0,5 [120 (NIR-G) -200 (RG)]
VARI	Indice visible de résistance à l'atmosphère	(Gitelson et al., 2002)	(GR) / (G + RB)
NDREI	Différence normalisée Red Edge Index	(Gitelson and Merzlyak, 1994)	(RE-R) (RE + R)

Parmi les techniques des indices de végétation, le NDVI est largement utilisé. Le NDVI est construit à partir du rouge (R) et du proche infrarouge (NIR) et met en évidence la différence entre la bande visible rouge et la bande visible proche infrarouge en utilisant l'équation standard: [(NIR-RED) / (NIR + RED)] (Tucker, 1979). Cet indice est sensible à la végétation et les valeurs vont de -1 à +1. Les valeurs négatives correspondent à des surfaces autres que les zones de végétation telles que la neige, l'eau ou les nuages pour lesquels la réflectance dans le rouge est supérieure à celle du proche infrarouge. Pour le sol nu, la réflectance est à peu près la même dans le rouge et le proche infrarouge, les valeurs du NDVI sont proches de 0. Les zones de végétation ont des valeurs positives de NDVI généralement comprises entre 0,1 et 0,7. Les valeurs au-delà de 0,7 correspondent à une végétation dense. Lorsqu'une série temporelle d'images est observée pour la même zone, on peut déduire l'état de stress de la végétation pour différentes valeurs de NDVI: en hiver ou en période de sécheresse, les valeurs seront plus basses qu'en été ou en saison des pluies.

Pour calculer le NDVI à partir d'une image Landsat, il suffit de combiner les trois premières bandes ; vous obtenez une image optique conventionnelle (RGB) et après ajouter la bande du proche infrarouge. Le calcul NDVI génère un fichier image ou une couche, et cela pourrait être réalisé par certains logiciels de traitement d'images (tels qu'ENVIE, IDRISI, Matlab, R, ERDAS Imagine) ou par des logiciels de systèmes d'information géographique (GIS) qui peuvent faire des calculs raster (tels que QGIS, ArcGIS).

3.3.3 Mesures de télédétection

Après avoir opté pour une mesure de télédétection (NDVI), il suffit maintenant de produire les images correspondantes par application des calculs d'indice. Cela permettra de mettre en exergue les environnements favorables au développement de l'épidémie. Par la suite, des techniques de data mining seront appliquées sur ces images afin d'extraire des informations notamment sur la quantité d'espace occupée par la zone d'étude et aussi suivre l'évolution dans le temps d'une zone d'étude. A cet effet, la majorité des algorithmes d'exploration de données ont des difficultés à traiter des valeurs continues. Par conséquent, les attributs continus doivent être divisés en intervalles pour s'adapter aux règles de structuration effectuées durant le prétraitement des données. Les données collectées sur le terrain doivent être préparées par des techniques de discrétisation.

3.3.4 Discrétisation des données

Dans cette étape, les données sont transformées ou consolidées dans des formats appropriés pour le processus de data mining. Nous avons deux types de données à discrétiser : les données épidémiologiques relatives au nombre de cas confirmés provenant des structures de santé et les données issues des mesures de télédétection. La discrétisation consiste à convertir des valeurs numériques en des valeurs catégoriques : les attributs continus doivent être divisés en plusieurs intervalles d'adaptation de l'algorithme d'extraction de données. Pour discrétiser les données de mesures de télédétection, l'algorithme de discrétisation basé sur l'entropie de l'information proposée par (Xie et al., 2016) est adopté. Ils définissent trois étapes pour discrétiser les données de détection : (i) Créer une table de décision (ii) déterminer les points de subdivision et les classes d'équivalence (iii) enfin dérouler l'algorithme de discrétisation (KNN) implémenté dans le logiciel Matlab. L'algorithme de discrétisation permet de subdiviser chaque bande en plusieurs intervalles (sections) (Figure 3.3). Il est important de redimensionner les images obtenues avec la méthode de rééchantillonnage du plus proche voisin. C'est la technique idéale si la nouvelle image doit subir une classification puisqu'elle se compose alors de la luminosité du pixel original et du simple réarrangement dans une position permettant de donner une géométrie d'image correcte (Richards and Richards, 1999). Le processus attribue une valeur à chaque pixel corrigé à partir des pixels voisins non corrigés. Les gains du plus proche voisin comprennent la simplicité et la capacité de préserver les valeurs originales dans les images.

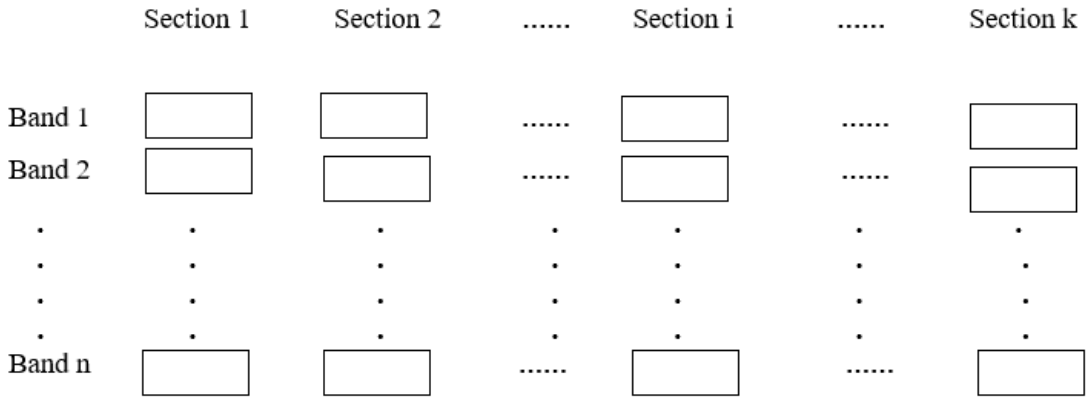


Figure 3.3 – Résultat de la discrétisation sur un échantillon (Xie et al., 2016)

Après avoir préparé la base de données par les techniques de discrétisation, l'étape suivante consiste à déterminer la corrélation entre la donnée discrétisée provenant des images de télédétection et la base de données épidémiologique relative au nombre de cas confirmés.

3.3.5 Traitement des données

La méthode de traitement des données est la classification supervisée. La classification est l'une des principales tâches de data mining qui est un processus inductif, itératif et interactif pour découvrir dans de grandes bases de données de modèles de données valides, nouveaux, utiles et compréhensibles (Han et al., 2011).

Les données pour la classification sont divisées en deux ensembles : un ensemble de données pour les tests d'entraînement et un ensemble de données pour les essais afin de consolider le modèle trouvé. Nous utiliserons ce modèle pour classer de nouvelles zones en deux catégories : zone à risque d'épidémie ou zone sans risque d'épidémie. Les méthodes de classification supervisées sont utilisées pour générer une image satellitaire en assignant chaque pixel à sa classe soit la classe de zone à risque épidémique ou à la classe de zone sans risque d'épidémie. Les classes sont déterminées en fonction de la composition spectrale des sites d'entraînement définis par l'utilisateur. Dans ce travail, la classification et l'analyse des images numériques ont été réalisées à l'aide du logiciel de traitement d'images Envi.

Description de l'algorithme de classification utilisé : Maximum de vraisemblance - ML (Abkar et al., 2000; Ahmad and Quegan, 2012; Richards and Richards, 1999)

Nous avons utilisé la méthode de classification supervisée Maximum de vraisemblance sur les images de télédétection basée sur l'approche générale de Bayes, qui indique qu'une distribution a posteriori P (i | ω), c'est-à-dire la probabilité qu'un pixel sachant le vecteur de caractéristiques ω appartienne à la classe i, est donnée par :

$$P(i|w) = \frac{P(w|i)P(i)}{P(w)} \qquad (3.1)$$

où P (ω|i) est la fonction de vraisemblance, P (i) est l'information à priori, c'est-à-dire la probabilité que la classe i apparaisse dans la zone d'étude et P (ω) la probabilité que ω soit observée, ce qui peut être écrit comme :

$$P(w) = \sum_{i=1}^{M} P\,(w|i)P(i) \qquad (3.2)$$

où M est le nombre de classes. P (ω) est souvent traité comme une constante de normalisation pour assurer que la somme $\sum_{i=1}^{M} P(i|w)$ est égale à 1. Le pixel x est affecté à la classe i par la règle :

$$i = argmax(P(i|w)) \qquad (3.3)$$

ML suppose souvent que la distribution des données au sein d'une classe donnée i obéit à une distribution gaussienne à plusieurs variables. Il est alors pratique de définir la vraisemblance logarithmique (ou la fonction discriminante) :

$$g_i(w) = lnP(i|w) = -\frac{1}{2}(w-\mu_i)^t C_i^{-l}(w-\mu_i) - \frac{N}{2}\ln(2\pi) - \frac{1}{2}\ln(|C_i|) \qquad (3.4)$$

Chaque pixel est assigné à la classe ayant le plus de vraisemblance ou étiqueté comme non classé si les valeurs de probabilité sont toutes inférieures à un seuil défini par l'utilisateur. Les procédures générales en ML sont les suivantes :

1. Le nombre de types de couverture terrestre dans la zone d'étude est déterminé.
2. Les pixels d'apprentissage pour chacune des classes souhaitées sont choisis en utilisant les informations de couverture terrestre pour la zone d'étude. Dans ce but, la distance Jeffries- Matusita (JM) peut être utilisée pour mesurer la séparabilité de classes des pixels d'entraînement choisis. Pour les classes normalement distribuées, la mesure de séparabilité de JM pour deux classes, J_{ij}, est définie comme suit :

$$J_{ij} = \sqrt{2(1 - e^{-\alpha})} \qquad (3.5)$$

où α est la distance de Bhattacharyya et est donnée par:

$$\alpha = \frac{1}{8}(\mu_i - \mu_j)^t \left[\frac{(C_i + C_j)}{2}\right]^{-1} (\mu_i - \mu_j) + \frac{1}{2}\ln\left(\frac{\left|\frac{(C_i + C_j)}{2}\right|}{\sqrt{|C_i||C_j|}}\right) \qquad (3.6)$$

J_{ij} varie de 0 à 2.0, où $J_{ij} > 1.9$ indique une bonne séparabilité des classes, une séparabilité modérée pour $1{,}0 \leq J_{ij} \leq 1.9$ et une médiocre séparabilité pour $J_{ij} < 1.0$

3. Les pixels d'apprentissage sont ensuite utilisés pour estimer le vecteur moyen et la matrice de covariance de chaque classe.
4. Enfin, chaque pixel de l'image est classé dans l'un des types de couvertures terrestres souhaités ou étiquetés comme inconnu.

Dans la classification ML, chaque classe est incluse dans une région dans un espace multispectral où sa fonction discriminante est plus grande que celle de toutes les autres classes. Ces régions de classes sont séparées par des limites de décision, où la limite de décision entre les classes i et j se produit lorsque:

$$g_i(w) = g_j(w) \qquad (3.7)$$

Pour les distributions normales multi variées, cela devient :

$$-(w - \mu_i)^t C_i^{-l}(w - \mu_i) - \ln(|C_i|) + (w - \mu_j)^t C_j^{-l}(w - \mu_j) - \ln(|C_j|) = 0 \qquad (3.8)$$

C'est une fonction quadratique en N dimensions. Par conséquent, si nous considérons seulement deux classes, les limites de décision sont des sections coniques (c.-à-d. Paraboles, cercles, ellipses ou hyperboles).

Au-delà des techniques de la classification supervisée et non supervisée, il y a d'autres travaux qui s'intéressent au classificateur temporel (Jacquin, 2010) et au classificateur orienté objet (Blaschke, 2010; Volker, 2004).

3.3.6 Interprétation des résultats

Enfin, la dernière étape consiste à interpréter les cartes de risque épidémique qui doivent être mises à la disposition des structures de santé ou des autorités compétentes pour une meilleure utilisation. Après le traitement des données de télédétection par la méthode de classification supervisée, nous pouvons obtenir une précision détaillée par classe. Ces mesures de précision donnent le taux de réussite en pourcentage de la classification appliquée. En effet, c'est un ensemble de cartes à différentes périodes (mensuellement ou annuellement) qui sont produites. Pareillement, à la même période d'obtention des images satellitaires, les données de terrain des différents cas de contamination de l'épidémie (mensuellement ou annuellement) sont également collectées.

A partir des résultats produits par la classification et les données de terrain disponibles, la suite consiste à faire une analyse comparative permettant de :

- Mieux expliquer les tendances observées à partir des indicateurs de terrain,
- Préciser la nature du lien entre l'environnement et l'épidémie,
- Clarifier l'impact des changements environnementaux sur l'épidémie.

3.4 Étude de cas : Épidémie de choléra à Mopti (Mali)

L'étude de cas porte sur l'épidémie de choléra dans la région de Mopti, au Mali. Ce choix est fait sur la base des informations fournies par la Direction Nationale de la Santé (DNS) du Mali qui stipulent que Mopti est la région la plus touchée au Mali.

3.4.1 Collecte des données

3.4.1.1 Collecte des informations sur le terrain à propos de l'épidémie de choléra

a. Les caractéristiques de l'épidémie de choléra

Le choléra est une épidémie contagieuse causée par la bactérie *Vibrio choléra*, ou *bacille de Comma* découvert par Pacini en 1854 et redécouvert par Robert Koch en 1884 (Harris et al., 2012). Il est strictement limité à l'espèce humaine, et se caractérise par une diarrhée soudaine et abondante (gastro-entérite) entraînant des déshydratations sévères. La surveillance et la sensibilisation permettent de maîtriser rapidement les épidémies de choléra ("WHO | Cholera," 2018).

Facteurs de propagation de l'épidémie de choléra (Dao et al., 2009):

- 78% des cas de contamination se baignaient dans les rivières,
- Les régions les plus touchées (Mopti, Ségou et Koulikoro) sont situées le long du fleuve Niger et constituent le noyau (point de départ) des épidémies de choléra au cours des trois dernières décennies,
- La période la plus favorable au Mali est la saison des pluies (juin, juillet, et août).

Par conséquent, le facteur géographique (environnement) retenu dans le cercle de Mopti est le fleuve Niger et aussi la période de l'étude est la période couvrant la saison des pluies (juin, juillet, et août). D'ailleurs, c'est ce qui explique le choix de la période de collecte des données des images satellitaires. La période de collecte des images satellitaires doit correspondre à la période de collecte des données épidémiologiques en termes du nombre de cas confirmés. De plus, il faut signaler le manque de données en dehors de ces périodes choisies.

Donc, il s'agit de déterminer l'impact du fleuve Niger sur l'épidémie de choléra pendant la saison des pluies.

b. Nombre de cas confirmés de l'épidémie de choléra

Le tableau ci-dessous présente la répartition du nombre de cas confirmés de l'épidémie de choléra par régions au Mali. Ces données ont été collectées auprès du ministère de la Santé (Dao et al.,

2009). Au cours de la décennie «1995-2004 », le choléra a frappé toutes les régions du Mali avec des variations selon les régions et les années (Tableau 3.3 et Figure 3.4). En effet, les régions de Mopti (3 412 cas) et de Ségou (1 500 cas) ont été atteintes par presque toutes les épidémies de choléra de 2004-1995 avec une accalmie en 2002 à Ségou.

Tableau 3.3 – Répartition des cas de choléra au Mali de 1995 à 2004 par région

(A noter que de 1997 à 2000, il n'y a eu aucun cas de choléra au Mali.)

Les régions	**Nombre de cas confirmés**						**Total**
	1995	1996	2001	2002	2003	2004	
Kayes	0	3 497 (63,1%)	0	0	0	0	3 497 (28,7%)
Koulikoro	57 (2,5%)	33 (0,6%)	0	0	99 (6,9%)	1 (0,0%)	190 (1,6%)
Sikasso	0	0	0	0	45 (3,1%)	5 (0,2%)	50 (0,4%)
Ségou	166 (7,4%)	453 (8,2%)	16 (23,5%)	0	645 (44,9%)	220 (7,7%)	1 500 (12,3%)
Mopti	**987 (43,8%)**	**735 (13,3%)**	**52 (76,5%)**	**16 (100%)**	**597 (41,5%)**	**1 025 (35,8%)**	**3 412 (28%)**
Tombouctou	776 (34,4%)	616 (11,1%)	0	0	36 (2,5%)	1 258 (44%)	2 686 (22,1%)
Gao	234 (10,4%)	205 (3,7%)	0	0	0	351 (12,3%)	790 (6,5%)
Bamako *	34 (1,5%)	2 (0,1%)	0	0	15 (1%)	0	51 (0,4%)
Total	2 254 (100%)	5 541 (100%)	68 (100%)	16 (100%)	1 437 (100%)	2 860 (100%)	12 176 (100%)

* District

Les régions de Kayes et de Sikasso ont connu moins de flambées de choléra entre 1995 et 2004. Pendant cette période, la région de Kayes a connu une seule flambée avec 3497 cas et la région de Sikasso deux flambées avec 50 cas.

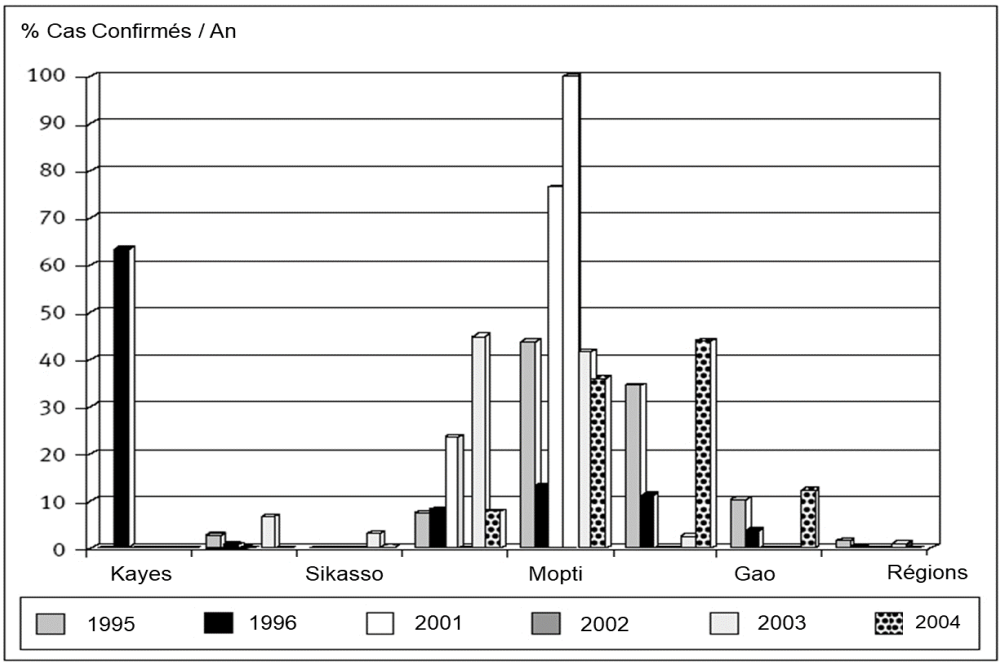


Figure 3.4 – Répartition des cas de choléra au Mali de 1995 à 2004 par région

Enfin, l'étude de cas concerne la région de Mopti, car c'est la région la plus touchée par l'épidémie de choléra au Mali de 1995 à 2004.

3.4.1.2 Collecte des images satellitaires de la zone d'étude

a. Zone d'étude :

De 1995 à 2004, la région de Mopti a été la seule à être touchée par toutes les épidémies de choléra au Mali, elle est située sur le fleuve Niger (Dao et al., 2009). La zone d'étude concerne la ville de Mopti soit une superficie de 7262 km² (Figure 3.5). Mopti est situé au centre du Mali, il est donc un point culminant entre le Nord et le Sud, avec une population de plus de 368 512 personnes (INSAT, 2009). Mopti est composé de plusieurs groupes ethniques exerçant plusieurs activités autour du fleuve Niger qui traverse la région notamment le commerce, le transport, la pêche, l'élevage et l'agriculture.



Figure 3.5 – Zone d'étude : Région de Mopti (département de Mopti en bande rouge)

b. Données satellites

Nous avons opté pour les images satellitaires d'archives du programme Landsat de la zone d'étude, et elles peuvent être téléchargées sur http://glovis.usgs.gov/ ou http://earthexplorer.usgs.gov/ à l'adresse suivante : Path = 197 et Row = 50 ou Lat = 14.5 et Long = 4.3 (Figure 3.6). Bien que le

cycle de couverture périodique du programme Landsat soit de 16 jours pour la même zone, les données satellitaires recueillies sont découpées en scènes successives et une vingtaine d'images sont produites chaque année par scène (Fu and Weng, 2016). Cependant, dans notre recherche, nous avons pris en compte les connaissances des experts du domaine qui stipulent que la saison des pluies est la période propice à l'épidémie de choléra, en particulier de juillet à août. Par conséquent, nous avons utilisé une image satellite par année tout en respectant les périodes indiquées. Car nous ne disposons pas de données en termes de nombre de cas d'épidémie en dehors de la période de la saison des pluies. Donc, les périodes de collecte des images satellitaires doivent correspondre aux périodes de collecte des données épidémiologiques.

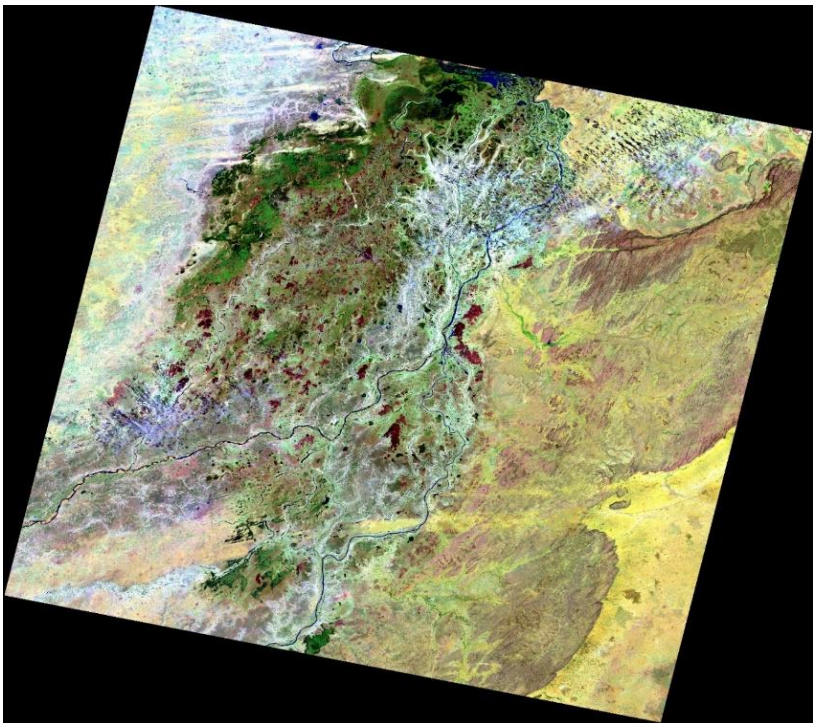


Figure 3.6 – Image satellite de Mopti (Chemin = 197 et Ligne = 50)

3.4.2 Sélection et transformation des données

Après les images satellitaires téléchargées, les bandes sélectionnées pour afficher les images sont SWIR2 (Band 7) Band, NIR (Band 5) et BLUE (Band 2) avec le stretch (0,5% Clip) et le gamma (0,8). Nous avons appliqué quelques corrections sur des images satellitaires telles que des rectifications atmosphériques dans le but d'améliorer la qualité des images.

3.4.3 Mesures de télédétection

A ce stade, nous générons le fichier NDVI, car NDVI est le meilleur indice pour faire apparaître la différence visuelle entre la végétation et la surface de l'eau (Figure 3.7) en utilisant la formule NDVI comme suit (NIR-R) / (NIR + R).

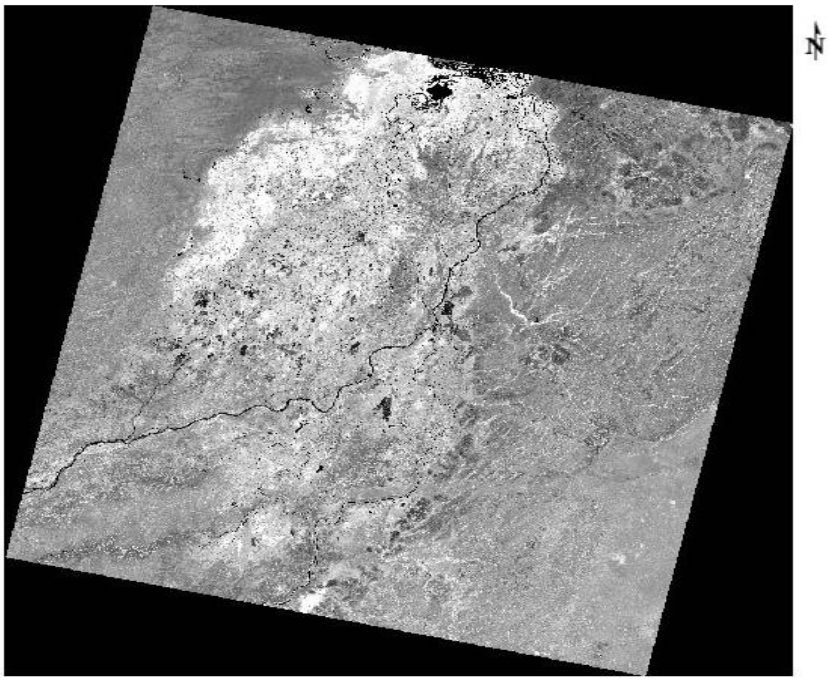


Figure 3.7 – Fichier NDVI

3.4.4 Discrétisation des données

Avant la classification des images en deux groupes (zones à risque d'épidémie et zones sans risque d'épidémie), il est nécessaire de préparer les images satellitaires par une technique de discrétisation et après cela les redimensionner avec la méthode de rééchantillonnage du plus proche voisin (Richards and Richards, 1999), tout ceci à l'aide du Logiciel Envi. Pour notre étude, la ville de Mopti est considérée, ce qui signifie que toutes les images traitées (Tableau 3.4) ont les mêmes coordonnées cartographiques comme indiqué ci-dessous :

Coordonnée supérieure gauche :

Lat : 14 31 52,16

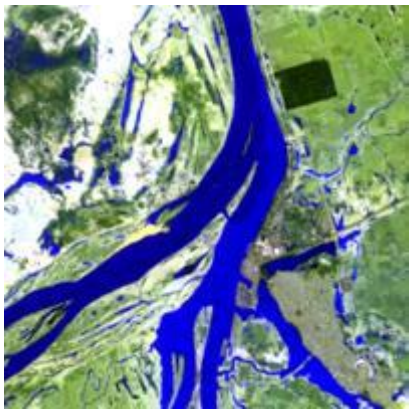
Lon : -4 14 2,64

Coordonnée inférieure droite :

Lat : 14 28 38,42

Lon : -4 10 41,67

Tableau 3.4 – Images satellitaires traitées



Date : 1995/08/05
Landsat 4

AUCUNE IMAGE DISPONIBLE DANS LES ARCHIVES LANDSAT DANS CETTE ZONE DE 05/08/1995-02/18/1998

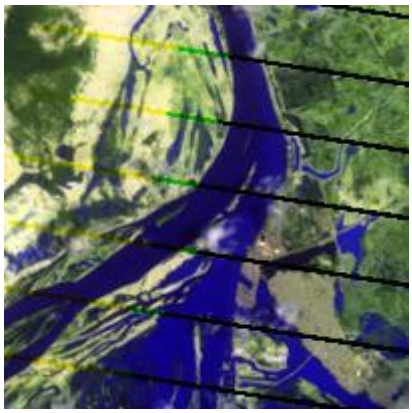
******



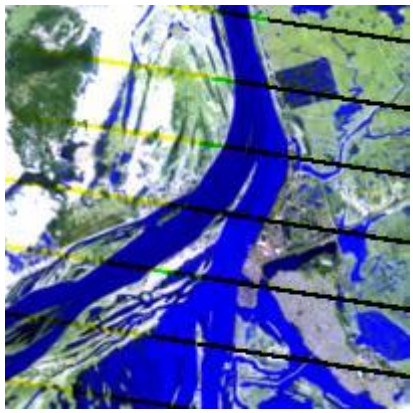
Date : 2001/07/28
Landsat 7 SLC-on (1999-2003)



Date : 2002/07/31
Landsat 7 SLC-on (1999-2003)



Date : 2003/08/19
Landsat 7 SLC-off (2003->)



Date : 2004/08/21
Landsat 7 SLC-off (2003->)

3.4.5 Traitement des images : classification

Pendant la phase de manipulation, nous avons testé plusieurs algorithmes à savoir le classificateur de distance minimale, classificateur de parallélépipèdes, les réseaux de neurone classique, Machine à vecteurs de support (SVM)… mais l'Algorithme Maximum de vraisemblance nous a donné de bons résultats en termes de précision de la classification. Cette précision est obtenue de la matrice de confusion après la classification. Il s'agit de la vérification du pourcentage de pixels de chaque classe de l'image classés correctement et/ou erronément A cet effet, les échantillons d'entraînement ont été étiquetés en deux classes : zones à risque d'épidémie (partie en bleu représentant le fleuve Niger) et tout le reste en zone sans risque d'épidémie. Ces aires d'entraînement (ou Région d'intérêt ROI) ont été établies suite à notre connaissance du terrain principalement sur l'étendue du fleuve Niger tout en bleu. Dans le souci de mettre en exergue l'environnement d'étude, nous nous sommes limités à deux classes sinon on aurait pu aller au-delà de deux classes pour faire référence aux niveaux de risques (Faible, Moyen, Elevé).

Dans le Tableau 3.5 ci-dessous, nous présentons les précisions de la classification des différentes images satellitaires de l'Algorithme Maximum de vraisemblance (Abkar et al., 2000).

Tableau 3.5 – Précisions de la classification des différentes images satellitaires

Année / Paramètres	1995	2001	2002	2003	2004
Précision de la classification	99.6350%	99.9448%	99.8540%	97.3451%	98.8845%
Coefficient Kappa	0.9924	0.9986	0.9970	0.9451	0.9760

Le Tableau 3.6 ci-dessous présente les différents résultats en termes de la distribution de chaque classe sur la totalité de l'image.

Tableau 3.6 – Résultats après classification



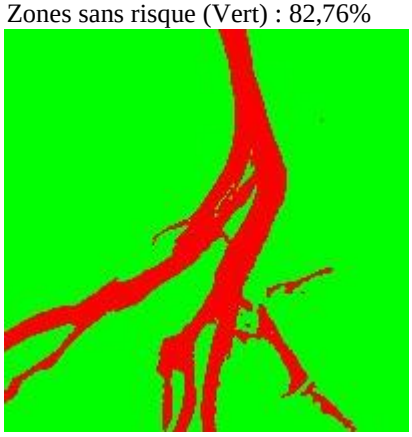
Date : 1995/08/05
Réalité du terrain (pourcentage)
Zones à risque (Rouge) : 17,24%
Zones sans risque (Vert) : 82,76%

AUCUNE IMAGE DISPONIBLE DANS LES ARCHIVES LANDSAT DANS CETTE ZONE DE 05/08/1995-02/18/1998

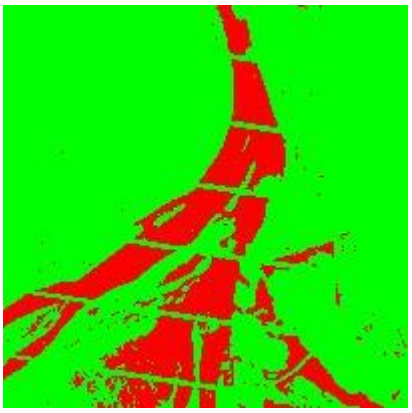
******



Date : 2001/07/28
Réalité du terrain (pourcentage)
Zones à risque (Rouge) : 16,75%
Zones sans risque (Vert) : 83,25%



Date : 2002/07/31
Réalité du terrain (pourcentage)
Zones à risque (Rouge) : 14,73%
Zones sans risque (Vert) : 85,27%



Date : 2003/08/19
Réalité du terrain (pourcentage)
Zones à risque (Rouge) : 19,69%
Zones sans risque (Vert) : 80,31%



Date : 2004/08/21
Réalité du terrain (pourcentage)
Zones à risque (Rouge) : 23,93%
Zones sans risque (Vert) : 76,07%

3.4.6 Interprétation des résultats

Les résultats issus de l'étape de la classification notamment la valeur des zones à risque multipliée par cent et les données collectées sur le terrain sont présentés dans le Tableau 3.7. La même base de données est représentée sous forme d'histogramme dans la Figure 3.8.

Tableau 3.7 – Base de données des cas confirmés et des zones à risque

Années	**Cas confirmés**	**Zones à risque (%*100)**
1995	987	1724
2001	52	1675
2002	16	1473
2003	597	1969
2004	1025	2393

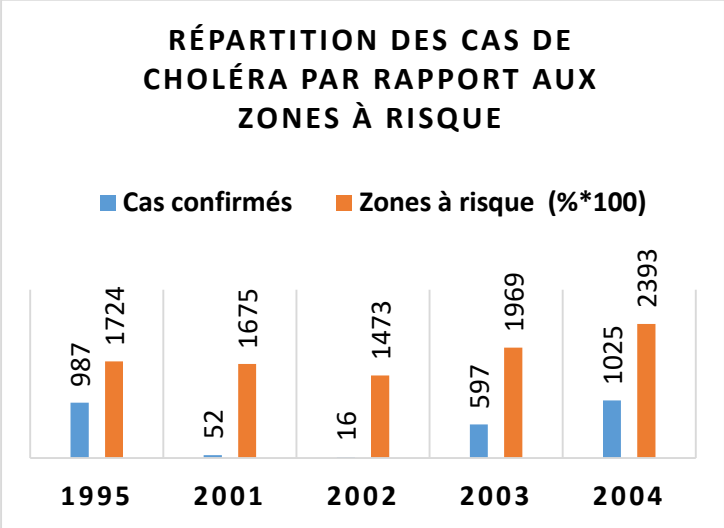


Figure 3.8 – Graphique correspondant à la base de données

Cette base de données sera codifiée par les techniques de discrétisation. Pour ce faire, nous avons utilisé le système de codification présenté dans le Tableau 3.8. Ainsi, le Tableau 3.9 représente la base de données codifiée par application du système de codification présenté. Les données sont maintenant utilisables par l'Algorithme Apriori pour la génération des règles d'association pour une bonne interprétation des résultats.

Tableau 3.8 – Système de codification de la base de données

Cas confirmés		
Code	Intervalle	Description
1	[0-350]	Faible
2	[351-700]	Moyen
3	[701-1050]	Elevé

Zones à risque		
Code	Intervalle	Description
11	[09%-16%]	Faible
22	[17%-24%]	Elevé

Tableau 3.9 – Base de données codifiée

Base de données codifiée		
Années	**Cas confirmés**	**Zones à risque (% * 100)**
1995	3	22
2001	1	11
2002	1	11
2003	2	22
2004	3	22

Tout ceci constitue le prétraitement des données. La base de données étant prête, il ne reste plus qu'à générer toutes les règles d'association qui ont un support et une confiance supérieurs au support minimum (appelé minsup) et à la confiance minimum (appelé minconf) spécifiés par l'utilisateur. De nombreux algorithmes existent pour découvrir les règles d'association, le plus connu étant l'algorithme d'Apriori (Srikant and Agrawal, 1996). Contrairement aux autres algorithmes, Apriori trouve toutes les règles d'association entre les ensembles d'éléments fréquents en les ajoutant à de grands ensembles, et en élaguant de petits ensembles. Ceci est cohérent avec notre objectif d'extraire tous les types de règles possibles pour une évaluation ultérieure des résultats par l'utilisateur. La combinaison des connaissances produites par les règles d'association avec celles du domaine provenant de l'expertise humaine est pertinente pour la résolution de problèmes (Kamsu-Foguem et al., 2013).

Les règles d'association générées par Apriori Association Rules Algorithm (Srikant and Agrawal, 1996) du logiciel SPMF (Fournier-Viger et al., 2014) sont présentées dans le Tableau 3.10.

Tableau 3.10 – Interprétation des résultats obtenus

Règles d'association générées	Interprétation
1 ==> 11 #SUP : 2 #CONF : 1	Si le nombre de cas confirmés est faible ([0-350]) alors la zone à risque est également faible ([09% -16%]) avec une confiance totale (100%)
22 ==> 3 #SUP : 2 #CONF : 0,66667	Si la zone à risque est élevée ([17% -24%]) alors le nombre de cas confirmés est élevé ([701-1050]) avec une bonne confiance (66%)
3 ==> 22 #SUP : 2 #CONF : 1	Si le nombre de cas confirmés est élevé ([701-1050]), alors la zone à risque est également élevée ([17% -24%]) avec une confiance totale (100%)

Parmi les règles générées, la deuxième est la plus intéressante. Elle s'interprète comme suite : une estimation de la probabilité conditionnelle P(Cas confirmés Elevé/Zones à risque Elevé) égale à à 66% le taux de contamination est élevé. L'interprétation sous-jacente concerne les facteurs sociétaux durant la période d'hivernage : le déversement des ordures et des excréments dans le fleuve Niger. Ce qui permet d'engendrer la bactérie Vibrio dans le fleuve. Ainsi, l'utilisation de cette eau est à la base de l'épidémie de choléra.

Les résultats obtenus suite au traitement des images satellitaires combinés aux données de terrains collectées auprès des structures de santé nous permettent de conclure que le niveau du fleuve Niger a un impact sur la variation du taux de l'épidémie de choléra dans la ville de Mopti. Ces résultats seront mis à la disposition des autorités politiques et administratives comme outil de prévention d'épidémie de choléra, d'information et de sensibilisation de la population de Mopti. De plus, Mopti étant une ville surpeuplée et située au confluent du Niger et de son affluent le Bani, il serait nécessaire d'éviter l'utilisation de l'eau et de la nourriture contaminées spécifiquement pendant la période hivernale. Ces actions s'inscrivent dans le cadre des mesures de sensibilisation, d'interdiction de l'utilisation de l'eau du fleuve sur des périodes bien déterminées.
L'usage des images satellitaires dans la lutte contre l'épidémie de choléra à Mopti est une perspective intéressante pour mieux appréhender cette épidémie et ses facteurs de propagation. Cela s'inscrit dans le cadre de la télésurveillance spécifiquement en périodes de crues à travers des systèmes d'alerte. Et aussi, il serait intéressant de comparer les résultats obtenus aux réalités de terrain de la zone spécifique d'étude puis que l'étude a concerné une partie de la région de Mopti sur laquelle nous ne disposons de données détaillées et suffisantes. Donc cette validation concernera et le nombre de cas confirmés de l'épidémie dans la localité concernée auprès des structures de santé et aussi le niveau du fleuve (ou la superficie occupée par le fleuve) auprès des services hydrauliques.

3.5 Discussions

Les taux élevés de mortalité de l'épidémie de choléra au Mali, particulièrement à Mopti constituent un défi pour les structures de santé. Il est nécessaire de les doter des techniques de surveillance épidémiologique. L'une des techniques privilégiées de nos jours est l'usage de la télédétection pour prévenir et faciliter la gestion d'une crise épidémique par des observations terrestres. C'est dans ce sens que (Kazansky et al., 2016) ont étudié le rôle de la télédétection dans la campagne de lutte contre le paludisme. Le but consiste à fournir des cartes de risque de malaria au gouvernement pour une meilleure gestion de l'épidémie. La méthodologie proposée est basée sur des indicateurs provenant de satellites, permettant de déterminer les régions les plus exposées au paludisme. Le principal inconvénient de cette méthode est qu'il est difficile de mettre en place un système de prévision des zones à risque sans considérer les données in situ relatives à l'épidémie en question.

Treize classes de zones à risque d'épidémie ont été identifiées en Inde par la technique de classification par arbres de décisions (Punia et al., 2011). Pour cela, ils ont utilisé les images Resourcesat-1 (aussi connu sous le nom d'IRS-P6), un satellite d'observation de la Terre construit par l'ISRO (Indian Space Research Organization). Toutefois, ils n'ont pas pris en compte les données du terrain provenant des experts du domaine pour valider les paramètres retenus dans la méthodologie. L'approche proposée par (Pitarch et al., 2015) combine le traitement des images satellitaires et des données de terrain pour la recherche des modèles et la cartographie des terres agricoles afin d'améliorer les systèmes d'alerte précoce pour la sécurité alimentaire. Cependant, cette approche a des limitations dans sa phase de discrétisation, car les intervalles de discrétisation choisis étaient trop grands pour capturer les variations temporelles de texture et de plus les auteurs suggèrent une discrétisation plus fine comme piste d'amélioration. L'avantage de notre approche par rapport à ceux de (Bicout et al., 2015) est que nous traitons plusieurs images satellites de la même zone d'étude corrélées avec des données de terrain collectées sur les mêmes périodes que les images satellitaires, permettant ainsi d'avoir une meilleure interprétation. (Conţiu and Groza, 2016) ont proposé un système hybride pour la classification des cultures à partir d'imageries satellitaires

à l'aide d'une méthode basée sur le système de vote pour la résolution des conflits dans la base d'apprentissage. Nous avons le même avantage en associant les connaissances d'experts du domaine avec des règles d'extraction générées à partir de la base de données afin d'améliorer la précision de la classification.

Les méthodes existantes de traitement des images satellitaires sont de plus en plus matures et efficaces, ainsi le défi consiste aujourd'hui à fournir les moyens les plus appropriés permettant une meilleure interprétation des résultats obtenus. Pour cela, nous nous sommes appuyés sur l'algorithme de classification supervisée (Maximum Likelihood) pour traiter un ensemble d'images satellitaires provenant de la même zone, mais à des périodes différentes. L'algorithme de maximum de vraisemblance est la méthode de classification supervisée la plus couramment utilisée avec les données d'images de télédétection (Richards and Richards, 1999). Il est basé sur la probabilité qu'un pixel appartienne à une classe particulière. Cependant, il existe d'autres méthodes de classification d'images avec certaines caractéristiques intéressantes. L'algorithme de distance minimale est également attrayant, mais n'utilise pas les données de covariance. Une autre variante de la Distance Minimum est la distance de Mahalanobis et la différence réside au niveau de l'utilisation de la matrice de covariance. La méthode de l'algorithme parallélépipède est un simple classificateur pour les tâches d'apprentissage automatique, mais il peut y avoir des écarts importants entre les parallélépipèdes ; les pixels de ces régions ne seront pas classés. Toutefois dans notre cas précis, l'Algorithme Maximum de vraisemblance nous a donné de meilleurs résultats pendant la phase de manipulation du logiciel Envi par rapport aux autres algorithmes de classification.

Par ailleurs, l'une des plus grandes contributions de notre approche consiste à une meilleure interprétation des résultats obtenus en tenant compte des données de terrain. Cette étude a également démontré que le fleuve Niger a un impact sur l'évolution de l'épidémie de choléra dans la région de Mopti, à l'aide des techniques de télédétection. Cependant, en ce qui concerne ses limites, il serait souhaitable d'avoir des images de meilleure qualité, mais de 2003 à 2004, seul Landsat 7 ETM + SLC-off (depuis 2003) a fourni des images du programme Landsat. C'est ce qui explique les rayures (zigzag) sur ces images produites à cette période. Le 31 mai 2003, Scan Line Corrector (SLC) de l'instrument ETM + a échoué, causant des bandes sur les images (2003/08/19 et 2004/08/21) d'où son nom SLC-off (Wulder et al., 2012). Pourtant, les rayures ont moins d'impact sur les résultats. Les bandes vertes au-dessus de la rivière classée zones à risque (vertes) sont faussement positives et doivent donc être classées dans les zones à risque (rouge), donc la croissance relative en pourcentage des zones à risque fournit une argumentation raisonnable et justifie les résultats décrits. Dans une perspective à moyen terme, il est nécessaire d'améliorer le traitement des images satellitaires à savoir les images de mauvaise qualité pour l'optimisation des résultats souhaités.

3.6 Conclusion

La téléépidémiologie est un nouvel axe de recherche pour la surveillance des maladies infectieuses. Elle consiste à mettre l'espace au service de la lutte contre les épidémies transmises par l'eau, l'air, les vecteurs, etc. en mettant en relation le climat, l'environnement et la santé pour traquer et prévenir les épidémies. Dans ce chapitre, nous avons étudié en détail l'extraction des connaissances à l'aide des techniques de data mining sur les images satellitaires afin de découvrir des zones à risques d'épidémie, dont l'étude de cas a porté sur l'épidémie de choléra dans la région de Mopti, au Mali. Ainsi, une méthodologie comportant six phases (prétraitement, data mining et post-traitement) a été présentée. Dans un premier temps, nous avons appliqué la classification supervisée

(l’algorithme du maximum de vraisemblance) sur les images satellitaires de la même zone, mais à différentes périodes (ou à différentes dates). En second lieu, nous avons créé une base de données regroupant les données collectées sur le terrain et les données de la classification des images satellitaires afin de trouver leur corrélation par la technique des règles d’association. Et enfin, les règles générées ont été interprétées pour produire de la connaissance compréhensible par l’utilisateur. La méthodologie proposée a été évaluée avec un certain succès relativement aux opportunités d’exploitation offertes. Les résultats nous indiquent qu’à 66% le taux de contamination est lié au fleuve Niger, en plus de certains facteurs sociétaux comme le jet des ordures en période hivernale. Par conséquent, nous avons pu établir le lien entre l’épidémie et son environnement d’évolution, ce qui permettra aux décideurs de mieux gérer une éventuelle crise d’épidémie. Les principales contributions de la méthodologie développée sont : la discrétisation en phase de prétraitement améliorant la qualité des résultats obtenus, l'établissement du lien entre l'environnement et l'épidémie, et l'identification des zones les plus à risque pour la propagation et l'émergence de l'épidémie.

Dans ce travail, nous avons utilisé le même format de données c’est-à-dire les images provenant d’un même type de satellite. Comme travaux futurs, il serait intéressant de considérer les images provenant de différents satellites. Le traitement d'images satellitaires de sources hétérogènes constitue également un défi considérable pour la communauté scientifique. Ces récentes années, l’apprentissage profond a permis une nouvelle révolution dans l'intelligence artificielle notamment dans le traitement des images et particulièrement dans la reconnaissance des formes. Il est basé sur les réseaux de neurones et a fait ses preuves dans plusieurs domaines, notamment : l’analyse de signaux auditifs ou visuels, la reconnaissance faciale, la reconnaissance des désastres, la reconnaissance vocale, la vision assistée par ordinateur, traitement automatique des langues. Une fois dans une situation de crise d’épidémie, il devient urgent de procéder aux analyses microscopiques des agents pathogènes responsables des épidémies pour confirmer ou non leur présence dans les prélèvements des cas suspects. Ainsi, le prochain chapitre sera consacré à l’analyse médicale des images d'agents pathogènes épidémiques à l’aide des techniques de l’apprentissage profond.

Chapitre 4

Classification des images d'agents pathogènes épidémiques à l’aide des réseaux de neurones convolutifs

4.1 Introduction 85
4.2 Travaux connexes 86
4.2.1 Définition du Deep learning et travaux connexes 86
4.2.2 Les réseaux de neurones convolutifs (CNN) 87
4.2.3 Les différentes couches constitutives d'un réseau de neurones convolutifs 88
4.2.3.1 La couche de convolution 88
4.2.3.2 Couche ReLU 90
4.2.3.3 Couche de Pooling 90
4.2.3.4 Couche entièrement connectée 90
4.2.4 Les paramètres des différentes couches 91
4.2.5 Apprentissage du réseau de neurones convolutifs 92
4.3 Méthodologie et logiciel utilisés 93
4.3.1 Méthodologie pour la classification des images des agents pathogènes 94
4.3.1.1 Constitution de la base d’apprentissage 95
4.3.1.2 Constitution de la base de validation 96
4.3.1.3 Application de l’apprentissage par transfert : Extraction de caractéristiques 96
4.3.1.4 Modèle généré par apprentissage par transfert 100
4.3.1.5 Soumission de la base de test 101
4.3.1.6 Evaluation d'images classées 101
4.3.2 Logiciel utilisé : framework Keras 104
4.4 Discussions 104
4.5 Conclusion 105

4. Classification des images d'agents pathogènes épidémiques à l'aide des réseaux de neurones convolutifs

4.1 Introduction

Dans le chapitre précédent, nous avons appliqué des techniques de data mining sur les images satellites pour la découverte des zones à risque d'épidémie. Cette étude s'inscrit dans le cadre de la téléépidémiologie et de la prévention des épidémies. La téléépidémiologie est l'étude de la propagation de maladies humaines ou animales, transmises par l'eau, l'air ou des vecteurs liés à des facteurs climatiques, météorologiques et environnementaux, à l'aide des technologies de l'information et de la communication (Marechal et al., 2008). L'étude de cas concernait l'épidémie de choléra dans la région de Mopti, au Mali (Afrique de l'Ouest) (Traore et al., 2017).

Cependant, une fois dans une situation de crise d'épidémie, il devient urgent de procéder aux analyses microscopiques des agents pathogènes responsables des épidémies pour confirmer ou non leur présence dans les prélèvements des cas suspects. Ainsi, afin de surpasser nos résultats précédents et de renforcer la capacité de gestion des crises épidémiques, ce chapitre est consacré à la reconnaissance d'images à l'aide des techniques de Deep learning pour classer automatiquement les images microscopiques des agents pathogènes des épidémies. A cet effet, nous nous sommes orientés vers les techniques des réseaux de neurones convolutifs du deep learning, car selon la littérature, elle est la plus adaptée pour la classification des images. Il est indéniable que de nombreuses recherches ont été menées dans ce domaine, mais à notre connaissance, rares sont celles qui ont vraiment mis l'accent sur l'identification des agents pathogènes des épidémies à travers des analyses microscopiques particulièrement en temps de crise d'épidémie. Le constat est qu'il y a souvent un manque de spécialistes en manipulation microscopique surtout dans les pays en voie de développement, d'où la nécessité de réaliser les analyses microscopiques à l'étranger occasionnant ainsi des coûts supplémentaires (Rosado et al., 2016) comme ce fut le cas de l'épidémie de la maladie à virus Ebola. Egalement, cela implique une perte de temps considérable et aussi favorise entre temps la propagation de l'épidémie. Pour gagner du temps dans l'analyse des échantillons, nous proposons de rendre les microscopes intelligents afin qu'ils puissent confirmer par eux-mêmes l'existence ou non des agents pathogènes d'une épidémie dans un échantillon ou un prélèvement.

Ainsi, ce chapitre est organisé comme suit. Dans un premier temps, nous allons aborder en détail les travaux connexes sur les réseaux de neurones particulièrement les réseaux de neurones convolutifs. En second lieu, nous présenterons notre approche pour classer les images des agents pathogènes des épidémies. Elle est basée sur l'apprentissage par transfert l'une des techniques des réseaux de neurones convolutifs, les différents résultats seront également présentés à chaque étape de l'approche et aussi nous présenterons le logiciel utilisé. En dernier lieu, nous mènerons une discussion et finir avec une conclusion.

4.2 Travaux connexes

Nous assistons aujourd'hui au remplacement progressif de nombreuses applications basées sur les technologies classiques d'apprentissage automatique dans le domaine de la vision par ordinateur par les techniques émergentes du deep learning (Grinblat et al., 2016; LeCun et al., 2015). Egalement, il est nécessaire d'améliorer la précision des méthodes de diagnostic des agents pathogènes axées sur les manipulations manuelles impliquant certaines erreurs humaines (Rosado et al., 2016; Tangpukdee et al., 2009). A cet effet, les réseaux de neurones convolutifs constituent aujourd'hui le standard en matière de reconnaissance d'images, telles que la reconnaissance des écritures manuscrites des chiffres à l'aide d'un réseau de rétropropagation (LeCun et al., 1990), la reconnaissance faciale (Russakovsky et al., 2015), la reconnaissance vocale, la vision par ordinateur (Karpathy et al., 2014), la reconnaissance des catastrophes (Liu and Wu, 2016) et le traitement automatique du langage naturel (Hinton et al., 2006).

4.2.1 Définition du Deep learning et travaux connexes

Le Deep learning est un sous domaine de l'apprentissage automatique. C'est un ensemble de méthodes de l'apprentissage automatique permettant de modéliser avec un haut niveau d'abstraction des données grâce à des architectures articulées de différentes transformations non linéaires (Schmidhuber, 2015). D'après (LeCun et al., 2015), le succès du Deep Learning est dû à la disponibilité du matériel (comme des unités centrales de traitement (CPU), des cartes graphiques (GPU) et des disques durs de grande capacité de stockage, etc.), des algorithmes d'apprentissage automatiques, des grands volumes de données par exemple la base de données ImageNet et MNIST (Krizhevsky et al., 2012).

Dans les zones agricoles, les techniques des réseaux de neurones convolutifs sont utilisées pour prévenir les dommages causés par les ravageurs agricoles ; en l'occurrence ceci a permis d'améliorer la production agricole avec une précision de 98,67% pour 10 classes de ravageurs (Luo et al., 2017). Dans le même domaine, Grinbat et ses collègues (Grinblat et al., 2016) utilisent les réseaux neuronaux convolutifs profonds (DCNN) pour identifier les plantes par la classification de trois espèces de légumes différentes: le haricot blanc, le haricot rouge et le soja. Pour l'identification des maladies du riz, Lu et ses collègues (Lu et al., 2017) utilisent également les techniques DCNN grâce à la reconnaissance d'images, avec une précision de 95,48% sur 500 images naturelles de feuilles et de tiges de riz malades et saines. En reconnaissance visuelle, ils explorent DCNN pour classer une image hyperspectrale avec une précision améliorée. Pour la classification des scènes de télédétection, Nogueira et ses collègues (Nogueira et al., 2017) explorent trois stratégies possibles, à savoir: l'apprentissage complet, le fine tuning et l'utilisation de ConvNets comme extracteurs de caractéristiques.

Dans la reconnaissance des formes par la méthode traditionnelle (Figure 4.1), l'expert du domaine ou le programmeur procède à l'extraction des caractéristiques importantes de l'image, puis ses paramètres sont soumis au réseau de neurones classique pour la classification des données en entrée. Cependant, la phase d'extraction manuelle des caractéristiques est fastidieuse, nécessite une connaissance des données en entrée et nécessite également beaucoup de temps. Le deep learning permet de se passer de cette douloureuse phase qui est faite automatiquement à l'aide de plusieurs couches cachées et interconnectées (Figure 4.2).

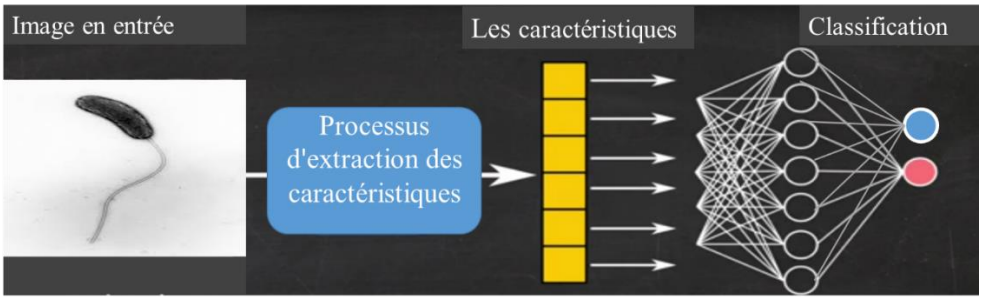


Figure 4.1 – Ancienne méthode d'extraction des caractéristiques

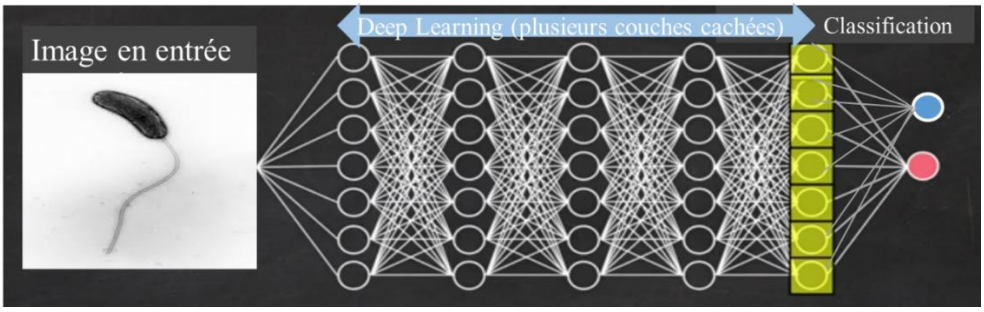


Figure 4.2 – Deep Learning avec plusieurs couches cachées

4.2.2 Les réseaux de neurones convolutifs (CNN)

Dans ce travail de classification d'images, nous avons opté pour les réseaux de neurones convolutifs, car ils sont plus adaptés à la classification d'images avec de bonnes performances et aussi à la rapidité d'entraînement du réseau (Barbedo, 2013; Krizhevsky et al., 2012). En effet, les pixels constitutifs des images ne sont pas complètement indépendants, et il est souvent important de considérer des zones de l'image.

Par ailleurs, les notions générales de l'architecture CNN ont été présentées dans le chapitre état de l'art. Cependant, quelques notions mathématiques seront abordées ci-dessous. En mathématiques, en particulier dans la topologie algébrique, la convolution est une transformation mathématique de deux fonctions (u et v) ; il génère une troisième fonction, qui est normalement considérée comme une version transformée d'une des fonctions initiales. Dans le cas de deux fonctions réelles ou complexes, *u* et *v* la convolution est une autre fonction, qui est généralement notée $u * v$ et qui est défini par (Hirschman and Widder, 2012) comme suit:

$$(u * v)(x) = \int_{-\infty}^{+\infty} u(x-t)v(t)dt \qquad (4.1)$$

La convolution satisfait certaines propriétés algébriques (la commutativité, l'associativité, la distributivité et l'identité multiplicative) afin de préserver les caractéristiques des images géométriques correspondantes.

Commutativité

$$u * v = v * u \qquad (4.2)$$

Associativité

$$u * (v * h) = (u * v) * h \qquad (4.3)$$

Associativité avec multiplication scalaire

$$a(u * v) = (au) * v \qquad (4.4)$$

La distributivité

$$u * (v + h) = (u * v) + (u * h) \qquad (4.5)$$

Formellement, l'opération de filtrage effectuée par une carte de caractéristique est une convolution discrète. La convolution discrète peut être vue comme une multiplication de matrice. Cependant, la matrice a plusieurs entrées avec des contraintes égales aux autres entrées. Pour les fonctions définies sur l'ensemble des entiers, la convolution de deux séquences finies est déterminée en étendant les séquences à des fonctions finement applicables sur l'ensemble des entiers (par exemple, une sommation finie peut être utilisée pour un support fini). La convolution discrète est définie par la formule suivante (Hirschman and Widder, 2012) :

$$(u * v)(n) = \int_{m=-\infty}^{\infty} u(n-m)v(m)dt \qquad (4.6)$$

4.2.3 Les différentes couches constitutives d'un réseau de neurones convolutifs

En général, l'architecture CNN se subdivise en quatre parties à savoir (Figure 4.3): la couche de Convolution responsable du traitement des parties de l'image c'est-à-dire des extracteurs de caractéristiques, la couche de Pooling est chargée de réduire la taille des images intermédiaires et aussi combine les sorties des couches de convolution pour en détecter des fonctionnalités de plus haut niveau, la couche de correction ReLU est la fonction d'activation des neurones, la couche complètement connectée vers toutes les sorties de la couche précédente sert à prédire la bonne classe parmi d'autres.

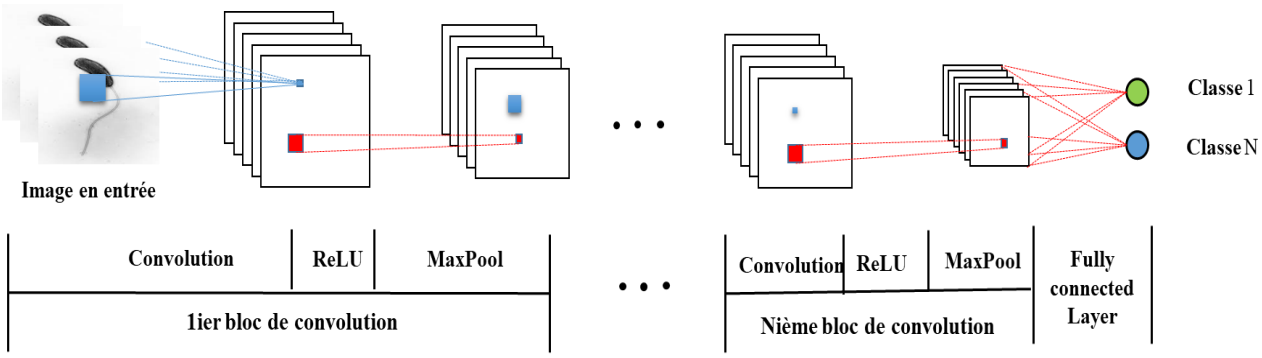


Figure 4.3 – L'architecture générale de CNN

4.2.3.1 La couche de convolution

C'est la première couche et la plus importante. Il s'agit de repérer la présence d'un ensemble de caractéristiques dans les images reçues en entrée. A cet effet, on réalise un filtrage par convolution : le principe est de faire glisser une fenêtre représentant la caractéristique sur l'image, et de calculer le produit de convolution entre cette caractéristique et chaque portion de l'image balayée. La couche de convolution reçoit donc en entrée plusieurs images, et calcule la convolution de chacune d'entre elles avec chaque filtre. Les filtres correspondent exactement aux caractéristiques que l'on souhaite retrouver dans les images. On obtient pour chaque paire (image, filtre) une carte d'activation ou

une carte de caractéristique, qui nous indique où se situent les caractéristiques dans l'image : plus la valeur est élevée, plus l'endroit correspondant dans l'image ressemble à la caractéristique. Les caractéristiques sont apprises par le réseau lors la phase d'entraînement. Elles sont initialisées puis mises à jour par rétropropagation du gradient. C'est ce qui constitue la force des réseaux de neurones convolutifs : ceux-ci sont capables de déterminer tout seuls les éléments discriminants d'une image, en s'adaptant au problème posé. Par exemple, si la question est de distinguer les chats des chiens, les caractéristiques automatiquement définies peuvent décrire la forme des oreilles ou des pattes (Monasse and Nadjahi, 2018).
Nous recensons trois hyperparamètres pour contrôler la taille du volume de sortie de la couche de convolution: la profondeur, le pas et la marge (Shahin et al., 2017).

- Profondeur de la couche : nombre de noyaux de convolution (ou nombre de neurones associés à un même champ récepteur) (Figure 4.4).

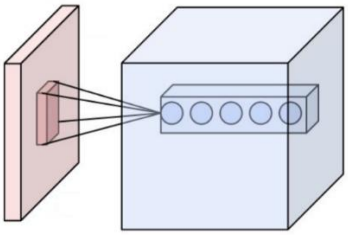


Figure 4.4 – Profondeur de la couche

- Pas de contrôle (le chevauchement des champs récepteurs) : Le pas indique de combien de pas on se déplace entre deux calculs de la convolution (Figure 4.5). Plus le pas est petit, plus les champs récepteurs se chevauchent et plus le volume de sortie sera grand.

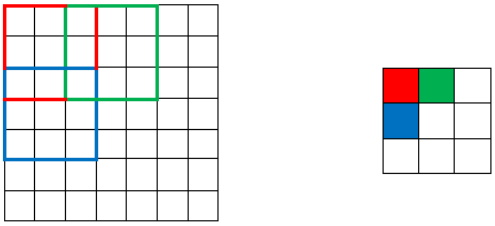


Figure 4.5 – Le pas de déplacement du filtre

- Marge (à 0) : parfois, il est commode de mettre des zéros à la frontière du volume d'entrée. La taille de cette marge est le troisième hyperparamètre (Figure 4.6). Cette marge permet de contrôler la dimension spatiale du volume de sortie. En particulier, il est parfois souhaitable de conserver la même surface que celle du volume d'entrée.

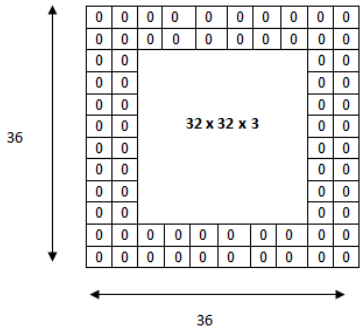


Figure 4.6 – La marge

4.2.3.2 Couche ReLU

La couche ReLU (ou Rectified Linear Units) est très utilisée. En effet, elle est à la fois simple à calculer et simple à dériver, et sa dérivée est suffisamment importante pour faciliter l'apprentissage. La fonction la plus utilisée est la fonction Max car elle est simple et efficace. Elle remplace donc toutes les valeurs négatives reçues en entrées par des zéros. Elle joue le rôle de fonction d'activation. Mathématiquement, elle se présente sous la forme suivante :

$$f(x) = max(0, x) \qquad (4.7)$$

En outre, nous pouvons aussi utiliser la fonction *Tanh* qui se présente comme suit :

$$f(x) = tanh(x) \qquad (4.8)$$

Et également, la fonction sigmoïde qui utilise une exponentielle. Elle est également appelée la fonction logistique.

$$f(x) = \frac{1}{(1 + e^{-x})} \qquad (4.9)$$

4.2.3.3 Couche de Pooling

Elle est souvent placée entre deux couches de convolution, elle reçoit en entrée plusieurs cartes de caractéristiques, et applique à chacune d'entre elles l'opération de pooling. Cette opération consiste à réduire la taille des images, tout en préservant leurs caractéristiques importantes. Parmi les opérations de Pooling, il y a le calcul de la moyenne, le second est le L2-norm Pooling et le troisième, le plus utilisé est le Max pooling à cause de sa rapidité et de sa convergence (Figure 4.7) (Scherer et al., 2010). Pour ce faire, l'image est découpée en cellules régulières, puis on garde au sein de chaque cellule la valeur maximale. En pratique, on utilise souvent des cellules carrées de petite taille pour ne pas perdre trop d'informations. Les choix les plus communs sont des cellules adjacentes de taille 2 × 2 pixels.

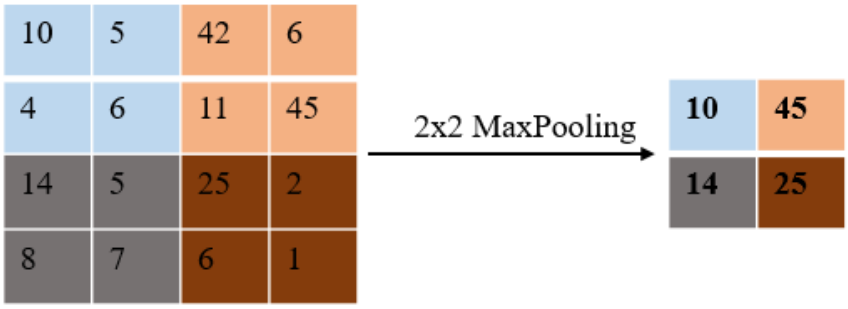


Figure 4.7 – Processus de Max pooling

4.2.3.4 Couche entièrement connectée

Chacune des entrées est connectée à chacune des sorties, d'où le terme "entièrement connecté" (Figure 4.8). C'est la dernière couche d'un réseau de neurones convolutifs. La Couche entièrement connectée permet de classifier l'image en entrée du réseau : elle renvoie un vecteur de taille N, où N est le nombre de classes pour un problème de classification d'images. Chaque élément du vecteur indique la probabilité pour l'image en entrée d'appartenir à une classe (Krizhevsky et al., 2012). Par exemple, si le problème consiste à distinguer les chats des chiens, le vecteur final sera de taille 2 : le premier élément indique la probabilité d'appartenir à la classe Chat et le deuxième la classe

Chien. Ainsi, le vecteur [0.9 0.1] signifie que l'image a 90% de chances de représenter un chat. Pour calculer les probabilités, la couche entièrement connectée multiplie donc chaque élément en entrée par un poids, fait la somme, puis applique une fonction d'activation (logistique si N=2, softmax si N>2).

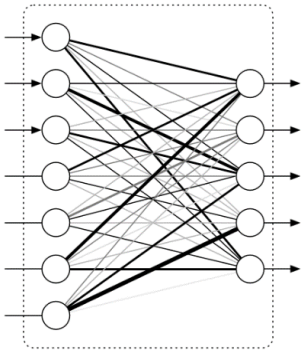


Figure 4.8 – Couche entièrement connectée

La fonction Softmax est un peu particulière, car elle permet d'agréger différentes sorties et de normaliser les valeurs de façon à ce que la somme des probabilités fasse 1. De cette façon, chaque valeur de sortie correspond directement au pourcentage d'appartenance à une classe. La valeur normalisée de chaque sortie est calculée comme suit :

$$f(x_i) = \frac{x_i}{(\sum_j x_j)} \qquad (4.10)$$

4.2.4 Les paramètres des différentes couches

Les paramètres des différentes couches d'un réseau de neurones convolutifs sont présentés dans le Tableau 4.1 comme suit (Torch GitHub, 2018).

Tableau 4.1 – Les paramètres des différentes couches d'une architecture CNN

Couches	Description des paramètres
Contribution	Fichiers d'image (jpeg, gif, ...) décrits avec un format RVB pouvant être modélisé en tant que matrice 3D.
Conv (x, y, z, u, v, h, i, j)	x = nombre de canaux d'entrée attendus dans l'image donnée (3 pour RVB) y = nombre de canaux de sortie après la phase de convolution z = La largeur du noyau de la convolution u = Hauteur du noyau de la convolution v = Le pas de la convolution dans la dimension de la largeur h = Le pas de la convolution dans la dimension de la hauteur i = zéros supplémentaires ajoutés aux données du plan d'entrée des deux côtés de l'axe de la largeur j = zéros supplémentaires ajoutés aux données du plan d'entrée des deux côtés de l'axe de la hauteur
ReLU (x)	Une unité linéaire rectifiée (fonction d'activation) a la sortie 0 si l'entrée est inférieure à 0 et la sortie brute sinon. x = vrai ou faux.

MaxPooling (x, y, z, u)	Une opération de pooling maximal qui examine les fenêtres X*Y et trouve la longueur de foulée max par Z*U. x = la largeur du filtre du regroupement y = hauteur du filtre du pooling z = le pas de la largeur de mise en commun u = le pas de la hauteur de mise en commun
FullyConnected (x, y)	x = taille de l'image saisie, par exemple 3 * 256 * 256 (3 canaux de couleur, 256 pixels de hauteur et 256 pixels de largeur) y = nombre de classes de sortie inférieur à la taille de l'image en entrée

Donc finalement, on peut résumer un CNN comme un empilement de plusieurs couches de convolution, pooling, correction ReLU et une couche entièrement connectée. Chaque image reçue en entrée va donc être filtrée, réduite et corrigée plusieurs fois, pour finalement former un vecteur. Dans le problème de classification, ce vecteur contient les probabilités d'appartenance aux classes.

4.2.5 Apprentissage du réseau de neurones convolutifs

Selon (Mathivet, 2017) le réseau de neurones convolutifs apprend les valeurs des poids de la même manière qu'il apprend les filtres de la couche de convolution : lors de la phase d'entraînement, par rétro propagation du gradient. Il s'agit en ce moment de corriger les poids entre les neurones de sortie et les neurones cachés, puis propager l'erreur en arrière (d'où le nom) et corriger les poids entre les neurones cachés et les entrées. Cette correction se fera exemple après exemple, et il faudra plusieurs passes (avec des ordres différents si possible) pour converger vers un optimum. Par souci de simplification, on suppose que nous n'avons qu'une couche de neurones cachée. Le principe est le même pour plusieurs couches, il suffit de propager l'erreur en arrière sur toutes les couches.

Posons **m** le nombre d'exemples, **h** le nombre de neurones cachés et **n** le nombre de sorties. La première étape consiste donc à calculer la sortie de chaque neurone caché nommée O_i pour un exemple donné. On fait ensuite de même pour les neurones de sortie, dont la sortie est calculée à partir de la somme pondérée des sorties de la couche cachée. C'est la phase de propagation vers l'avant. L'erreur commise est toujours :

$$Erreur = Y_i - S_i \qquad (4.11)$$

avec Y_i la sortie attendue et S_i la sortie obtenue.

L'évaluation du réseau sera basée sur l'erreur quadratique. C'est l'évaluation classique pour une tâche de régression, ce que nous supposerons ici. On note **n** le nombre d'entrées, et **m** le nombre d'exemples utilisés pour l'apprentissage. Généralement, pour les tâches de classification, on calculera plutôt le coût noté J, ce qui modifiera sa dérivée et donc les équations :

$$J = \frac{1}{m}\sum_{i=1}^{m}[-y_i \log(s_i) - (1 - y_i)\log(1 - s_i)] \qquad (4.12)$$

On passe ensuite à la phase de propagation vers l'arrière. La première étape consiste à calculer les deltas sur les neurones de sortie. Pour cela, on calcule pour chacun la dérivée de la fonction d'activation multipliée par l'erreur, qui correspond à notre delta.

$$\delta_i = S_i.(1 - S_i).(Y_i - S_i) \qquad (4.13)$$

Il faut ensuite faire la même chose pour les neurones cachés, reliés chacun aux K neurones de sortie. Le calcul est alors :

$$\delta_i = O_i.(1 - O_i).\sum_k \delta_k * W_{i \to k} \qquad (4.14)$$

En effet, ce calcul prend en compte la correction sur les neurones de sortie (δ_k) et le poids qui les relie. De plus, plus un poids est important et plus la correction à appliquer sera forte.
Lorsque tous les deltas sont calculés, les poids peuvent être modifiés en faisant le calcul suivant, où seule la dernière valeur change selon qu'il s'agit d'un neurone caché (on prend l'entrée) ou d'un neurone de sortie (on prend son entrée, c'est-à-dire la sortie du neurone caché) :

$$W_i = W_i + \tau.\delta_i.(X_i \; OU \; O_i) \qquad (4.15)$$

La modification d'un poids est donc plus importante si l'erreur est forte et/ou si l'entrée est importante. Le facteur τ s'appelle le taux d'apprentissage. Le taux d'apprentissage dépend du problème à résoudre : trop faible, il ralentit énormément la convergence. Trop grand, il peut empêcher de trouver la solution optimale (voire faire diverger l'algorithme). Il y a quelques années, on utilisait souvent un taux adaptatif, fort au début de l'apprentissage, et qui diminuait ensuite au cours du temps. Actuellement, on préfère un taux fixe, bien choisi, quitte à faire plusieurs essais sur des sous-ensembles de données pour trouver des taux qui semblent optimaux.

L'algorithme de l'apprentissage du réseau de neurones convolutifs est présenté en annexe B.1 (Page 117).

Après cette description des principaux éléments théoriques des réseaux de neurones et particulièrement le CNN, nous allons maintenant aborder leur domaine d'application. Comme énoncé précédemment, les réseaux de neurones convolutifs sont adaptés aux traitements des images. Ainsi, le domaine de la santé n'est pas resté en marge (Litjens et al., 2017) notamment dans: (i) la classification des images d'analyses médicales (ii) la détection des organes, des régions particulières dans une image (iii) la segmentation des organes, des sous-structures et des lésions (iv) l'enregistrement automatique des images par catégories. Dans ce chapitre, nous nous focalisons sur la classification des images des agents pathogènes des épidémies. Des améliorations dans la précision sont nécessaires dans les méthodes actuelles de diagnostic des agents pathogènes, car le processus d'analyse est axé sur la manipulation humaine, ce qui peut entraîner des erreurs. De plus, la détection des agents pathogènes épidémiques par traitement d'images à travers les microscopes est pertinente surtout pour les pays en voie de développement (Rosado et al., 2016). C'est pourquoi il devient nécessaire de rendre les microscopes plus intelligents pour plus de rapidité et pour combler le manque de spécialistes dans la manipulation microscopique.

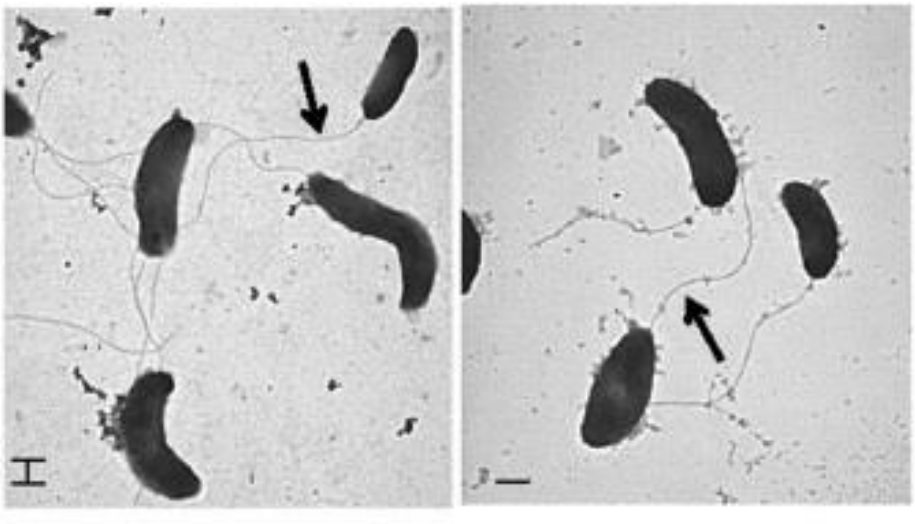
4.3 Méthodologie et logiciel utilisés

L'objectif de cette méthodologie consiste à déterminer si les images microscopiques contiennent soit des virus vibrions cholériques (Figure 4.9) responsables de l'épidémie de choléra ou soit des virus du paludisme appelés plasmodium falciparum (Figure 4.10) responsables de l'épidémie du paludisme. A cet effet, nous avons opté pour l'approche CNN plus précisément l'apprentissage par

transfert, car nous ne disposons pas d'un jeu de donnée assez conséquent pour entraîner un réseau CNN à partir de zéro. C'est pourquoi généralement pour les tâches de classification d'images, on utilise un réseau connu préentraîné sur une grande base de données telle qu'ImageNet avec plus de 1,4 million d'images labellisées et classifiées en 1000 classes. La base de données ImageNet dispose de 5 images de vibrion choléra et 10 images de Plasmodium falciparum. Donc, finalement, il s'agit d'utiliser un réseau existant, préentraîné tel que le VGG-16, VGG-19, ResNet, Inception, etc. Pour réaliser l'apprentissage par transfert, nous avons opté pour l'architecture VGG-16 entrainée sur la base de données ImageNet (Simonyan and Zisserman, 2014) et implémentée dans la bibliothèque Keras. Ainsi, le logiciel utilisé est Keras qui est une bibliothèque open source de haut niveau de réseaux de neurones (Gullì and Pal, 2017).

4.3.1 Méthodologie pour la classification des images des agents pathogènes

L'analyse microscopique consiste à confirmer ou non si l'échantillon est contaminé par le pathogène Vibrio choléra ou par le Plasmodium falciparum (Harris et al., 2012). Le choléra est une épidémie contagieuse transmise principalement par la consommation de l'eau de boisson ou des aliments souillés contenant l'agent pathogène Vibrio choléra (Davies et al., 2017). Le choléra se caractérise par une diarrhée soudaine et abondante (gastro-entérite) entraînant une déshydratation sévère. Dans le monde, il y a une prévalence de 2,8 millions de cas et les victimes critiques sont près de 100 000 morts par an. Visuellement, le Vibrio choléra est une bactérie en forme de virgule, mobile chez l'homme (Wang et al., 2017)(Figure 4.9).



a) b)

Figure 4.9 – Vibrio choléra (Wang et al., 2017)

Le paludisme est une maladie infectieuse et reste l'une des principales causes de mortalité et de morbidité, avec environ 212 millions de cas et 429 000 décès en 2015 dans le monde. Le paludisme est causé par un parasite du genre Plasmodium falciparum, visuellement en forme de petit cercle, propagé par la piqûre de certaines espèces de moustiques anophèles (Walker et al., 2018) (Figure 4.10).

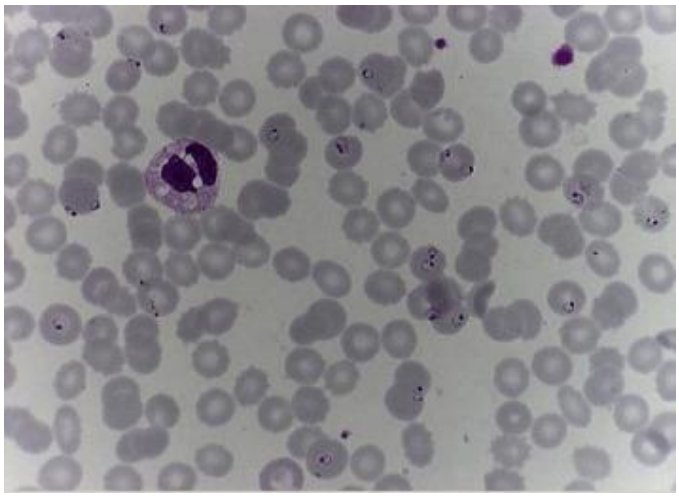


Figure 4.10 – Plasmodium falciparum (Walker et al., 2018)

C'est dû aux raisons évoquées ci-dessus que nous avons opté pour ces deux types d'épidémie à savoir le choléra et le paludisme comme domaine d'application. Ainsi, notre méthodologie comprend six phases, décrites dans la figure ci-dessous (Figure 4.11) : (1) Constitution de la base d'apprentissage (2) Constitution de la base de validation (3) Application de l'apprentissage par transfert (4) Génération du modèle final (5) Soumission de la base de test au modèle généré et enfin (6) Evaluation du processus de classification avant l'intégration de la solution dans les futurs microscopes.

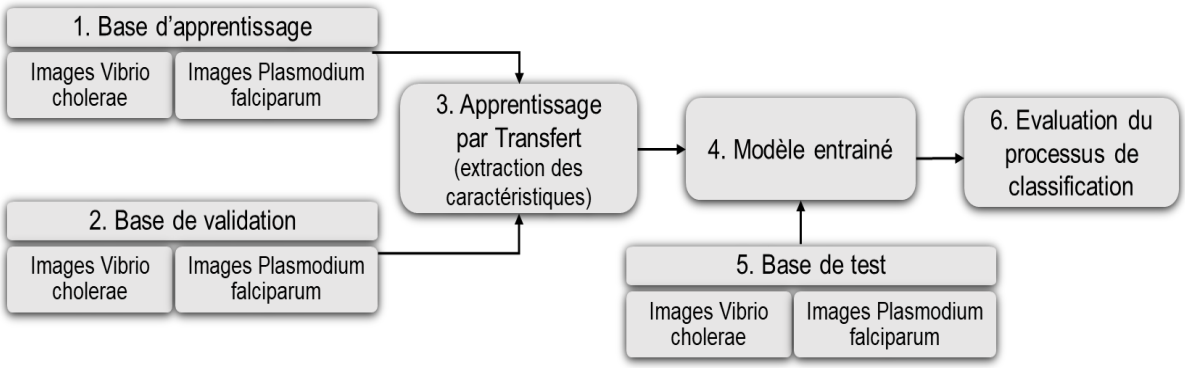


Figure 4.11 – Apprentissage par transfert sur les images d'agents pathogènes épidémiques

4.3.1.1 Constitution de la base d'apprentissage

Pour distinguer les agents pathogènes du choléra des images des agents pathogènes du paludisme, nous avons besoin de constituer une base d'apprentissage des données des images vibrio choléra dans un dossier et des images plasmodium falcparum dans un autre dossier. Ces images ont été téléchargées sur Google image avec comme mot-clé les noms des deux agents pathogènes épidémiques respectivement : Choléra Vibrio et Plasmodium falciparum, comme le montre la Figure 4.12. Par la suite, elles sont redimensionnées à la même taille afin d'obtenir de meilleurs résultats. L'ensemble des images de la base d'apprentissage est correctement labélisé comme indiqué dans la Figure 4.12. Pour notre étude, la base d'apprentissage est constituée de 200 images soit 100 images de Vibrio choléra et 100 images de Plasmodium falciparum.

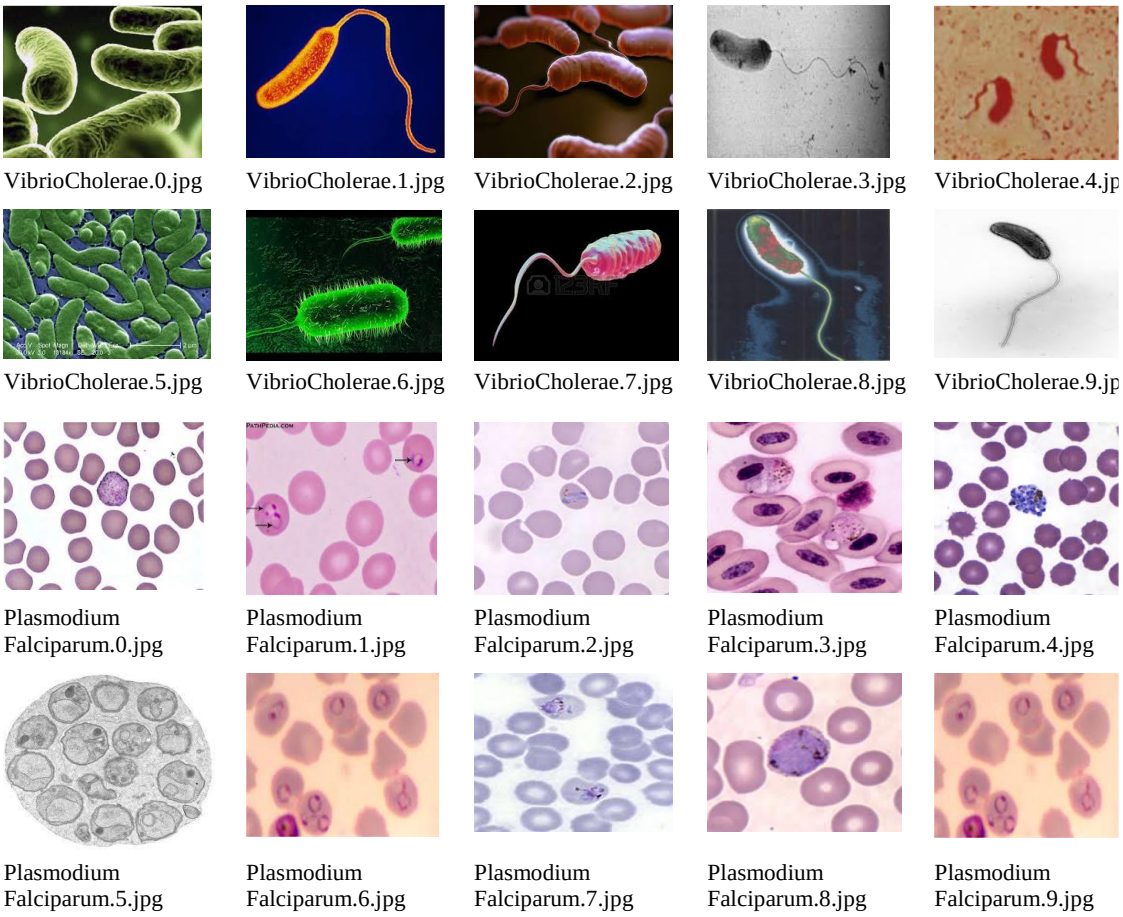


Figure 4.12 – Quelques images de notre base d'apprentissage

4.3.1.2 Constitution de la base de validation

Cette base de validation est constituée de la même manière que la base d'apprentissage, c'est-à-dire un premier dossier pour les images vibrio choléra et un deuxième dossier pour les images plasmodium falcparum. Egalement, les images sont labélisées de la même manière. Cependant, la base de validation est constituée de 100 images soit 50 images de Vibrio choléra et 50 images de Plasmodium falciparum.

4.3.1.3 Application de l'apprentissage par transfert : Extraction de caractéristiques

L'entraînement d'un réseau de neurones convolutifs CNN à partir de zéro est très coûteux en matériels (GPU, Disque dur de grande capacité de stockage…) et en temps d'exécution. De plus, lorsqu'on ne dispose pas d'un grand volume de données ; il est plus conseillé d'utiliser l'apprentissage par transfert afin d'éviter le sur-apprentissage caractérisé par un nombre de paramètres à entraîner très élevé par rapport au volume de données disponibles (Chollet, 2018). Donc l'Apprentissage par transfert nous permet de gagner du temps et de profiter de l'expérience des autres, tout en veillant à réduire le sur-apprentissage. L'Apprentissage par transfert consiste à se servir des connaissances acquises par un réseau de neurones (un réseau préentraîné) lors de la résolution d'un problème pour résoudre un autre plus ou moins similaire (Monasse and Nadjahi, 2018). Ci-dessous, nous allons présenter l'architecture VGG-16.

- **Architecture VGG-16**

L'architecture VGG-16 est constituée de plusieurs couches, dont 13 couches de convolution et 3 couches entièrement connectées (Figure 4.13). Elle prend en entrée une image en couleur de taille 224 × 224 pixel et la classifie dans une des 1000 classes d'Image Net. Ensuite, elle renvoie donc un vecteur de taille 1000 représentants les différentes classes. Ce vecteur contient les probabilités d'appartenance à chacune des classes.

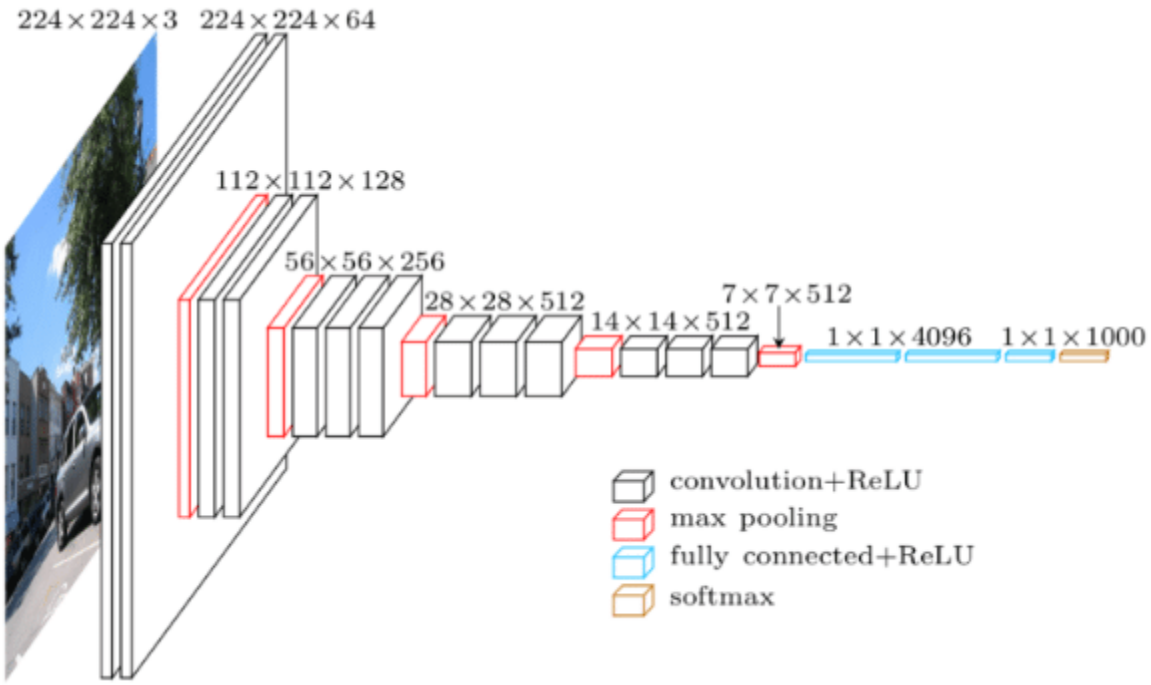


Figure 4.13 – Architecture VGG-16 (Monasse and Nadjahi, 2018)

Chaque couche de convolution utilise des filtres en couleurs de taille 3 ×3 pixel, déplacés avec un pas de 1 pixel. La marge à 0 vaut 1 pixel afin que les volumes en entrée aient les mêmes dimensions en sortie. Le nombre de filtres varie selon le "bloc" dans lequel la couche se trouve. De plus, un paramètre de biais est introduit dans le produit de convolution pour chaque filtre. Chaque couche de convolution a pour fonction d'activation une ReLU. Autrement dit, il y a toujours une couche de correction ReLU après une couche de convolution. L'opération de pooling est réalisée avec des cellules de taille 2 × 2 pixel et un pas de 2 pixel, donc les cellules ne se chevauchent pas. Les deux premières couches entièrement connectées calculent chacune un vecteur de taille 4096, et chacune est suivie d'une couche ReLU. La dernière renvoie le vecteur de probabilités de taille 1000 (le nombre de classes) en appliquant la fonction softmax. De plus, ces trois couches utilisent un paramètre de biais pour chaque élément du vecteur en sortie.

Il y a deux façons d'utiliser un réseau préentraîné : l'extraction de caractéristiques et le fine-tuning (Chollet, 2018; Fotso Kamga et al., 2018; Shahin et al., 2017). L'extraction de caractéristiques consiste à entraîner uniquement le nouveau classifieur adapté au nouveau problème en lieu et place du classifieur de VGG-16, mais par contre le fine-tuning entraîne de plus le nouveau classifieur et aussi tout le réseau (c'est ce qu'on appelle le fine tuning total) ou entraîne les couches hautes (appelé fine-tuning partiel). L'implémentation du fine-tuning est gourmant en ressources matérielles et en temps d'exécution, et nécessite un plus grand volume de données que l'extraction

de caractéristiques. Ainsi, l'extraction de caractéristiques est une technique précieuse pour travailler avec de petits ensembles de données d'images. C'est pourquoi par la suite nous avons opté pour l'extraction de caractéristiques pour implémenter notre apprentissage par transfert.

- **Extraction de caractéristiques**

L'extraction de caractéristiques consiste à utiliser des caractéristiques du réseau préentraîné pour représenter les images du nouveau problème. Pour cela, on retire la dernière couche entièrement connectée et on fixe tous les autres paramètres (Figure 4.14).

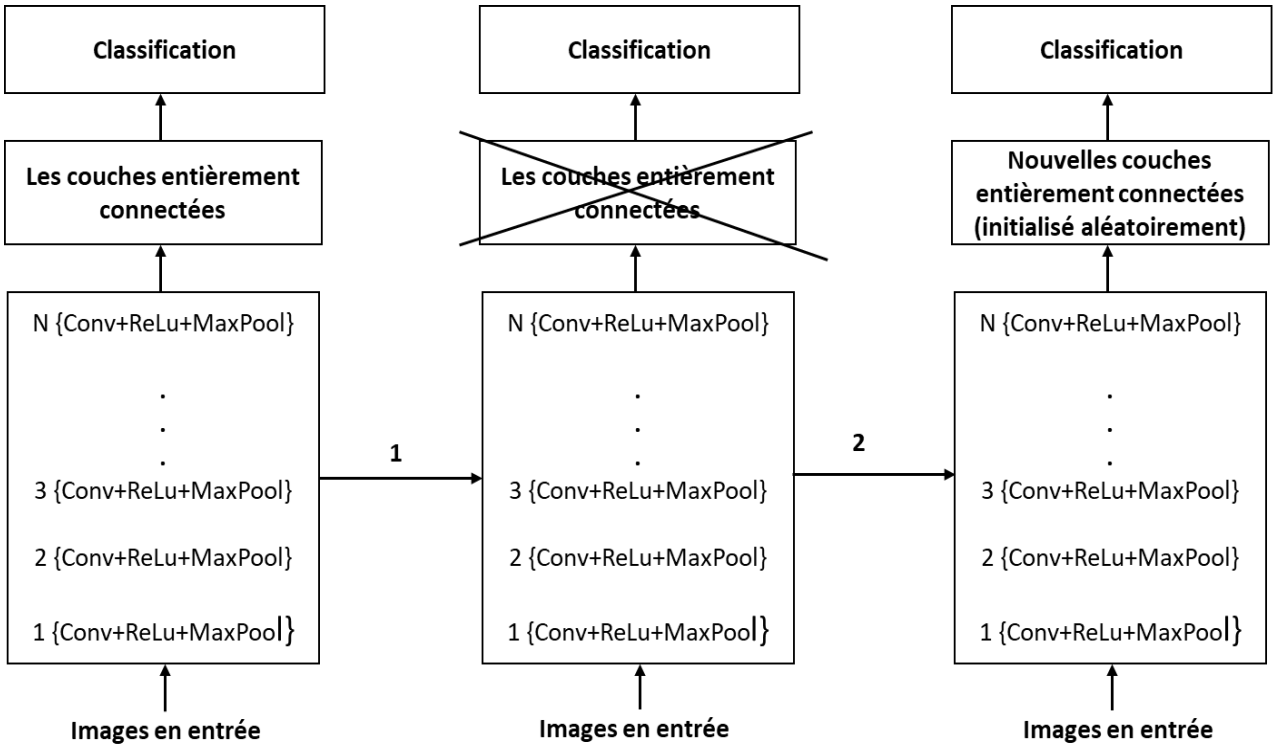


Figure 4.14 – Processus d'extraction de caractéristiques (Chollet, 2018)

Ce réseau tronqué va ainsi calculer la représentation de chaque image en entrée à partir des caractéristiques déjà apprises lors du préentraînement (Tableau 4.2). On entraîne alors un classifieur, initialisé aléatoirement, sur ces représentations pour résoudre le nouveau problème. Par exemple, pour le cas de VGG-16, ça consiste tout simplement à remplacer les couches complètement connectées par les couches complètement connectées du nouveau classifieur adapté au nouveau problème.

Tableau 4.2 – Représentation de l'architecture VGG-16 sans les dernières couches entièrement connectées

Layer (type)	**Output Shape**	**Param #**
input_1 (InputLayer)	(None, 150, 150, 3)	0
block1_conv1 (Conv2D)	(None, 150, 150, 64)	1792

block1_conv2 (Conv2D)	(None, 150, 150, 64)	36928
block1_pool (MaxPooling2D)	(None, 75, 75, 64)	0
block2_conv1 (Conv2D)	(None, 75, 75, 128)	73856
block2_conv2 (Conv2D)	(None, 75, 75, 128)	147584
block2_pool (MaxPooling2D)	(None, 37, 37, 128)	0
block3_conv1 (Conv2D)	(None, 37, 37, 256)	295168
block3_conv2 (Conv2D)	(None, 37, 37, 256)	590080
block3_conv3 (Conv2D)	(None, 37, 37, 256)	590080
block3_pool (MaxPooling2D)	(None, 18, 18, 256)	0
block4_conv1 (Conv2D)	(None, 18, 18, 512)	1180160
block4_conv2 (Conv2D)	(None, 18, 18, 512)	2359808
block4_conv3 (Conv2D)	(None, 18, 18, 512)	2359808
block4_pool (MaxPooling2D)	(None, 9, 9, 512)	0
block5_conv1 (Conv2D)	(None, 9, 9, 512)	2359808
block5_conv2 (Conv2D)	(None, 9, 9, 512)	2359808
block5_conv3 (Conv2D)	(None, 9, 9, 512)	2359808
block5_pool (MaxPooling2D)	(None, 4, 4, 512)	0

Total des paramètres : 14,714,688
Paramètres entraînés : 14,714,688
Paramètres non entraînés : 0

Dans ce Tableau 4.2, on peut constater l'absence des trois dernières couches entièrement connectées qui seront remplacées par le nouveau classifieur pour résoudre notre problème de deux classes après la phase d'extraction des caractéristiques.
Le code de la fonction d'extraction des caractéristiques en python est présenté en annexe B.2 (Page 117.
Ensuite, cette fonction d'extraction des caractéristiques est utilisée pour extraire les caractéristiques des bases données d'apprentissage, de validation et de test comme. Le code python pour application

de la fonction d'extraction des caractéristiques sur les bases d'apprentissage, de validation et de test est présenté en annexe B.3 (Page 117).
Après la phase d'extraction des caractéristiques, il ne reste plus qu'à ajouter le nouveau classifieur au modèle tronqué. Pour la compilation du nouveau modèle, nous avons utilisé un optimiseur, une fonction de perte et des métriques pour évaluer la précision du modèle.
Dans un premier temps la fonction de perte est utilisée pour mesurer la variation entre les valeurs prédites et les valeurs réelles. Pour cela nous avons utilisé le binary_crossentropy puisqu'il s'agit d'un problème de classification binaire (Ketkar, 2017). En deuxième lieu, il faut réajuster le modèle en fonction des valeurs de la fonction de perte. Ce réajustement consiste à mettre à jour les paramètres tels que les poids du réseau pendant la phase d'entraînement. Ce travail est fait par l'optimiseur. Parmi les optimiseurs, nous avons l'optimiseur RMSProp (Tieleman and Hinton, 2012), l'optimiseur descente de gradient stochastique (SGD) et aussi SGD avec moment (Ruder, 2016), l'optimiseur Adam (Kingma and Ba, 2014), l'optimiseur AdaGrad (Duchi et al., 2011), etc. Egalement comme il s'agit d'une classification binaire, l'optimiseur RMSProp s'avère plus performant avec une convergence plus rapide.

Le code Python pour l'ajout du nouveau classifieur est présenté en annexe B.4 (Page 117). Ce nouveau classifieur entièrement connecté est initialisé de manière aléatoire (Tableau 4.3). Il doit faire référence au nombre de catégories à classifier. Comme, il s'agit d'un problème de classification binaire. Nous avons utilisé la classe Dense de Keras pour créer les couches entièrement connectées qui reçoivent en entrée un vecteur d'une dimension. De plus, nous avons choisi la fonction d'activation Sigmoid pour le calcul des probabilités d'appartenance à une classe dont les probabilités sont entre 0 et 1. Donc, en dessous de 0.5 c'est la première classe le Vibrio choléra et en dessus de 0.5 c'est la deuxième classe Plasmodium falciparum.

Tableau 4.3 – Représentation de la couche entièrement connectée

Layer (type)	Output Shape	Param #
dense_1 (Dense)	(None, 256)	2097408
dropout_1 (Dropout)	(None, 256)	0
dense_2 (Dense)	(None, 1)	257

Total des paramètres : 2,097,665
Paramètres entraînés : 2,097,665
Paramètres non entraînés : 0

4.3.1.4 Modèle généré par apprentissage par transfert

Après l'étape d'apprentissage, le modèle généré est la fusion du modèle VGG-16 et le nouveau classifieur entièrement connecté. Ce modèle généré par l'apprentissage par transfert est utilisé pour évaluer la précision du modèle et également pour classifier les données de la base de test. Ainsi, le nouveau classifieur pour notre problème se limite à 2 classes (Vibrio choléra et Plasmodium falciparum) au lieu de 1000 pour ImageNet.

4.3.1.5 Soumission de la base de test

Egalement, nous avons constitué la base de test afin de la soumettre au modèle généré de l'étape 4 par les techniques l'apprentissage par transfert. Cette base de test est constituée des images d'agent pathogène Vibrio choléra et d'agent pathogène Plasmodium falciparum (Figure 4.15). Ces images seront classifiées à la sortie du déroulement de la méthodologie soit en images d'agent pathogène de Vibrio choléra ou soit en images d'agent pathogène de Plasmodium falciparum. Pour notre étude, la base de test est constituée de 80 images. Cette base de test est évaluée par le modèle obtenu. Le code python pour l'évaluation de la base de test est présenté en annexe B.5 (Page 117).

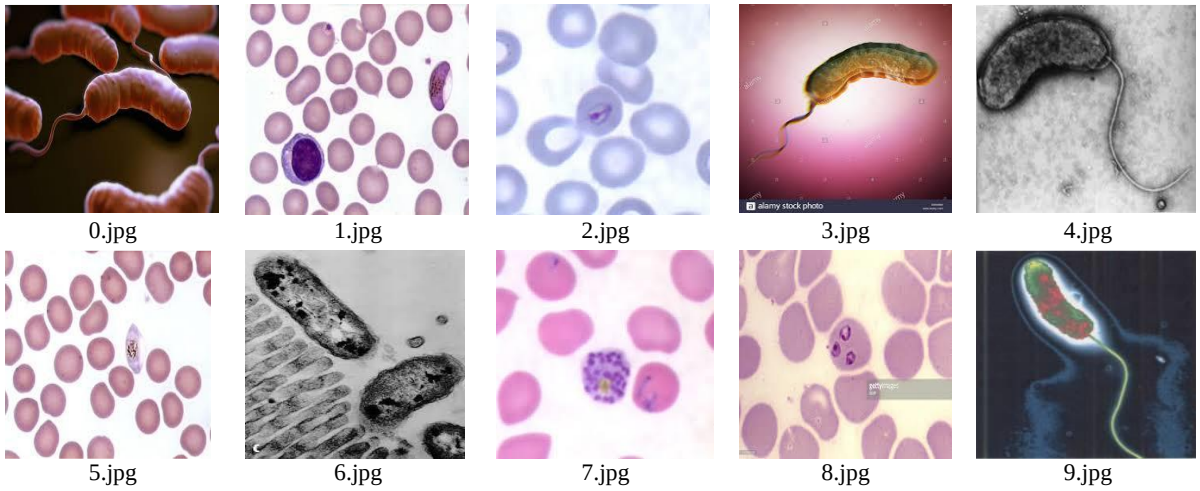


Figure 4.15 – Quelques images de la base de test

4.3.1.6 Evaluation d'images classées

La dernière étape consiste à évaluer la performance du modèle généré notamment à travers sa précision et sa perte sur les images classées. Pour ce faire, les données de validation et les données de test ont été utilisées. La Figure 4.16 représente les trois dernières étapes de l'exécution du programme et aussi le résultat du taux de réussite de classification sur les données de test. Nous avons atteint un taux de réussite de classification sur les données de validation (val_acc) de 97%, comme c'est indiqué dans la Figure 4.16 et dans le Tableau 4.4. De plus, nous avons obtenu un taux de réussite de classification sur les données de test (Test accuracy) de 97% indiqué dans la Figure 4.16.

```
Epoch 28/30
200/200 [==============================] - 1s 4ms/step - loss: 0.0299 - acc: 0.9950 - val_loss: 0.1294 - val_acc: 0.9700
Epoch 29/30
200/200 [==============================] - 1s 3ms/step - loss: 0.0281 - acc: 1.0000 - val_loss: 0.1377 - val_acc: 0.9600
Epoch 30/30
200/200 [==============================] - 1s 3ms/step - loss: 0.0231 - acc: 0.9950 - val_loss: 0.1287 - val_acc: 0.9700
50/50 [==============================] - 1s 27ms/step
Test accuracy: 0.97000002861
```

Figure 4.16 – Evaluation du modèle sur la base d'évaluation et la base de test

Egalement, il est à noter qu'il existe d'autres paramètres pour évaluer la performance d'un modèle notamment la précision, le rappel, la moyenne harmonique appelée le fmeasure ou F-score et le fbeta_score (Chollet, 2015; Sokolova and Lapalme, 2009). Toutefois, ces paramètres sont plus significatifs pour une tache multi label et sur les données déséquilibrées.

Considérons une matrice de confusion avec les éléments suivants : Vrais Positifs (VP), Vrais Négatifs (VN), Faux Positifs (FP) et Faux Négatifs (FP). L'exactitude de la classification étant la proportion d'éléments correctement prédits est donnée par la formule suivante :

$$Accuracy = \frac{VP + VN}{VP + VN + FP + FN} \qquad (4.16)$$

La précision détermine la proportion de prédictions correctes parmi les prédictions réalisées :

$$Précision = \frac{VP}{VP + FP} \qquad (4.17)$$

En ce qui concerne le rappel, il détermine la proportion de prédictions correctes réalisées parmi toutes prédictions correctes réalisables.

$$Rappel = \frac{VP}{VP + FN} \qquad (4.18)$$

Le fmeasure est la combinaison de la précision et du rappel comme suit.

$$\text{Fmeasure } = 2 * \frac{Précision * Rappel}{Précision + Rappel} \qquad (4.19)$$

Le fbeta_score est la moyenne harmonique pondérée de la précision et du rappel.

$$\text{Fbeta_score } = \frac{(1 + \beta)^2 * (Précision * Rappel)}{(\beta^2 * Précision + Rappel)} \qquad (4.20)$$

En résumé, plus ces paramètres sont proches de 1 plus le modèle est bon.

Tableau 4.4 – La dernière étape de l'entraînement du modèle

Epoch 30/30 - Dernière itération
Epoch 30/30 200/200 [==========] - 1s 3ms/step - loss: 0.0231 - acc: 0.9950 - val_loss: 0.1287 - **val_acc: 0.9700**

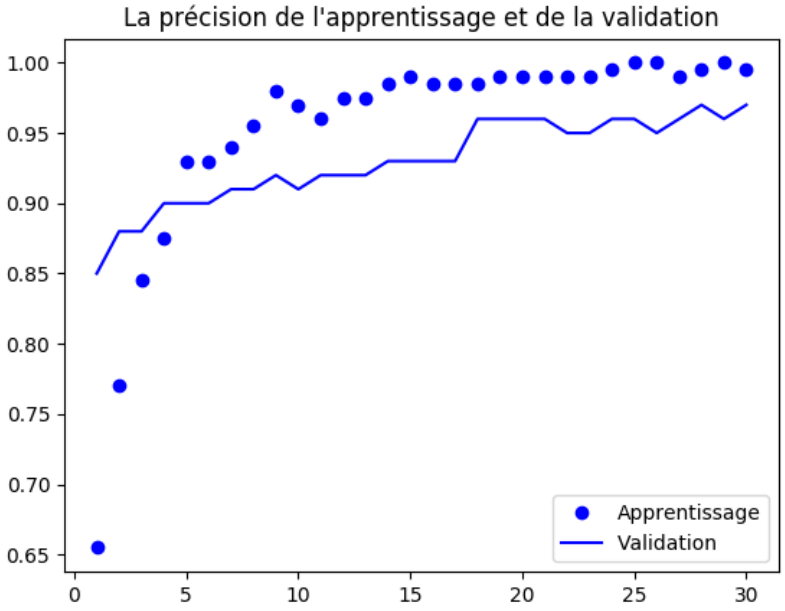


Figure 4.17 – La courbe représentative de la précision

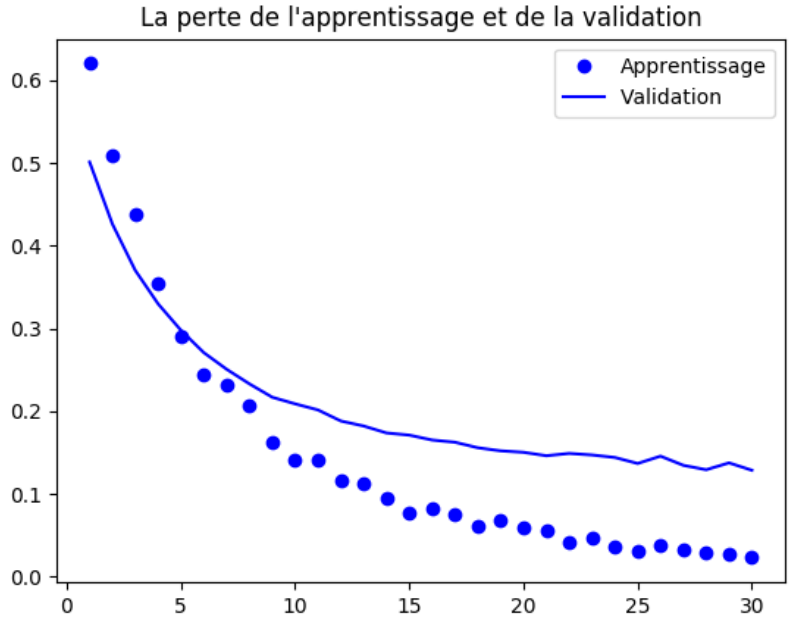


Figure 4.18 – La courbe représentative de la perte

La Figure 4.17 et la Figure 4.18 nous indiquent respectivement un taux de réussite de classification pour la base de validation de 97% et une perte de validation de 0. 1287 suite à la classification des images d'agent pathogène du choléra et des images d'agent pathogène du paludisme. Le taux de réussite du modèle d'apprentissage est de 99% donc légèrement supérieur au taux de réussite sur la base de validation qui est 97%. Donc, on peut est aussi performant sur les données d'entraînement que sur les données de validation. Ce qui est le cas favorable en apprentissage par

transfert, signifiant que le modèle reconnaît aussi bien les images qu'il connaît que celles qu'il n'a jamais vues. On peut également constater que la précision sur la base de validation et la précision sur la base de test sont égales car les jeux de données en réalité sont assez similaires.
Nous constatons que l'approche apprentissage par transfert permet d'améliorer considérablement la précision de la classification des images des agents pathogènes. Car nous avions opté dans un premier temps pour la mise en place de notre propre architecture qui nous a donné 94% de réussite de classification or avec l'approche de l'apprentissage par transfert nous sommes à 97%.
Ci-dessous dans la Figure 4.19, nous présentons le résultat de classification de quinze images aléatoires prises dans la base de test.

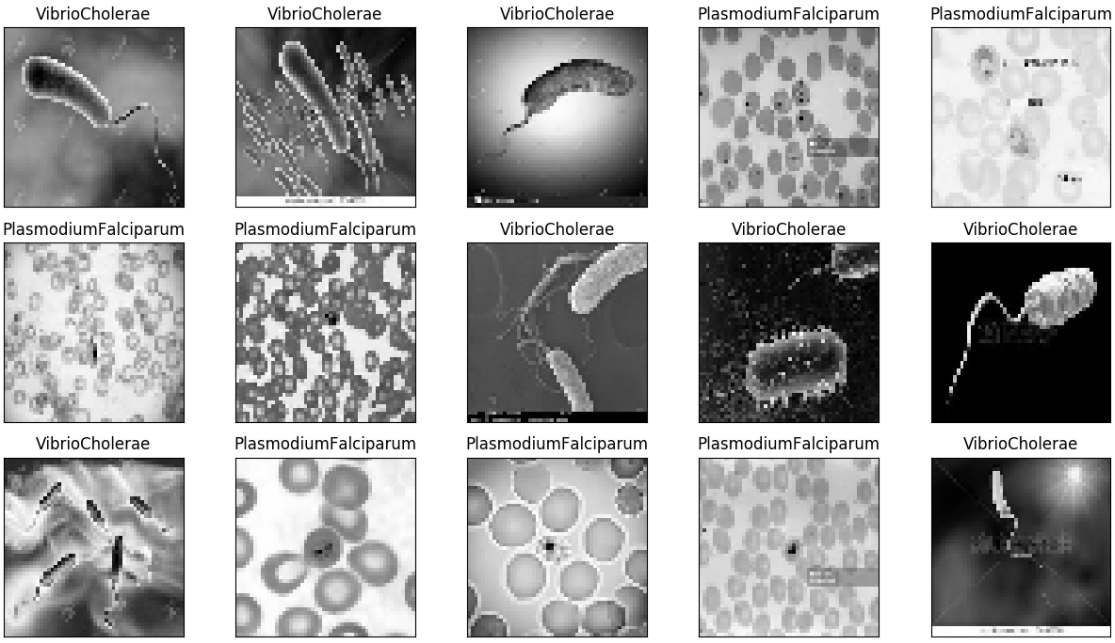


Figure 4.19 – Résultat de classification de quelques images de test

4.3.2 Logiciel utilisé : framework Keras

Keras est une librairie open source populaire des réseaux de neurones, développée en Python par François Chollet. Keras peut s'exécuter sur TensorFlow, Microsoft Cognitive Toolkit et Theano. Keras a été conçu afin de permettre une implémentation rapide des réseaux de neurones particulièrement les réseaux de neurone profond. Son point fort est la disponibilité des fonctionnalités prêtes à l'emploi, sa modularité et son extensibilité. Donc Keras est beaucoup plus une interface sur les programmes de l'apprentissage automatique, car elle offre un haut niveau d'abstraction afin de faciliter la programmation des réseaux de neurones. Keras supporte l'implémentation des réseaux de neurones convolutifs, des réseaux de neurones récurrents et peut s'exécuter sur un CPU tout comme un GPU.

4.4 Discussions

Les réseaux de neurones convolutifs sont devenus une référence pour l'analyse des images médicales notamment les images des agents pathogènes. Dans ce chapitre, nous avons opté pour l'extraction de caractéristiques l'une des techniques de l'apprentissage par transfert, car nous ne disposons pas d'un grand volume de données pour suffisamment apprendre le réseau et aussi pour éviter le sur-apprentissage. Cette approche nous a permis d'atteindre les 97% de taux de réussite

de classification des images des agents pathogènes. Toutefois, on aurait également pu utiliser la méthode de fine-tuning partiel c'est-à-dire entraîner les couches hautes avec les données à notre disposition, car la base de données d'ImageNet ne dispose pas suffisamment d'images d'agents pathogènes pour extraire assez de caractéristiques semblables. Par ailleurs, nous avons constaté que les images de mauvaise qualité ont un impact sur la précision de la classification. C'est pourquoi, lors de la phase de collecte du jeu de données des images à partir Google image, nous avons supprimé les images non significatives, en particulier celles qui ne représentent pas une vue microscopique. Par la suite, il s'agit d'intégrer cette approche dans les microscopes afin de les doter d'une certaine intelligence notamment être en mesure de reconnaître et classer automatiquement les agents pathogènes. En résumé, nous pouvons noter que l'objectif initial a été largement atteint dans la mesure où nous avons pu obtenir les résultats escomptés c'est-à-dire un taux de réussite de classification de 97% par application de l'apprentissage par transfert.

En outre, d'autres travaux ont été faits dans le domaine de l'imagerie médicale à l'aide des réseaux de neurones convolutifs, mais la plupart n'utilisent pas l'apprentissage par transfert. Par exemple, en diagnostic des parasites intestinaux des êtres humains, Peixinho et ses collègues ont proposé une approche d'analyse d'images optiques microscopiques pour découvrir les caractéristiques des parasites dans les images à partir d'une petite base de données (Peixinho et al., 2015). Un état de l'art intéressant sur la détection automatique des parasites du paludisme a été présenté par Rosado et ses collègues à l'aide des techniques de segmentation des images microscopiques. La segmentation d'image est le processus de partitionnement de l'image entière en plusieurs sous-parties (Rosado et al., 2016). De même, Quinn et ses collègues ont évalué la performance des réseaux neuronaux convolutifs profonds sur quelques tâches des microscopes (Quinn et al., 2016). Leur architecture CNN ne comprend que deux couches de convolution et s'entraîne également sur les images segmentées. Egalement, Shahin et ses collègues ont proposé un nouveau système d'identification des images contenant des globules blancs à l'aide des réseaux de neurones profonds avec une précision de classification de 96.1% (Shahin et al., 2017).
Notre approche basée sur l'apprentissage par transfert permet de classer plusieurs types de pathogènes épidémiologiques à condition de lui fournir la base de données des images correspondantes. Les techniques de la vision assistée par ordinateur en l'occurrence leur déploiement sur les microscopes constituent un enjeu majeur pour les pays en voie de développement. Cela permettra de combler non seulement le manque de techniciens qualifiés en manipulation microscopique, mais aussi de faciliter les analyses médicales (microscopiques) surtout en période de crises épidémiques.

4.5 Conclusion

Les réseaux de neurones convolutifs sont le choix idéal pour les tâches de vision par ordinateur. Cependant, il n'est recommandé d'entraîner le ConvNet à partir de zéro sur les petites bases de données au risque de faire du sur-apprentissage. Pour éviter cela, le mieux c'est de réutiliser des réseaux existants préentraînés sur de grandes bases de données telles qu'ImageNet, MNIST, etc. C'est ce qu'on appelle l'apprentissage par transfert. Dans ce chapitre, nous avons présenté une approche basée sur l'extraction de caractéristiques, l'une des deux techniques de l'apprentissage par transfert, pour la classification des images d'agents pathogènes épidémiques. Il s'agissait de distinguer une image microscopique contenant le virus de l'épidémie de choléra appelé Vibrio choléra d'une image microscopique contenant le virus de l'épidémie du paludisme appelé

Plasmodium falciparum. Des résultats intéressants ont été obtenus soit un taux de réussite de classification de 97%, sur 100 images de données de validation et 80 images de données de test.

Dans les CNN, la transférabilité multi sources reste un problème ouvert. Par conséquent, une perspective future intéressante consisterait à mener une réflexion sur l'extension du cadre méthodologique proposé pour incorporer simultanément plusieurs CNN sources pendant l'apprentissage par transfert (Afridi et al., 2018). Dans ce contexte, il sera également intéressant d'étudier comment différentes approches de sélection de caractéristiques peuvent être comparées théoriquement et expérimentalement avec des sources multiples.

Les principales contributions de ce chapitre sont les suivantes :

- premièrement, un soutien au processus de prise de décision, car il contribue à améliorer la gestion des crises épidémiques en économisant du temps et de l'argent dans les analyses microscopiques ;
- deuxièmement, l'intégration de cette solution dans les futurs microscopes pourrait permettre d'avoir des microscopes intelligents ;
- enfin, les microscopes intelligents peuvent aider à combler les lacunes des spécialistes des manipulations microscopiques et le manque de plateaux techniques.

Conclusion générale

Bilan.....109
Contributions 109
Perspectives 111

Conclusion générale

Bilan

Notre étude sur la modélisation des informations et l'extraction des connaissances pour la gestion des crises nous a permis d'aborder trois grandes thématiques. Ces thématiques portent notamment sur la Modélisation des informations pour la Gestion de crise avec la télémédecine, les techniques de Data Mining sur les images satellitaires pour découvrir les zones à risque d'épidémie et la Classification des images d'agents pathogènes épidémiques à l'aide des réseaux de neurones convolutifs. Pour mener à bien ce travail de recherche, nous avons commencé par un état de l'art sur les thèmes clés à savoir la Télémédecine, la Télédétection, le Data Mining et le Deep learning. Ce qui nous a permis de comprendre d'une part, l'ensemble des mots clés en détail et d'autre part, de délimiter le sujet.

Contributions

Les trois principales contributions de ce travail de recherche sont résumées ci-dessous relativement aux trois thématiques présentées ci-dessus :

1. Modélisation des informations pour la Gestion de crise avec la télémédecine

Cette première contribution concerne le chapitre 2 de la thèse. En effet, nous avons proposé une méthodologie en trois grandes phases et principalement axée sur l'interopérabilité des actes de télémédecine. L'objectif général consistait à mettre à la disposition des gestionnaires de crise un système d'aide à la décision informatisé à l'aide d'une interopérabilité des systèmes d'information de télémédecine (permettant la réalisation des différents actes de télémédecine) dans le but de favoriser la coopération des professionnels de la santé et aussi aider la décision médicale. Il s'agit d'une interopérabilité conceptuelle syntaxique (format des données) ou sémantique (signification des données) et d'une interopérabilité technologique (plateformes et infrastructures de stockage, traitement et transport des données). De plus, la méthodologie proposée repose sur une formalisation mathématique de la gestion de crise. Cette formalisation concerne tous les actes de télémédecine en spécifiant les paramètres d'entrée, de sortie, d'entrée-sortie, les conditions préalables à remplir avant d'effectuer un acte de télémédecine et les résultats éventuels après chaque acte de télémédecine. Par conséquent, un algorithme générique a été présenté pour permettre une implémentation concrète d'une plateforme informatique pour la gestion de crise. Cette plateforme informatique favorisera l'échange, le partage et le traitement d'informations pour l'efficacité de la collaboration à distance des professionnels de la santé pendant la phase de transport et les soins médicaux des patients. Les principaux résultats de cette étude sont: (i) une approche méthodologique pour une meilleure protection des villes contre les grandes vagues généralement provoquées par des tsunamis (ii) des systèmes d'information avec des composants interopérables pour une meilleure prise en charge médicale des victimes (iii) un cadre formel de gestion des crises pour capitaliser les expériences et exploiter les leçons apprises.

2. Les techniques de Data Mining sur les images satellitaires pour découvrir les zones à risque d'épidémie.

La deuxième contribution fait référence au chapitre 3 de la thèse. Cette contribution s'inscrit dans le cadre de la prévention d'une crise d'épidémie par les techniques d'extraction de connaissances à partir de données (ECD) notamment le Data mining sur les images satellitaires pour la découverte des zones à risques d'épidémie. Elle porte principalement sur la méthodologie de six phases que nous avons proposée. Cette méthodologie peut se résumer comme suit. Dans un premier temps, nous avons appliqué la classification supervisée (l'algorithme du maximum de vraisemblance) sur les images satellitaires de la même zone, mais à différentes périodes (ou à différentes dates). En second lieu, nous avons créé une base de données regroupant les données collectées sur le terrain et les données de la classification des images satellitaires afin de trouver leurs corrélations par la technique des règles d'association. Et enfin, les règles générées ont été interprétées pour produire de la connaissance compréhensible par l'utilisateur. La méthodologie proposée a été appliquée sur la problématique de l'épidémie de choléra dans la région de Mopti, au Mali. Les résultats nous indiquent qu'à 66% le taux de contamination est lié au fleuve Niger, en plus de certains facteurs sociétaux comme le jet des ordures en période hivernale. Par conséquent, nous avons pu établir le lien entre l'épidémie et son environnement d'évolution, ce qui permettra aux décideurs de mieux gérer une éventuelle crise d'épidémie. Les résultats spécifiques de la méthodologie sont : la discrétisation en phase de prétraitement améliorant la qualité des résultats obtenus, l'établissement du lien entre l'environnement et l'épidémie, et l'identification des zones les plus à risque pour la propagation et l'émergence de l'épidémie.

3. La classification des images d'agents pathogènes épidémiques à l'aide des réseaux de neurones convolutifs

Lors de l'apparition d'une situation de crise d'épidémie, il devient urgent de procéder aux analyses microscopiques des agents pathogènes responsables des épidémies pour confirmer ou non leur présence dans les prélèvements des cas suspects. Cette troisième contribution s'inscrit dans ce cadre et est relatée dans le chapitre 4 de la thèse. Pour ce faire, nous avons proposé une méthodologie en six étapes qui se focalise sur la reconnaissance d'images à l'aide des techniques du Deep learning. Il s'agit de distinguer une image microscopique contenant le virus de l'épidémie du choléra appelé Vibrio choléra d'une image microscopique contenant le virus de l'épidémie du paludisme appelé Plasmodium falciparum. A cet effet, compte tenu de la petite quantité de données disponible, nous nous sommes intéressés aux techniques de l'apprentissage par transfert qui ont été spécifiquement appliquées avec les réseaux de neurones convolutifs. Ainsi, nous avons opté pour l'architecture du réseau de neurones à 16 couches VGG-16 implémentée dans la bibliothèque Keras. Des résultats intéressants ont été obtenus, soit un taux de réussite de classification de 97% sur la base de validation et de même sur la base de test. La suite de ce travail consiste à déployer cette approche sur les microscopes. Ceci permet de rendre les microscopes plus intelligents afin de gagner non seulement en temps d'analyse, mais aussi de combler les lacunes des spécialistes en manipulation microscopique.

Perspectives

Ces travaux de recherche nous ont permis d'acquérir une vision globale sur la modélisation des informations et l'extraction des connaissances pour la gestion des crises, à travers les différentes thématiques qui ont été abordées dans le souci d'avoir une plateforme unifiée de gestion de crise, afin d'améliorer le temps de gestion et aussi le coût d'investissement (Figure 5.1).

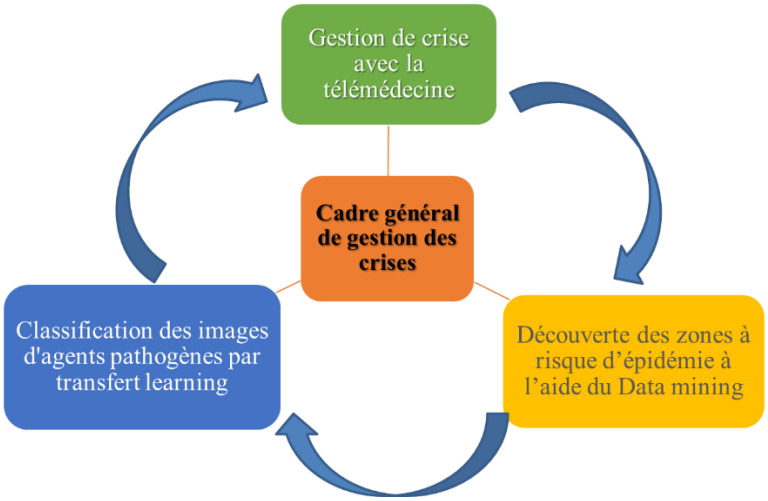


Figure 5.1 - Cadre général de gestion des crises

Cependant, certaines perspectives de recherche peuvent être identifiées en rapport aux trois thématiques abordées.

1. En ce qui concerne la modélisation des informations pour la Gestion de crise avec la télémédecine, nous envisageons d'implémenter le système d'aide à la décision que nous avons modélisé tout en décrivant les algorithmes des différentes fonctionnalités.
2. Par rapport aux techniques de Data Mining sur les images satellitaires pour découvrir les zones à risque d'épidémie, nous comptons améliorer nos résultats en utilisant des données de terrain mensuelles, hebdomadaires et aussi de granularités différentes. Au-delà de notre approche de classification basée pixel, nous envisageons d'investiguer la classification basée objet. Egalement, nous avons utilisé le même format de données c'est-à-dire les images provenant d'un même type de satellite. Il serait intéressant de considérer les images provenant de différents satellites. Le traitement d'images satellitaires de sources hétérogènes constitue également un défi considérable. En dernier lieu, nous comptons approcher les services techniques de gestion du fleuve Niger pour valider les résultats de la classification par rapport aux réalités du terrain.
3. La transférabilité multi sources reste un problème ouvert dans les CNN. Par conséquent, une perspective future intéressante à la classification des images d'agents pathogènes épidémiques consisterait à mener une réflexion sur l'extension du cadre méthodologique proposé pour incorporer simultanément plusieurs CNN sources pendant l'apprentissage par transfert (Afridi et al., 2018). Dans ce contexte, il sera également intéressant d'étudier comment différentes approches de sélection de caractéristiques peuvent être comparées théoriquement et expérimentalement avec des sources multiples.

Annexe A : Algorithme de gestion de crise avec les actes de télémédecine

Nous présentons les principales étapes de l'algorithme comme suit : (i) l'étape 1 fournit des informations de définition de crise (ii) l'étape 2 fournit des informations sur l'analyse de crise et (iii) l'étape 3 présente le processus de prise en charge des patients de la phase de transport du patient jusqu'à son rétablissement à l'aide des actes de télémédecine (téléconsultation, télésurveillance, téléassistance, téléexpertise, et réponse médicale). Les différentes fonctions et les paramètres des actes de télémédecine sont alors définis. La complexité de l'analyse théorique des étapes 1 et 2 est négligeable face aux deux boucles imbriquées dans l'étape 3, ce qui ramène la complexité de cet algorithme à l'ordre de (O (N^2)) pour N entrées dans la gestion de crise considérée.

```
Fonction Teleconsultation (MedecinTraitant, Patient, DossierMedicalPatient) /* Définition de la
  fonction de téléconsultation. */
Début
    Si (Patient.casVulnerable=oui && PlateformeTM.operationnelle=oui &&
        DossierMedicalPatient.disponible=oui) alors // Les préconditions doivent être satisfaites
            PrescriptionMedicalePatient ← Realisation_Acte_de_Teleconsultation() ; /* Le médecin
            traitant réalise la téléconsultation sur le patient à l'aide de la Plateforme TIC */
            AutreActedeTelemedecine←Telesurveillance||Teleassistance||Teleexpertise||Null;
            // Indiquer si un autre acte de télémédecine est nécessaire.
            RapportTC←Rapport ; // Rapport du processus de téléconsultation
    Fin si
    Retourne (PrescriptionMedicalePatient, AutreActedeTelemedecine, RapportTC) ;

Fin
```

```
Fonction Telesurveillance (MedecinTraitant, Patient, DonneesSurveillanceMedicale) /* Définition de la
  fonction de télésurveillance*/
Début
    Si (Patient.casVulnerable=oui && PlateformeTM.operationnelle=oui) alors /* Les préconditions
        doivent être satisfaites */
            InterpretationResultat←Realisation_Acte_de_Telesurveillance(); /*Le médecin traitant
            effectue la télésurveillance sur le patient à l'aide de la Plateforme TIC et de données de
            surveillance */
            AutreActedeTelemedecine ←Teleconsultation||Teleassistance||Teleexpertise||Null;
            // Indiquer si un autre acte de télémédecine est nécessaire.
            RapportTM←Rapport ; // Rapport du processus de télésurveillance.

    Fin si
     Retourne (InterpretationResultat, AutreActedeTelemedecine, RapportTM) ;
Fin
```

```
Fonction Teleexpertise (MedecinRequerant, MedecinSollicite[], DossierMedicalPatient) // Définition de
  la fonction de téléexpertise
Début
    Si (DossierMedicalPatient.disponible=oui && PlateformeTM.operationnelle=oui) alors /* Les
        préconditions doivent être satisfaites */
            AvisExpert ← Realisation_Acte_de_Teleexpertise () ; /*Le médecin demandeur et les
            médecins requis effectuent une téléexpertise sur le dossier médical du patient à l'aide de la
            Plateforme TIC. */
            AutreActedeTelemedecine ←Telesurveillance||Teleassistance||Teleconsultation||Null ;
            // Indiquer si un autre acte de télémédecine est nécessaire.
            RapportTE ←Rapport ; // Rapport du processus de Téléexpertise
        Fin si

   Retourne (AvisExpert, AutreActedeTelemedecine, RapportTE) ;
Fin
```

```
Fonction Teleassistance (MedecinRequerant, MedecinSollicite[], ActeMedical) /*Définition de la fonction de téléassistance */
Début
    Si (DossierMedicalPatient.disponible=oui && PlateformeTM.operationnelle=oui) alors
        // Les préconditions doivent être satisfaites //
        AssistanceExpert ← Realisation_Acte_de_Teleassistance () ; /* Le médecin demandeur et les médecins requis effectuent une téléassistance lors de l'acte médical à l'aide de la Plateforme TIC. */
        AutreActedeTelemedecine ←Telesurveillance||Teleconsultation||Teleexpertise||Null ;
        // Indiquer si un autre acte de télémédecine est nécessaire.
        RapportTA ←Rapport ; // Rapport du processus de Téléassistance
    Fin si
    Retourne (AssistanceExpert, AutreActedeTelemedecine, RapportTA) ;
Fin
```

Algorithme : ***Gestion_de_crise_à_distance***

Etape 1 : Définition_de_Crise // Fournir les informations de définition de crise

Localisation ← où ;
Type_de_crise ← de_quoi_il_s'agit ;
Niveau_de_Gravite ← les_enjeux ;

Fin

Etape 2 : Analyse_de_Crise // Fournir des informations d'analyse de crise

Objectifs ← définir_les_objectifs ;
Ressources_disponibles ← lister_les_Ressources_disponibles ;
Limites ← déterminer_les_limites ;

Fin

Etape 3 : Gestion_de_Crise

Pour tout Patient Pi ϵ Pn **faire // Pi** représente un objet au point de vue structure des données

Structure_de_Santé← RéponseMédicale(Pi) ; // phase de transport

Pi←Téléconsultation (MédecinTraitant, Pi, Pi.DossierMédical) ; /* Le patient subit une téléconsultation pour confirmer ou non s'il a besoin de soins médicaux. */

Pi.casConfirme←oui||non ; // résultat de la téléconsultation pour des soins médicaux

Tantque (Pi. casConfirme =oui && Pi.retabli=non) **faire /*** pour chaque cas confirmé et pas encore rétabli, le même processus est appliqué : d'abord le soin médical et en second lieu, le patient peut subir d'autres actes de télémédecine */

Pi←Soin_Medical (*Pi.PrescriptionMedicale);*

Cas_Où (Pi.Besoin_autre_acte_Telemedecine) /* Si c'est nécessaire de réaliser un autre acte de Télémédecine selon l'état du patient*/

Cas Telesurveillance **alors**
Pi← Telesurveillance (MedecinTraitant, Pi, Pi.DonneesSurveillance) ;

Cas Teleexpertise **alors**
Pi←Teleexpertise (MedecinRequerant, MedecinSollicite[], Pi.DossierMedical);

Cas Teleassistance **alors**
Pi←Teleassistance(MedecinRequerant, MedecinSollicite[], ActeMedical);

ParDefaut : Pi.retabli←oui|non|decede ;

FinCasOù

Si (Pi.retabli=non) **alors**
Pi←Teleconsultation (MedecinTraitant, Pi, Pi. DossierMedical) ;
Pi. casConfirme ←oui ;

FinSi

FinTantque /* La boucle est arrêtée lorsque le patient est rétabli ou s'il est décédé.*/

FinPour

Fin

Annexe B : Codes de l'apprentissage par transfert sur les images d'agents pathogènes épidémiques

B.1 Algorithme de la phase d'apprentissage d'un réseau de neurones convolutifs

```
Tantque critère d'arrêt non atteint
    Initialiser les di

    Pour chaque exemple :
          Calculer la sortie si

          Pour chaque poids des neurones de sortie :
                di = si * (1 - si) * (yi - si)
          FinPour

          Pour chaque poids des neurones cachés :
                sum = 0
                Pour chaque lien vers le neurone de sortie k :
                      sum += dk * wi vers k
                FinPour
                di = oi * (1 - oi) * sum
          FinPour

          Pour chaque poids du réseau :
                Si lien vers neurone de sortie :
                      wi += taux * di * oi
                Sinon
                      wi += taux * di * si
                FinSi
         FinPour
    FinPour

    Si besoin, modification du taux
FinTantque
```

B.2 : Code python de la fonction d'extraction des caractéristiques

```
def extract_features(chemin, nombre_exemple):
      features = np.zeros(shape=( nombre_exemple, 4, 4, 512))
      labels = np.zeros(shape=( nombre_exemple))
      generator = datagen.flow_from_directory(
            chemin,
            target_size=(150, 150),
            batch_size=batch_size,
            class_mode='binary')
      i = 0
      for inputs_batch, labels_batch in generator:
            features_batch = conv_base.predict(inputs_batch)
            features[i * batch_size : (i + 1) * batch_size] = features_batch
            labels[i * batch_size : (i + 1) * batch_size] = labels_batch
            i += 1
            if i * batch_size >= nombre_exemple:
                  break
      return features, labels
```

B.3 : Code Python de l'application de la fonction d'extraction des caractéristiques sur les bases d'apprentissage, de validation et de test

```
train_features, train_labels = extract_features(base_apprentissage,
nombre_exemple)

validation_features, validation_labels =
extract_features(base_validation_dir, nombre_exemple)

test_features, test_labels = extract_features(base_test, nombre_exemple)
```

B.4 : Code Python pour l'ajout du nouveau classifieur

```
model = models.Sequential()
model.add(layers.Dense(256, activation='relu', input_dim=4 * 4 * 512))
model.add(layers.Dropout(0.5))
model.add(layers.Dense(1, activation='sigmoid'))

model.compile(optimizer=optimizers.RMSprop(lr=2e-5),
loss='binary_crossentropy', metrics=['accuracy', precision, recall,
fmeasure, fbeta_score])

history = model.fit(train_features, train_labels, epochs=30, batch_size=20,
      validation_data=(validation_features, validation_labels))
```

B.5 : Code python de l'évaluation de la base de test

```
test_loss, test_acc = model.evaluate(test_features, test_labels, steps=50)
print('Test accuracy:', test_acc)
```

Bibliographie

Abkar, A.-A., Sharifi, M.A., Mulder, N.J., 2000. Likelihood-based image segmentation and classification: a framework for the integration of expert knowledge in image classification procedures. Int. J. Appl. Earth Obs. Geoinformation 2, 104–119.

Aboueljinane, L., Sahin, E., Jemai, Z., 2013. A review on simulation models applied to emergency medical service operations. Comput. Ind. Eng. 66, 734–750. https://doi.org/10.1016/j.cie.2013.09.017

Afridi, M.J., Ross, A., Shapiro, E.M., 2018. On automated source selection for transfer learning in convolutional neural networks. Pattern Recognit. 73, 65–75. https://doi.org/10.1016/j.patcog.2017.07.019

Ahmad, A., Quegan, S., 2012. Analysis of maximum likelihood classification on multispectral data. Appl. Math. Sci. 6, 6425–6436.

Alrababah, M.A., Alhamad, M.N., 2006. Land use/cover classification of arid and semi-arid Mediterranean landscapes using Landsat ETM. Int. J. Remote Sens. 27, 2703–2718. https://doi.org/10.1080/01431160500522700

Ando, M., Kitamura, A., Tu, Y., Ohashi, Y., Imai, T., Nakamura, M., Ikuta, R., Miyairi, Y., Yokoyama, Y., Shishikura, M., 2018. Source of high tsunamis along the southernmost Ryukyu trench inferred from tsunami stratigraphy. Tectonophysics 722, 265–276. https://doi.org/10.1016/j.tecto.2017.11.007

Arentze, T.A., 2009. Spatial Data Mining, Cluster and Pattern Recognition, in: Kitchin, R., Thrift, N. (Eds.), International Encyclopedia of Human Geography. Elsevier, Oxford, pp. 325–331. https://doi.org/10.1016/B978-008044910-4.00524-1

Arnott, D., Pervan, G., 2005. A critical analysis of decision support systems research. J. Inf. Technol. 20, 67–87. https://doi.org/10.1057/palgrave.jit.2000035

ASIP, S., 2012. Cadre d'Interopérabilité des Systèmes d'Information de Santé (CI-SIS) | esante.gouv.fr, le portail de l'ASIP Santé [WWW Document]. URL http://esante.gouv.fr/services/referentiels/referentiels-d-interoperabilite/cadre-d-interoperabilite-des-systemes-d (accessed 12.18.17).

Asokan, G.V., Asokan, V., 2015. Leveraging "big data" to enhance the effectiveness of "one health" in an era of health informatics. J. Epidemiol. Glob. Health 5, 311–314. https://doi.org/10.1016/j.jegh.2015.02.001

Balogh, W., Canturk, L., Chernikov, S., Doi, T., Gadimova, S., Haubold, H., Kotelnikov, V., 2010. The United Nations Programme on Space Applications: Status and direction for 2010. Space Policy 26, 185–188. https://doi.org/10.1016/j.spacepol.2010.03.008

Barbedo, J.G.A., 2013. Digital image processing techniques for detecting, quantifying and classifying plant diseases. SpringerPlus 2, 660.

Barge, H., Boutteau, B., Calinaud, C., Catz, S., Chamberland, G., 2012. Plan national de déploiement de la télémédecine, Recommandations pour la mise en ø euvre d'un projet de télémédecine. Rapp. DGOS.

Ben Othman, S., Zgaya, H., Dotoli, M., Hammadi, S., 2017. An agent-based Decision Support System for resources' scheduling in Emergency Supply Chains. Control Eng. Pract. 59, 27–43. https://doi.org/10.1016/j.conengprac.2016.11.014

Bénaben, F., 2016. A Formal Framework for Crisis Management describing information flows and functional structure. Procedia Eng. 159, 353–356.

Benaben, F., Truptil, S., Lauras, M., Salatge, N., 2015. Management of collaborative behavior through a service-oriented mediation system: the case of crisis management, in: Services Computing (SCC), 2015 IEEE International Conference On. IEEE, pp. 554–561.
Bicout, D.J., Vautrin, M., Vignolles, C., Sabatier, P., 2015. Modeling the dynamics of mosquito breeding sites vs rainfall in Barkedji area, Senegal. Ecol. Model. 317, 41–49. https://doi.org/10.1016/j.ecolmodel.2015.08.027
Binacchi, F., 2017. Cannes: La cité des festivals en «alerte tsunami» (pour un exercice), c'est une première en France [WWW Document]. URL http://www.20minutes.fr/planete/2160695-20171031-cannes-cite-festivals-alerte-tsunami-exercice-premiere-france (accessed 12.18.17).
Blaschke, T., 2010. Object based image analysis for remote sensing. ISPRS J. Photogramm. Remote Sens. 65, 2–16. https://doi.org/10.1016/j.isprsjprs.2009.06.004
Broge, N.H., Leblanc, E., 2001. Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. Remote Sens. Environ. 76, 156–172.
Brunet, G., Shapiro, M., Moncrieff, M., Dole, R., Kiladis, G.N., Kirtman, B., Lorenc, A., Mills, B., Morss, R., Polavarapu, S., 2009. Toward a seamless process for the prediction of weather and climate: The advancement of sub-seasonal to seasonal prediction. Bull. Am. Meteorol. Soc.
Chollet, F., 2018. Deep learning with Python. Manning Publications Co, Shelter Island, New York.
Chollet, F., 2015. Keras: The Python Deep Learning library [WWW Document]. URL https://keras.io/ (accessed 9.14.18).
Church, P., Goscinski, A., Lefèvre, C., 2015. Exposing HPC and sequential applications as services through the development and deployment of a SaaS cloud. Future Gener. Comput. Syst. 43–44, 24–37. https://doi.org/10.1016/j.future.2014.10.001
COMITAAS, 1988. Glossaire des termes officiels de la teledetection aerospatiale, Société française de photogrammétrie et de télédétection bulletin, 0244-6014. Societe Francaise de Photogrammetrie et de Teledetection.
Conţiu, Ş., Groza, A., 2016. Improving remote sensing crop classification by argumentation-based conflict resolution in ensemble learning. Expert Syst. Appl. 64, 269–286. https://doi.org/10.1016/j.eswa.2016.07.037
Dao, S., Konaté, I., Oumar, A.A., Sacko, M., Maiga, I., Toure, K., Diarra, S., Bougoudogo, F., 2009. Les épidémies de choléra au Mali de 1995 à 2004, Cholera epidemics in Mali between 1995 and 2004. Santé Publique 21, 263–269. https://doi.org/10.3917/spub.093.0263
Davies, H.G., Bowman, C., Luby, S.P., 2017. Cholera – management and prevention. J. Infect., Hot Topics in Infection and Immunity in Children 74, S66–S73. https://doi.org/10.1016/S0163-4453(17)30194-9
Dejasmin, J., 2018. Deep learning [WWW Document]. Agence Web Mob. À Marseille Nat. Solut. URL http://www.natural-solutions.eu/blog/histoire-du-deep-learning (accessed 6.9.18).
Devlin, E.S., 2006. Crisis management planning and execution. CRC Press.
Duchi, J., Hazan, E., Singer, Y., 2011. Adaptive subgradient methods for online learning and stochastic optimization. J. Mach. Learn. Res. 12, 2121–2159.

El Merabet, Y., Meurie, C., Ruichek, Y., Sbihi, A., Touahni, R., 2015. Building roof segmentation from aerial images using a lineand region-based watershed segmentation technique. Sensors 15, 3172–3203.
Emmanuel, N.N., Loha, N., Okolo, M.O., Ikenna, O.K., 2011. Landscape epidemiology: An emerging perspective in the mapping and modelling of disease and disease risk factors. Asian Pac. J. Trop. Dis. 1, 247–250. https://doi.org/10.1016/S2222-1808(11)60041-8
Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., 1996. From Data Mining to Knowledge Discovery in Databases. AI Mag. 17, 37–37. https://doi.org/10.1609/aimag.v17i3.1230
Fotso Kamga, F.K.G., Akram, T., Laurent, B., Naqvi, S.R., Alex, M.M., Muhammad, N., 2018. A deep heterogeneous feature fusion approach for automatic land-use classification. Inf. Sci. 467, 199–218. https://doi.org/10.1016/j.ins.2018.07.074
Fournier-Viger, P., Gomariz, A., Gueniche, T., Soltani, A., Wu, C.-W., Tseng, V.S., 2014. SPMF: a Java open-source pattern mining library. J. Mach. Learn. Res. 15, 3389–3393.
Freden, S.C., Mercanti, E.P., Friedman, D.B., 1974. Volume III Discipline Summary Reports 160.
Fu, P., Weng, Q., 2016. Consistent land surface temperature data generation from irregularly spaced Landsat imagery. Remote Sens. Environ. 184, 175–187. https://doi.org/10.1016/j.rse.2016.06.019
Gagné, C., 2016. Apprentissage profond.
Girard, J.-F., Lalande, F., Salmi, L.-R., Le Bouler, S., Delannoy, L., 2006. Rapport de la mission d'évaluation et d'expertise de la veille sanitaire en France. Doc. Fr.
Gitelson, A., Merzlyak, M.N., 1994. Quantitative estimation of chlorophyll-a using reflectance spectra: Experiments with autumn chestnut and maple leaves. J. Photochem. Photobiol. B 22, 247–252. https://doi.org/10.1016/1011-1344(93)06963-4
Gitelson, A.A., Kaufman, Y.J., Stark, R., Rundquist, D., 2002. Novel algorithms for remote estimation of vegetation fraction. Remote Sens. Environ. 80, 76–87. https://doi.org/10.1016/S0034-4257(01)00289-9
González, F.E., Ruiz, J.M., Acosta, F.M., 2013. MANUEL de TÉLÉDÉTECTION SPATIALE, TELECAN. ed.
Grinblat, G.L., Uzal, L.C., Larese, M.G., Granitto, P.M., 2016. Deep learning for plant identification using vein morphological patterns. Comput. Electron. Agric. 127, 418–424.
Guiddir, M., Perez, D., Faggion, J., 2010. TIC & Santé : Au-délà de l'innovation technologique ?
Gullì, A., Pal, S., 2017. Deep learning with Keras: implement neural networks with Keras on Theano and TensorFlow. Packt, Birmingham Mumbai.
Gupta, A.K., Suresh, I.V., Misra, J., Yunus, M., 2002. Environmental risk mapping approach: risk minimization tool for development of industrial growth centres in developing countries. J. Clean. Prod. 10, 271–281. https://doi.org/10.1016/S0959-6526(01)00023-3
Halimi, M., Farajzadeh, M., Delavari, M., Arbabi, M., 2015. Developing a climate-based risk map of fascioliasis outbreaks in Iran. J. Infect. Public Health 8, 481–486. https://doi.org/10.1016/j.jiph.2015.04.024
Han, J., Pei, J., Kamber, M., 2011. Data mining: concepts and techniques. Elsevier.
Hand, D.J., 2007. Principles of Data Mining: Drug Saf. 30, 621–622. https://doi.org/10.2165/00002018-200730070-00010
Harris, J.B., LaRocque, R.C., Qadri, F., Ryan, E.T., Calderwood, S.B., 2012. Cholera. Lancet Lond. Engl. 379, 2466–2476. https://doi.org/10.1016/S0140-6736(12)60436-X

Hengl, T., Heuvelink, G.B.M., Stein, A., 2004. A generic framework for spatial prediction of soil variables based on regression-kriging. Geoderma 120, 75–93. https://doi.org/10.1016/j.geoderma.2003.08.018
HERVEG, J., 2007. Confidentialité et sécurité pour les applications de télémédecine en droit européen.
Hinton, G.E., Osindero, S., Teh, Y.-W., 2006. A fast learning algorithm for deep belief nets. Neural Comput. 18, 1527–1554.
Hirschman, I.I., Widder, D.V., 2012. The Convolution Transform. Courier Corporation.
Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sens. Environ., The Moderate Resolution Imaging Spectroradiometer (MODIS): a new generation of Land Surface Monitoring 83, 195–213. https://doi.org/10.1016/S0034-4257(02)00096-2
Huete, A.R., 1988. A soil-adjusted vegetation index (SAVI). Remote Sens. Environ. 25, 295–309. https://doi.org/10.1016/0034-4257(88)90106-X
INSAT, 2009. Institut national de la statistique - Quatrième recensement général de la population et de l'habitat 2009 du mali.
Jacquin, A., 2010. Dynamique de la végétation des savanes en lien avec l'usage des feux à Madagascar. Analyse par série temporelle d'images de télédétection (phd).
Jamison, A., Tuttle, E., Jensen, R., Bierly, G., Gonser, R., 2015. Spatial ecology, landscapes, and the geography of vector-borne disease: A multi-disciplinary review. Appl. Geogr. 63, 418–426.
Jutla, A.S., Akanda, A.S., Islam, S., 2012. Satellite remote sensing of space–time plankton variability in the Bay of Bengal: Connections to cholera outbreaks. Remote Sens. Environ. 123, 196–206. https://doi.org/10.1016/j.rse.2012.03.005
Kamsu-Foguem, B., Rigal, F., Mauget, F., 2013. Mining association rules for the quality improvement of the production process. Expert Syst. Appl. 40, 1034–1045. https://doi.org/10.1016/j.eswa.2012.08.039
Kansakar, P., Hossain, F., 2016. A review of applications of satellite earth observation data for global societal benefit and stewardship of planet earth. Space Policy 36, 46–54. https://doi.org/10.1016/j.spacepol.2016.05.005
Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., Fei-Fei, L., 2014. Large-scale Video Classification with Convolutional Neural Networks. Presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1725–1732.
Kash, T.J., Darling, J.R., 1998. Crisis management: prevention, diagnosis and intervention. Leadersh. Organ. Dev. J. 19, 179–186.
Kazansky, Y., Wood, D., Sutherlun, J., 2016. The current and potential role of satellite remote sensing in the campaign against malaria. Acta Astronaut. 121, 292–305. https://doi.org/10.1016/j.actaastro.2015.09.021
Keramitsoglou, I., Kiranoudis, C.T., Sarimvels, H., Sifakis, N., 2004. A Multidisciplinary Decision Support System for Forest Fire Crisis Management. Environ. Manage. 33, 212–225. https://doi.org/10.1007/s00267-003-0092-4
Ketkar, N., 2017. Introduction to keras, in: Deep Learning with Python. Springer, pp. 97–111.
Khan Academy, 2015. La lumière : ondes électromagnétiques, spectre électromagnétique et photons [WWW Document]. Khan Acad. URL https://fr.khanacademy.org/science/physics/light-waves/introduction-to-light-waves/a/light-and-the-electromagnetic-spectrum (accessed 6.7.18).

Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. ArXiv Prepr. ArXiv14126980.
Kotze, P., Foster, R., Van Greunen, D., Adebesin, F., 2013. A review of interoperability standards in e-Health and imperatives for their adoption in Africa. South Afr. Comput. J. 50, 55–72.
Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet Classification with Deep Convolutional Neural Networks, in: Pereira, F., Burges, C.J.C., Bottou, L., Weinberger, K.Q. (Eds.), Advances in Neural Information Processing Systems 25. Curran Associates, Inc., pp. 1097–1105.
Lacaux, J.P., Tourre, Y.M., Vignolles, C., Ndione, J.A., Lafaye, M., 2007. Classification of ponds from high-spatial resolution remote sensing: Application to Rift Valley Fever epidemics in Senegal. Remote Sens. Environ. 106, 66–74. https://doi.org/10.1016/j.rse.2006.07.012
LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444.
LeCun, Y., Boser, B.E., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W.E., Jackel, L.D., 1990. Handwritten digit recognition with a back-propagation network, in: Advances in Neural Information Processing Systems. pp. 396–404.
Lioult, C., 2017. Exercice de simulation d'un tsunami à Cannes [WWW Document]. URL https://france3-regions.francetvinfo.fr/provence-alpes-cote-d-azur/alpes-maritimes/cannes/exercice-simulation-tsunami-cannes-1357539.html (accessed 12.18.17).
Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A.W.M., van Ginneken, B., Sánchez, C.I., 2017. A survey on deep learning in medical image analysis. Med. Image Anal. 42, 60–88. https://doi.org/10.1016/j.media.2017.07.005
Liu, Q., Wang, G., Liu, X., Peng, T., Wu, J., 2017. Achieving reliable and secure services in cloud computing environments. Comput. Electr. Eng. 59, 153–164. https://doi.org/10.1016/j.compeleceng.2016.10.005
Liu, S., Hodgson, M.E., 2016. Satellite image collection modeling for large area hazard emergency response. ISPRS J. Photogramm. Remote Sens. 118, 13–21. https://doi.org/10.1016/j.isprsjprs.2016.04.007
Liu, Y., Hu, J., Snell-Feikema, I., VanBemmel, M.S., Lamsal, A., Wimberly, M.C., 2015. Software to facilitate remote sensing data access for disease early warning systems. Environ. Model. Softw. 74, 247–257. https://doi.org/10.1016/j.envsoft.2015.07.006
Liu, Y., Wu, L., 2016. Geological Disaster Recognition on Optical Remote Sensing Images Using Deep Learning. Procedia Comput. Sci. 91, 566–575.
Lu, Y., Yi, S., Zeng, N., Liu, Y., Zhang, Y., 2017. Identification of rice diseases using deep convolutional neural networks. Neurocomputing 267, 378–384. https://doi.org/10.1016/j.neucom.2017.06.023
Luo, X., Shen, R., Hu, J., Deng, J., Hu, L., Guan, Q., 2017. A Deep Convolution Neural Network Model for Vehicle Recognition and Face Recognition. Procedia Comput. Sci. 107, 715–720.
Mabaso, M.L.H., Ndlovu, N.C., 2012. Critical review of research literature on climate-driven malaria epidemics in sub-Saharan Africa. Public Health 126, 909–919. https://doi.org/10.1016/j.puhe.2012.07.005
Mak, H.-Y., Mallard, A.P., Bui, T., Au, G., 1999. Building online crisis management support using workflow systems. Decis. Support Syst. 25, 209–224. https://doi.org/10.1016/S0167-9236(99)00007-X

Marechal, F., Ribeiro, N., Lafaye, M., Güell, A., 2008. Satellite imaging and vector-borne diseases: the approach of the French National Space Agency (CNES). Geospatial Health 3, 1–5.

Mathivet, V., 2017. L'intelligence artificielle pour les développeurs: concepts et implémentations en C#.

Moges, S.M., Raun, W.R., Mullen, R.W., Freeman, K.W., Johnson, G.V., Solie, J.B., 2005. Evaluation of green, red, and near infrared bands for predicting winter wheat biomass, nitrogen uptake, and final grain yield. J. Plant Nutr. 27, 1431–1441.

Monasse, P., Nadjahi, K., 2018. Découvrez les différentes couches d'un CNN [WWW Document]. OpenClassrooms. URL https://openclassrooms.com/fr/courses/4470531-classez-et-segmentez-des-donnees-visuelles/5083336-decouvrez-les-differentes-couches-dun-cnn (accessed 9.10.18).

Mosomtai, G., Evander, M., Sandström, P., Ahlm, C., Sang, R., Hassan, O.A., Affognon, H., Landmann, T., 2016. Association of ecological factors with Rift Valley fever occurrence and mapping of risk zones in Kenya. Int. J. Infect. Dis. 46, 49–55. https://doi.org/10.1016/j.ijid.2016.03.013

Najeeb, A.-S., 2012. Tele-epidemiology and tele-health 7-8 February 2012, Vienna, Austria.

Nogueira, K., Penatti, O.A.B., dos Santos, J.A., 2017. Towards better exploiting convolutional neural networks for remote sensing scene classification. Pattern Recognit. 61, 539–556. https://doi.org/10.1016/j.patcog.2016.07.001

Nokhbeh Foghahaayee, H., Menhaj, M.B., Torbati, H.M., 2014. Fuzzy decision support software for crisis management in gas transmission networks. Appl. Soft Comput. 18, 82–90. https://doi.org/10.1016/j.asoc.2014.01.019

Padawangi, R., Turpin, E., Herlily, Prescott, M.F., Lee, I., Shepherd, A., 2016. Mapping an alternative community river: The case of the Ciliwung. Sustain. Cities Soc. 20, 147–157. https://doi.org/10.1016/j.scs.2015.09.001

Peixinho, A.Z., Martins, S.B., Vargas, J.E., Falcao, A.X., Gomes, J.F., Suzuki, C.T.N., 2015. Diagnosis of human intestinal parasites by deep learning, in: Computational Vision and Medical Image Processing V: Proceedings of the 5th Eccomas Thematic Conference on Computational Vision and Medical Image Processing (VipIMAGE 2015, Tenerife, Spain. p. 107.

Piarroux, R., 2002. [Cholera: epidemiology and transmission. Experience from several humanitarian interventions in Africa, Indian Ocean and Central America]. Bull. Soc. Pathol. Exot. 1990 95, 345–350.

Pitarch, Y., Ienco, D., Vintrou, E., Bégué, A., Laurent, A., Poncelet, P., Sala, M., Teisseire, M., 2015. Spatio-temporal data classification through multidimensional sequential patterns: Application to crop mapping in complex landscape. Eng. Appl. Artif. Intell. 37, 91–102. https://doi.org/10.1016/j.engappai.2014.09.001

Prabhakara, K., Hively, W.D., McCarty, G.W., 2015. Evaluating the relationship between biomass, percent groundcover and remote sensing indices across six winter cover crop fields in Maryland, United States. Int. J. Appl. Earth Obs. Geoinformation 39, 88–102. https://doi.org/10.1016/j.jag.2015.03.002

Punia, M., Joshi, P.K., Porwal, M.C., 2011. Decision tree classification of land use land cover for Delhi, India using IRS-P6 AWiFS data. Expert Syst. Appl. 38, 5577–5583. https://doi.org/10.1016/j.eswa.2010.10.078

Qin, Y., Zhang, S., Zhu, X., Zhang, J., Zhang, C., 2009. POP algorithm: Kernel-based imputation to treat missing values in knowledge discovery from databases. Expert Syst. Appl. 36, 2794–2804. https://doi.org/10.1016/j.eswa.2008.01.059

Quinn, J.A., Nakasi, R., Mugagga, P.K., Byanyima, P., Lubega, W., Andama, A., 2016. Deep convolutional neural networks for microscopy-based point of care diagnostics, in: Machine Learning for Healthcare Conference. pp. 271–281.

Richards, John A., Richards, J. A., 1999. Remote sensing digital image analysis. Springer.

Röbke, B.R., Vött, A., 2017. The tsunami phenomenon. Prog. Oceanogr. 159, 296–322. https://doi.org/10.1016/j.pocean.2017.09.003

Rosado, L., Correia da Costa, J.M., Elias, D., S Cardoso, J., 2016. A Review of Automatic Malaria Parasites Detection and Segmentation in Microscopic Images. Anti-Infect. Agents 14, 11–22.

Ruder, S., 2016. An overview of gradient descent optimization algorithms. ArXiv Prepr. ArXiv160904747.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., 2015. Imagenet large scale visual recognition challenge. Int. J. Comput. Vis. 115, 211–252.

Saidi, S., Kattan, L., Jayasinghe, P., Hettiaratchi, P., Taron, J., 2018. Integrated infrastructure systems—A review. Sustain. Cities Soc. 36, 1–11. https://doi.org/10.1016/j.scs.2017.09.022

Scheer, S.J., Varela, V., Eftychidis, G., 2012. A generic framework for tsunami evacuation planning. Phys. Chem. Earth Parts ABC, Innovative modelling approaches for disaster risk reduction 49, 79–91. https://doi.org/10.1016/j.pce.2011.12.001

Schmidhuber, J., 2015. Deep learning in neural networks: An overview. Neural Netw. 61, 85–117.

Shahin, A.I., Guo, Y., Amin, K.M., Sharawi, A.A., 2017. White blood cells identification system based on convolutional deep neural learning networks. Comput. Methods Programs Biomed. https://doi.org/10.1016/j.cmpb.2017.11.015

Simon, P., Williatte Pellitteri, L., 2012. Le décret français de télémédecine : une garantie pour les médecins. Eur. Res. Telemed. Rech. Eur. En Télémédecine 1, 70–75. https://doi.org/10.1016/j.eurtel.2012.05.003

Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. ArXiv Prepr. ArXiv14091556.

Slam, N., Wang, W., Xue, G., Wang, P., 2015. A framework with reasoning capabilities for crisis response decision–support systems. Eng. Appl. Artif. Intell., Innovative Artificial Intelligence Solutions for Crisis Management 46, 346–353. https://doi.org/10.1016/j.engappai.2015.06.017

Sokolova, M., Lapalme, G., 2009. A systematic analysis of performance measures for classification tasks. Inf. Process. Manag. 45, 427–437. https://doi.org/10.1016/j.ipm.2009.03.002

Son, N.-T., Thanh, B.-X., 2018. Decadal assessment of urban sprawl and its effects on local temperature using Landsat data in Cantho city, Vietnam. Sustain. Cities Soc. 36, 81–91. https://doi.org/10.1016/j.scs.2017.10.010

Song, K.-Y., Oh, H.-J., Choi, J., Park, I., Lee, C., Lee, S., 2012. Prediction of landslides using ASTER imagery and data mining models. Adv. Space Res. 49, 978–993. https://doi.org/10.1016/j.asr.2011.11.035

Srikant, R., Agrawal, R., 1996. Mining quantitative association rules in large relational tables, in: Acm Sigmod Record. ACM, pp. 1–12.
Takahata, C., Amin, R., Sarma, P., Banerjee, G., Oliver, W., Fa, J.E., 2010. Remotely-Sensed Active Fire Data for Protected Area Management: Eight-Year Patterns in the Manas National Park, India. Environ. Manage. 45, 414–423. https://doi.org/10.1007/s00267-009-9411-8
Talbi, E.-G., 2018. Fouille de données (Data Mining): Un tour d'horizon. Lab. D'informatique Lille.
Tan, P.-N., Steinbach, M., Kumar, V., 2006. Introduction to data mining, 1st ed. ed. Pearson Addison Wesley, Boston.
Tangpukdee, N., Duangdee, C., Wilairatana, P., Krudsood, S., 2009. Malaria diagnosis: a brief review. Korean J. Parasitol. 47, 93.
Tieleman, T., Hinton, G., 2012. Divide the gradient by a running average of its recent magnitude. COURSERA: Neural networks for machine learning. Technical Report. Available online: https://zh. coursera. org/learn ….
Torch GitHub, 2018. Contribute to torch/nn development. Torch.
Traore, B.B., Kamsu-Foguem, B., Tangara, F., 2017. Data mining techniques on satellite images for discovery of risk areas. Expert Syst. Appl. 72, 443–456. https://doi.org/10.1016/j.eswa.2016.10.010
Traore, B.B., Kamsu-Foguem, B., Tangara, F., 2016. Integrating MDA and SOA for improving telemedicine services. Telemat. Inform. 33, 733–741. https://doi.org/10.1016/j.tele.2015.11.009
Tsai, F., Lai, J.-S., Chen, W.W., Lin, T.-H., 2013. Analysis of topographic and vegetative factors with data mining for landslide verification. Ecol. Eng., Soil Bio- and Eco-Engineering: The Use of Vegetation to Improve Slope Stability 61, 669–677. https://doi.org/10.1016/j.ecoleng.2013.07.070
Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sens. Environ. 8, 127–150. https://doi.org/10.1016/0034-4257(79)90013-0
Tucker, C.J., Sellers, P.J., 1986. Satellite remote sensing of primary production. Int. J. Remote Sens. 7, 1395–1416.
Vescoukis, V., Doulamis, N., Karagiorgou, S., 2012. A service oriented architecture for decision support systems in environmental crisis management. Future Gener. Comput. Syst. 28, 593–604. https://doi.org/10.1016/j.future.2011.03.010
Voigt, S., Kemper, T., Riedlinger, T., Kiefl, R., Scholte, K., Mehl, H., 2007. Satellite Image Analysis for Disaster and Crisis-Management Support. IEEE Trans. Geosci. Remote Sens. 45, 1520–1528. https://doi.org/10.1109/TGRS.2007.895830
Volker, W., 2004. Object-based classification of remote sensing data for change detection. ISPRS J. Photogramm. Remote Sens. 58, 225–238. https://doi.org/10.1016/j.isprsjprs.2003.09.007
Walker, N.F., Nadjm, B., Whitty, C.J.M., 2018. Malaria. Medicine (Baltimore) 46, 52–58. https://doi.org/10.1016/j.mpmed.2017.10.012
Wang, H., Silva, A.J., Benitez, J.A., 2017. 3-Amino 1,8-naphthalimide, a structural analog of the anti-cholera drug virstatin inhibits chemically-biased swimming and swarming motility in vibrios. Microbes Infect. 19, 370–375. https://doi.org/10.1016/j.micinf.2017.03.003
WHO | Cholera [WWW Document], 2018. . WHO. URL http://www.who.int/cholera/en/ (accessed 9.26.18).

Wulder, M.A., Masek, J.G., Cohen, W.B., Loveland, T.R., Woodcock, C.E., 2012. Opening the archive: How free data has enabled the science and monitoring promise of Landsat. Remote Sens. Environ., Landsat Legacy Special Issue 122, 2–10. https://doi.org/10.1016/j.rse.2012.01.010

Xie, L., Li, G., Xiao, M., Peng, L., 2016. Novel classification method for remote sensing images based on information entropy discretization algorithm and vector space model. Comput. Geosci. 89, 252–259. https://doi.org/10.1016/j.cageo.2015.12.015

Xu, Z., Liu, Y., Ma, Z., Li, S., Hu, W., Tong, S., 2014. Impact of temperature on childhood pneumonia estimated from satellite remote sensing. Environ. Res. 132, 334–341. https://doi.org/10.1016/j.envres.2014.04.021

Zhang, S., Zhang, C., Yang, Q., 2003. Data preparation for data mining. Appl. Artif. Intell. 17, 375–381. https://doi.org/10.1080/713827180

Zhou, L., Wu, X., Xu, Z., Fujita, H., 2017. Emergency decision making for natural disasters: An overview. Int. J. Disaster Risk Reduct. https://doi.org/10.1016/j.ijdrr.2017.09.037

Bibliographie de l'auteur

Traore, B.B., Kamsu-Foguem, B. et Tangara, F. 2018. Deep convolution neural network for image recognition. Ecological Informatics, 48, 257-268 https://doi.org/10.1016/j.ecoinf.2018.10.002

Traore, B.B., Kamsu Foguem, B., Tangara, F., Desforges, X., 2018. Service-Oriented Computing for intelligent train maintenance. Enterp. Inf. Syst. 1–24. https://doi.org/10.1080/17517575.2018.1501818

Traore, B.B., Kamsu-Foguem, B., Tangara, F., Tiako, P., 2018. Software services for supporting remote crisis management. Sustain. Cities Soc. 39, 814–827. https://doi.org/10.1016/j.scs.2018.02.029

Traore, B.B., Kamsu-Foguem, B., Tangara, F., 2017. Data mining techniques on satellite images for discovery of risk areas. Expert Syst. Appl. 72, 443–456. https://doi.org/10.1016/j.eswa.2016.10.010

Traore, B.B., Kamsu-Foguem, B., Tangara, F., 2016. Integrating MDA and SOA for improving telemedicine services. Telemat. Inform. 33, 733–741. https://doi.org/10.1016/j.tele.2015.11.009