



HAL
open science

Intégration d'ontologies dans la classification parallèle de données médicales pour le diagnostic de lésions du foie

Rim Messaoudi

► To cite this version:

Rim Messaoudi. Intégration d'ontologies dans la classification parallèle de données médicales pour le diagnostic de lésions du foie. Imagerie médicale. Université Clermont Auvergne; Université de Sfax (Tunisie), 2021. Français. NNT: 2021UCFAC029 . tel-03578338

HAL Id: tel-03578338

<https://theses.hal.science/tel-03578338v1>

Submitted on 17 Feb 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**THESE EN COTUTELLE
L'UNIVERSITE CLERMONT AUVERGNE ET L'UNIVERSITE DE SFAX**

Ecoles doctorales :

Ecole Doctorale Sciences de l'Ingénieur de l'Université Clermont Auvergne

Ecole Doctorale de la Faculté des Sciences Économiques et de Gestion de Sfax

Présentée par

Rim MESSAOUDI

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ

Discipline : **Informatique**

soutenue le 2 juin 2021

**Intégration d'ontologies dans la classification
parallèle de données médicales pour le diagnostic
de lésions du foie**

devant le Jury composé de :

M. Nicolas PASSAT	Professeur, URCA, France	Rapporteur
M. Chokri BEN AMAR	Professeur, ENIS, Sfax, Tunisie	Rapporteur
M. Imed Riadh FARAH	Professeur, UMA, Tunisie	Examineur
Mme. Alice OTHMANI	Maître de conférences, UPEC, France	Examinatrice
M. Faïez GARGOURI	Professeur, ISIM Sfax, Tunisie	Co-directeur
M. Antoine VACAVANT	Maître de conférences, HDR, UCA, France	Co-directeur
M. Achraf MTIBAA	Maître assistant, ENET'COM, Tunisie	Co-encadrant
M. Faouzi JAZIRI	Docteur en informatique, UCA, France	Co-encadrant

*Je dédie ce travail
Aux plus chers du monde*

A mes Très Chers Parents MOULDI et ZINA

Tous les mots du monde ne sauraient exprimer l'immense amour que je vous porte, ni la profonde gratitude que je vous témoigne pour tous les efforts et les sacrifices que vous n'avez jamais cessé de consentir pour mon instruction et mon bien-être. J'espère avoir répondu aux espoirs que vous avez fondés en moi. Je vous rends hommage par ce modeste travail en guise de ma reconnaissance éternelle et de mon amour.

Mes dédicaces vont aussi

À Ma Chère sœur IMEN

qui m'a toujours souhaité le meilleur. Je la désire la réussite et le bonheur dans sa vie professionnelle

À la mémoire de mes grands parents

***À tous les membres de ma famille
Et à tous mes amis***

Remerciements

Je voudrais tout d'abord, exprimer ma profonde et respectueuse reconnaissance à mes co-directeurs de thèse Monsieur Faïez GARGOURI, professeur à L'Institut Supérieur d'Informatique et de Multimédia de Sfax, et Monsieur Antoine VACAVANT, Maître de conférences HDR à l'Université Clermont Auvergne, pour la confiance qu'ils m'ont accordée en acceptant de diriger mes travaux de thèse de doctorat. Qu'ils trouvent dans ce travail l'expression de ma profonde gratitude.

Je remercie aussi mes co-encadrants Monsieur Achraf MTIBAA, maître assistant à l'École Nationale d'Électronique et des Télécommunications de Sfax, et Monsieur Faouzi JAZIRI, docteur en informatique à l'Université Clermont-Auvergne, pour les efforts fournis, les conseils prodigués, leurs patiences, leurs encouragements inlassables et leurs persévérances dans le suivi ainsi que le meilleur accueil, leurs directives et leurs conseils qui m'ont guidé à l'élaboration de ce mémoire

Je tiens à mentionner le plaisir et l'honneur que m'ont fait Monsieur Nicolas PASSAT, professeur à l'Université de Reims-Champagne-Ardenne et Monsieur Chokri BEN AMAR, professeur à l'École Nationale d'Ingénieurs de Sfax, pour avoir accepté d'être les rapporteurs de ma thèse et d'évaluer mon travail.

J'adresse aussi mes remerciements à Monsieur Imed Riadh FARAH, professeur à l'Université de la Manouba, et Madame Alice OTHMANI, maître de conférences à l'Université Paris-Est Créteil, d'avoir accepté avec beaucoup d'élan d'être les examinateurs de ce travail et pour le temps et l'attention qu'ils ont consacrée pour examiner cette thèse

Ma reconnaissance s'adresse aussi à tous ceux qui m'ont aidée à réaliser ce travail. Je voudrais, également, adresser mes remerciements à toute l'équipe du laboratoire MIRACL et toute l'équipe du laboratoire Institut Pascal.

Finalement, j'adresse mes remerciements à tous mes enseignants et tous mes collègues.

Résumé

Le diagnostic des lésions hépatique est une tâche complexe surtout lorsque les nodules détectés sont de petites tailles. Dans ce cas, il devient très difficile de connaître leurs natures (tumeur bénigne ou maligne, type de lésion, *etc*). Dans des cas similaires, il faut répéter des examens cliniques pendant plusieurs mois pour voir l'évolution des masses hépatiques. Afin de mieux répondre à ces problèmes, il faut trouver des solutions informatiques qui servent à l'optimisation du diagnostic des tumeurs du foie.

Dans le contexte de la classification des lésions hépatiques, nous avons développé une première approche ontologique (OntHCC) pour l'aide au diagnostic, à la stadification et au choix de traitement des tumeurs CHC (Carcinome Hépatocellulaire). Cette approche est fondée sur l'analyse d'images IRM de foies infectés et sur des rapports radiologiques. Par la suite, nous avons proposé une deuxième approche ontologique (MROnt) pour la modélisation de l'information médicale contenue dans les rapports radiologiques, dans le cadre du diagnostic et de suivi de tumeurs du foie. La détection automatique des tumeurs du foie nécessite un processus de diagnostic primaire en utilisant obligatoirement les images médicales (par exemple IRM ou scanner). Pour ce faire, nous avons intégré l'apprentissage profond dans la classification d'images IRM avec prise de contraste.

Dans la suite de la thèse et afin d'accroître la performance du processus de classification des images, nous avons intégré les connaissances sémantiques. L'objectif est de profiter de la base de connaissances offerte par les ontologies pour décrire les images médicales et fournir des informations sur les tumeurs détectées (par exemple, le type, la taille et le stade). En outre, notre approche consiste à développer un CNN *multi-label* afin de supporter les ontologies développées (OntHCC et MROnt). Nous montrons l'efficacité des approches et prototypes proposés dans ces travaux de thèse à travers des évaluations numériques comparatives et des études de cas.

Mots clés : Carcinome Hépatocellulaire (CHC), Ontologies, Deep Learning, Classification de données médicales, Traitement d'images médicales.

Abstract

Diagnosis of liver cancer is a complex task, especially when the nodules to be detected are small. Moreover, it becomes very difficult to extract the nature of the detected nodules (benign or malignant, type of lesion, etc.). A possible long period of several months may be necessary to study the evolution of the hepatic nodules. To solve these problems, IT (Information Technology) solutions must be found in order to ensure optimization of the diagnosis process of liver tumors.

In the context of hepatic lesions classification, we develop in this PhD a first ontological approach (OntHCC) that aids in the diagnosis, staging and treatment of HCC (Hepatocellular Carcinoma) lesions. This approach is based on information extracted from liver MRI images and the related radiological reports. Then, we propose a second ontological approach (MROnt) that aims to model and analyze clinical data from these reports. Automatic detection of liver tumors requires a primary diagnosis process using medical images (e.g. MRI or CT). To do so, we propose a deep learning based approach that offers a parallel preprocessing algorithm. It allows HCC detection and localization in MRI images with contrast enhancement via a CNN (Convolutional Neural Network) algorithm.

In the next part of this thesis and in order to improve the performance of the classification process, we integrate the developed ontologies into the CNN approach. The aim is to provide more information about the detected tumors (e.g. type, size and stage). In addition, our approach consists in developing a multi-label CNN that supports the developed ontologies (OntHCC and MROnt). We show that we improve the accuracy of the proposed methods through a set of comparative numerical evaluations and case studies.

Keywords : Hepatocellular Carcinoma(HCC), Ontologies, Deep Learning, Medical Data Classification, Medical Image Processing.

Table des matières

I Introduction générale	xvi
I.1 Cadre de travail	1
I.2 Partenaires du projet	1
I.2.1 France : Institut Pascal	1
I.2.2 Tunisie : Laboratoire MIRACL	2
I.3 Contexte du projet	2
I.3.1 Contexte médical	2
I.3.2 Contexte Informatique	3
I.4 Motivation et définition du problème	4
I.5 Objectifs	4
I.6 Contributions	5
I.7 Organisation de la thèse	6
II Concepts de base : Ontologies et deep learning	8
II.1 Introduction	9
II.2 Ontologie	9
II.2.1 Définitions de l'ontologie	10
II.2.2 Objectifs de l'ontologie	11
II.2.3 Caractéristiques des ontologies	11
II.2.4 Classifications des ontologies	11
II.2.5 Composants de l'ontologie	13
II.2.6 Cycle de vie d'une ontologie	14
II.2.7 Méthodes de développement d'une ontologie	15
II.2.7.1 Méthode de Gruninger et Fox (TOVE)	15
II.2.7.2 Méthodologie de Uschold et King	15
II.2.7.3 Méthode de Bernaras et al.	15
II.2.7.4 Terminae	16
II.2.7.5 Methontology	16
II.2.7.6 La méthodologie NeOn	17

II.2.8	Langages de modélisation des ontologies	17
II.2.8.1	RDF	18
II.2.8.2	RDF Schéma	18
II.2.8.3	OWL	18
II.2.8.4	Logique de description	19
II.2.8.5	Raisonnement ontologique et inférence	19
II.2.9	SWRL : Langage de représentation de règles	20
II.2.10	Les domaines d'application des ontologies	20
II.3	Apprentissage automatique	22
II.3.1	Définition	22
II.3.2	Types d'apprentissage automatique	23
II.3.2.1	L'apprentissage supervisé	23
II.3.2.2	L'apprentissage non supervisé	23
II.3.2.3	L'apprentissage semi-supervisé	23
II.3.2.4	Apprentissage par renforcement	23
II.3.3	RNA (Réseaux de neurones artificiels)	23
II.3.3.1	Définition	24
II.3.3.2	Architecture	25
II.3.4	Deep learning	28
II.3.4.1	Définition	28
II.3.4.2	Architecture et fonctionnement	29
II.3.5	Les réseaux de neurones convolutifs (CNN)	30
II.4	La programmation parallèle	31
II.4.1	Définition	31
II.4.2	Architectures	32
II.4.3	Grille de calcul	33
II.5	Conclusion	33
III	Etat de l'art sur les ontologies dédiées à la représentation des lésions hépatiques	35
III.1	Introduction	36
III.2	Définition du CHC	36
III.2.1	Imagerie Médicale des tumeurs CHC	37
III.2.2	Détection du CHC <i>via</i> l'imagerie médicale	38
III.2.3	Les systèmes de stadification des tumeurs hépatiques	41
III.2.3.1	BCLC (<i>Barcelona Clinic Liver Cancer</i>)	41
III.2.3.2	TNM (<i>Tumor, Nodes, and Metastasis</i>)	43
III.2.3.3	Li-RADS (<i>Liver Imaging Reporting and Data System</i>)	45
III.3	Ontologies pour la représentation des maladies hépatiques	46
III.3.1	Interprétation des images médicales <i>via</i> les ontologies	47
III.3.1.1	Les ontologies pour l'analyse des images TDM	47
III.3.1.2	Les ontologies pour l'analyse des images IRM	48
III.3.1.3	Les ontologies pour l'analyse d'autres types d'images	49
III.3.2	Modélisation des données médicales <i>via</i> les ontologies	55

III.3.2.1	Stadification des tumeurs hépatiques	55
III.3.2.2	Résolution des problèmes cliniques	58
III.3.3	Modélisations des rapports médicaux	63
III.3.3.1	Modélisation des rapports en lien avec le cancer CHC	63
III.3.3.2	Modélisation d'autres types de rapports médicaux	64
III.4	Classification des données médicales <i>via</i> les techniques d'apprentissage automatique et des ontologies	70
III.4.1	Application des techniques d'apprentissage automatique	70
III.4.2	Application des ontologies avec les techniques d'apprentissage automatique	70
III.5	Discussion	72
IV	Représentation des connaissances médicales à travers les ontologies	74
IV.1	Introduction	75
IV.2	Approche de détection et de classification du CHC <i>via</i> une ontologie : OntHCC	75
IV.2.1	Phase A : Extraction des informations médicales à partir de l'image médicale du foie	76
IV.2.2	Phase B : La représentation sémantique des connaissances médicales	77
IV.2.3	Présentation de l'ontologie OntHCC	78
IV.2.3.1	Les concepts de l'ontologie OntHCC	79
IV.2.3.2	Les propriétés de l'ontologie OntHCC	81
IV.2.3.3	Visualisation de l'ontologie OntHCC	83
IV.2.4	Phase C : Diagnostic du CHC	84
IV.2.5	Raisonnement dans notre modèle ontologique	85
IV.3	Développement d'une ontologie pour la modélisation des rapports radiologiques du CHC (MROnt)	86
IV.3.1	Phase 1 : Population de la base de connaissances	87
IV.3.2	Phase 2 : Analyse des rapports radiologiques	88
IV.3.3	Présentation de l'ontologie MROnt	88
IV.3.3.1	Les concepts de l'ontologie MROnt	88
IV.3.3.2	Les propriétés de l'ontologie MROnt	90
IV.3.3.3	Visualisation de l'ontologie MROnt	91
IV.3.4	Les règles SWRL	91
IV.4	Conclusion	94
V	Approche de classification d'images IRM basée sur le couplage ontologies-Deep Learning	95
V.1	Introduction	96
V.2	Approche de détection et de localisation des tumeurs hépatiques	96
V.2.1	Phase de prétraitement et découpage des images IRM en petits fragments	97
V.2.2	Segmentation du foie	98
V.2.3	Récupération des images d'un fichier format RAW	100
V.2.4	Extraction et ajustement des images IRM/CT	101
V.2.5	Création de la base de données	101

V.3	Description de l’algorithme deep learning	102
V.3.1	Les paramètres de l’algorithme <i>deep learning</i>	102
V.3.2	Parallélisation de prétraitement	103
V.3.3	Phase de prédiction	104
V.3.4	Validation des résultats	104
V.4	Intégration d’ontologies dans la classification des tumeurs hépatiques	104
V.4.1	Les ontologies appliquées	105
V.4.2	Collection des labels	106
V.4.3	Création d’un CNN multi-label	107
V.5	Conclusion	110
VI	Prototypes et Expérimentations	111
VI.1	Introduction	112
VI.2	Outils de développement	112
VI.2.1	Développement des ontologies	112
VI.2.1.1	Protégé	112
VI.2.1.2	Eclipse	112
VI.2.1.3	Jena	113
VI.2.1.4	SWRL	113
VI.2.1.5	SPARQL	113
VI.2.1.6	SQWRL	114
VI.2.1.7	Weasis	114
VI.2.2	Développement des modèles deep learning	114
VI.2.2.1	Python	114
VI.2.2.2	Microsoft Visual Studio	114
VI.2.2.3	Google Colaboratory Notebook	115
VI.2.2.4	TensorFlow	115
VI.2.2.5	Keras	115
VI.2.2.6	Numpy	116
VI.2.2.7	OpenCV	116
VI.3	Prototypes développés	116
VI.3.1	Prototypes ontologiques	116
VI.3.1.1	Prototype OntHCC	116
VI.3.2	Module A : Création de l’ontologie	117
VI.3.3	Module B : Construction des règles SWRL	117
VI.3.4	Module C : Affichage du résultat	118
VI.3.5	Prototype MROnt	120
VI.3.6	Prototype de détection et de localisation de la tumeur de foie	123
VI.4	Evaluation et validation des résultats	127
VI.4.1	Evaluation du Prototype OntHCC	127
VI.4.2	Evaluation des algorithmes de deep learning	130
VI.4.2.1	Métriques de performance	130
VI.4.2.2	Résultats sans couplage	131

VI.4.2.3 Résultats avec couplage	133
VI.5 Conclusion	135
VII Conclusion Générale	136
VII.1 Bilan des contributions	137
VII.2 Perspectives	138
Bibliographie	139
Annexe	151
Annexe A	152

Table des figures

II.1	Visualisation des sous-domaines de l'Intelligence Artificielle [126]	10
II.2	Classification des ontologies [5]	12
II.3	Cycle de vie d'une ontologie [11]	14
II.4	Vue sur la méthodologie MethOntology d'après [17]	17
II.5	Architecture du Web sémantique [120]	18
II.6	Visualisation d'un neurone biologique	24
II.7	Visualisation d'un neurone artificiel	25
II.8	Architecture d'un réseau de neurones à couches	25
II.9	Illustration d'un neurone artificiel et notations	26
II.10	Différentes fonctions d'activation pour un réseau de neurones [147]	27
II.11	Visualisation des couches d'une architecture <i>deep learning</i> [143]	29
II.12	Architecture d'un réseau CNN [W10]	30
II.13	Fonctionnement d'un CNN [134]	31
III.1	Visualisation du foie et ses vaisseaux [148]	37
III.2	Diagnostic de CHC <i>via</i> les modalités radiologiques [63]	39
III.3	Critères de Barcelone pour la détection du CHC [65]	40
III.4	Visualisation des catégories du système Li-RADS [149]	46
IV.1	Approche ontologique pour le diagnostic, la stadification et le traitement du CHC	76
IV.2	Un exemple d'une IRM hépatique	77
IV.3	Visualisation de l'IRM hépatique <i>via</i> Weasis	77
IV.4	Visualisation de l'ontologie OntHCC <i>via</i> OntoGraf	84
IV.5	Ontologie pour la modélisation des rapports radiologiques du CHC (MROnt)	87
IV.6	Visualisation de l'ontologie MROnt <i>via</i> le plugin OntoGraf	91
IV.7	Processus de changement d'état et classification du nodule détecté	93
V.1	Approche de détection et de localisation de la tumeur de foie	97
V.2	Les trois phases IRM appliquées (a) phase 1 : avant injection (b) entre 30 s et 1 min après injection (c) phase 3 : phase portale ou tardive (> 1 min) en fonction des séquences	98

V.3 Les images IRM des différents patients	98
V.4 Les étapes de segmentation du foie dans une image IRM	99
V.5 Exemples d'examens IRM avant et après la segmentation du foie	100
V.6 Visualisation d'une image IRM à partir du logiciel ImageJ	100
V.7 Visualisation d'un fragment avec recouvrement	101
V.8 Les étapes de l'algorithme <i>deep learning</i>	102
V.9 Visualisation du modèle CNN développé	103
V.10 Description de la phase de prédiction	104
V.11 Approche de couplage ontologies/ <i>deep learning</i>	105
V.12 Les concepts appliqués dans l'algorithme de classification	107
V.13 Vecteur des labels appliqués	107
V.14 Visualisation des fragments d'une image IRM contenant une tumeur	108
V.15 Visualisation des fragments d'une image IRM contenant une tumeur	109
VI.1 Architecture fonctionnelle du prototype OntHCC	117
VI.2 Interface de détection des tumeurs CHC	118
VI.3 Interface de stadification des tumeurs hépatiques	119
VI.4 Calcul du score Child-Pugh	119
VI.5 Exemples de requêtes SPARQL : (a) Règle SPARQL pour la détection de la tumeur hépatique, (b) Règle SPARQL pour la stadification BCLC, (c) Règle SPARQL pour la stadification TNM, (d) Règle SPARQL pour le type de traitement à appliquer	120
VI.6 Interface de traitement du CHC	120
VI.7 Architecture fonctionnelle du prototype MROnt	121
VI.8 Exécution d'une requête SPARQL	122
VI.9 Extraction des observations radiologiques d'un patient	123
VI.10 Architecture fonctionnelle de détection et localisation de la tumeur de foie	123
VI.11 Exécution de l'algorithme pour l'extraction des fragments	124
VI.12 Résultat de l'algorithme d'extraction exécuté sur tous les images	124
VI.13 Exemple d'extraction et d'enregistrement des fragments d'une image IRM	125
VI.14 Exécution de l'algorithme de <i>deep learning</i> sur une machine	126
VI.15 Exécution de l'algorithme de <i>deep learning</i> sur Google Colab	126
VI.16 Exemple d'exécution d'un modèle CNN multi-label	127
VI.17 Évaluation du prototype OntHCC	129
VI.18 Le modèle CNN : (a) taux de précision, (b) taux d'erreur	131
VI.19 Le modèle CNN multi-label : (a) taux de précision, (b) taux d'erreur	133

Liste des tableaux

III.1 Calcul du score Child-Pugh	42
III.2 Stadification du CHC selon le système BCLC	43
III.3 Les valeurs possibles des paramètres du système TNM	44
III.4 Stadification du CHC selon le système TNM	45
III.5 Ontologies pour la modélisation des images médicales. Les signes '+' et '-' désignent dans chaque ligne du tableau si le travail étudié (1) visualise l'ontologie proposée, (2) s'il a créé un prototype à réutiliser, (3) s'il a fait l'implantation de l'ontologie dans une application, et (4) s'il a fait la validation de l'ontologie proposée	51
III.6 Propriétés de l'ontologie proposée par [92]	55
III.7 Ontologies pour la modélisation des données médicales. Les signes '+' et '-' désignent dans chaque ligne du tableau si le travail étudié (1) visualise l'ontologie proposée, (2) s'il a créé un prototype à réutiliser, (3) s'il a fait l'implantation de l'ontologie dans une application, et (4) s'il a fait la validation de l'ontologie proposée	59
III.8 Les relations de l'ontologie proposée par [35]	64
III.9 Ontologies pour la modélisation des rapports médicaux. Les signes '+' et '-' désignent dans chaque ligne du tableau si le travail étudié (1) visualise l'ontologie proposée, (2) s'il a créé un prototype à réutiliser, (3) s'il a fait l'implantation de l'ontologie dans une application, et (4) s'il a fait la validation de l'ontologie proposée	67
IV.1 Les concepts définis dans l'ontologie OntCHC et leurs significations	79
IV.2 Les propriétés d'objets	81
IV.3 Les propriétés du type de données	82
IV.4 Les concepts définis dans l'ontologie MROnt et leurs significations	89
IV.5 Les propriétés définies dans l'ontologie MROnt	90
VI.1 Evaluation Quantitative	128
VI.2 Evaluation Qualitative	128
VI.3 Description de la base d'apprentissage	130
VI.4 Métrique de performance	130
VI.5 Evaluation durant la phase 2 (avec contraste)	132
VI.6 Evaluation durant la phase 3 (après contraste)	132

VI.7 Comparaison des modèles CNN et CNN multi-label	134
VI.9 Taux des TP, TN, FP et FN des modèles CNN et CNN multi-label	134
VI.11 Taux de précision des modèles CNN et CNN multi-label	134

Liste des abréviations

AFP	Alpha-fœtoprotéine
BCLC	Barcelona Clinic Liver Cancer
CHC	Carcinome Hépatocellulaire
CT	Computerized Tomography
CO-ODE	Collaborative Open Ontology Development Environment
DOLCE	Descriptive Ontology for Linguistic and Cognitive Engineering
FMA	Foundational Model of Anatomy
GO	Gene Ontology
HCC	Hepato Cellular Carcinoma
Li-RADS	Liver Imaging Reporting and Data System
IRM	Imagerie par Résonance Magnétique
MeSH	Medical Subject Headings
OBO	Open Biomedical Ontology
OWL	Web Ontology Language
RDF	Resource Description Framework
RDFS	RDF Schema
SPARQL	Query Language for RDF
SWRL	Semantic Web Rule Language
SUMO	Standard Upper Ontology
SQWRL	Semantic Query-Enhanced Web Rule Language
SNOMED	The Systematized Nomenclature of Medicine
TAMBIS	Transparent Access to Multiple Bioinformatics Information Sources
TNM	Tumeur, Nodules, et Métastases
UMLS	Unified Medical Language System
URIs	Universal Resource Identifiers
XML	eXtensible Markup Language
TDM	Tomodensitométrie
DAO	Diagnostic Assisté par Ordinateur
GPU	Graphic Processing Unit

CPU	Central Processing Unit
CNN	Convolutional Neural Network
ReLU	Rectified Linear Unit
API	Application Programmable Interface
OpenCV	Open Source Computer Vision Library

INTRODUCTION GÉNÉRALE

Sommaire

I.1 Cadre de travail	1
I.2 Partenaires du projet	1
I.2.1 France : Institut Pascal	1
I.2.2 Tunisie : Laboratoire MIRACL	2
I.3 Contexte du projet	2
I.3.1 Contexte médical	2
I.3.2 Contexte Informatique	3
I.4 Motivation et définition du problème	4
I.5 Objectifs	4
I.6 Contributions	5
I.7 Organisation de la thèse	6

I.1 Cadre de travail

Cette thèse de doctorat intitulée « Intégration d'ontologies dans la classification parallèle de données médicales pour le diagnostic de lésions du foie » est menée dans le cadre des travaux de recherche du projet franco-tunisien « Classification d'Images Répartie pour la Reconnaissance de lésions Hépatiques, avec des Ontologies, et à grande Echelle » (CIRRHOSE), au sein des laboratoires MIRACL (Sfax, Tunisie) et l'Institut Pascal (Clermont-Ferrand, France). Ce projet est financé par la bourse du programme franco-tunisien PHC (Partenariat Hubert Curien) UTIQUE. Des praticiens hospitaliers du CHU Hedi Chaker (Sfax - Tunisie) et du CHU de Clermont-Ferrand participent à ce projet. Le mésocentre de Clermont-Ferrand est également membre partenaire du projet CIRRHOSE. Les travaux de cette thèse visent à proposer une plateforme de classification de grandes masses de données image, basée sur des connaissances médicales par le biais d'ontologies, dans le cadre de l'identification et du traitement des tumeurs CHC (Carcinome Hépatocellulaire) du foie. Cette plateforme englobera de nouvelles approches informatiques qui utilisent les ontologies avec leur grande adaptabilité et robustesse et leurs capacités de raisonnement. L'utilisation d'une base de connaissances élaborées à partir des ontologies peut également aider le médecin dans son diagnostic et ses pratiques cliniques et le guider dans sa prise de décision.

I.2 Partenaires du projet

Cette thèse s'est déroulée au sein du laboratoire MIRACL à l'Université de Sfax, Tunisie et le laboratoire Institut Pascal à l'Université Clermont Auvergne, France.

I.2.1 France :Institut Pascal

L'Institut Pascal ¹, UMR 6602, est une unité mixte de recherche et de formation interdisciplinaire composée de 370 personnes placées sous la triple tutelle de l'Université Clermont Auvergne (UCA) ², du Centre national de la recherche scientifique (CNRS) ³ et de SIGMA ⁴ Clermont. Le CHU de Clermont-Ferrand est également partenaire du laboratoire. L'Institut développe des connaissances et des technologies contribuant à trois domaines d'application : l'usine et les écosystèmes, les transports et l'hôpital du futur. L'équipe CaVITI (« *Cardio-Vascular Interventional Therapy and Imaging* ») ⁵ s'appuie sur une forte expérience clinique en radiologie, cardiologie interventionnelle et chirurgie cardiaque pour mener des projets en liaison avec l'imagerie de haute résolution, méthodes non invasives, fusion multimodale et algorithmes de traitement d'images. Son effectif est constitué de plusieurs enseignants chercheurs et de praticiens hospitaliers. Ils s'intéressent aux thématiques (1) des endoprothèses, (2) de l'embolisation et (3) de la fonction et ischémie myocardique. Le thème des endoprothèses vise à proposer un *framework* de prise en

1. <http://www.institutpascal.uca.fr/>

2. <https://www.uca.fr/fr>

3. <https://www.cnrs.fr/>

4. <https://www.sigma-clermont.fr/fr>

5. <http://www.institutpascal.uca.fr/index.php/fr/tgi-caviti>

charge des planifications, de suivi des endoprothèses, piloté par l'imagerie avec application aux artères coronaires, à l'aorte et aux vaisseaux digestifs. Le thème de l'embolisation, qui participe à notre projet, porte sur la réalisation d'outils logiciels assistant les médecins pour le traitement des tumeurs hépatiques par la chimio-embolisation transartérielle (ou TACE pour « *transarterial chemo-embolization* ») ainsi que dans d'autres activités telles que la chirurgie et la biopsie. Il est animé par des membres de l'équipe installés à l'IUT du Puy-en-Velay. Ces outils s'appuient sur des environnements techniques et informatiques, des mathématiques et des sciences de l'ingénieur. Le troisième thème est conçu pour deux volets principaux : (a) la détection et le suivi de l'ischémie myocardique et sur (b) le bilan et le suivi des asynchronismes cardiaques.

I.2.2 Tunisie : Laboratoire MIRACL

Le laboratoire MIRACL ⁶ « *Multimedia, InfoRmation systems and Advanced Computing Laboratory* » a été créé en 2005. Il est réparti sur les sites de l'Université de Sfax en Tunisie. Les principales activités de MIRACL s'articulent autour des missions suivantes : l'encadrement des doctorants et des étudiants en Master, l'ouverture sur des institutions et les unités de recherche nationales et internationales (par exemple l'encadrement de thèses en cotutelles, les conventions de recherche), l'ouverture sur l'environnement industriel et la réalisation de projets de coopération internationale (CMCU, *etc*). Les recherches au sein de MIRACL se focalisent essentiellement sur les thèmes des systèmes d'information, du langage naturel, du Web sémantique et des systèmes multimédias. Il est organisé en trois pôles de compétences complémentaires : 1) « Ingénierie des systèmes d'information et de décision », 2) « TALN (Traitement Automatique du Langage Naturel) et Web sémantique » et 3) « Ingénierie des systèmes Multimédia ». En particulier, le pôle « Ingénierie des systèmes d'information et de décision », qui participe à CIRRHOSE s'intéresse aux thématiques de systèmes décisionnels, d'ingénierie ontologique, de gestion des connaissances, et des systèmes d'informations complexes, *etc*.

I.3 Contexte du projet

I.3.1 Contexte médical

Le cancer du foie survient après le développement d'une ou plusieurs masses cancéreuses dans le foie. Il est considéré comme l'un des cancers les plus agressifs dans le monde. C'est le sixième cancer le plus fréquent au monde et le troisième le plus meurtrier. Dans environ 75 à 80 % des cas ⁷, ces lésions apparaissent dans un foie infecté par une maladie chronique qui a engendré une cirrhose, notamment l'hépatite B ou C. Ce cancer touche environ 8 000 Français chaque année, le plus souvent des hommes plus de 40 ans [Simonetti 91]. Selon l'InVS ⁸ et l'INCa⁹, le nombre de nouveaux cas de cancer primitif du foie en France était passé de 1 800 en 1980 à 7 100 en 2008

6. <http://www.miracl.rnu.tn/>

7. <https://www.snfge.org/content/7-carcinome-hepatocellulaire-cancer-primitif-du-foie>

8. www.invs.sante.fr

9. www.e-cancer.fr

et 8 723 en 2012. Il peut rester longtemps sans provoquer de symptômes ; son diagnostic précoce est un réel enjeu. Le CHC (Carcinome Hépatocellulaire), lésion que nous traitons dans ces travaux de thèse, est l'un des plus critiques et plus fréquents. Il se développe sur une cirrhose à partir des cellules spécialisées du foie (les hépatocytes). Le CHC survient suite à de nombreux facteurs de risques (par exemple l'obésité, l'alcool, des antécédents médicaux, diabète, *etc.*). La chirurgie ou la greffe du foie, l'ablation partielle du foie, la chimiothérapie localisée sont considérés comme les principaux types de traitement du CHC.

I.3.2 Contexte Informatique

Le projet CIRRHOSE porte sur l'aide au diagnostic et au traitement du CHC. Les approches que nous souhaitons développer doivent traiter de grandes masses de données dans un temps limité. Aussi, les ontologies se comportent ici comme un outil de classification des lésions hépatiques en se basant sur des séquences IRM reliées à des patients différents et pendant des moments différents. Nous devons trouver un moyen de communication entre l'ontologie et l'imagerie médicale en tenant compte de la quantité de données traitée, pour optimiser le diagnostic de tumeurs CHC. Par la suite, l'élargissement de l'ontologie créée permettra de supporter plus d'informations et de données médicales décrivant l'état du patient et de la maladie. Ce projet est une continuité des projets réalisés et en cours de réalisation. Parmi ces travaux nous citons :

- **« Descripteurs de texture pour la classification des lésions hépatiques de type CHC »** : une contribution réalisée dans le cadre d'un stage de Master 2 Technologie pour la Médecine en 2016 par Hawa MOHAMED ALI.
- **« Vers une approche à base d'ontologie pour l'assistance au diagnostic des tumeurs du foie CHC »** : Une contribution réalisée dans le cadre d'un stage de Master de Recherche en Science, Technologie de l'Information et Communication-Services Informatiques (STIC-SI) à l'ENET'COM en 2016 par Rim MESSAOUDI.
- **« Classification des données médicales en intégrant une approche parallèle (cas d'études (Détection de la tumeur du foie)) »** : Une contribution réalisée dans le cadre d'un stage d'un projet de fin d'études d'ingénieur en Informatique Technologies Web et Multimédia à l'ISIMS en 2016 par Fadwa KRAINI.
- **« Segmentation interactive et visualisation des composants du foie sur des images médicales (CT-IRM) »** : une contribution réalisée en 2017 par Marie-Ange LEBRE lors de son post-doctorat.
- **« Vers un *framework* parallèle pour la détection et la localisation de tumeurs CHC du foie par apprentissage automatique »** : Une contribution réalisée dans le cadre d'un stage d'un projet de fin d'études d'ingénieur en Réseaux et Télécommunication à l'ENET'COM en 2017 par Marwa BEN ABDALLAH.

- «**Développement d'un algorithme parallèle basé sur le *deep learning* pour la détection du cancer de foie**» : Une contribution réalisée dans le cadre d'un stage d'un projet de fin d'études d'ingénieur en Réseaux et Télécommunication à l'ENET'COM en 2018 par Omar AYADI.
- «**Approche de classification d'images médicales du cancer de foie guidée par les ontologies et le *deep learning***» : Une contribution réalisée dans le cadre d'un stage de Master de recherche en informatique à l'ISIMS en 2021 par Nibras Mizouri.

I.4 Motivation et définition du problème

Le diagnostic des nodules hépatiques est une tâche complexe et sensible surtout lorsque les anomalies détectées sont de petites tailles. Il devient très difficile de savoir leurs natures (bénigne ou maligne). Dans des cas similaires, il faut répéter des examens cliniques après quelques mois pour voir l'évolution des masses hépatiques. Lorsque les anomalies sont de taille moyenne (<3 cm de diamètre), le diagnostic reste encore difficile à réaliser. Le diagnostic du CHC ne peut être affirmé que dans environ un tiers des cas. Certains CHC infiltrant (10-15 % des cas) sont également difficiles à diagnostiquer, en l'absence de lésion focale en imagerie, et sont peu accessibles au traitement. Nous avons besoin tout d'abord de trouver des moyens qui servent à l'optimisation du diagnostic des tumeurs du foie à travers l'introduction des ontologies dans l'analyse d'images IRM. Cette solution permettra aux médecins et aux radiologues d'accélérer et de faciliter le diagnostic et la détection du CHC et d'autres lésions hépatiques.

I.5 Objectifs

Ce sujet de thèse porte sur l'aide au diagnostic et au traitement du CHC assisté par ordinateur. En premier lieu, nous proposons d'élaborer une synthèse générale et complète sous forme d'un état de l'art sur la classification de données médicales pour le diagnostic de tumeurs du foie, et l'utilisation des ontologies pour l'optimisation du diagnostic. Nous souhaitons ensuite développer une plateforme capable d'intégrer des grandes masses de données médicales, à travers la proposition de solutions cohérentes, extensibles et exploitables dans l'approche d'analyse d'images médicales pour la détection des tumeurs du foie de type. La détection et la localisation de ces tumeurs hépatiques est faite à travers le développement d'un algorithme de type *deep learning*. Aussi, nous abordons l'extension des ontologies développées pour le traitement d'images médicales et issues de différents types d'examen (scanner et IRM) pour la détection et le diagnostic des tumeurs du foie. Ces ontologies sont utilisées initialement pour le diagnostic et l'aide au traitement de tumeurs CHC du foie ainsi que pour l'analyse des rapports radiologiques reliées à des patients infectés. En troisième lieu, nous proposons de coupler le *deep learning* et les ontologies à travers une approche d'analyse d'images IRM du foie dédiée au diagnostic du CHC. L'objectif est d'adapter les modélisations sémantiques avec les algorithmes d'apprentissage automatique pour enrichir l'architecture *deep learning* et améliorer les résultats de classification. Cette approche devrait être capable de

traiter une grande masse d'images médicales. A l'avenir, il faudra passer à la validation de l'ensemble des solutions proposées dans le cadre du projet CIRRHOSE avec des praticiens hospitaliers et des experts du domaine. Finalement, le travail sur ce sujet sera finalisé par une étude critique et un positionnement par rapport aux autres approches.

I.6 Contributions

Les principales contributions présentées dans cette thèse sont les suivantes :

1. La proposition d'une approche ontologique pour l'assistance au diagnostic, classification et traitement du CHC (OntHCC) permettant l'aide au diagnostic, à la classification et au choix de traitement des tumeurs CHC. Cette approche est fondée sur l'analyse d'images IRM de foies infectés et sur des rapports radiologiques. Elle permet de rassembler les informations nécessaires et relatives à la maladie concernée (CHC) sur la base de cas réels. Les informations tirées à partir de ces images seront introduites par la suite dans l'ontologie. Elle est basée également sur les directives des différents systèmes médicaux standards, utilisés à l'échelle internationale, pour la classification de tumeurs CHC, à savoir les systèmes BCLC (*Barcelona Clinic Liver Cancer*), TNM (*Tumor, Node, and Metastasis*) et LI-RADS (*Liver Imaging Reporting and Data System*).
2. La proposition d'une ontologie MROnt (« *Medical Report Ontology*») pour la modélisation de l'information médicale contenue dans les rapports radiologiques du CHC, dans le cadre du diagnostic et de suivi de tumeurs du foie. L'interprétation radiologique décrit l'état du foie (par exemple la dysmorphie, les contours irréguliers) et de la lésion hépatique (taille, localisation, nombre de nodules). De plus, notre système vérifie s'il existe une anomalie liée au système hépatique et autres organes, y compris les observations de pancréas ou des reins par exemple. Cette approche permet un meilleur échange de connaissances et de données cliniques. Nous avons introduit des aspects sémantiques pour analyser les informations contenues dans les rapports radiologiques. En outre, nous avons inclus des règles de raisonnement pour extraire des observations cliniques à partir de rapports de cas réels de tumeurs du foie.
3. Le développement d'une approche de classification d'images IRM basée sur le *deep learning* vise à intégrer l'apprentissage profond dans la classification d'images IRM. L'objectif de ce travail est de : (1) développer un algorithme parallèle de prétraitement permettant de fournir l'information nécessaire à l'entrée de l'algorithme *deep learning*. Il prend comme entrée des images médicales IRM et donne en sortie des fragments de tailles identiques, (2) créer un algorithme CNN convolutif (*Convolutional Neural Network*) qui permet la détection et la localisation des tumeurs dans une image IRM du foie.

4. La proposition d'une approche de couplage basée sur le *deep learning* et guidée par les ontologies. Dans cette approche, nous avons développé un CNN *multi-label* permettant d'intégrer l'information sémantique extraite à partir des ontologies développées dans le processus de classification des images IRM.

I.7 Organisation de la thèse

Ce mémoire de thèse est constitué de sept chapitres :

Chapitre 2 « **Concepts de base : Ontologies et *deep learning*** » : dans ce chapitre, nous nous intéressons aux notions de base des deux principaux concepts de notre thèse : ontologie et *deep learning*. Nous présentons en premier lieu une vue générale sur les ontologies : leurs caractéristiques, types, objectifs et méthodes de développement. En deuxième lieu, nous exposons les techniques de l'apprentissage automatique en se concentrant principalement sur les aspects liés à l'apprentissage profond.

Chapitre 3 « **Etat de l'art : Classification de lésions hépatiques via l'apprentissage automatique et les ontologies** » : Ce chapitre présente un état de l'art sur l'intégration d'ontologies dans la classification des lésions hépatiques. Il étudie également les travaux utilisant les techniques d'apprentissage automatique pour la détection des lésions hépatiques à partir des images médicales ainsi que leur couplage avec les techniques d'ingénierie de connaissances pour améliorer le processus de classification.

Chapitre 4 « **Représentation des connaissances médicales à travers les ontologies** » : Ce chapitre détaille la construction des ontologies pour la représentation des lésions hépatiques. en se basant sur la méthodologie « MethOntology ». Nous présentons dans ce chapitre les prototypes ontologiques OntHCC (« *HCC Ontology* ») et MROnt. L'ontologie OntHCC est destinée à la stadification des tumeurs CHC en utilisant trois différents systèmes de classification de tumeurs, à savoir : BCLC [67], TNM [Bertero 18] et Li-RADS [I35] [I36]. La deuxième ontologie MROnt permet l'analyse de rapports radiologiques et la représentation des observations cliniques notées par les radiologues. Dans ce chapitre, nous présentons également les règles de raisonnement créées à travers le langage SWRL¹⁰ (« *Semantic Web Rule Language* »).

Chapitre 5 « **Approche de classification d'images IRM basée sur le couplage ontologies-deep learning** » : Ce chapitre étudie les algorithmes basés sur le *deep learning* pour la détection des tumeurs hépatiques à partir des séquences IRM. Nous détaillons dans la première partie de ce chapitre les étapes de la construction d'un modèle CNN permettant la détection des lésions de type CHC observés dans les fragments de chaque image. Dans la deuxième partie, nous intégrons les connaissances sémantiques dans le processus de classification pour améliorer la performance de l'algorithme développé à travers un modèle CNN multi-label.

10. <http://www.institutpascal.uca.fr/>

Chapitre 6 « **Prototypes et expérimentations** » : Ce chapitre est consacré à la présentation des environnements logiciels utilisés pour le développement des prototypes présentés dans les chapitres précédents. Nous détaillons aussi dans ce chapitre les évaluations effectuées pour la validation des approches proposées. De plus, nous montrons les expérimentations réalisées dans le cadre d'une étude de cas pour l'ontologie OntHCC. Les approches proposées ont été évaluées en se basant sur des données réelles issues d'examens cliniques anonymisés au préalable.

Nous clôturons ce mémoire par un bilan de nos travaux tout en donnant une vue sur les perspectives de recherche envisagées et les nouveaux axes de recherche qui nous paraissent les plus pertinents.

CONCEPTS DE BASE : ONTOLOGIES ET DEEP LEARNING

Sommaire

II.1 Introduction	9
II.2 Ontologie	9
II.2.1 Définitions de l'ontologie	10
II.2.2 Objectifs de l'ontologie	11
II.2.3 Caractéristiques des ontologies	11
II.2.4 Classifications des ontologies	11
II.2.5 Composants de l'ontologie	13
II.2.6 Cycle de vie d'une ontologie	14
II.2.7 Méthodes de développement d'une ontologie	15
II.2.8 Langages de modélisation des ontologies	17
II.2.9 SWRL : Langage de représentation de règles	20
II.2.10 Les domaines d'application des ontologies	20
II.3 Apprentissage automatique	22
II.3.1 Définition	22
II.3.2 Types d'apprentissage automatique	23
II.3.3 RNA (Réseaux de neurones artificiels)	23
II.3.4 Deep learning	28
II.3.5 Les réseaux de neurones convolutifs (CNN)	30
II.4 La programmation parallèle	31
II.4.1 Définition	31
II.4.2 Architectures	32
II.4.3 Grille de calcul	33
II.5 Conclusion	33

II.1 Introduction

L'ontologie a connu un succès croissant depuis son apparition. L'utilisation fréquente de cette technique d'ingénierie des connaissances est expliquée par son comportement comme un référentiel conceptuel développé sous la forme d'une modélisation informelle. Elle est caractérisée par une représentation explicite et sous format compréhensible d'un domaine donné. Elle reflète également la description du monde réel à travers des langages de formalisation spécifiques. Dans de multiples domaines de recherche, l'ontologie est devenue un outil pour l'achèvement des tâches complexes et la réalisation de solutions performantes pour la recherche d'information, le Web sémantique, le génie logiciel, ou encore l'informatique médicale. Dans le domaine médical, les ontologies ont été appliquées non seulement pour la modélisation et l'analyse des données mais également pour la classification des images médicales. En effet, elles sont liées à plusieurs techniques de classification à savoir l'apprentissage automatique ou le *deep learning* (ou apprentissage profond). Le fait de combiner les ontologies aux techniques issues des domaines de l'apprentissage automatique caractérise un domaine récent permettant de concevoir de nouvelles stratégies de raisonnement. Dans ce chapitre, nous explorons en premier lieu les notions liées aux ontologies et en deuxième lieu les techniques de l'apprentissage automatique tout en mettant l'accent sur les approches de *deep learning*.

II.2 Ontologie

L'ontologie est définie comme étant une discipline de la philosophie qui a pour rôle la définition de l'être ou de l'existence. Elle caractérise un constituant principal du domaine des sciences humaines. Historiquement, Le terme « ontologie », construit à partir des racines grecques, est la composition des deux mots : « Ontos » qui désigne l'être et « Logia » qui désigne le discours. Dans ce contexte, Aristote a décrit cette technique comme étant « la science de l'être en tant qu'être ». Plus précisément, l'ontologie se préoccupe de déterminer quelles catégories de l'être sont fondamentales. Elle est liée à l'intelligence artificielle et aux techniques d'extraction de données de manière qu'elle a investi le cadre des sciences de l'information et des humanités numériques (voir Figure [II.1](#)).

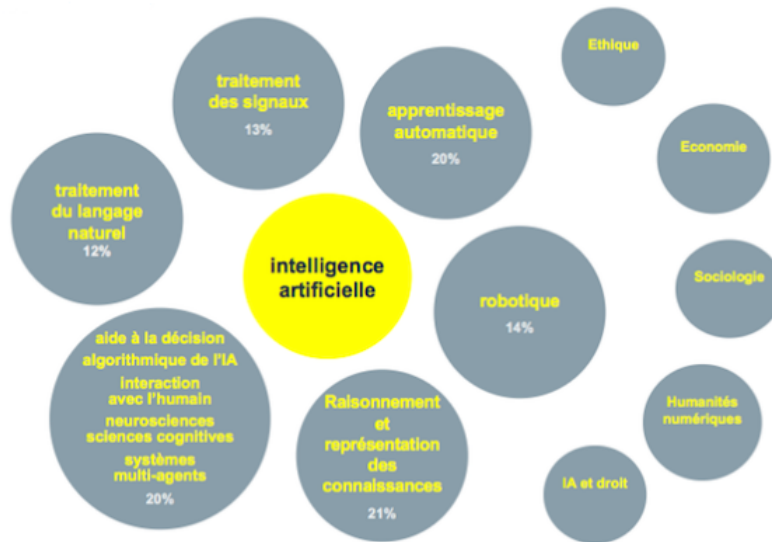


FIGURE II.1 – Visualisation des sous-domaines de l'Intelligence Artificielle [126]

II.2.1 Définitions de l'ontologie

L'ontologie est une représentation partagée, consensuelle, et sous format compréhensible d'un domaine donné. Elle caractérise un moyen pour fournir un vocabulaire spécifique à ce domaine. Par ailleurs, elle permet d'offrir tous les moyens aidant à comprendre les constituants d'un domaine. En effet, elle facilite la modélisation du monde réel à travers des formalismes et des langages de représentation très spécifiques.

Dans la littérature, nous trouvons plusieurs définitions de l'ontologie. Selon Gruber « une ontologie est une spécification explicite d'une conceptualisation » [1]. Elle caractérise une représentation d'un domaine sous la forme de concepts, de propriétés et de relations entre les concepts. Elle est considérée également comme étant « une base de connaissances particulière » [2]. Dans [3], elle est construite à travers des termes et des relations de base du vocabulaire ainsi que des règles qui indiquent la manière de combiner ces termes et ces relations de façon à pouvoir élargir le vocabulaire. Elle peut être référencée aussi par des identificateurs universels [131]. Elle est définie également comme « une convention de vocabulaire et de relations entre ses mots » [137]. De plus, l'ontologie est devenue une référence pour les systèmes d'information à travers un vocabulaire vaste et riche en informations. Elle offre un corpus de connaissances supportable et compréhensible par la machine. Selon [138], une ontologie peut prendre des formes multiples, mais elle doit inclure obligatoirement un vocabulaire de termes et une spécification de leur signification. Cette dernière inclut des définitions et une indication de la façon dont les concepts sont reliés entre eux, les liens imposant collectivement une structure sur le domaine et contraignant les interprétations possibles des termes. Dans le même contexte, [117] considère l'ontologie comme étant un outil qui englobe tout ce qui constitue un domaine donné. [4] mentionne qu'une ontologie est une « spécification rendant partiellement compte d'une conceptualisation ». Dans la section suivante, nous abordons les différents rôles possibles des ontologies à travers les objectifs que ces dernières peuvent remplir.

II.2.2 Objectifs de l'ontologie

L'ontologie permet en premier lieu de réaliser un monde de partage commun entre les systèmes hétérogènes et les personnes ainsi qu'une structure d'information compréhensible entre eux, ce qui permet d'éviter l'ambiguïté entre les termes et clarifier leurs sens. Elle permet également d'offrir tous les outils aidant à comprendre les constituants d'un domaine particulier. En effet, elle facilite la modélisation du monde réel à travers des formalismes et des langages de représentation très spécifiques. En deuxième lieu, l'ontologie est devenue une référence pour les systèmes d'informations à travers un vocabulaire riche en informations. Dans [4], l'auteur considère que l'ontologie vise à réaliser : la spécification dont elle permet d'analyser les besoins des systèmes informatiques, la réutilisation et dans ce cas, l'ontologie devient un composant réutilisable qui peut être traduit à tout moment pour atteindre la compréhension commune des termes. De plus, l'auteur mentionne que la maintenance constitue aussi les objectifs principaux de l'ontologie. Elle assure l'interopérabilité tout en faisant coopérer différents types de formats de systèmes d'information.

II.2.3 Caractéristiques des ontologies

Les ontologies possèdent généralement des caractéristiques fondamentales communes. Elles sont formelles, ce qui signifie qu'elles sont écrites par un langage ayant une syntaxe à base mathématique. Elles sont lisibles, c'est-à-dire elles doivent être compréhensibles par les machines et par les personnes. De plus, les ontologies sont vastes de façon à couvrir tous les domaines. En outre, elles sont partageables sous la forme de bibliothèques communes de concepts fondamentaux. Ceci facilite la réutilisation et la combinaison des ontologies développées, ce qui assure la bonne communication entre les systèmes d'information. En se basant sur le travail de [1], il existe quelques critères pour la mise en évidence des aspects d'une ontologie. Parmi ces critères, la clarté consiste à définir les concepts d'une manière aussi objective que possible, indépendante du contexte, complète et écrite en langage naturel. Nous trouvons également la cohérence pour éviter les contradictions lors de la phase de la création des concepts. L'extensibilité permet d'enrichir l'ontologie avec de nouveaux concepts sans altérer la stabilité de l'ontologie. Et finalement, l'ontologie doit être ciblée à un domaine d'application précis dès sa création pour achever ses objectifs attendus.

II.2.4 Classifications des ontologies

Plusieurs critères de classification ont été proposés pour catégoriser une ontologie. Parmi ces classifications, on trouve l'objet de conceptualisation, le niveau de détail et le niveau de formalisation (informel, semi-formel, ou formel). La classification de [2] est basée sur le niveau de généralité et de dépendance de l'ontologie. Nous distinguons quatre types d'ontologies (voir Figure II.2) : les ontologies de haut niveau, les ontologies de domaine, les ontologies de tâche et les ontologies d'application.

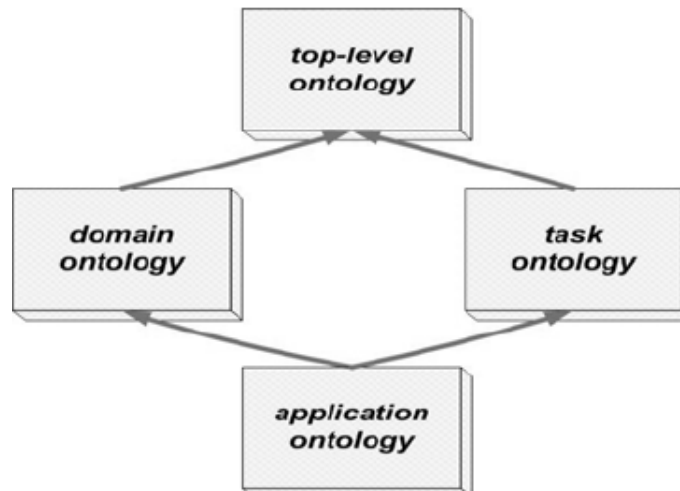


FIGURE II.2 – Classification des ontologies [5]

- **Les ontologies de haut niveau (ou « *top-level ontology* »)** : Ce type concerne une ontologie générique, globale et de plus haut niveau. Il est fondé principalement sur la théorie de l'identité et la théorie de la dépendance. De plus, il traite des sujets et des catégories universels tels que les entités, les concepts, les états, les événements les actions, le temps, les processus, les relations, et les propriétés. SUMO (*Standard Upper Ontology*) et DOLCE (*Descriptive Ontology for Linguistic and Cognitive Engineering*) [119] sont les ontologies de haut niveau les plus connues [6] [7].
- **Les ontologies de domaine (ou « *domain ontology* »)** : Cette ontologie s'intéresse à la création des concepts relatifs à un domaine d'application cible tels qu'ils sont manipulés par les experts. Elle s'agit d'un méta-modèle de connaissance où les concepts et les propriétés dans ce cas sont de type déclaratif. La majorité des ontologies existantes sont des ontologies de domaine. Parmi les ontologies de domaine, nous pouvons citer TOVE (Ontologie sur la notion d'activité) [119].
- **Les ontologies de tâche (ou « *task ontology* »)** : Ce type d'ontologie traite des tâches spécifiques à savoir les tâches de planification, de diagnostic, d'éducation et de configuration. Une ontologie de tâche crée un ensemble de vocabulaires servant à la résolution des problèmes liés aux tâches et indépendante du domaine. Cette catégorie d'ontologie est basée sur une architecture computationnelle d'un système qui exerce une tâche. L'ontologie de formation par ordinateur (« *Computer Based Training Ontology* ») et l'ontologie des objectifs d'apprentissage (« *Learning Goal Ontology* ») sont deux exemples d'ontologies de tâches [8] [9].
- **Les ontologies d'application (ou « *application ontology* »)** : Elle s'intéresse non seulement aux concepts du domaine mais aussi aux tâches. Elle caractérise l'ontologie la plus spécifique permettant d'appliquer les concepts pour l'exécution d'une certaine tâche ou une activité [10].

II.2.5 Composants de l'ontologie

L'ontologie, ayant pour but la conceptualisation d'un domaine, offre un vocabulaire commun permettant la signification des termes et les relations entre elles. Elle est formalisée en utilisant cinq composants essentiels : les concepts, les propriétés, les relations, les axiomes et les instances [139].

- **Concepts** : nommés aussi classes ou termes, ils constituent les notions, les fonctions ou les tâches de l'ontologie. Ils caractérisent l'unité de base qui forme le vocabulaire de l'ontologie. Ils sont des notions permettant la description d'un processus de raisonnement qui constitue le centre d'intérêt d'une ontologie. Les concepts sont liés par des relations et sont décrits par des propriétés. Un concept peut prendre plusieurs formes : abstrait ou concret, élémentaire ou composé, réel ou non [125] [139].
- **Propriétés** : (ou attributs) ce sont des caractéristiques ou fonctionnalités que portent les concepts d'une ontologie. Elles constituent des restrictions des concepts ou des relations. Au sein d'une représentation ontologique, on affecte des attributs (concepts ou instances) à une instance d'un concept [125]. Plusieurs liens conceptuels peuvent s'exprimer à l'aide de propriétés portées par les concepts, d'autres doivent être représentés à l'aide de relations autonomes. Par exemple le concept « Personne » a comme attribut « Nom ».
- **Relations** : une relation définit un lien interactif entre les concepts et traduit les associations entre eux. Elle assure la structuration et l'interrelation des concepts. Elle est caractérisée par un terme et une signature qui précise le nombre d'instances de concepts que la relation lie, leurs types et l'ordre des concepts [124]. Les relations peuvent également être spécifiées par des attributs. Parmi les propriétés liées à une relation, on peut citer : les propriétés algébriques (la symétrie, la réflexivité, et la transitivité), les propriétés liant deux relations (l'incompatibilité, l'inverse et l'exclusivité) et les associations (partie de, sous classe de, associé à, instance de, *etc.*).
- **Axiomes** : les axiomes décrivent des descriptions sémantiques des classes ou des relations sans oublier leurs propriétés. Ils désignent des assertions ou des contraintes exprimées d'une façon formelle. Ils donnent la signification des composants de l'ontologie. Les axiomes sont écrits globalement en utilisant les prédicats de l'ontologie et sous la forme de formules bien structurées.
- **Instances** : les instances sont des groupes d'individus qui représentent les connaissances à propos du domaine étudié. Ils représentent l'extension des concepts qui sont déjà créés. L'enrichissement (ou le peuplement) d'une ontologie s'effectue toujours à travers ces composants.

II.2.6 Cycle de vie d'une ontologie

Le développement d'une ontologie, comme tout autre logiciel ou système qui répond à des objectifs particuliers, s'appuie sur des principes bien précis. En effet, l'ontologie doit passer par un processus de construction évolutif et équilibré. Les activités liées au cycle de vie d'une ontologie sont des activités de gestion de projet et des activités transversales de support. Les phases sont les suivantes : la spécification de besoins, la conception, l'évaluation, l'évolution, la diffusion et l'utilisation [11] (voir Figure II.3).

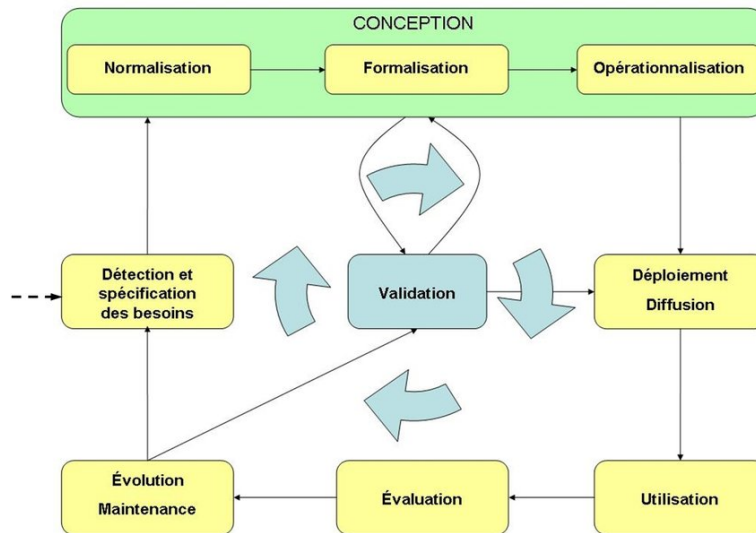


FIGURE II.3 – Cycle de vie d'une ontologie [11]

Le cycle de vie d'une ontologie comporte tout d'abord une phase initiale, c'est la spécification, qui consiste à extraire les besoins et identifie le domaine à traiter. Elle s'effectue généralement à travers des entretiens, l'étude de l'ergonomie, des questionnaires, et à travers l'identification (par exemple, la modélisation par scénarios). Ensuite, nous avons la phase de conception. Elle est basée sur le processus de construction d'une ontologie selon une méthodologie bien déterminée. Cette phase est composée de trois étapes principales qui sont la normalisation (conceptualisation), la formalisation (ontologisation) et l'opérationnalisation. L'étape de diffusion consiste à la mise en place de l'ontologie. C'est une phase de déploiement et de détection des problèmes évoqués. La phase de l'utilisation est une interaction directe avec l'utilisateur. Cette étape sert à effectuer une annotation de ressources telles que la rétro-ingénierie de base de données, la résolution de requête, la déduction de connaissances et l'aide à la décision. Durant la phase de gestion, la modification de l'ontologie et l'évolution de ses performances sont visées. C'est une phase itérative qui s'applique tout au long du processus de développement. L'évaluation de l'ontologie développée se fait à partir de tests répondant à l'objectif opérationnel de l'ontologie en se basant essentiellement sur des questionnaires. Il s'agit d'une phase très délicate car elle emmène par la suite à la validation de l'ontologie. En effet, le contenu d'une ontologie doit être évalué avant d'être utilisé ou réutilisé [127]. Dans ce contexte, [12] indique que l'évaluation caractérise une série de tests établie et exécutée. Les résultats sont analysés par la suite en se basant sur trois états possibles : la pré-modélisation, la modélisation ou la libération [13] [132]. En outre, l'ontologie peut être évaluée

en s'appuyant seulement sur des critères spécifiques tels que le vocabulaire utilisé, la structure taxonomique, la consistance et la non redondance, *etc.* [140]. Tous ces moyens servent à évaluer une ontologie sans oublier que les experts du domaine peuvent aussi donner leurs avis sur sa consistance. En ce qui concerne la phase de l'évolution, elle se réalise lorsqu'il y a une modification au niveau des objectifs attendus ou lorsque le domaine de connaissances est élargi. Dans ce cas, il faut ajouter de nouvelles connaissances [17].

II.2.7 Méthodes de développement d'une ontologie

L'application des méthodes d'ingénierie ontologique n'est pas suffisante pour assurer le bon développement d'une ontologie. En effet, il faut des directives bien structurées et communes à suivre pour ne pas retarder son développement et sa réutilisation par la suite. Une méthodologie caractérise les démarches ou les étapes à suivre pour la construction d'un travail ou d'un projet. Dans la suite de cette section, nous allons présenter quelques méthodes utiles pour le développement d'une ontologie.

II.2.7.1 Méthode de Gruninger et Fox (TOVE)

Cette méthodologie est fondée sur le développement du projet TOVE qui a comme objectif la modélisation des processus et des activités [14]. Elle est basée sur la construction d'un modèle logique de connaissance qui doit être spécifié et adéquat au moyen de l'ontologie. L'élaboration de ce modèle se fait en suivant un enchaînement particulier :

- dégager les spécifications ;
- capturer les scénarios et écrire les questions de compétences d'une manière informelle ;
- spécifier formellement la terminologie de l'ontologie ;
- déterminer les conditions et les contraintes pour accomplir la complétude de l'ontologie.

II.2.7.2 Méthodologie de Uschold et King

Cette méthode est fondée sur l'expérience de développement dans le domaine de la gestion d'une entreprise, qui caractérise une ontologie présentant le fonctionnement des processus au sein de l'entreprise [15]. Elle offre une stratégie de construction en fonction des besoins et des objectifs attendus. Cette méthodologie comporte les étapes suivantes :

- l'identification des buts ;
- la création de l'ontologie en se basant sur trois phases, à savoir la capture des connaissances, le codage, et la réutilisation des ontologies existantes ;
- l'évaluation de l'ontologie ;
- la documentation de l'ontologie.

II.2.7.3 Méthode de Bernaras et al.

Cette méthodologie s'intègre dans le cadre du projet Esprit KACTUS, elle met l'accent sur la faisabilité des ontologies et leur réutilisation pour supporter les systèmes complexes [16]. Elle est

conditionnée par le développement d'une application, et basée essentiellement sur trois points :

- la spécification de l'application permettant d'extraire les tâches à modéliser ;
- l'organisation et la conception des termes (concepts, relations, attributs, *etc.*) ;
- l'affinement et la structuration de l'ontologie en se basant sur des modèles d'organisation hiérarchiques.

II.2.7.4 Terminae

Terminae ¹¹ est une méthodologie qui s'inscrit dans le cadre de l'ingénierie des connaissances linguistiques [121]. C'est une succession des étapes suivantes :

- la collection des données pour présenter le corpus d'un domaine ;
- l'extraction de la terminologie du thème traité à travers l'outil LEXTER [121] ;
- la normalisation et la conceptualisation des termes ;
- la formalisation des données.

II.2.7.5 Methontology

Cette méthodologie (voir Figure II.4), proposée par [17], vise à clarifier les étapes de construction d'une ontologie en respectant les activités de gestion de projets. Elle est basée sur les étapes suivantes :

- la spécification : indique l'objectif de l'ontologie à développer ainsi que les utilisateurs visés ;
- la conceptualisation : consiste à déterminer les connaissances (concepts, propriétés, relations et instances) du domaine étudié et assurer le passage de la création de diagrammes abstraits à la phase de l'implémentation ;
- la formalisation : sert à transformer l'ontologie en un modèle formel en utilisant des traducteurs. Durant cette étape, l'ontologie est écrite avec un langage spécifique (logique de premier ordre par exemple) ;
- l'implémentation : il s'agit de l'expression du modèle ontologique formel à l'aide d'un langage d'implémentation tel que l'OWL ¹² (« *Web Ontology Language* ») ;
- la maintenance : améliore et raffine l'ontologie avec des modifications.

11. <http://lipn.univ-paris13.fr/fr/article/1314-terminae/166>

12. <http://www.w3.org/TR/owl-features/>

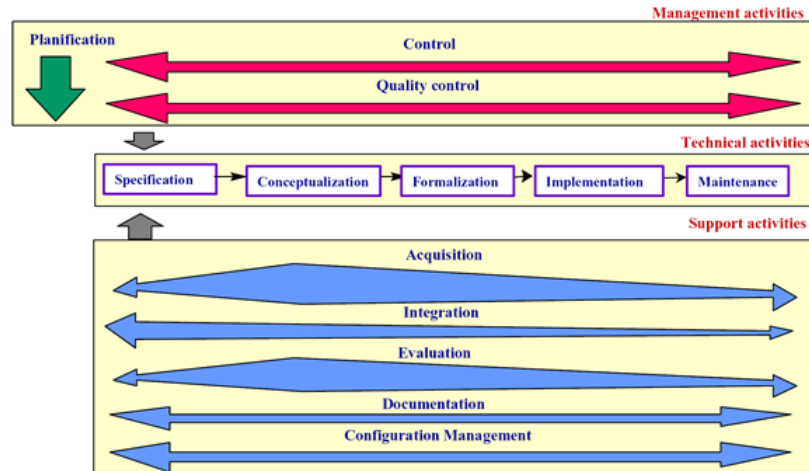


FIGURE II.4 – Vue sur la méthodologie MethOntology d’après [17]

II.2.7.6 La méthodologie NeOn

Cette méthode a été lancée en 2006 à travers un projet européen nommé NeOn¹³ et a pour objectif la mise en place de différentes ontologies dans un seul réseau. Les ontologies concernées sont généralement reliées par divers types de relations comme les dépendances, les importations, les alignements, et la modulation [133].

II.2.8 Langages de modélisation des ontologies

Un langage de modélisation d’ontologie est basé sur une sémantique formelle et une structure syntaxique particulière. Il assure une représentation ontologique compréhensible et de forte puissance. Dans ce contexte, OWL, RDF¹⁴ (« *Resource Description Framework* »), RDF Schéma, et LD (Logique de description) ont été fondés par l’organisme W3C¹⁵ (« *World Wide Web Consortium* »). Ces langages sont présentés sous la forme de couches interconnectées en se basant sur l’architecture de [18] (voir Figure II.5). Dans la section suivante, nous détaillons chacune de ces couches.

13. <http://neon-project.org>

14. <http://www.w3.org/TR/rdf-syntax-grammar>

15. <https://www.w3.org/>

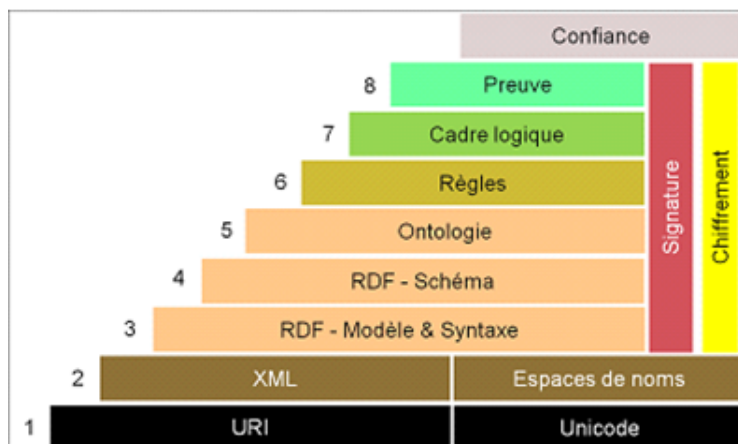


FIGURE II.5 – Architecture du Web sémantique [120]

II.2.8.1 RDF

RDF (« *Resource Description Framework* ») a été développé par W3C, il caractérise un modèle standard permettant l'échange de données sur le Web. Il assure également la structuration des ressources ainsi que leurs métadonnées. Principalement, il permet de faire l'annotation sémantique des documents non structurés à travers un modèle de graphe. En outre, ce langage est lié directement au Web sémantique en utilisant la syntaxe XML (« *eXtended Markup Language* »). Il est utilisé pour exprimer formellement les liaisons entre les ressources sous la forme d'un triplet <Sujet, Prédicat, Objet>. Le sujet et l'objet définissent les ressources à décrire, et le prédicat représente une propriété applicable à une ressource. Les composants de ces triplets peuvent être des URIs (« *Universal Resource Identifiers* ») [120], des variables ou des littéraux.

II.2.8.2 RDF Schéma

RDF Schéma¹⁶ (ou RDFS) est un langage extensible permettant la représentation des connaissances et la structuration des documents écrits en RDF. Il utilise un vocabulaire pour représenter des ressources RDF. Il détermine une sémantique de généralisation/hiéarchisation formée de concepts et de leurs propriétés, et permet de définir des triplets de graphes RDF. Il caractérise une des bases du Web sémantique puisqu'il permet de représenter des concepts liés à d'autres concepts définis et partagés sur le Web. À titre d'exemple *rdfs :subClassOf* définit une hiérarchie entre deux classes. Les classes de base reliées à ce langage sont *rdfs :Class*, *rdf :Property*, *rdfs :Ressource*, *rdfs :Literal* et *rdf :Statement*. Les propriétés RDFS présentant les relations sont *rdf :type*, *rdfs :subClassOf* et *rdfs :subPropertyOf*.

II.2.8.3 OWL

OWL (« *Web Ontology Language* ») est un langage universel standardisé par W3C et basé sur la syntaxe RDF/XML. Il permet de représenter les classes et leurs relations d'une façon compatible

16. <http://www.w3.org/TR/PR-rdf-schema>

au raisonnement logique des systèmes d'information. En plus, ce langage fournit la sémantique nécessaire appliquée pour la formalisation des connaissances avec une représentation plus fine. La particularité de ce langage par rapport à RDF et RDFS est sa capacité à insérer des outils de comparaison des propriétés et des classes, et des relations sémantiques entre ces classes telles que l'équivalence, la disjonction, la cardinalité, la symétrie, *etc.* En se basant sur le niveau d'expressivité et en fonction des besoins de représentation ou d'inférence, trois sous-langages ont été proposés par W3C avec une expressivité croissante, OWL-Lite, OWL DL et OWL Full, qui sont décrits ci-après.

- **OWL Lite** : c'est le sous-langage le plus simple. Il a été conçu pour caractériser les ontologies assez légères. Il est moins expressif mais comprend un aspect de raisonnement rapide. Il est appliqué pour exprimer des relations et des contraintes simples entre les classes. OWL Lite permet de définir de nouvelles classes, d'assurer l'équivalence entre les instances, et d'ajouter des contraintes simples à savoir celles de cardinalité 0 et 1.
- **OWL DL** (« *OWL Description Logics* ») : ce langage est basé sur la logique de description et comprend toutes les structures d'OWL. Il est destiné aux ontologies plus complexes qui décrivent plusieurs axiomes. Il assure également une expressivité importante tout en gardant une complétude de raisonnement et de calcul. Ce langage permet d'une part de gérer le secteur existant de la logique de description, et d'autre part d'offrir les propriétés de calcul désirées par les systèmes de raisonnement.
- **OWL Full** : c'est le sous-langage d'OWL le plus complexe, comprenant un état d'expressivité maximal. Il permet d'augmenter le sens du vocabulaire prédéfini d'une ontologie. Ce langage permet le traitement simultané des instances. Il assure une expressivité maximale et une liberté syntaxique de RDF.

II.2.8.4 Logique de description

La LD (logique de description) est un formalisme qui permet de modéliser essentiellement deux composants d'une base de connaissance : la TBox (pour les Terminologies) et la ABox (pour les Assertions). La TBox permet de traduire la terminologie, tandis que l'ABox comporte des assertions sur les individus. Il offre une inférence décidable et complète selon [118].

II.2.8.5 Raisonnement ontologique et inférence

L'ontologie possède l'avantage d'intégrer les notions de raisonnement et d'inférence au modèle créé. Par ailleurs, l'application d'un raisonneur permet de garantir une formalisation hiérarchique des concepts ainsi que la cohérence de l'ontologie. Les ontologies construites à l'aide du langage OWL utilisent des règles logiques qui sont définies par la sémantique du langage. Les mécanismes de raisonnement permettent de déterminer certaines actions relatives à l'ontologie et à sa hiérarchie. Elles permettent également de déduire des conséquences logiques liées aux classes d'une ontologie. Elles sont implantées directement dans des moteurs d'inférence, où sont exécutées des règles SWRL.

La notion d'inférence consiste essentiellement à déduire des informations explicites à partir des données implicites. Elle se réalise souvent à travers des règles associées à l'ontologie et exécutées par un raisonneur. La construction de ces règles se fait après la modélisation et l'implémentation du modèle ontologique. Cette définition est basée sur trois composants :

1. un langage pour la définition des règles et la déduction des faits simples à partir des faits déjà existants ;
2. un outil pour éditer l'ontologie développée et les règles associées ;
3. un raisonneur.

En outre, les règles d'inférence permettent l'enrichissement et le peuplement automatique de l'ontologie par les données de base, à savoir les instances. Ceci permet d'éviter la saisie manuelle de certaines tâches fastidieuses. Ces règles peuvent être déductives, c'est-à-dire qu'elles assurent la déduction des faits en se basant sur des conditions bien déterminées, ou réactives, et ainsi elles permettent l'exécution des actions après la vérification des conditions.

II.2.9 SWRL : Langage de représentation de règles

SWRL est un langage fondé par W3C qui fournit des capacités de raisonnement déductif permettant aux utilisateurs de rédiger des règles exprimées en utilisant des concepts OWL. Il combine les deux langages OWL-DL pour la construction des ontologies et RuleML (« *Rule Markup Language* ») pour la construction des règles. SWRL est destiné essentiellement au Web sémantique. Il est construit à travers des méthodes appelées *built-ins*, qui sont définies dans la spécification de SWRL, comme les opérateurs mathématiques et les fonctions de manipulations de chaînes de caractères et de dates. Les règles SWRL sont écrites selon le schéma suivant : Antécédent (*body*) -> Conséquent (*head*). L'antécédent et le conséquent sont des conjonctions d'atomes qui peuvent être soit une instance de concept, c'est-à-dire un prédicat unaire, soit une relation OWL. La première partie (Antécédent) spécifie les conditions qui doivent être vérifiées et la deuxième partie (Conséquent) spécifie les actions à faire. Une règle SWRL fonctionne selon le principe de satisfaction de l'antécédent ou du conséquent. Ce langage intègre plusieurs moteurs d'inférences à savoir Jess [19], Hoolet ¹⁷, Pellet ¹⁸, KAON2 ¹⁹, RacerPro ²⁰, et R2ML ²¹(« *REVERSE Rule Markup Language* »).

II.2.10 Les domaines d'application des ontologies

Les ontologies sont devenues très populaires dans de nombreux domaines d'application et pour le traitement de divers thèmes de recherche tels que l'intelligence artificielle, la médecine, le Web sémantique, la biologie, le traitement d'images, l'industrie, *etc.* En outre, elles permettent une

17. <http://owl.man.ac.uk/hoolet/>

18. <https://www.w3.org/2001/sw/wiki/Pellet>

19. <http://kaon2.semanticweb.org/>

20. <https://www.w3.org/2001/sw/wiki/RacerPro>

21. <https://www.w3.org/2005/rules/wg/wiki/R2ML.html>

modélisation claire, structurelle et partageable d'un domaine. Elles assurent également l'automatisation des données en se basant sur la sémantique des données. Dans cette section, nous abordons l'intégration des ontologies dans quelques domaines de recherche. En ce qui concerne le traitement d'images, l'ontologie se concentre sur le contenu sémantique de l'image. L'objectif est d'éviter tout « fossé sémantique » représentant le manque de concordance entre les données visuelles et leurs interprétations dans une situation déterminée [141]. Le raisonnement sémantique de l'image est établi en termes d'objets et de contenus et non pas seulement à travers les caractéristiques des données images à savoir la couleur, la texture, la forme, *etc.* Par ailleurs, la sémantique de l'image dépend de plusieurs aspects à savoir le niveau de connaissances, la perception de l'observateur, et l'objectif poursuivi par l'utilisateur de cette image. Sur le volet algorithmique, l'ontologie joue un rôle important pour enrichir des méthodes et des fonctions dans un système d'application en introduisant la notion de raisonnement. Dans ce contexte, le travail de [129] vise à annoter les images en se basant sur une modélisation unifiée des caractéristiques de l'image, la description visuelle des relations entre ces caractéristiques tels que les lignes, les régions, *etc.*, ainsi que l'utilisation des informations contextuelles tout en assurant un aspect de rapprochement entre le niveau visuel et le niveau sémantique de l'image à traiter. L'ingénierie linguistique ou le traitement automatique du langage naturel englobe les aspects de l'intelligence artificielle et de l'informatique pour assurer le traitement du langage humain. Dans ce contexte, le travail de [123] met l'accent sur l'importance de l'application des ontologies pour accéder à différents types de ressources d'un tel système. Il s'intéresse plus particulièrement aux notions de base de l'ontologie et aux concepts. Un autre travail de [20] propose également une approche ontologique pour le traitement d'images. Cette ontologie contient 279 concepts, 42 rôles et 192 restrictions. Les concepts y sont classés par niveau : physique, perception, sémantique, tâche et contrainte. [130] présente une modélisation sémantique permettant de capturer la structure fonctionnelle de la phrase, c'est-à-dire les rôles des termes les uns par rapport aux autres.

Dans le domaine médical, les ontologies offrent un environnement de travail compréhensible, commun et partageable entre médecins et radiologues. Elles permettent d'une part de faciliter le processus de diagnostic et d'autre part d'aider à la prise de décision. En effet, les notions des réseaux sémantiques encouragent à exploiter en premier lieu les données médicales (images médicales, données reliées au patients, *etc.*) et les automatiser en deuxième lieu pour trouver des résultats pertinents. Depuis que cette technique a été intégrée dans la médecine, elle devient le sujet principal pour de nombreux travaux de recherche notamment les projets GALEN, UMLS ²² (« *Unified Medical Language System* »), MeSH ²³ (« *Medical Subject Headings* »), TAMBIS ²⁴ (« *Transparent Access to Multiple Bioinformatics Information Sources* »), FMA ²⁵ (« *Foundational Model of Anatomy* »), SNOMED ²⁶ (« *The Systematized Nomenclature of Medicine* »), Gene Ontology ²⁷,

22. <https://www.nlm.nih.gov/research/umls>

23. <https://www.nlm.nih.gov/mesh/>

24. <http://www.cs.man.ac.uk/~stevensr/tambis/details.html>

25. <http://si.washington.edu/projects/fma>

26. [ioportal.bioontology.org/ontologies/SNOMEDCT/?p=summary](http://portal.bioontology.org/ontologies/SNOMEDCT/?p=summary)

27. <http://geneontology.org/>

MENELAS²⁸, LinkBase²⁹, CO-ODE (« *Collaborative Open Ontology Development Environment* »), OBO (« *Open Biomedical Ontology* »), etc.)

L'utilité d'une ontologie médicale s'appuie sur deux phases complémentaires qui sont la modélisation ou la construction des concepts fondamentaux permettant la prise en charge d'une maladie ou le suivi hospitalier (diagnostic des patients, dossiers médicaux, etc.) et l'analyse terminologique des corpus qui permet d'enrichir et de structurer ce noyau ontologique. Dans ce cadre, certains travaux de recherche ont intégré les connaissances sémantiques avec les techniques d'apprentissage automatique pour améliorer le processus de diagnostic et de classification des lésions hépatiques. Dans la section suivante, nous étudions la notion de l'apprentissage automatique et ses différentes architectures.

II.3 Apprentissage automatique

Dans cette section, nous identifions l'apprentissage automatique et nous détaillerons ses architectures, et son mode de fonction.

II.3.1 Définition

L'apprentissage automatique (ou « *machine learning* ») fait partie de l'intelligence artificielle. Il a connu une évolution remarquable ces dernières années et reste toujours en développement continu. Il s'intéresse essentiellement à la conception, le développement, l'optimisation et l'implémentation de plusieurs méthodes de calcul qui visent à analyser et traiter les structures de données et les transformer sous forme de modèles. Ces modèles sont ensuite appliqués pour la résolution de différentes tâches (reconnaissance de formes, classification de données médicales, etc.) [21]. Ce qui caractérise l'apprentissage automatique par rapport aux autres approches classiques de classification c'est le fait d'entraîner les ordinateurs à calculer ou analyser des problèmes et améliorer leurs capacités à apprendre. En effet, les modèles basés sur l'apprentissage automatique sont construits grâce à différentes modélisations. Ils sont issus de courants pluridisciplinaires. Ils font référence à l'analyse, au traitement et à l'implémentation de méthodes qui permettent à une machine d'évoluer et de réaliser des tâches difficiles à remplir par des approches algorithmiques plus classiques. De plus, l'apprentissage automatique intègre l'analyse statistique pour former des données ou des valeurs. Il permet une construction des modèles à partir des données d'entrée, ce qui aide par la suite à l'automatisation des opérations de prise de décision.

L'apprentissage automatique a été intégré dans de très nombreux champs de recherche et de développement, notamment les sciences naturelles, la reconnaissance faciale, les jeux vidéos, la biologie ou encore la robotique. En ce qui concerne le domaine de l'imagerie médicale, l'apprentissage automatique est considéré comme un outil puissant pour l'analyse de différents types d'images médicales. Par exemple, il fournit des outils permettant de reconnaître et analyser des tissus anormaux dans les images médicales, dans le cadre du diagnostic assisté par ordinateur [22] [23].

28. <https://biportal.bioontology.org/ontologies/TOP-MENELAS>

29. <http://www.nuance.com/for-healthcare/resources/clinical-languageunderstandingontology/index.htm>

II.3.2 Types d'apprentissage automatique

Nous distinguons plusieurs types d'apprentissage automatique, qui sont décrits ci-après.

II.3.2.1 L'apprentissage supervisé

L'apprentissage supervisé consiste à guider l'algorithme avec des entrées qui sont étiquetées préalablement avec des sorties prédéterminées. Dans ce cas, les classes sont connues à l'avance. L'objectif de cette méthode est d'entraîner l'ordinateur à trouver des erreurs pour ajuster le modèle de classification en modifiant ses paramètres (par exemple les poids pour un réseau de neurones artificiel).

II.3.2.2 L'apprentissage non supervisé

L'apprentissage non supervisé est basé sur des entrées non étiquetées et les sorties ne sont pas déterminées *a priori*. Dans ce cas, l'algorithme d'apprentissage va déterminer les classes d'une façon autonome. L'objectif de cette méthode consiste à permettre à la machine de révéler des modèles cachés assurant la classification automatique des données brutes. Ce type d'apprentissage est utile pour traiter des données complexes et les organiser d'une manière significative.

II.3.2.3 L'apprentissage semi-supervisé

Ce type d'apprentissage utilise une combinaison des données étiquetées et non étiquetés et se situe entre les deux types d'apprentissage précédents. Cette combinaison a montré une précision de classification performante mais les problèmes du temps de calcul et des coûts de mémoire sont encore ouverts.

II.3.2.4 Apprentissage par renforcement

Ce type d'apprentissage se positionne également entre l'apprentissage supervisé et non supervisé. L'objectif de cette méthode est d'apprendre l'algorithme à trouver la meilleure solution à travers la construction de modèles successifs. L'apprentissage par renforcement s'intéresse à l'interaction avec l'environnement, ce qui permet de renforcer les comportements produisant les réponses adéquates. Il tient en compte quelques concepts clés basés sur le fait que l'agent intelligent observe les effets de ses actions, déduit de ses observations la qualité de ses actions et améliore ses futures actions. Lors de la phase de transmission au réseau, l'information peut être réduite : elle indique si la réponse générée du réseau est correcte ou non.

II.3.3 RNA (Réseaux de neurones artificiels)

Dans cette section, nous définissons en premier lieu le principe des réseaux de neurones artificiels (RNA), ensuite nous exposons différentes architectures de ce type.

II.3.3.1 Définition

Un réseau de neurones, prenant sa conception de la biologie, constitue une famille particulière d'algorithmes d'apprentissage. Il s'agit d'une technique de traitement connectionniste qui s'inspire du fonctionnement du cerveau humain. Il est constitué de plusieurs neurones formellement interconnectés assurant la résolution de plusieurs problèmes complexes tels que le traitement du langage naturel, la reconnaissance des formes, le traitement du signal, l'optimisation et la prédiction de données. Cette résolution s'effectue à travers l'ajustement des coefficients de pondération dans un algorithme d'apprentissage.

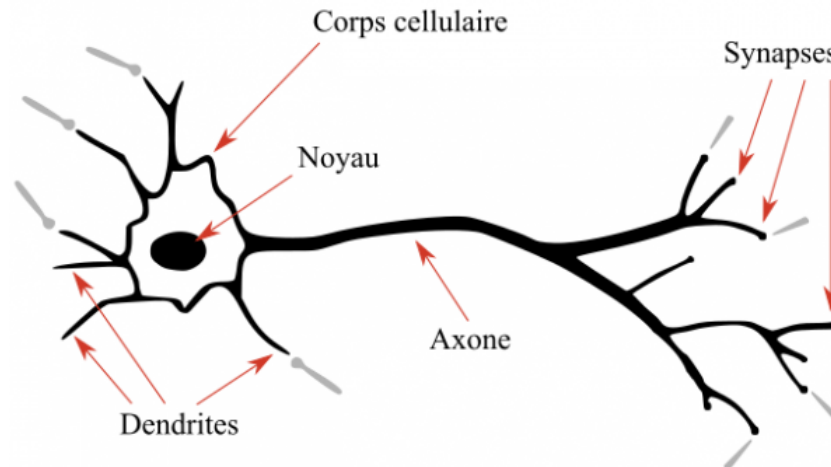


FIGURE II.6 – Visualisation d'un neurone biologique

Les réseaux de neurones sont définis comme étant des structures de traitement parallèle massif. Ils sont capables de stocker des connaissances expérimentales et de les rendre utilisables. Ils sont généralement optimisés à travers des méthodes d'apprentissages probabilistes. Ces réseaux appartiennent d'une part aux applications statistiques basées sur des paradigmes qui permettent la création des classifications rapides. D'autre part, ils appartiennent aux méthodes de l'intelligence artificielle qui offrent un mécanisme indépendant des idées propres du développeur, et des données d'entrée au raisonnement formel. Les réseaux de neurones artificiels et biologiques ont plusieurs points en communs ; (a) les neurones qui sont des unités simples et fortement connectés et (b) les connexions entre les neurones, appelées poids ou poids synaptiques, servent à stocker les connaissances acquises (voir Figure II.6)³⁰. Le fonctionnement des réseaux de neurones informatiques sont basés sur une modélisation simplifiée des neurones biologiques réels (voir Figure II.7). En effet, les neurones reçoivent des signaux (ou des impulsions électriques) à travers des unités appelés dendrites et envoient l'information par les axones. Le contact entre deux neurones se fait par des synapses.

30. <https://deeplylearning.fr/>

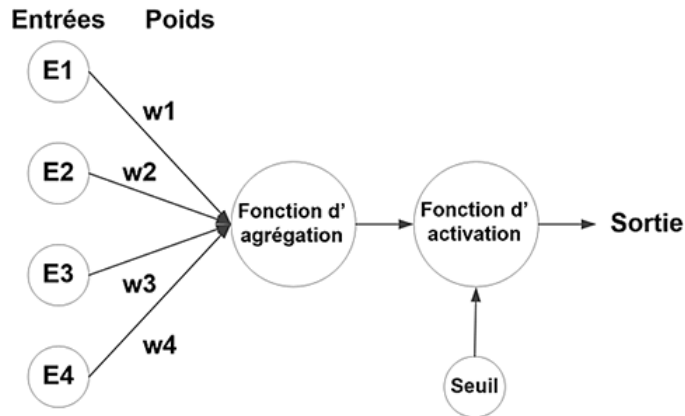


FIGURE II.7 – Visualisation d'un neurone artificiel

Le neurone artificiel reçoit les entrées d'informations (valeurs numériques) qui vont être attachées à un poids. Ces entrées correspondent aux dendrites, et les poids qui leur sont associés, correspondent aux actions des synapses, ils vont pouvoir amplifier ou minimiser un signal d'entrée. Le neurone dans sa forme basique, va effectuer une somme de l'ensemble de ces variables en fonction de leurs poids, correspondant au soma. Cette valeur passe ensuite par une fonction d'activation, qui en sera l'unique sortie. Celle-ci correspond au point de départ de l'axone qui est le cône d'émergence.

II.3.3.2 Architecture

Le réseau de neurones désigne un système logiciel inspiré des neurones humains. Il affecte des modifications et des améliorations à lui-même en fonction des résultats générés. Il y a plusieurs topologies de réseaux de neurones. La topologie la plus simple et la plus courante est le réseau à couches (voir Figure II.8), elle définit la base des autres architectures. Elle est constituée d'une couche d'entrée, une couche de sortie et de plusieurs couches cachées. Trois paramètres essentiels pour cette topologie : (a) le nombre de couches cachées, (b) le nombre de neurones dans chaque couche cachée et (c) les connexions entre les nœuds. Une description détaillée de ces entités est donnée dans la section suivante.

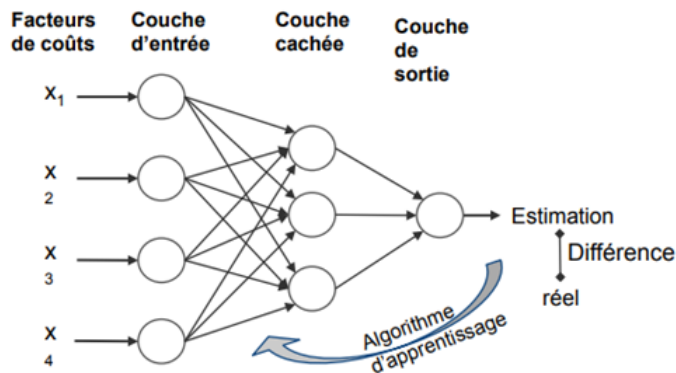


FIGURE II.8 – Architecture d'un réseau de neurones à couches

- **Neurone** : (ou nœud) contient une série d'entrées issues des sorties d'une couche précédente de sortie. Toutes les données d'entrée x_j sont pondérées d'une manière indépendante ω_j avant que leur somme et la valeur de seuillage θ_i soient utilisées pour calculer la sortie à travers le biais de la fonction d'activation (F). Chaque signal en entrée est multiplié par une valeur de poids. L'illustration d'un neurone est présentée dans la Figure [II.9](#).

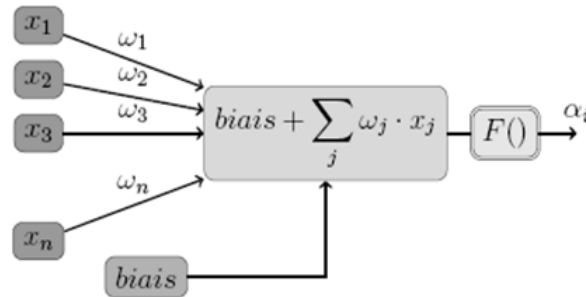


FIGURE II.9 – Illustration d'un neurone artificiel et notations

La formulation de la sortie est donnée par :

- **Poids** : représente la force de connexion entre les neurones. Il détermine l'influence des entrées sur les sorties. Plus sa valeur est élevée, plus l'intensité du signal entrant est importante, plus l'entrée est influente sur le calcul de la sortie.
- **Biais** : caractérise un neurone où la fonction d'activation est égale à 1 en permanence. Cela garantit qu'il y aura toujours une activation dans le réseau. Il se connecte aux autres neurones de la couche précédente à travers le poids.
- **Couche d'entrée** : caractérise la première couche du réseau. Elle transmet les signaux d'entrée (issus des données) à la couche suivante. Dans cette couche, aucune opération sur les signaux d'entrée n'est appliquée.
- **Couche cachée** : constitue une collection de neurones empilés verticalement. Elle est composée par des neurones qui appliquent des transformations aux données d'entrée.
- **Couche de sortie** : caractérise la dernière couche, elle reçoit l'entrée de la dernière couche cachée et a en sortie un nombre de neurones égal aux sorties attendues du système. Par exemple, si on attend une catégorisation en 2 classes, cette couche aura 2 neurones.
- **Fonction de perte** (ou « *loss function* ») : est une fonction associée à la dernière couche. Elle calcule l'erreur pour chaque opération de classification. On peut citer par exemple l'entropie croisée.

— **Fonction d'activation** : Elle représente une opération mathématique appliquée à un signal en sortie d'un neurone artificiel. Elle assure souvent la non-linéarité dans le réseau de neurones. Par exemple, une fonction d'activation sigmoïde normalise les valeurs entre 0 et 1, Heaviside renvoie 1 si le signal en entrée est positif et 0 s'il est négatif. La Figure II.10 expose les différentes fonctions d'activation classiques d'un réseau de neurone.

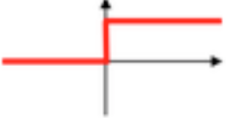
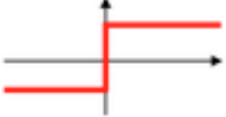
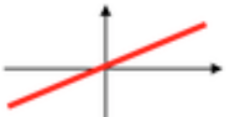



Activation function	Equation	Example	1D Graph
Unit step (Heaviside)	$\phi(z) = \begin{cases} 0, & z < 0, \\ 0.5, & z = 0, \\ 1, & z > 0, \end{cases}$	Perceptron variant	
Sign (Signum)	$\phi(z) = \begin{cases} -1, & z < 0, \\ 0, & z = 0, \\ 1, & z > 0, \end{cases}$	Perceptron variant	
Linear	$\phi(z) = z$	Adaline, linear regression	
Piece-wise linear	$\phi(z) = \begin{cases} 1, & z \geq \frac{1}{2}, \\ z + \frac{1}{2}, & -\frac{1}{2} < z < \frac{1}{2}, \\ 0, & z \leq -\frac{1}{2}, \end{cases}$	Support vector machine	
Logistic (sigmoid)	$\phi(z) = \frac{1}{1 + e^{-z}}$	Logistic regression, Multi-layer NN	
Hyperbolic tangent	$\phi(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$	Multi-layer NN	

FIGURE II.10 – Différentes fonctions d'activation pour un réseau de neurones [147]

Le réseau de neurones contient deux types de propagations : La propagation vers l'avant (ou « *Forward Propagation* ») et la propagation arrière (ou « *Back-Propagation* »). Dans la propagation vers l'avant, les données d'entrée sont acheminées vers l'avant et dans une seule direction. Chaque couche cachée reçoit ces données, elle les traite en utilisant la fonction d'activation et par la suite elle les envoie à la couche suivante jusqu'à arriver à la sortie. Dans ce type de propagation, il n'y a pas de cycles dans le réseau [24]. Il s'agit également d'un processus d'alimentation des données en

entrée dans le réseau afin de prédire une valeur de sortie. En premier lieu, il y aura la transmission des valeurs d'entrée à la première couche du réseau neuronal, et dans ce niveau il n'y a aucune opération à effectuer. La deuxième couche prend les valeurs de celle la première et applique une fonction donnée ci-après, et transmet le résultat de ce calcul à la couche suivante. Ce processus se répète pour toutes les couches afin de donner une valeur de sortie finale. La propagation vers l'avant est déterminée à travers la fonction suivante :

$$x_j^{(n)} = g^{(n)}(h_j^{(n)}) = g^{(n)}(\sum_h w_{jk}^{(n)} x_k^{(n-1)})$$

Avec g : la fonction d'activation, h : la fonction d'agrégation et des poids synaptiques w_{jk} entre le neurone $x_k^{(n-1)}$ et le neurone $x_j^{(n)}$. La propagation arrière sert à calculer l'erreur à partir de la valeur de sortie prédite dans la dernière phase du processus d'apprentissage. La valeur prédite est donc comparée à la valeur de sortie réelle en utilisant la fonction de perte. Ensuite, il faut calculer la dérivée de la valeur d'erreur par rapport à chaque poids du réseau. Cette propagation utilise la règle de chaîne du calcul différentiel : il faut calculer les dérivées et les gradients d'une couche donnée pour calculer les gradients de la précédente. Cette opération sera répétée jusqu'à obtenir des gradients de toutes les fonctions de pondération du réseau neuronal. Par la suite, il faut appliquer des opérations de soustraction entre les valeurs pour réduire à chaque fois le taux d'erreur. La formule de cette propagation se fait à travers la formule de dérivée suivante :

$$e_j^{(n-1)} = g'^{(n-1)}(h_j^{(n-1)}) \sum_i w_{ij}^{(n)} e_i^{(n)}$$

Les valeurs des poids des couches sont apprises par rétro propagation du gradient : il faut calculer progressivement pour chaque couche, en commençant par la fin du réseau les paramètres qui minimisent la fonction de perte régularisée. Cette optimisation se fait généralement avec une descente du gradient stochastique, mais de nombreuses autres approches existent. Le taux d'apprentissage (ou « *learning rate* ») est un paramètre permettant le réglage de l'entraînement des réseaux de neurones profonds. Le contrôle de modèle se fait par rapport à l'erreur estimée quand il y a des mises à jour des poids du modèle, ce qui permet de contrôler la vitesse à laquelle les pondérations sont ajustées [42][44]. Le réseau de neurones peut prendre plusieurs formes en se basant sur les données à traiter et leurs complexités. Nous pouvons distinguer trois types de réseaux parmi d'autres : les réseaux de neurones « *Feed forwarded* » [45], les réseaux de neurones récurrents (RNN), et les réseaux de neurones à résonance [46]. Chacune de ces formes a ses avantages et inconvénients. Ils peuvent être combinés pour optimiser et améliorer les résultats. Le choix de l'architecture s'avère ainsi crucial et il est déterminé principalement par l'objectif de l'application à développer.

II.3.4 Deep learning

Dans cette section, nous définissons tout d'abord la notion du *deep learning* et ensuite nous détaillons son architecture et son fonctionnement.

II.3.4.1 Définition

Le « *deep learning* » (ou l'apprentissage profond), est un sous-domaine de l'intelligence artificielle. Il s'agit d'une technique récente qui a révolutionné l'apprentissage automatique. Il a été

introduit pour la première fois dans les années 1980 [47]. C'est un type spécial des RNA qui ressemble à un système cognitif humain multicouche. Il gagne actuellement beaucoup d'attention grâce à sa capacité d'extraire lui-même les caractéristiques des données à traiter, sans avoir besoin d'une catégorisation préalable. Il se diffère d'un RNA standard par un nombre plus important de couches, et un mode de calcul tensoriel. C'est une technologie clé pour la résolution des problèmes de complexité élevée. Cette technique contient un modèle informatique qui effectue principalement des tâches de classification [51] [52]. Les modèles d'apprentissage profond peuvent atteindre une précision de pointe qui surmonte souvent les capacités humaines. Ils sont formés à travers une énorme quantité de données étiquetées et des architectures de nombreuses couches [53] (voir Figure II.11).

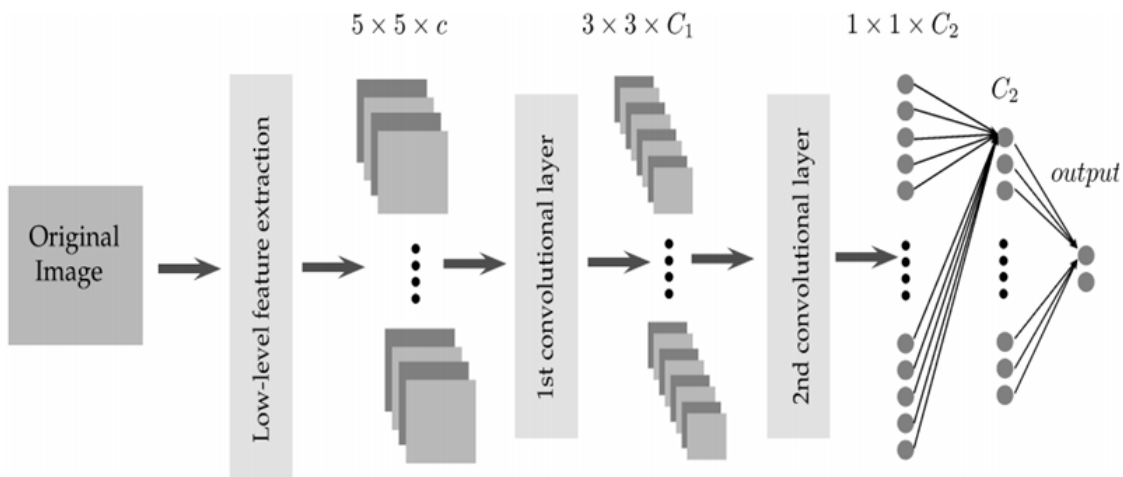


FIGURE II.11 – Visualisation des couches d'une architecture *deep learning* [143]

Le *deep learning* est très gourmand en temps de calcul et en mémoire, et nécessite de grandes masses de données d'apprentissage pour être performant. La nécessité des architectures de calcul parallèle est donc indispensable. Ces architectures permettent l'accélération des calculs complexes ou coûteux en temps d'occupation CPU (« *Central Processing Unit* ») à travers des simulations numériques, le calcul matriciel, *etc.* Parmi les architectures de calcul parallèle les plus adaptées pour le bon fonctionnement d'un algorithme *deep learning*, nous trouvons les GPUs (« *Graphic Processing Unit* ») caractérisant une architecture parallèle performante. Il peut être appliqué sur des clusters de calcul ou bien à l'aide du *Cloud Computing*, ce qui permet de réduire le temps d'apprentissage d'un tel réseau profond.

II.3.4.2 Architecture et fonctionnement

Le *deep learning* fait appel à des familles particulières de réseaux de neurones artificiels et se caractérise par une capacité d'apprentissage autonome. Le terme profond fait référence à la grande quantité de couches cachées dans le réseau. En outre, les modèles de ce type d'apprentissage sont formés en utilisant de grands ensembles de données étiquetées. Les réseaux de neurones convolutifs, les réseaux de neurones récurrents (« *Recurrent Neural Networks* » ou RNNs), les

réseaux antagonistes génératifs (« *Generative Adversarial Networks* » ou GANs), et les réseaux de neurones résiduels (ResNet) sont considérés parmi les architectures les plus répandues de *deep learning* dans plusieurs champs d'application. Dans la section suivante, nous explorons les réseaux de neurones convolutifs (CNN).

II.3.5 Les réseaux de neurones convolutifs (CNN)

Le réseau de neurones convolutif ou à convolution (appelés aussi ConvNet ou CNN pour « *Convolutional Neural Networks* ») fait partie des systèmes neuronaux profonds les plus utilisés surtout pour la vision par ordinateur, dans la reconnaissance la détection des objets dans les images notamment. Il représente une approche puissante et performante qui s'appuie sur l'apprentissage profond pour réaliser des tâches descriptives et génératives (voir Figure II.12). Le nom « convolution » est dérivé d'une opération mathématique de convolution impliquée dans le calcul de propagation avant.

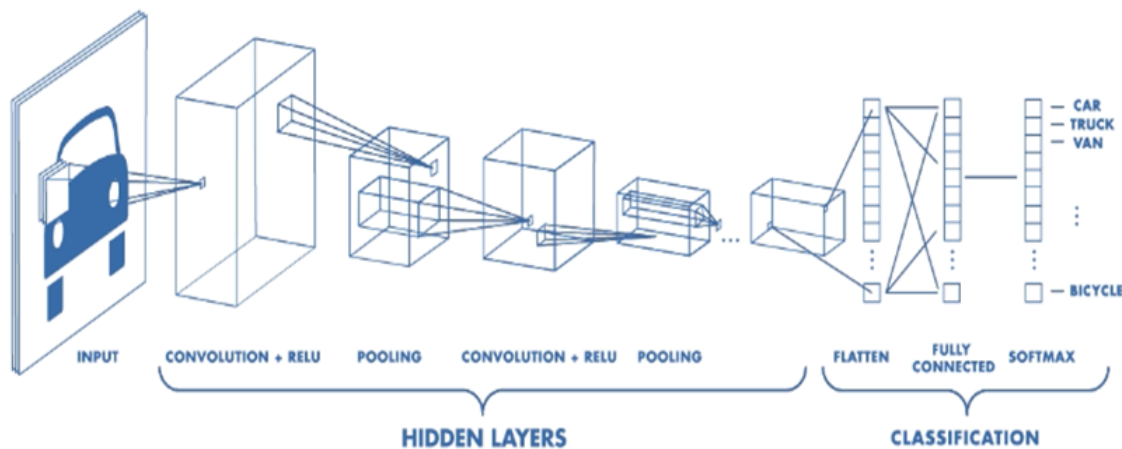


FIGURE II.12 – Architecture d'un réseau CNN [W10]

Les CNNs comportent deux blocs bien distincts. Le premier bloc caractérise la partie convolutive qui fonctionne comme un extracteur de caractéristiques des images. L'extraction automatique de caractéristiques utilise seulement la partie convolutive d'un réseau pré-entraîné. L'entrée est une image fournie sous la forme d'une matrice de pixels. Elle est passée à travers une succession de filtres générant de nouvelles images appelées cartes de convolutions. La sortie fournie du premier bloc convolutif est ensuite connectée en entrée d'un deuxième bloc, constitué de couches entièrement connectées. Le rôle de ce bloc est de fusionner les caractéristiques du code CNN pour faire la classification de l'image. La sortie désigne la dernière couche du réseau comportant un neurone par catégorie. Un CNN applique un ensemble de perceptrons multicouches (PMC), présentant une famille de réseau neuronal artificiel, caractérisés par des combinaisons de séparateurs linéaires permettant de produire un séparateur global non-linéaire [144], conçus pour réduire le temps de calcul. En effet une architecture de CNN est formée par un empilement de couches de traitement (voir Figure II.13) ; la couche de convolution (CONV) permettant le traitement des données d'un champ

récepteur par un filtre de convolution ; la couche de *pooling* (POOL) permettant la compression d'information en réduisant la taille de l'image intermédiaire (souvent par sous-échantillonnage) ; la couche de correction (ReLU), souvent appelée par abus « ReLU » en référence à la fonction d'activation (Unité de rectification linéaire) ; la couche entièrement connectée (« *fully connected* »), qui est une couche de type perceptron et finalement LOSS présentant la couche de calcul de la fonction d'erreur.

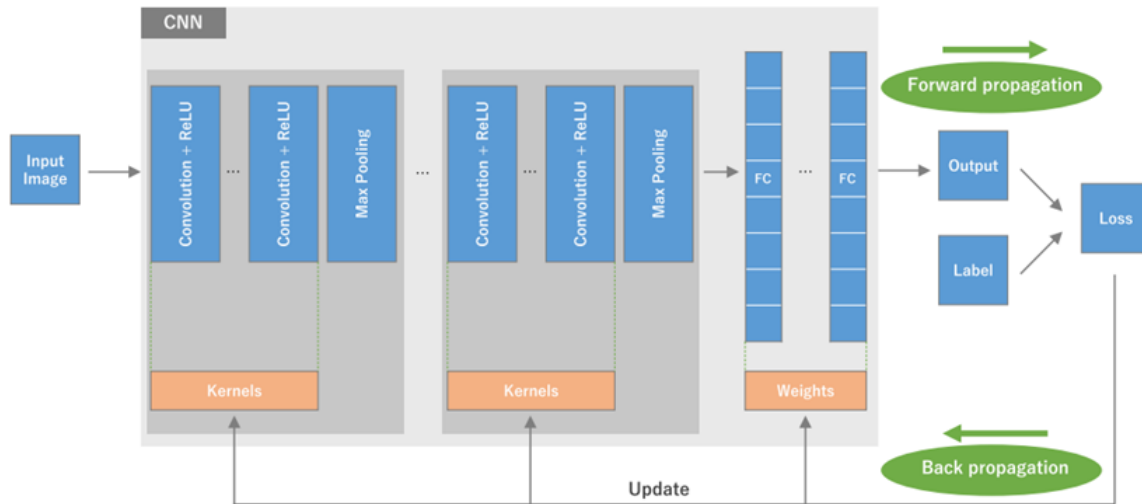


FIGURE II.13 – Fonctionnement d'un CNN [134]

II.4 La programmation parallèle

Les réseaux neuronaux profonds entraînent généralement de grandes quantités de données (souvent référencé comme *big data*). Par conséquent, pour assurer le bon fonctionnement des algorithmes basés sur les techniques d'apprentissage profonds, il faut trouver des outils et des moyens permettant de réduire le temps de calcul et traiter le maximum d'opération. La programmation parallèle est l'une des méthodes efficaces et puissantes pour la réalisation d'un modèle d'apprentissage performant et consistant. Dans la section suivante, nous abordons la notion du calcul parallèle et ses architectures.

II.4.1 Définition

Le calcul parallèle consiste à la mise en œuvre des techniques logicielles et matérielles pour le traitement simultané de séquences d'instructions indépendantes. L'objectif de cette technique est de réaliser un grand nombre d'opérations en un temps le plus court possible sur plusieurs cœurs de calcul. Les techniques de parallélisation sont appliquées afin d'améliorer la performance des applications et les exécuter en un temps raisonnable. En plus, il assure la coopération de plusieurs processeurs pour réaliser le traitement rapide des instructions de l'algorithme. Le calcul parallèle permet d'améliorer les performances des applications et les rendre capables de tourner en un temps

raisonnable et réduit. L'objectif est de gagner en efficacité, performance et en précision, mais aussi de gérer l'augmentation exponentielle des données à traiter. La programmation parallèle est largement utilisée dans différents domaines d'application, tels que la biologie, la climatologie, la physique des particules, l'imagerie médicale, *etc.*

II.4.2 Architectures

Les architectures parallèles sont caractérisées par des modèles d'interconnexions entre les processeurs et la mémoire. Plusieurs approches de classification de ces architectures ont été proposées à savoir la classification de Flynn, de Kuck, de Treleaven, de Hwang and Briggs, de Skillicorn, *etc.* La classification de Flynn est l'approche la plus ancienne et la plus connue. Elle a été proposée par Michael Flynn en 1966 [54]. Elle classe les différentes architectures en 4 catégories en se basant sur deux flux : le flux d'instructions et le flux de données [55] :

- **SISD** (« *Single Instruction Single Data Stream* ») : Dans cette architecture, une seule instruction est exécutée et une seule donnée est traitée à un instant donné. Ce type d'architecture fonctionne sans le principe de parallélisme, ni au niveau des instructions ni au niveau des données. Il caractérise des anciennes machines avec des processeurs séquentiels sans parallélisme.
- **SIMD** (« *Single Instruction Multiple Data Stream* ») : Ici, le même flux d'instructions est appliqué à différents ensembles de données. Dans ce mode, une instruction est appliquée simultanément avec plusieurs données pour générer plusieurs résultats. Cette machine définit un mode de fonctionnement des ordinateurs caractérisés avec des capacités de parallélisme. Elle fait l'exécution d'une seule instruction à tout instant donné, mais elle permet le fonctionnement parallèle de plusieurs données. Dans ce cas, il y a application des processeurs vectoriels ou graphiques avec des mémoires indépendantes.
- **MISD** (« *Multiple Instruction Single Data Stream* ») : Dans cette architecture, plusieurs instructions arithmétiques et logiques peuvent s'exécuter en mode parallèle et par plusieurs processeurs. Ce modèle caractérise les micro-processeurs modernes qui sont les processeurs vectoriels et les architectures en pipelines. Cette structure englobe certaines architectures tolérant les pannes par réplication de calcul.
- **MIMD** (« *Multiple Instruction Multiple Data Stream* ») Dans ce modèle, les processeurs traitent plusieurs instructions indépendantes simultanément. Dans cette structure, chaque processeur fonctionne d'une façon autonome et asynchrone et il dispose de sa propre unité de contrôle. Les techniques de synchronisation dépendent souvent de l'organisation de la mémoire qui peut être partagée ou distribué. Cette structure caractérise généralement les architectures distribuées tel que le multiprocesseur symétrique (SMP). Il consiste à multiplier les processeurs similaires de manière à accroître la puissance de calcul, tout en conservant une seule mémoire. C'est la forme parallèle est la plus répandue. Pour cette architecture, nous distinguons 3 classes en fonction de la gestion de la mémoire :

1. **Machines à mémoire partagée SM (« *Shared Memory* »)** : cette machine permet de réaliser le parallélisme de données et de contrôle. En effet, elle peut être accédée par plusieurs processus d'une manière simultanée. Nous distinguons 3 types de mémoire partagée : UMA (« *Uniform Memory Access* »), NUMA (« *Non Uniform Memory Access* ») et COMA (« *Cache Only Memory Access* »).
2. **Machine à mémoire distribuée DM (« *Distributed Memory* »)** : Dans ce cas, une mémoire propre locale à chaque processeur ainsi l'interconnexion de systèmes indépendants. Chaque processeur est indépendant donc il possède sa propre mémoire pour exécuter son propre flux d'instructions. La notion de parallélisme s'effectue par échange de messages.
3. **Machine à mémoire virtuelle partagée VSM (« *Virtual Shared Memory* »)** : cette machine contient une mémoire physiquement distribuée. Ici, chaque processeur peut accéder aux mémoires spécifiques des autres processeurs.

II.4.3 Grille de calcul

L'architecture de grille de calcul s'est développée au début des années 2000. Elle permet de traiter des données stockées à distance avec une performance comparable à un traitement local. Cette approche distribuée permet donc de relier et partager des ressources hétérogènes de calcul et/ou de stockage de plusieurs centres de calcul distants. L'objectif est de passer d'une puissance de calcul de quelques centaines de processeurs pour un cluster à une capacité de calcul et de stockage quasi illimitée dans le cadre d'une grille de calcul. A part sa grande capacité en ressources informatiques, un autre aspect important de la grille de calcul est son interopérabilité. En effet, une grille doit être capable de fonctionner avec différents produits ou systèmes informatiques, sans restriction d'accès ou de mise en œuvre. La mise en place d'une grille de calcul nécessite l'intégration d'un middleware (intergiciel) qui s'interface entre les ressources et les applications pour faire collaborer tous les éléments de la grille. Le middleware assure la gestion des jobs (soumission, ordonnancement), des données (transfert, stockage, réplication), de la sécurité et des utilisateurs. Plusieurs *middlewares* existent : « *Globus Toolkit* », « *Glite* », « *EMI Grid Middleware* », « *Advanced Resource Connector* » ARC, « *UNICORE* », « *iRODS* », etc.

II.5 Conclusion

Dans ce chapitre, nous avons présenté les piliers techniques de cette thèse, qui sont issus de deux domaines de l'intelligence artificielle : les ontologies et l'apprentissage automatique. Ils caractérisent des outils fréquemment appliqués à l'imagerie médicale. En premier lieu, nous avons exposé une vue générale sur les ontologies en détaillant ses caractéristiques fondamentales, les langages de modélisations, les méthodes développement, le cycle de vie et nous avons clôturé

cette première partie par citer quelques domaines d'application. En deuxième lieu, nous avons exploré les notions de base liées aux techniques de l'apprentissage automatique à savoir les réseaux de neurones et le *deep learning* en donnant leurs architectures et leurs modes de fonctionnement ainsi que leurs impacts sur les algorithmes de classification. Nous avons exposé ensuite une vue sur la programmation parallèle. Dans le chapitre suivant, nous exposons un état de l'art exhaustif sur l'emploi des ontologies et des techniques d'apprentissage automatique pour la détection et la classification des tumeurs hépatiques en mettant l'accent sur les lésions de type carcinome hépatocellulaire (CHC). Nous allons voir également l'impact d'appliquer les systèmes de stadification de tumeurs pour extraire le stade des lésions détectées dans une image IRM.

ETAT DE L'ART SUR LES ONTOLOGIES DÉDIÉES À LA REPRÉSENTATION DES LÉSIONS HÉPATIQUES

Sommaire

III.1 Introduction	36
III.2 Définition du CHC	36
III.2.1 Imagerie Médicale des tumeurs CHC	37
III.2.2 Détection du CHC <i>via</i> l'imagerie médicale	38
III.2.3 Les systèmes de stadification des tumeurs hépatiques	41
III.3 Ontologies pour la représentation des maladies hépatiques	46
III.3.1 Interprétation des images médicales <i>via</i> les ontologies	47
III.3.2 Modélisation des données médicales <i>via</i> les ontologies	55
III.3.3 Modélisations des rapports médicaux	63
III.4 Classification des données médicales <i>via</i> les techniques d'apprentissage au- tomatique et des ontologies	70
III.4.1 Application des techniques d'apprentissage automatique	70
III.4.2 Application des ontologies avec les techniques d'apprentissage automatique	70
III.5 Discussion	72

III.1 Introduction

Les ontologies représentent un mécanisme de représentation de données largement appliqué pour la mise en place d'un formalisme commun et des ressources terminologiques. Ces ressources sont créées pour répondre à des besoins particuliers tout en permettant l'interopérabilité des données et en facilitant l'ajout de nouvelles connaissances. Les ontologies sont utiles également pour faire la sauvegarde des connaissances du domaine sous la forme d'un vocabulaire partageable. Elles offrent la possibilité de garder la sémantique du domaine et de fournir des outils d'inférence. Dans le domaine médical, l'utilisation des ontologies et leur puissance de raisonnement sémantique pour le diagnostic des maladies hépatiques, en particulier les tumeurs CHC fait l'intérêt de plusieurs travaux de recherche.

Dans ce chapitre, nous présentons dans un premier temps le carcinome hépatocellulaire, et les systèmes de classification standards des tumeurs CHC, qu'on appelle aussi « stadification tumorale ». Dans un deuxième temps, nous exposons une synthèse, sous forme d'un état de l'art, sur la classification de données médicales pour le diagnostic de tumeurs du foie, et l'utilisation des ontologies pour l'optimisation du diagnostic et l'aide à la décision clinique. Nous nous sommes essentiellement focalisés sur trois thèmes : i) l'utilisation des ontologies pour la formalisation des rapports radiologiques, ii) pour la représentation des données médicales et iii) pour l'interprétation des images médicales. Nous abordons par la suite comment coupler les ontologies avec les techniques de l'apprentissage automatique pour améliorer le processus de classification des tumeurs hépatiques.

III.2 Définition du CHC

Le cancer du foie est l'une des causes majeures de décès dans le monde, il se développe à partir des cellules de cet organe nommées « hépatocytes », qui assurent le bon fonctionnement du foie. Ce dernier possède un rôle important qui se caractérise par la synthèse de la plupart des protéines du sang, le stockage et la répartition des nutriments issus de la digestion et la dégradation des substances toxiques. Il aide également à contrôler le métabolisme et aide à combattre les cellules et les substances nocives qui menacent l'organisme à travers le phénomène appelé « phagocytose ». La Figure III.1 montre une visualisation du foie et de ses vaisseaux.

Le CHC est un cancer primitif qui survient dans environ 80% des cas sur un foie endommagé par une maladie chronique : la cirrhose [56] [57]. Les virus de l'hépatite B et de l'hépatite C, l'obésité et le diabète, les antécédents médicaux, et les maladies héréditaires peuvent conduire à la cirrhose, qui engendre par la suite à l'apparition du CHC. Selon l'organisation mondiale de la santé ³¹(OMS), le CHC est le cancer primitif du foie le plus grave. C'est le sixième cancer le plus fréquent au monde [58]. Il conduit à plus de 700 000 décès chaque année selon la Société américaine du cancer ³². Les statistiques de l'institut national du cancer ³³ estiment que 42 810 nouveaux cas (30 170 hommes et 12 640 femmes) ont été diagnostiqués en 2020 dans le monde.

31. <https://www.who.int/fr>

32. <https://www.cancer.org/>

33. <https://www.e-cancer.fr/>

Elles estiment également un nombre de décès d'environ 30 160 personnes (20 020 hommes et 10 140 femmes) au États-Unis en 2020. Une étude utilisant des données du programme de surveillance SEER (« *Surveillance, Epidemiology and End Results* ») estime qu'il y aurait plus de 38000 cas de CHC en 2020 et plus de 56000 cas de CHC en 2030 aux États-Unis [59]. En France, le CHC est responsable d'environ 8000 décès par an [61] avec une incidence liée essentiellement à l'alcool [60]. Les symptômes les plus fréquents du CHC sont : la distension de l'abdomen, la perturbation des fonctions digestives, la jaunisse, une fatigue excessive, *etc.*

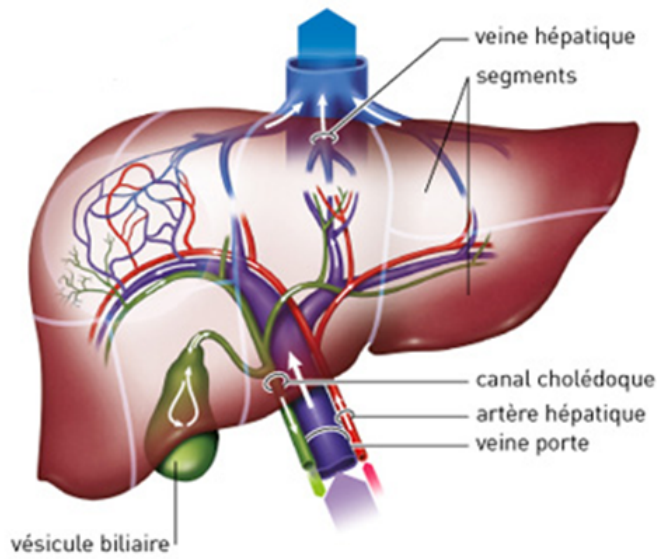


FIGURE III.1 – Visualisation du foie et ses vaisseaux [148]

Dans la section suivante, nous présentons les outils et les modalités radiologiques permettant le diagnostic des tumeurs hépatiques de type CHC.

III.2.1 Imagerie Médicale des tumeurs CHC

L'imagerie est un outil qui a révolutionné le domaine médical qui s'avère importante dans le dépistage de plusieurs maladies et la définition de nouvelles techniques de traitement. Elle est utilisée par les radiologues pour comprendre le fonctionnement de l'organisme humain et vérifier l'existence ou non d'une maladie. Elle visualise la physiologie et l'anatomie du corps humain en appliquant plusieurs phénomènes physiques comme la résonance magnétique nucléaire, les rayons X, ou encore les ondes ultrasons. Elle s'intègre dans le choix du traitement à appliquer et intervient pour interpréter l'évolution de la tumeur. L'imagerie médicale a connu des évolutions impressionnantes ces dernières années avec diverses modalités :

- **Échographie** : c'est une technique qui repose sur l'utilisation de l'ultrason et permet la visualisation des organes internes. Elle est employée pour faire le diagnostic de diverses pathologies et pour guider les autres examens médicaux tels que la biopsie. Elle détecte

souvent les tumeurs de taille supérieure à 2cm et précise sa nature.

- **Scanner** : nommé aussi Tomodensitométrie (TDM) est un type d'examen qui génère des images en coupe. Son déroulement nécessite l'injection d'un produit de contraste. Il donne des résultats plus précis et détecte les anomalies difficiles à visualiser dans la radiographie standard ou dans l'échographie. C'est la méthode la plus performante pour l'évaluation et l'interprétation des tumeurs, des infections, des kystes et des hémorragies.
- **Imagerie par résonance magnétique (IRM)** : cet examen médical permet de produire une imagerie par coupes du corps humain en utilisant des champs magnétiques et des ondes radios. Il s'effectue après l'injection du produit de contraste. Avec sa grande sensibilité, il donne des informations sur les lésions invisibles par les autres modalités, notamment dans les tissus mous. Il s'agit donc d'une méthode précise pour la détection des tumeurs du foie.

III.2.2 Détection du CHC *via* l'imagerie médicale

L'imagerie médicale regroupe un ensemble d'images capturées suite à une radiographie, un scanner, une IRM, ou une échographie appliquées pour visualiser l'existence des corps étrangers dans l'organisme. Elle est désormais bien ancrée dans la pratique clinique. Afin d'améliorer encore le diagnostic de certaines pathologies *via* l'imagerie médicale, l'injection d'un produit de contraste s'avère nécessaire [62]. Il s'agit d'une substance qui permet d'observer la structure anatomique ou pathologique d'un organe (par exemple, une masse cancéreuse). Le diagnostic du CHC se fait fréquemment par application des modalités radiologiques. Cette lésion hépatique est découverte lors de la surveillance d'une maladie chronique du foie (voir Figure III.2)³⁴.

34. <https://www.snfge.org/>

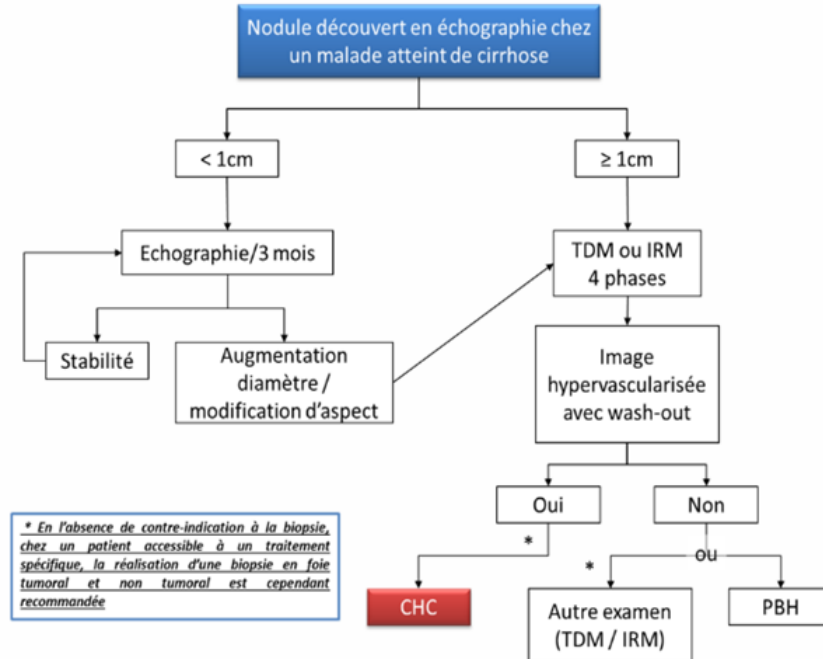


FIGURE III.2 – Diagnostic de CHC *via* les modalités radiologiques [63]

La particularité des nodules CHC par rapport aux autres tumeurs hépatiques se caractérise par l'hyper-vascularité. En effet, elles se remplissent et se vident d'une manière rapide de sang artériel. Le dépistage du CHC s'effectue en premier lieu par l'échographie, lors d'un suivi pour une cirrhose notamment. Un scanner abdominal est appliqué par la suite pour confirmer les résultats observés dans l'échographie. En cas de doute des résultats de scanner, l'IRM intervient comme solution pour mieux caractériser les nodules hépatiques. Elle devient une modalité de plus en plus utilisée en première intention pour le diagnostic et surtout pour le suivi du patient. L'IRM offre des informations physiologiques supplémentaires caractérisant surtout le tissu hépatique et les nodules hépatiques détectés. Elle génère différentes séquences permettant d'observer les tissus avec des contrastes plus élevés qu'avec la tomodensitométrie. Elle est appliquée également dans l'exploration des pathologies cérébrales, osseuses, digestives, articulaires, cardiaques, *etc.* La durée d'un examen IRM dure en moyenne entre 15 et 35 minutes incluant la préparation du patient, le post-traitement et l'interprétation des images capturées [64]. Au cas où les nodules sont de très petite taille, il est assez compliqué d'identifier les nodules CHC. Dans ce cas, il faut planifier un examen tous les 3 mois pour suivre l'évolution des tumeurs. Si la taille de ce nodule augmente, il faut faire une biopsie pour pouvoir décider la démarche thérapeutique à suivre. Pour une meilleure détection, on se base sur le taux d'AFP (*alpha feto-protéine*), nommé aussi « fétuine ». C'est un très bon indicateur de diagnostic de CHC. Il s'agit d'une protéine sanguine qui peut être trouvée dans le sang des patients atteints de cancer du foie ou par certains autres cancers. Le CHC est caractérisé généralement par un taux d'AFP supérieur à 400 ng/ml.

Le diagnostic du CHC chez un patient à travers la TDM ou l'IRM se fait par l'injection d'un produit de contraste. En se basant sur les critères de Barcelone BCLC (« *Barcelona Clinic Liver Cancer* ») (voir Figure III.3; une description détaillée de ce système sera expliquée dans la section

suivante), la taille du nodule et son profil vasculaire sont de bons indicateurs de caractérisation du CHC. Dans le cas général, le profil vasculaire peut être typique avec un wash-in (phase artérielle), et un wash-out (lavage portal et/ou tardif. Pour le cas de CHC, le profil vasculaire est caractérisé par le wash-in avec hyper-intensité et le wash-out avec hypo-intensité. Comme les tumeurs CHC sont hypervascularisées, elles réagissent davantage au produit de contraste, comparativement aux autres tissus du foie. Tout autre profil vasculaire est considéré comme atypique. De plus, chez un patient cirrhotique, la présence en TDM ou en IRM, d'une prise de contraste artérielle associée à un wash-out au sein d'un nodule de 2 cm suffit pour affirmer le diagnostic du CHC. Cependant, une prise de contraste artérielle associée à un wash-out, au sein d'une masse de taille comprise entre 1 cm et 2 cm, ne permet un diagnostic formel que si elle est constatée sur deux techniques d'imagerie distinctes (échographie de contraste, TDM ou IRM). En cas de « profil vasculaire atypique » ou « profil vasculaire typique » mais sur une seule imagerie, une biopsie devrait être réalisée.

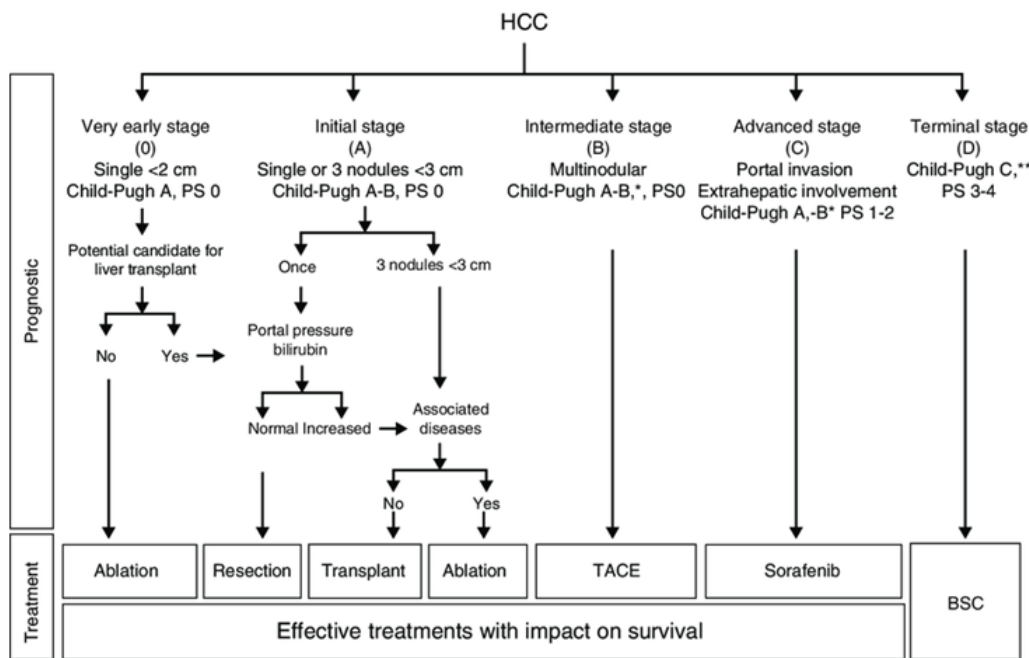


FIGURE III.3 – Critères de Barcelone pour la détection du CHC [65]

La nature des signaux IRM est également un paramètre important pour la détection du CHC. En effet, la séquence IRM d'un tissu spécifique dépend de plusieurs paramètres physio-chimiques. Parmi ces caractéristiques, nous distinguons les deux signaux modélisés par T1 et T2. Une séquence IRM est caractérisée par un ensemble d'impulsions excitatrices dont les paramètres à savoir TE (temps d'écho), TR (temps de répétition) sont ajustés pour avoir des images ayant un contraste donné T1 ou T2 par exemple. Dans une image pondérée en T1, la graisse apparaît hyperintense (couleur claire) et l'eau hypointense, tandis que dans une image pondérée en T2, l'eau apparaît hyperintense (couleur claire) et la graisse un peu plus sombre que l'eau. En cas de découverte du CHC, ces deux signaux sont caractérisés par un caractère « hyperintense » pendant le wash-out.

III.2.3 Les systèmes de stadification des tumeurs hépatiques

Les systèmes de stadification servent à savoir le stade de la tumeur, c'est-à-dire s'il est précoce ou avancé. Ils décrivent sa classe en fonction de la propagation des nodules cancéreux et de leurs emplacements lors des phases de diagnostic initial et de suivi thérapeutique [116]. Ils sont utiles pour comprendre le degré de gravité du cancer, connaître les chances de survie, planifier le meilleur traitement à suivre et même identifier les essais cliniques compatibles avec les patients infectés. L'identification du stade repose sur la taille de la tumeur observée dans l'imagerie médicale, les parties de l'organe atteintes par cette tumeur, la propagation du cancer et l'emplacement où elle s'est propagée (métastases). Il est important de faire plusieurs types d'examen tels que les analyses sanguines, les acquisitions d'imagerie ou encore la biopsie. La chirurgie est aussi un moyen pour observer le tissu atteint par la tumeur, en pratiquant une analyse anatomopathologique des tissus retirés.

Dans ce travail, nous nous intéressons à étudier trois systèmes de stadification standards et très connus dans les routines cliniques : BCLC (« *Barcelona Clinic Liver Cancer* »), TNM (« *Tumor, Nodes, and Metastasis* ») et Li-RADS (« *Liver Imaging Reporting and Data System* »).

III.2.3.1 BCLC (*Barcelona Clinic Liver Cancer*)

BCLC est un système de stadification standard, utilisé par les radiologues pour identifier le stade de la tumeur hépatique [66] [67]. Il prend en compte plusieurs paramètres :

- le nombre et la (les) taille(s) de(s) nodule(s) détecté(s) dans l'imagerie médicale ;
- l'indice fonctionnel du foie : il permet d'évaluer à quel point cet organe est capable de fonctionner normalement. Il correspond à une valeur numérique déterminée à l'aide de l'échelle de l'*Eastern Cooperative Oncology Group*³⁵ (ECOG). Cette valeur est comprise entre 0 et 4 et varie selon l'avancement de la tumeur dans l'organisme [68]. Plus elle est élevée, plus le foie a des difficultés à effectuer ses tâches ;
- le taux de Bilirubine : il provient de la dégradation de destruction de l'hémoglobine dans le sang. Il possède habituellement une valeur de moins de 12 mg/l. Une analyse sanguine de la bilirubine est nécessaire au cas où il y a des troubles hépatobiliaires ;
- la pression de la veine portale : caractérise l'évolution de la pression dans la veine porte et ses branches, et notamment l'hypertension portale (HTP) ;
- le score Child-Pugh : il définit un grade (A, B ou C) à calculer en se basant sur cinq mesures différentes : encéphalopathie, ascite, bilirubine totale, albumine, et taux de prothrombine. Lorsque le score est entre 5-6 points, c'est le grade A, lorsque le score est entre 7-9 points, c'est le grade B et lorsque le score est entre 10-15 points, c'est le grade C.

35. <https://ecog-acrin.org/resources/ecog-performance-status>

TABLE III.1 – Calcul du score Child-Pugh

Points par critère	1	2	3
Encéphalopathie	0	confusion	Coma
Ascite	0	discrète	abondante
Albumine	>35	28-35	<28
Bilirubine	<20	20-30	>30
Prothrombine	>50	40-50	<40

Le Tableau [III.1](#) récapitule les détails de calcul de ce score.

Le système BCLC classe la tumeur en cinq stades : très précoce, précoce, intermédiaire, avancé et terminal (voir Tableau [III.2](#)) :

- **Très précoce** : noté A0, un seul nodule de taille inférieure à 2 cm est détecté, l'indice fonctionnel est souvent de l'ordre de 0. La fonction du foie est caractérisée par une pression de la veine portale et un taux de bilirubine normaux. A ce niveau, la tumeur peut être retirée par une chirurgie. L'ablation partielle du foie nécessite parfois une laparotomie où le chirurgien ouvre l'abdomen pour accéder à cet organe.
- **Précoce** : peut être l'un de ces cas : A1, A2, A3, ou A4 en se basant sur la gravité de la maladie :
 1. Le stade A1 est caractérisé par la présence d'un seul nodule de taille inférieure à 5 cm, l'indice fonctionnel est égal à 0. La fonction du foie est exprimée par une pression veine portale normale. Dans ce cas, la chirurgie ou l'ablation par ARF (radiofréquence) peuvent être appliquées comme type de traitement.
 2. Le stade A2 est caractérisé par la présence d'une masse hépatique de taille inférieure à 5 cm. L'indice fonctionnel est égal à 0. La fonction du foie est exprimée par une haute pression de la veine portale et un taux normal de bilirubine. Le type du traitement à suivre peut être une chirurgie, une ARF ou une greffe du foie.
 3. Le stade A3 a les mêmes caractéristiques que le stade A2. La différence est un taux élevé de bilirubine. L'ARF ou la TACE (« *Transarterial Chemoembolization* » ou chimoembolisation transartérielle) sont les traitements qui peuvent être appliqués dans ce stade.
 4. Le stade A4 est caractérisé par la présence d'au moins trois masses hépatiques dans le foie. Chacune est de taille inférieure à 3 cm. L'indice fonctionnel est 0. La pression portale et le taux de bilirubine ne sont plus applicables dans ce stade. Le type du traitement à suivre peut être soit la greffe de foie soit la TACE.

TABLE III.2 – Stadification du CHC selon le système BCLC

Stades	OMS		Morphologie tumorale	Fonction hépatique
Stade très précoce	0	0	Unique, <2 cm	Pas d’HTP, bilirubinémie N
Stades précoces	A1	0	Unique, <5 cm	Pas d’HTP, bilirubinémie N
	A2	0	Unique, <5 cm	HTP, bilirubinémie N
	A3	0	Unique, <5 cm	HTP, hyper bilirubinémie
	A4	0	3 lésions, <3 cm	Child-Pugh A-B
Stade intermédiaire	B	0	Multinodulaire	Child-Pugh A-B
Stade évolué	C	1-2	Invasion vasculaire, métastases	Child-Pugh A-B
Stade terminal	D	3-4	Indifférente	Child-Pugh C

- **Intermédiaire** : noté par B, il est caractérisé par une grosse tumeur multifocale. Ce dernier prend naissance dans différentes régions du foie simultanément. Il est défini par la présence d’au moins deux lésions séparées par un tissu sain. L’indice fonctionnel est égal à 0. Dans ce niveau, le fonctionnement du foie est mesuré par le score Child-Pugh qui pourrait être A ou B. La TACE est appliquée comme traitement dans ce cas.
- **Avancé** : noté C, un seul nodule envahit les vaisseaux sanguins ou bien le cancer s’est propagé vers d’autres organes du corps humain. L’indice fonctionnel peut être 1 ou 2 et le score Child-Pugh est en stade A ou B. À ce niveau, la chimiothérapie ciblée est le seul traitement préconisé.
- **Terminal** : noté D, ce stade est caractérisé par la présence de plusieurs nodules de différentes tailles. L’indice fonctionnel est alors égal à 3 ou 4. Le score Child-Pugh est C. Seuls des soins palliatifs sont proposés au patient infecté.

III.2.3.2 TNM (*Tumor, Nodes, and Metastasis*)

La classification TNM est appliquée fréquemment dans plusieurs types de maladies. Ce système est basé sur les paramètres T, N et M. Le Tableau III.3 montre les valeurs possibles de ces

TABLE III.3 – Les valeurs possibles des paramètres du système TNM

Valeur T, N, M	Signification
TX	L'évaluation de la taille n'est plus possible
T0	Pas de signe sur la tumeur hépatique
T1	Cas d'une seule tumeur qui n'a pas atteint les vaisseaux sanguins
T2	Une tumeur unique ou multiple de taille <5 cm qui atteint les vaisseaux sanguins
T3a	Plusieurs tumeurs dont chacune est de taille 5 cm
T3b	La tumeur envahit les veines portales ou hépatiques
T4	La tumeur se propage vers les autres parties du corps
NX	L'évaluation des métastases n'est plus possible
N0	Cas d'absence des métastases
N1	Cas de présence des métastases
M0	Cas d'absence des métastases
M1	Cas de présence des métastases

paramètres :

- **(T)** : exprime la taille de la tumeur primitive. Il possède l'une des valeurs suivantes : TX, T0, T1, T2, T3a, T3b ou T4.
- **(N)** : présence ou absence de métastases dans les ganglions lymphatiques, elle peut prendre l'une des valeurs : NX, N0 ou N1.
- **(M)** : présence ou absence de métastases à distance, elle peut prendre l'une de ces valeurs : M0 ou M1. Ce système dépend donc de la taille de la tumeur, dans les ganglions lymphatiques, sa propagation dans les vaisseaux sanguins, ainsi que sa propagation dans un organe autre que le foie.

Le système TNM permet la classification du CHC en sept stades :

- Le stade I : c'est un stade primaire, caractérisé par la présence d'une seule tumeur qui a atteint uniquement les vaisseaux sanguins.
- Le stade II : caractérisé par l'observation d'une ou plusieurs tumeurs mesurant chacune une taille inférieure ou égale à 5 cm. De plus, aucune de ces tumeurs n'a été propagée vers les ganglions lymphatiques ou vers une autre partie du corps.
- Le stade IIIA : pour ce stade, nous observons une ou plusieurs tumeurs dont au moins une possède une taille supérieure à 5 cm. A ce stade, il n'y a pas de propagation vers d'autres organes du corps.
- Le stade IIIB : ici, le cancer a envahi les ganglions lymphatiques mais elle n'est pas propagée vers un autre organe du corps humain.

TABLE III.4 – Stadification du CHC selon le système TNM

Stade	Valeur T	Valeur N	Valeur M
I	T1	N0	M0
II	T2	N0	M0
IIIA	T3a	N0	M0
IIIB	T3b	N0	M0
IIIC	T4	N0	M0
VIA	Tout T	N1	M0
VIC	Tout T	Tout N	M1

- Le stade IIIC : une tumeur s’est propagée vers d’autres organes voisins du foie autres que la vésicule biliaire. De plus, elle a envahi les ganglions lymphatiques mais elle n’est pas propagée vers une autre partie plus éloignée de l’organisme humain.
- Le stade VIA : une tumeur a réussi à envahir les ganglions lymphatiques mais elle n’est pas propagée vers d’autres organes.
- Le stade VIC : le cancer s’est développée jusqu’à atteindre les ganglions lymphatiques et autres parties de l’organisme humain.

Le Tableau III.4 expose les valeurs fondamentales des paramètres T, N et M, que peut prendre chaque stade de la tumeur, selon ce système de stadification.

III.2.3.3 Li-RADS (*Liver Imaging Reporting and Data System*)

Li-RADS est un système normalisé basé sur des terminologies standardisées de diagnostic. Il a été approuvé par l’ACR (« *American College of Radiology* »). La première version a été lancée en 2011. Deux autres versions de ce système ont été proposées respectivement en 2014 et 2017. LI-RADS est appliqué pour représenter et interpréter les tests d’imagerie hépatique reliés à des patients cirrhotiques ou des patients atteints de CHC. Il repose sur plusieurs critères permettant d’extraire des informations sur la tumeur détectée (par exemple le diamètre, l’état de la phase artérielle, *etc.*). Ce système est utile pour réduire ou éliminer les erreurs d’interprétation [135]. La dernière mise-à-jour de Li-RADS a définit sept catégories de lésions hépatiques : LR-NC, LR-1, LR-2, LR-3, LR-M, LR-TR, et LR-TIV.

Li-RADS est utilisé également pour guider les radiologues et les médecins à travers des observations radiologiques standardisées. Il se base sur le type des tumeurs CHC détectées pour offrir une base de connaissances utilisée à des fins cliniques. De plus, il interprète plusieurs types d’imageries médicales telles que l’IRM et la TDM. Il peut être inclus dans les algorithmes destinés au diagnostic des tumeurs hépatiques [69]. La Figure III.4 présente les catégories du système Li-RADS³⁶. Les critères de ce système prennent en compte plusieurs paramètres : le rehaussement

36. <https://radiologyassistant.nl/abdomen/liver/li-rads>

à la phase artérielle, la taille tumorale, le lavage et la croissance sur des examens successifs. Ces critères sont utiles principalement pour le diagnostic des nodules de petites tailles.

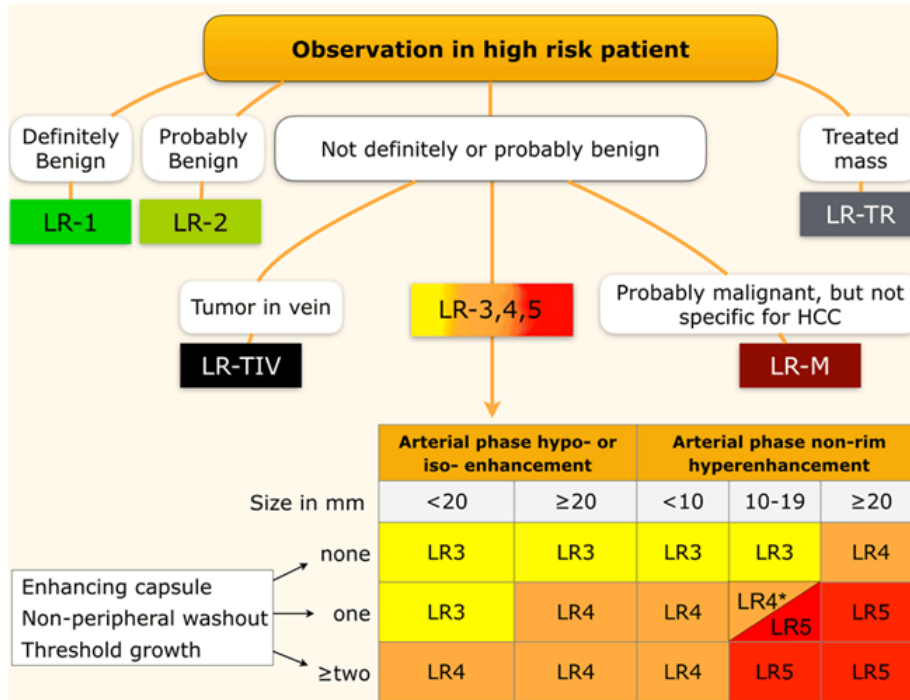


FIGURE III.4 – Visualisation des catégories du système Li-RADS [149]

Le terme *observation* est utilisé pour savoir si la masse détectée est une lésion ou pas. Les catégories LI-RADS reflètent la probabilité que la lésion hépatique soit un CHC en commençant par la catégorie LR-1, qui caractérise une tumeur certainement bénigne jusqu'à arriver à LR-5 qui caractérise une tumeur certainement CHC. Les catégories supplémentaires sont : LR-TIV qui présente une tumeur dans la veine, LR-M représente une tumeur maligne mais pas un CHC, LR-TR caractérise un CHC traité, LR-NC présente une tumeur où la qualité de l'imagerie ne permet pas d'évaluer les principales caractéristiques [70].

III.3 Ontologies pour la représentation des maladies hépatiques

Les ontologies médicales servent généralement à faire l'analyse qualitative des décisions cliniques au sein des centres de soins. En effet, l'utilité d'une ontologie médicale s'appuie sur deux phases complémentaires qui sont : i) la modélisation ou la construction des concepts fondamentaux permettant la prise en charge d'une maladie ou le suivi hospitalier (diagnostic des maladies, dossiers médicaux, etc.) et ii) l'analyse terminologique des corpus qui permet d'enrichir et de structurer ce noyau ontologique. Dans la section suivante, nous présentons une vue sur les ontologies dédiées à la classification des lésions hépatiques. De nombreuses approches ont été proposées pour la représentation des maladies hépatiques en exploitant les ontologies. Nous nous sommes

focalisés sur les travaux liés à l'utilisation des ontologies dans le traitement des images médicales, des données cliniques et des rapports radiologiques.

III.3.1 Interprétation des images médicales *via* les ontologies

L'intégration des techniques de l'ingénierie de connaissance dans le domaine de l'imagerie médicale, et plus particulièrement pour le traitement et l'interprétation des images médicales, a un impact appréciable, du fait que les ontologies sont intégrées comme outil d'aide à la décision clinique. Par conséquent, plusieurs travaux de recherche ont discuté l'utilisation des ontologies pour l'interprétation et la représentation sémantique des images médicales. Ils arrivent même à proposer des solutions techniques et informatiques pour améliorer à chaque fois cette interprétation clinique. Dans cette section, nous présentons certains de ces travaux en terminant à chaque fois par une étude comparative basée sur différents critères tels que les outils de développement, les objectifs, les prototypes ontologiques générées, la classification des tumeurs hépatiques, *etc.*

III.3.1.1 Les ontologies pour l'analyse des images TDM

Dans la littérature, plusieurs travaux ont été proposés pour le traitement des images scanner (TDM) en se basant sur des approches ontologiques. Dans ce contexte, les auteurs de [71] avaient comme objectif d'améliorer les méthodes CBIR (« *Content-Based Image Retrieval* ») permettant l'annotation automatique des images médicales TDM du foie. La méthode proposée est dédiée non seulement à faire une annotation simple des images au niveau des observations médicales mais elle interprète également ces images d'une manière sémantique. Cette méthode décrit les caractéristiques de l'image, des informations sur le foie (l'intensité, la taille, *etc.*), des informations sur la tumeur (l'intensité moyenne, la texture, la forme, la densité, la localisation et la taille) et des informations sur les vaisseaux (la taille). Les auteurs proposent une approche basée sur la technique WNN (« *Weighted Nearest-Neighbours* ») comme stratégie de récupération de données suite à des requêtes de recherche. Pour faire l'évaluation de cette approche, les auteurs ont utilisé des données TDM (3D) volumétriques extraites de « *ImageCLEF 2014 Liver Annotation Challenge* » [73] sous la forme de 73 annotations traitées *via* l'ontologie ONLIRA [72].

De même, [74] proposent un outil informatique permettant l'annotation sémantique des images médicales. Ce travail présente une ontologie médicale nommée MEDICO qui rassemble plusieurs données cliniques. Ces données sont extraites à partir de plusieurs ontologies telles que FMA (« *Foundational Model of Anatomy* »), et des terminologies telles ICD-10³⁷ (« *the International Classification of Diseases* ») et RadLex³⁸ (« *Radiology Lexicon* »). Cette approche vise aussi à développer un outil de métadonnées piloté par l'ontologie et utilisé pour simplifier l'annotation manuelle. Il fournit des visualisations intuitives formant les concepts. Les métadonnées extraites sont associées à des annotations anatomiques et à des résultats cliniques permettant de générer une vue unifiée des antécédents médicaux du patient. L'ontologie proposée permet de lier les patients, les images médicales et les annotations ensemble dans une liste complète de résultats. Les annotations médicales sont diffusées vers des sources externes comme Wikipédia.

37. <https://www.cdc.gov/nchs/icd/icd10cm.htm>

38. <http://radlex.org/>

Dans le même contexte, [75] présente un *framework* de modélisation des lésions hépatiques à partir d'images scanner. Ce système vise à prédire l'existence d'une tumeur hépatique dans le foie. Il se base sur les outils *Riesz wavelets* [77] et SVM (« *Support Vector Machines* ») [76] qui vont fournir par la suite des modèles VST (« *Visual Semantic Terms* ») [78]. La distance entre ces modèles est calculée pour faire l'annotation des termes synonymes et complémentaires tout en utilisant le vocabulaire RadLEX [79]. La première étape de la méthodologie appliquée consiste à faire l'extraction de chaque terme qui indique l'existence des nodules hépatiques dans l'image médicale. Ensuite, la distance entre tous les modèles VST est calculée pour construire une ontologie unissant tous les termes et les synonymes. L'évaluation de ce travail a été réalisée en se basant sur 74 cas de lésions hépatiques annotées avec 18 modèles VST.

Dans [80], les auteurs proposent une méthode basée sur l'annotation sémantique des images médicales. Cette méthode est utilisée pour l'aide à la décision clinique et l'élimination de la fausse interprétation au sein d'un processus de diagnostic médical. L'approche proposée vise à améliorer la qualité de la recherche d'images en intégrant l'aspect sémantique dans cette recherche. Elle introduit les aspects et les caractéristiques pathologiques et anatomiques communes entre les patients infectés par le lymphome (cancer qui attaque le système lymphatique), pour faciliter le traitement des cas similaires. Pour tester cette approche, 100 images TDM annotées sont introduites en utilisant les ontologies MEDICO, FMA et RadLex. Deux mécanismes d'interrogation de données sont intégrés : i) « *Query by Concept* » qui permet à l'utilisateur d'interroger la base de données d'images médicales à travers l'ontologie MEDICO et ii) « *Query by scribble* » permet à l'utilisateur de se référencer sur les traces visuelles des tumeurs à travers des requêtes.

III.3.1.2 Les ontologies pour l'analyse des images IRM

Le dépistage du cancer du foie s'effectue habituellement à travers l'échographie ou la TDM. Cependant, Malgré l'importance de ces deux techniques, elles sont parfois insuffisantes pour bien caractériser les lésions hépatiques. L'IRM est devenue une modalité de plus en plus privilégiée en routine clinique, car elle est plus sensible dans la détection de ces lésions. Dans ce contexte, [81] expose une représentation ontologique des relations spatiales permettant de guider l'interprétation de l'image IRM et des structures biomédicales. Cette ontologie est définie comme étant une contribution théorique originale. Elle est enrichie par des représentations conceptuelles floues qui définissent leurs aspects sémantiques. Elle permet également d'établir le lien entre les concepts (qui sont souvent exprimés sous forme de termes linguistiques) et les informations sémantiques extraites à partir des images médicales. De plus, l'intégration des représentations floues permet de gérer l'imprécision liée aux connaissances du domaine concerné. Cette ontologie constitue enfin un cadre adéquat pour le raisonnement sémantique de ces connaissances, tout en réduisant l'écart sémantique entre les concepts symboliques et l'information numérique. Ainsi, l'approche proposée assure la réduction des lacunes entre l'information numérique contenue dans l'image et son niveau conceptuel. Pour la représentation des concepts anatomiques cérébraux, ce travail fait appel à l'ontologie FMA. Le peuplement de l'ontologie proposée est fait à travers des images IRM 3D cérébrales. [82] développe une ontologie permettant la conceptualisation des modèles utilisés pour la simulation des images médicales. L'objectif est l'exploitation et la réutilisation des informations cliniques contenues dans d'autres ontologies pour assurer une interprétation complète des images

médicales. Cette approche importe quelques ontologies existantes telles que FMA, PATO (« *Phenotype And Trait Ontology* ») [83], MPATH (« *Mouse Pathology ontology* ») [84], RadLex [79] et ChEBI (« *Chemical Entities of Biological Interest ontology* ») [85]. Elle génère un prototype nommé OntoVIP (« *VIP Ontology* »).

III.3.1.3 Les ontologies pour l'analyse d'autres types d'images

Cette section présente les travaux qui intègrent les ontologies pour l'analyse d'autres types d'images médicales (par exemple la tomographie par émission de positrons PET ou les images histopathologiques). Dans ce contexte, [86] propose une approche ontologique permettant la représentation des données contenues dans une imagerie hispathologique qui sert à assurer le diagnostic en faisant l'étude microscopique des tissus. Ce travail cherche à fusionner ces données avec d'autres termes cliniques associés. Un prototype nommé QHIO (« *Quantitative Histopathological Imaging Ontology* ») a été développé. Il contient plusieurs images pathologiques, des processus techniques ainsi que des algorithmes de calcul. Il sert à raffiner les approches déjà développées en améliorant les outils d'annotation. Il comporte :

1. une entrée contenant un type (par exemple, image), un processus de préparation (résolution, autres paramètres d'annotation, *etc.*);
2. des paramètres (nombre d'itérations, *etc.*);
3. une sortie contenant aussi un type (par exemple image, mesure), des détails statistiques (par exemple la précision), une méthodologie d'évaluation statistique, (par exemple, courbe ROC (« *Receiver Operating Characteristic* ») et des logiciels requis.

Dans [87], les auteurs utilisent les techniques du raisonnement sémantique pour faire l'analyse des données médicales associées à l'image. Ce travail propose un système permettant la reconnaissance et l'extraction des objets biologiques à partir des images. En outre, il sert à dégager l'information visuelle par l'application d'un moteur d'analyse d'images. La méthodologie appliquée a permis d'améliorer l'expressivité des modèles cliniques, l'utilisation de la plateforme pour les domaines pathologiques et la sensibilité des algorithmes d'analyse d'images médicales. Dans [88], l'objectif est d'explicitier l'information que peut contenir une image médicale. L'approche proposée consiste à construire une base de données structurée d'images médicales nucléaires utilisant l'annotation AIM³⁹ (« *Annotation and Image Markup* »). Elle vise à faciliter le processus de recherche d'images. Ce travail a permis également de construire une méthode d'annotation de 100 images issues de la modalité combinée PET-CT (« *Positron Emission Tomography-Computed Tomography* »). Cette annotation est faite en se basant sur les vocabulaires RadLex, FMA et SNO-MED CT. Pour interroger cette base de données d'images, 20 requêtes SPARQL ont été intégrées pour tester l'efficacité et le niveau de performance du système proposé.

Le travail de [89] propose un outil permettant l'annotation sémantique des images médicales sous le format DICOM⁴⁰ (« *Digital imaging and communications in medicine* »). L'approche proposée exploite le contenu des images médicales pour créer une boîte de communication entre les

39. <https://rubinlab.stanford.edu/node/319>

40. <https://www.dicomstandard.org/>

experts du domaine. C'est un outil qui permet de collecter les annotations des chercheurs et des médecins concernant des observations d'images bien déterminées. Ce travail a également comme objectif de sauvegarder la sémantique d'images et de toutes les données scientifiques associées (signification biologique et physiologique). La méthode proposée est constituée essentiellement de trois phases : i) créer une ontologie pour contrôler les terminologies décrivant l'image médicale en assurant un modèle d'information standard, ii) développer un outil d'annotation d'image pour collecter les annotations des utilisateurs en tant qu'instances dans l'ontologie. Cette ontologie va informer l'utilisateur sur le type d'information qui doit être recueillie compte tenu de l'annotation contextuelle. Finalement, iii) sérialiser les données en utilisant HL7 ⁴¹(« *Health Level Seven* »), CDA (« *Clinical Document Architecture* ») [90] et OWL) comme langages pour permettre l'accès aux données sur le Web. Le Tableau II.2 donne un récapitulatif des travaux que nous avons décrits dans cette section.

41. <https://www.hl7.org/>

TABLE III.5 – Ontologies pour la modélisation des images médicales. Les signes ‘+’ et ‘-’ désignent dans chaque ligne du tableau si le travail étudié (1) visualise l’ontologie proposée, (2) s’il a créé un prototype à réutiliser, (3) s’il a fait l’implantation de l’ontologie dans une application, et (4) s’il a fait la validation de l’ontologie proposée

Approche	Sources de données médicales	Objectifs	Framework, vocabulaire utilisé	Représentation de l’ontologie	Nom de l’approche ontologique	Implémentation De l’ontologie	Validation de l’ontologie	Outils
[88]	IRB (the Institutional Review Board)	Annotation d’images médicales de type PET-CT pour faciliter et améliorer le processus de la recherche d’image dans une base de données.	ePAD, RadLex, FMA, SNOMED CT	+	-	+	+	Protégé, OWL, SPARQL, SWRL, Java
[75]	Rapports radiologiques interprétant le cancer du foie	Un <i>framework</i> pour la modélisation des lésions hépatiques interprétées dans des images médicales de type CT	VST, RadLex	-	-	+	+	Riesz wave-lets, SVM
[72]	Rapports radiologiques	Représentation sémantiques des images médicales CT du foie	RadLex	+	ONLIRA	+	+	OWL

[86]	NCI's Early Detection Research Network (EDRN)	Représentation ontologique des données médicales utilisée pour l'analyse des imageries pathologiques	GO OBO FMA, PATO, MPATH, RadLex, ChEBI, DOLCE RDF	+	QHIO	+	+	Protégé, OWL
[82]	Rapports radiologiques	Développement d'un <i>framework</i> dédié à l'annotation des images médicales	OntoVIP	+	OntoVIP	+	+	OWL- DL
[87]	Images histopathologiques	Proposer des formalisations ontologiques de haut niveau permettant d'annoter les images médicales d'une manière précise	RDF, JDOM	-	BCO (Breast Cancer Ontology)	+	+	Protégé, OWL- DL, SWRL, SPARQL, WFML, Ru- leML, Java

	National Cancer Institute (NCI) through the cancer Biomedical Informatics Grid (caBIG)	Annotation et sauvegarde des observations d'imageries médicales sous format DICOM	AIM (Annotation and Image Markup), RadLex, MIAME project	+	-	+	+	Protégé, DI-COM, HL7 CDA (XML), OWL
[122]	Images IRM	Représentation ontologique floue des relations spatiales permettant de guider l'interprétation de l'image et la reconnaissance des structures biomédicales associées	FMA	+	-	+	+	Protégé, OWL-DL
[81]								

	Amélioration des méthodes de recherche d'images (CBIR) en assurant des annotations et des interprétations automatiques de ces images ONLIRA	SVMs, WNN	-	-	+	+	+	-	-	-
[71]	ImageCLEF 2014 Liver Annotation Challenge									
[80]	Des experts du domaine médical	MEDICO ontology, FMA, RadLex	+	-	+	+	+	-	-	SPARQL, OWL, RDF, Jena (TDB)

TABLE III.6 – Propriétés de l’ontologie proposée par [92]

Relation	Domain	Range
hasCauses	Hepatocellular-Carcinoma	Risk-Factors
hasSymptoms	Hepatocellular-Carcinoma	Symptoms
hasStage	Hepatocellular-Carcinoma	Stage
diagnosedBy	Hepatocellular-Carcinoma	Diagnosis
treatedBy	Hepatocellular-Carcinoma	Treatment
stagedBy	Hepatocellular-Carcinoma	Staging
has-T	Stage	T
has-N	Stage	N
has-M	Stage	M

III.3.2 Modélisation des données médicales *via* les ontologies

Dans cette section, nous étudions les travaux d’intégration des ontologies pour la modélisation des données médicales, liés essentiellement à la stadification des tumeurs et la modélisation du processus de diagnostic médical.

III.3.2.1 Stadification des tumeurs hépatiques

Le travail de [91] expose les travaux réalisés par les mêmes auteurs dans le contexte de l’intégration de l’ingénierie de connaissance pour le diagnostic du cancer du foie. Ces travaux servent à la représentation des lésions hépatiques en se basant sur les ontologies. L’objectif principal de ces travaux est de montrer le rôle important des ontologies dans plusieurs domaines de recherche et plus précisément dans la médecine. En effet, ces travaux mettent l’accent sur la modélisation des pathologies hépatiques (*e.g.*, les virus hépatiques, le CHC, *etc.*). Parmi ces travaux, nous trouvons l’approche citée dans [92]. Elle vise à créer une application Web destinée aux utilisateurs (experts d’un domaine donné par exemple) pour faciliter l’extraction des informations associées au foie. Elle fournit une représentation sémantique du cancer du foie. La stadification des tumeurs hépatiques est aussi traitée en se basant sur le système de classification TNM. L’ontologie créée assure la modélisation du processus de diagnostic médical. Elle offre une base de connaissance sur cette maladie, accessible aux chercheurs du domaine médical. Cette ontologie contient 25 concepts. Parmi ces concepts, nous trouvons : « *Cancer* », « *LiverCancer* », « *Hepatocellular-carcinoma* », « *Staging* », « *Diagnosis* », « *Treatment* » et « *References* » « *Symptoms* », « *TNM-System* », *etc.* Le Tableau III.6 présente quelques propriétés utilisées par cette ontologie.

L’objectif de [93] est la représentation des virus hépatiques A, B, C et D à travers des connaissances médicales. Ce travail s’appuie sur le *framework* de construction des ontologies médicales OBR (« *Ontology of Biomedical Reality* »). La réalisation de ce travail repose sur trois phases principales : (1) l’extraction des termes relatifs au domaine médical, (2) la classification des maladies hépatiques en trois classes et enfin (3) une phase de validation. Cette approche collecte les données médicales

associées à la maladie étudiée en consultant différentes sources (par exemple les livres médicaux, les sites Web et les experts du domaine). Elle consiste à extraire les symptômes, les signes, et les résultats des tests médicaux de chaque virus hépatique. Ainsi, elle sert à classer les maladies causées par ces virus. La validation de ce travail est faite en consultant les experts du domaine. Ce travail possède comme finalité un arbre de classification des virus hépatiques.

Dans le même contexte, le travail de [94] présente un système de classification des maladies hépatiques en se basant sur des modélisations sémantiques proposées dans des travaux traitant le même sujet [93], [95], [96], et [97]. Ce travail donne une vue globale sur le système hépatobiliaire de l'organisme humain, incluant donc le foie. Il étudie ses fonctionnalités et les maladies pouvant affecter ce système dans le cas général. Le but principal était de faire une étude comparative entre les travaux existants permettant de classer les maladies reliées principalement au système hépatique. VHOSWS (« *Viral Hepatitis Ontology Sharing and Diagnosing* ») [98] et LIO (« *Liver Immunology Ontology* ») [95] sont deux systèmes ontologiques présentés dans cette étude et destinés à faire le diagnostic du cancer du foie.

[98] est aussi un travail réalisé par les mêmes auteurs. Il vise à la construction d'un service Web partageable entre les médecins et les physiciens. Ce service permet de traiter les maladies hépatiques, le diagnostic et les types de traitement possibles en se basant sur des modélisations sémantiques. Pour assurer ce modèle ontologique, il est obligatoire d'ajouter un ensemble d'opérations sous forme de requêtes répondant à quelques questions (par exemple 'ViewHierarchical-ViralHepatitisOntology' qui retourne un fichier OWL contenant l'ontologie décrivant les virus hépatiques, 'QueryForSymptomsSignsOrLaboratory-Findings' qui retourne l'ensemble des symptômes, les signes ou les tests médicaux de la maladie donnée comme entrée au système, *etc.*). L'implémentation du prototype VHOSWS repose sur la construction d'un service web développée en ASP.NET (« *Active Server Pages Technology* »). C'est une application accessible aux physiciens, médecins et étudiants.

[99] étudie les ontologies médicales permettant le traitement des maladies cancéreuses. Ce travail a pour objectif de déterminer le type de la tumeur et son stade en intégrant le système de stadification TNM, ainsi que le type de traitement à appliquer. L'architecture proposée est basée sur trois modules : i) un module de diagnostic, ii) un module de stadification et iii) un module de traitement. Ces modules utilisent des bases de données formées de plusieurs ontologies qui décrivent chaque type de cancer à travers les concepts et les propriétés qu'elles contiennent. Les signes, les symptômes et les stades de chaque cancer sont extraits à travers l'outil de raisonnement FACT++⁴².

Dans [100], les auteurs se concentrent sur le problème de la classification des tumeurs lors du processus de diagnostic. Ce travail propose un outil de classification des cancers en se basant sur les technologies du Web sémantique et l'ingénierie des connaissances. Il est destiné à améliorer et faciliter l'interprétation des images médicales. Les ontologies sont intégrées pour réaliser l'identification automatique des stades des lésions détectées dans les images médicales. En outre, elles permettent une annotation sémantique de ces lésions (la localisation, la taille, *etc.*). Cette approche utilise le système TNM ainsi que l'outil ePAD (« *the electronic Physician Annotation Device* ») pour assurer respectivement la représentation sémantique et le stockage des résultats sous le format

42. <http://owl.man.ac.uk/factplusplus/>

AIM. Le raisonnement est effectué à travers les règles d'inférence SWRL.

[101] présente une application dédiée à la classification des tumeurs en exploitant le système de classification TNM. Ce travail profite des capacités du calcul intelligent offert par les ontologies pour générer le prototype TNM-O (« *TNM-Ontology* »). Cette approche réutilise les ontologies FMA et BioTopLite2⁴³ [102] pour représenter les données médicales. Ces données sont incluses dans des interfaces graphiques ou représentées à travers des tableaux d'information. Le processus de la classification commence par la création d'un fichier RDF qui englobe tous les individus de l'ontologie. Par la suite, ces données sont classées en appliquant le raisonneur Hermit⁴⁵ [103]. Pour évaluer ce travail, une base de données contenant 362 données médicales, provenant du département de la pathologie relative au Centre médical universitaire de Fribourg en Allemagne a été utilisée dans l'application.

Dans [104], l'objectif est de proposer une représentation sémantique des tumeurs hépatiques. Cette approche décrit le processus de diagnostic du cancer du foie. Les paramètres inclus ici sont les symptômes, les facteurs de risques, le diagnostic à travers les tests cliniques et l'imagerie médicale (TDM ou IRM), l'existence ou non d'un cancer du foie et le stade associé (localisé, régional ou distant). Le processus de diagnostic présenté est identifié à travers les classes suivantes : « *TRIPLE Assessment* » qui contient trois sous classes : « *Clinical* », « *Imaging* » et « *Pathology* » :

- « *Clinical* » contient « *Tummy Abdomen* », « *Self Examination* », « *Skin* » et « *Eyes* »
- « *Imaging* » contient « *Utrasound* », « *CT* », « *MRI* », et « *BoneScan* »
- « *Pathology* » contient « *Fineneedleaspiration* », « *CoreBiopsy* », « *Laparoscopy* », et « *Surgical Biopsy* »

L'interrogation et la manipulation des données de l'ontologie (classes et propriétés) sont réalisées en utilisant SPARQL sous forme de requêtes et de réponses à ces requêtes. Les auteurs de [95] développent une modélisation sémantique nommée LIO basée sur le *framework* OBO⁴⁶ (« *Open Biological and Biomedical Ontologies* ») [25]. C'est une ontologie qui fournit des informations décrivant des réactions immunologiques dans le contexte du cancer du foie. La méthodologie de cette approche est décomposée en trois phases : i) extraction des concepts à partir des travaux connexes traitant le même sujet, ii) raffinement de ces concepts pour les rendre plus expressifs, et iii) construction d'un vocabulaire bien formé. Cette approche importe des ontologies existantes notamment FMA pour extraire des termes anatomiques, CO⁴⁷ (« *The Cell Ontology* ») pour présenter les types de cellules hépatiques, PRO⁴⁸ (« *The Protein Ontology* ») , GOCC⁴⁹ (« *Gene Ontology Cellular Component* ») et GOBP⁵⁰ (« *Gene Ontology Biological Processes* »). [26] offre une modélisation descriptive des connaissances qui prend en compte les critères de diagnostic. Ces critères sont utiles pour classifier les patients selon le type des maladies. Cette méthode re-

43. <https://biportal.bioontology.org/ontologies/BT>

45. <http://www.hermit-reasoner.com/>

46. <http://www.obofoundry.org/>

47. <http://www.obofoundry.org/ontology/cl.html>

48. <http://www.obofoundry.org/ontology/pr.html>

49. <https://rdrr.io/rforge/dcGOR/man/onto.GOCC.html>

50. <https://rdrr.io/cran/dcGOR/man/onto.GOBP.html>

pose sur des connaissances médicales spécifiques et définies par les sociétés savantes (médecins, radiologues, *etc.*). Ce travail a réussi à classifier 30 cas réels de patients en exploitant l'outil de raisonnement Hermit .

III.3.2.2 Résolution des problèmes cliniques

Le travail de [27] propose une méthode permettant la représentation sémantique des données images contenues dans un fichier DICOM. Il cherche à stocker les données extraites et les caractéristiques des images dans un fichier RDF. L'objectif est de trouver les patients ayant les plus grandes tumeurs. Les outils appliqués pour la réalisation de ce travail sont : SPARQL, (SeDI) (« *Semantic DICOM conversion service* »), PACS (« *Picture Archiving and Communication System* ») et dcm4che⁵¹. [73] décrit le projet CaReRa (« *Case Retrieval in Radiology* »). C'est un projet basé sur les concepts de CES (« *Clinical Experience Sharing* ») destinés à traiter le cancer du foie. Ce travail s'intéresse aux données issues de ImageCLEF⁵² qui fait partie du forum CLEF (« *Cross Language Evaluation Forum* »)[28]. Parmi les accomplissements de ce projet, nous trouvons le lancement de l'ontologie ONLIRA⁵³. Ce travail réalise des tests sur l'annotation sémantique des observations médicales en se basant sur des portions d'examens TDM du foie. Il applique le CoG (« *Computer generated features* ») qui génère des caractéristiques globales (le volume, l'intensité du foie, le nombre de lésions, le volume des lésions, la lésion la plus grande, la lésion la plus petite, *etc.*) et des caractéristiques individuelles (la texture, le niveau de gris, les frontières, *etc.*). Les observations médicales collectées sont affichées dans le site Web du projet CaReRa⁵⁴.

[29] propose le prototype LiCO (« *Liver Case Ontology* ») modélisant les tumeurs hépatiques. Pour la réalisation de ce travail, cette approche réutilise plusieurs terminologies médicales telles que SNOMED-CT, ONLIRA, RadLex et LOINC (« *the Logical Observation Identifier Names and Codes* »). Les auteurs de [30] et [31] fournissent respectivement LiverAtlas et DILIO (« *drug-induced liver injury* »), qui sont des bases de connaissances contenant des données cliniques sur les tumeurs hépatiques.

Dans le Tableau III.7, nous exposons une synthèse des différents travaux de recherche proposant des approches ontologiques pour la modélisation des données médicales.

51. <https://www.dcm4che.org/>

52. <https://www.imageclef.org/>

53. <https://bioportal.bioontology.org/ontologies/ONLIRA>

54. <https://vavlab.ee.boun.edu.tr:5904/CareraWeb2>

TABLE III.7 – Ontologies pour la modélisation des données médicales. Les signes ‘+’ et ‘-’ désigne dans chaque ligne du tableau si le travail étudié (1) visualise l’ontologie proposée, (2) s’il a créé un prototype à réutiliser, (3) s’il a fait l’implantation de l’ontologie dans une application, et (4) s’il a fait la validation de l’ontologie proposée

Approche	Classification des lésions hépatiques	Sources de données médicales	Objectifs	Framework, vocabulaire utilisé	Représentation de l’ontologie	Nom de l’approche ontologique	Implémentation de l’ontologie	Validation de l’ontologie	Outils
[91]	+	Cancer.Net, Medicine-Net, The National Cancer Institute(NCI)	Efficacité de l’utilisation des ingénieries de connaissances pour la modélisation des données médicales	-	+	-	+	+	Protégé, OWL
[101]	-	Données médicales, offertes par le département de la pathologie relative à un hôpital universitaire	Classification des tumeurs en se basant sur le système de classification TNM	FMA, Bio-TopLite2	+	TNM-O(TNM-Ontology)	+	+	OWL, Java, Hermit reasoner

[100]	+	Département de Radiologie de l'Université de Stanford	Classification des tumeurs malignes en se basant sur le système TNM	ePAD (web-based platform for quantitative imaging in the clinical workflow), AIM (Annotation and Image Markup)	-	-	+	+	+	OWL, SPARQL, SWRL
[94]	+	N/A	Présentation d'un système de classification des maladies reliées au système hépatobiliaire basée sur des travaux connexes	-	+	-	-	+	+	Protégé XML RDF(S) DAML-OIL OWL
[104]	+	Medecinenet, Cancer.org, Indiahospital.com, Healthline.com	Ontologie dédiée au diagnostic et au traitement du cancer de foie	-	+	-	+	+	+	SPARQL, OWL, Protégé

[92]	+		MedicineNet, CancerNet, NCI(The National Cancer Institute)	Création et mise en ligne d'une ontologie dédiée à la représentation du cancer foie	-	+	-	+	+	+	Protégé, OWL, OWL DL
[93]	+		Les experts du domaine, Livres médicaux	Ontologie présentant les virus hépatiques	OBR	+	-	+	+	+	Protégé, OWL
[98]	+		Web	Construction d'un service Web partageable entre les médecins et les physiciens	-	+	VHOSWS	+	+	+	OWL, C# ASP.NET
[26]	+		ASAS(The Assessment of SpondyloArthritis international Society)	Ontologie permettant la classification d'un patient selon le type de la maladie	-	+	-	+	+	+	OWL, Protégé, OWL-DL 2.0,Hermit 1.2.2, reasoner

<p>[95]</p>	<p>-</p>	<p>The National Center for Biomedical Ontology</p>	<p>Ontologie dédiée à la représentation des réactions immunologiques du foie contre les maladies hépatiques</p>	<p>OBO, FMA, CO, PRO, CO :CC et GO :BP</p>	<p>LIO</p>	<p>+</p>	<p>+</p>	<p>+</p>	<p></p>
-------------	----------	----------------------------------------------------	-----------------------------------------------------------------------------------------------------------------	--------------------------------------------	------------	----------	----------	----------	---------

III.3.3 Modélisations des rapports médicaux

Lors du processus de diagnostic médical, les cliniciens et les radiologues rédigent habituellement des rapports permettant de caractériser l'état clinique du patient. Ces rapports contiennent généralement des informations sur le patient, sa maladie, l'interprétation de l'image médicale, et le processus de diagnostic à appliquer. Par conséquent, au sein des centres médicaux (hôpitaux, cliniques, *etc.*), le stockage des données cliniques se fait à l'aide de systèmes informatiques dédiés (comme le PACS). Ces données peuvent être des dossiers médicaux, des résultats d'examen complémentaires d'images, des comptes-rendus radiologiques, *etc.* Le sauvegarde de ces informations permet d'y accéder d'une manière plus simple, d'améliorer la collecte de nouvelles informations, ou encore de faciliter l'aide à la décision clinique [32]. Par conséquent, le traitement des rapports textuels à travers des mécanismes sémantiques a constitué l'objectif principal de divers travaux de recherche. Dans cette section, nous étudions des exemples de ces travaux.

III.3.3.1 Modélisation des rapports en lien avec le cancer CHC

[33] proposent une approche basée sur le système EHR (« *Electronic Health Record* ») permettant de traiter les données médicales. Ce système est développé non seulement pour fournir des informations relatives à un patient, mais il se comporte aussi comme un outil d'aide à la décision clinique. Cette méthode fonctionne en transformant les termes cliniques des rapports radiologiques en des vecteurs numériques. L'objectif est de faire une comparaison entre ces concepts et les termes relatifs à l'ontologie médicale SNOMED-CT. La contribution de cette approche utilise 112 images abdominales de type TDM. Parmi ces images, 59 images présentant un cancer du foie ou des métastases hépatiques de type CHC et 53 présentant des cas sans ces pathologies. La méthodologie suivie consiste tout d'abord à collecter des rapports radiologiques réels d'une manière aléatoire. La deuxième étape sert à l'extraction des termes cliniques présentant le CHC. Il suffit de relier ces termes avec des concepts ontologiques extraits de l'ontologie médicale SNOMED. La méthode de recherche PubMed⁵⁵ et le coefficient de corrélation Pearson sont appliqués pour mesurer la similarité entre les termes. L'ontologie UMLS est intégrée dans cette approche pour faciliter l'extraction des termes cliniques présentant le CHC. Elle est utile aussi pour construire des dossiers de santé et des dictionnaires électroniques. Par la suite, et par application de SNOMED-CT, ce travail propose l'organisation de ces termes en ajoutant les relations qui associent chaque concept à un autre en incluant la hiérarchie «*is-a*» entre ces deux concepts. Puis, une comparaison est conduite entre tous les rapports pour permettre d'identifier ceux qui présentent le CHC. Cette représentation est importée du travail de [34]. Le système proposé génère à la fin 30 concepts présentant le cancer du foie. Parmi les concepts générés, nous pouvons citer : « *Abdominal organ finding* », « *Blood vessel finding* », « *Disorder of body cavity* », « *Disorder of digestive system* », « *Disorder of body system* », « *Liver finding* », « *Radiologic finding* », *etc.* Le système proposé offre des guides cliniques permettant l'aide à la décision médicale (par exemple dégager le traitement à appliquer).

Dans le même contexte, [35] proposent un système basé sur du machine learning. Ce système applique la technique de la résolution par référence qui permet l'extraction automatique des caractéristiques des lésions hépatiques à partir d'un rapport médical textuel. Ce travail décrit également

55. <https://pubmed.ncbi.nlm.nih.gov/>

TABLE III.8 – Les relations de l’ontologie proposée par [35]

Relation	Domain	Range
hadMeasurement	Tumor reference, Measurement	Measurement
hasCount	Tumor reference	Tumor count
hasMeasurement	Tumor reference	Measurement
hasTumEvid	Tumor reference, measurement	Tumorhood Evidence
isNegated	Tumor reference	Negation
locatedIn	Anatomy	Tumor reference, Measurement
refersTo	Measurement	Anatomy

un algorithme qui prend en entrée un modèle caractérisant une tumeur et fournit en sortie les caractéristiques de cette tumeur (par exemple le nombre de nodules, la taille, et le stade). Cette approche fait appel aux algorithmes reliés aux systèmes de classification des lésions hépatiques cancéreuses tels que AJCC (« *American Joint Committee on Cancer* ») [145], BCLC, et CLIP (« *Cancer of the Liver Italian Program* ») [65]. La base de données utilisée contient au total 3211 entités et 2283 relations. Les rapports sont annotés en utilisant 7 entités : « *Anatomy* » (par exemple "liver"), « *Measurement* » (e.g. "2.4 cm"), « *Negation* » (e.g. "no"), « *Tumor count* » (e.g. "2" ou "multiple"), « *Tumorhood Evidence* » (e.g. "cyst", "suspicious for HCC") et « *Tumor reference* » (e.g. "lesion"). Le tableau III.8 présente les relations proposées par ce modèle ontologique.

D’autres termes sont intégrés comme « *AnatomyMeas* » qui décrit les événements reliés avec la relation *refersTo*, *Negation* décrit les événements de la relation « *isNegated* », « *TumorSingleton* » décrit les événements des entités « *Tumor reference* » ou « *Measurement* », « *OtherSingleton* » décrit les événements reliés avec une seule entité autre que « *TumorSingleton* » et « *Tumor* » décrit les autres types d’événements autres que ceux qui sont cités précédemment. Pour annoter les rapports radiologiques, deux types d’expressions sémantiques sont indiquées qui sont : *coreference* qui exprime les expressions équivalentes dans un rapport radiologique et *particularization* qui exprime des entités qui se trouvent dans une première expression et contiennent des références d’autres entités qui se trouvent dans la deuxième expression. Pour annoter la résolution par référence, ce travail utilise : MUC (« *Message Understanding Conference* ») [36], B-cubed [37] et CEAF [38]. Pour annoter les caractéristiques des tumeurs, le système utilise les informations suivantes : la taille et le nombre des tumeurs ainsi que l’état d’envahissement tumoral au foie.

III.3.3.2 Modélisation d’autres types de rapports médicaux

Les auteurs de [39] proposent un modèle ontologique nommé MCI (« *Model for Clinical Information* ») basé sur les ontologies OBO . Ce modèle vise à représenter sémantiquement des données médicales extraites à partir des rapports radiologiques. Cette représentation consiste à transformer les notes textuelles écrites par les radiologues à une formalisation structurée bien détaillée. Dans ce contexte, les indications cliniques trouvées dans les images médicales permettent de détecter

le type de la tumeur et d'identifier la nature des entités anatomiques. Cette approche utilise ces indications pour développer le prototype *ReportViewer* qui sert à la visualisation des résultats de l'analyse d'images médicales et à la classification des rapports radiologiques. Le modèle clinique proposé fonde son architecture sur les ontologies OBO, RO ⁵⁶ (« *Relations Ontology* »), OGMS ⁵⁷ (« *Ontology for General Medical Science* »), IAO ⁵⁸ (« *Information Artifact Ontology* »), OBI ⁵⁹ (« *the Ontology for Biomedical Investigations* »), PATO ⁶⁰ (« *Phenotypic Quality Ontology* »), UO ⁶¹ (« *Units Ontology* ») et FMA. Le modèle MCI contient 551 classes, 107 propriétés d'objet, et 33 propriétés de données. Parmi les classes définies par le modèle MCI, nous trouvons Clinical Finding, Normal finding et Abnormal finding. L'évaluation de ce travail est réalisée *via* l'implémentation d'un outil nommé RECIST (« *Response Evaluation Criteria In Solid Tumors* »). Il utilise SPARQL pour interroger l'ontologie médicale créée.

Pour [72], l'objectif était de faire la modélisation des images médicales sous la forme d'une représentation sémantique. Cette représentation est basée sur le vocabulaire RadLex. Le principe de ce travail consiste à représenter les concepts radiologiques, extraire les propriétés individuelles relatives à ces concepts (par exemple la taille et la densité du foie) et ajouter les relations entre ces concepts. Les concepts de base dans cette ontologie sont : « *Liver* » qui contient des informations cliniques (contour, taille, densité, segments, régions, *etc.*), « *Lesion* » qui caractérise la tumeur présente dans le foie et définie par une forme, une densité, un contraste, *etc.*, et le concept « *HepaticVascularity* » contient des informations sur les veines hépatiques. Parmi les propriétés utilisées, nous pouvons citer : « *located* » qui désigne la localisation d'une tumeur dans un segment du foie, « *hasCraniocaudalDimension* » qui désigne une valeur numérique en millimètres décrivant la dimension du foie, et « *hasSizeChange* » qui est une propriété destinée à suivre l'évolution de la taille du foie. Une description sémantique d'une lésion hépatique est présentée par les caractéristiques suivantes : le nombre de nodules : *hasLesionQuantity"multiple"*, la forme : *hasLesionShape"Hypodense"*, la densité : *hasAreaDensity"hypodense"*, le contraste : *hasLesionContrastPattern"peripheric"*, *etc.* Cette ontologie a été mise en ligne sous le nom ONLIRA ⁶³ (« *Ontology of Liver for Radiology* »). L'évaluation de cette ontologie a été réalisée avec 30 rapports radiologiques issus de différents patients, en utilisant des requêtes qui manipulent les données et des résultats répondant à ces requêtes.

[40] présentent également un prototype ontologique nommé RadiO (« *Radio Ontology* ») permettant le traitement des rapports radiologiques. La méthode de développement consiste à collecter les observations faites par les radiologues. Ces observations sont utilisées pour créer une ontologie de domaine d'imagerie, représentant des connaissances sur les entités de l'image et leurs caractéristiques. Le développement de cette ontologie est basé sur l'ontologie FMA. Cette approche vise à améliorer l'identification des caractéristiques des entités anatomiques d'une image. Ces caractéristiques sont appliquées par la suite dans les interprétations radiologiques et dans le processus de

56. <http://www.obofoundry.org/ontology/ro.html>

57. <https://bioportal.bioontology.org/ontologies/OGMS>

58. <http://www.obofoundry.org/ontology/iao.html>

59. <http://obi-ontology.org/>

60. <https://bioportal.bioontology.org/ontologies/PATO>

62. <https://bioportal.bioontology.org/ontologies/UO>

63. <http://vavlab.ee.boun.edu.tr/pages.php?p=research/CARERA/OnliraDownload.php>

diagnostic médical. De plus ils fournissent une base de connaissances structurée dans le domaine de l'imagerie médicale.

Les auteurs de [41] proposent le système AMBIT (« *a processing framework for Aquiring Medical and Biological Information from Text* »). Il s'agit d'un outil d'analyse de texte destiné à extraire des informations cliniques à partir des documents médicaux. Les données extraites sont sauvegardées sous un format bien structuré pour assurer un accès facile aux informations par les utilisateurs. Ce système a été développé dans le cadre d'une collaboration entre deux projets : CLEF (« *Clinical E-Science Framework* ») [42] et myGrid [43]. Il fonctionne en employant (1) un moteur de terminologie nommé *Termino* qui à son tour est composé des données issues des terminologies UMLS, HUGO, et Go et d'un compilateur pour l'identification de ces termes, et (2) un moteur d'interrogation par requêtes. AMBIT contient 160 000 termes importés à partir d'UMLS. Des exemples, contenant 83 rapports radiologiques sur le cancer des poumons, ont été utilisés pour tester l'efficacité de ce système.

[105] traitent la question de la structuration des rapports radiologiques. L'objectif principal était de surmonter les obstacles liés à l'extraction des informations durant le traitement de la langue allemande. La méthodologie proposée consiste à résoudre les problèmes de recherche, dans les routines cliniques, liés au domaine du NLP (« *Natural Language Processing* »), d'ajuster les artefacts liés aux termes cliniques ainsi que d'intégrer les représentations sémantiques présentant des rapports textuels. L'ontologie est intégrée dans ce travail pour faire la reconnaissance et la normalisation des terminologies cliniques. Le Tableau III.9 récapitule les travaux cités précédemment.

TABLE III.9 – Ontologies pour la modélisation des rapports médicaux. Les signes ‘+’ et ‘-’ désigne dans chaque ligne du tableau si le travail étudié (1) visualise l’ontologie proposée, (2) s’il a créé un prototype à réutiliser, (3) s’il a fait l’implantation de l’ontologie dans une application, et (4) s’il a fait la validation de l’ontologie proposée

Ontologie	Classification des lésions hépatiques	Sources de données médicales	Objectifs	Framework, vocabulaire utilisé	Représentation de l’ontologie	Nom de l’approche ontologique	Implémentation de l’ontologie	Validation de l’ontologie	Outils
[33]	+	Rapports radiologiques, Hôpitaux de Hong Kong	Identification et détection des lésions hépatiques CHC à partir des rapports radiologiques	SNOMED, EHR, PubMed	-	-	+	+	UMLS, Brat Software
[35]	+	Rapports radiologiques abdominaux	Proposition d’un système pour l’extraction des caractéristiques d’une tumeur hépatique à partir d’un rapport radiologique	MetaMap, WordNet	-	-	+	+	UMLS

[39]	-	Rapports radiologiques	Classification des maladies en se basant sur des rapports radiologiques	OBO, RO, OGMS, IAO, the OBI, PATO, UO, et FMA	+	MCI, ReportViewer	+	+	SPARQL, RECIIST
[40]	-	Rapports radiologiques	Présentation d'une application à base d'ontologie permettant le traitement des rapports radiologiques	RadLex, FMA	+	Radio (Radio Ontology)	+	+	Protégé
[105]	-	Des rapports radiologiques issus des hôpitaux allemands	Résoudre les problèmes de recherche, dans les routines cliniques, reliés au domaine NLP	SNOMED-CT	-	-	+	+	CLEF, myGrid

<p style="text-align: right;">[4]</p>	-	Rapports radiologiques	Présentation du système AMBIT dédié à extraire des informations cliniques à partir des données médicales et biomédicales	UMLS, HUGO (Gene Nomenclature database)	-	AMBIT	+	+	-
----------------------------------------------	---	------------------------	--------------------------------------------------------------------------------------------------------------------------	-----------------------------------------	---	-------	---	---	---

III.4 Classification des données médicales *via* les techniques d'apprentissage automatique et des ontologies

Dans cette section, nous étudions quelques travaux qui intègrent les techniques d'apprentissage automatique pour la classification des données, éventuellement en lien avec le cancer du foie. La première sous-section discute le sujet d'appliquer les techniques d'apprentissage artificiel, tandis que la deuxième sous-section aborde les travaux combinant les ontologies avec l'apprentissage automatique.

III.4.1 Application des techniques d'apprentissage automatique

La détection du cancer du foie est une tâche complexe à traiter. Le diagnostic de CHC est habituellement envisagé après la découverte d'une lésion focale nodulaire hépatique dans l'image médicale (échographie en premier lieu), ou à l'occasion de symptômes en cas de tumeur évoluée. Le diagnostic des nodules CHC repose sur l'examen de leur vascularisation. Cette complexité se caractérise également par le fait que les nodules de petites tailles sont difficiles à détecter.

[106] propose une approche basée sur l'architecture CNN permettant l'extraction des caractéristiques des cellules du foie à partir des images pathologiques et l'intégration de ces caractéristiques pour améliorer le processus de détection. La réalisation de cette approche repose sur l'application des techniques de VGG16⁶⁴ et Inception V4 [108]. [107] discute également le sujet de la classification des lésions hépatiques à partir des images IRM de différentes phases. Dans le même contexte, [48] propose un algorithme parallèle pour la détection des tumeurs hépatiques de type CHC à partir des séquences IRM avec prise de contraste. Ces derniers nécessitent l'intervention des agents de contraste quand les images sont acquises avec une résolution temporelle élevée, pour capturer les changements d'intensité du signal IRM. L'approche proposée est basée sur les méthodes ondelettes (« *wavelets* ») et un apprentissage par SVM. Elle utilise un traitement parallèle des données médicales réparties en patchs d'images IRM caractérisant les trois phases du contraste.

Dans [49], l'objectif est d'appliquer l'apprentissage profond pour la détection des lésions CHC à partir de séquences IRM avec prise de contraste. Cette approche utilise l'architecture U-Net dans le processus de classification pour la détection des tumeurs CHC (*U-CatHCC*). L'architecture proposée utilise une succession de couches de convolution pour faire une segmentation rapide et précise des images médicales. Ces couches servent aussi à restaurer progressivement l'information spatiale, la prédiction en sortie devient plus précise [109].

III.4.2 Application des ontologies avec les techniques d'apprentissage automatique

Le couplage entre les ontologies et les techniques d'apprentissage automatique fait l'objet de plusieurs travaux de recherche. Dans le contexte de classification des images, [110] propose

64. <https://neurohive.io/en/popular-networks/vgg16/>

une approche *multi-label* basée sur les ontologies. Cette approche repose sur des connaissances de base sur les recettes, les catégories d'aliments et leur lien avec les maladies chroniques. Ces connaissances sont modélisées à travers un modèle ontologique. La méthode proposée vise à faire la classification des aliments tout en exploitant les réseaux de neurones profonds, où chaque image alimentaire est classée avec des étiquettes décrivant les ingrédients de chaque recette. Le modèle appliqué pour faire la prédiction des composants de chaque image est Inception-V3.

[111] présente un système appelé *DeepVisInterests* qui prédit les intérêts des utilisateurs à partir de données visuelles en se basant sur des représentations sémantiques. Ce système applique les architectures CNN pour les modules d'extraction de caractéristiques et de classification. Il est basé sur une méthode neuronale profonde appliquée à la construction de l'ontologie. *DeepVisInterests* utilise un ensemble de fonctionnalités visuelles approfondies pour apprendre la représentation sémantique des sujets d'intérêt populaires définis par Facebook. Ce système aborde le problème de la découverte des intérêts attribués et de l'analyse des performances de la prédiction automatique à travers des images sociales collectées auprès de 240 utilisateurs de Facebook.

Dans le domaine médical, [75] expose une modélisation sémantique des lésions hépatiques observées dans les images TDM. L'objectif est de prédire les termes notés par les radiologues. Les modèles VST sont appliqués pour créer une ontologie modélisant ces lésions. Ces modèles sont créés à partir des images médicales en utilisant des combinaisons linéaires *via* une technique SVM et les *Riesz wavelets*. Dans un premier temps, ces modèles sont utilisés pour prédire la présence de chaque terme sémantique décrivant les lésions hépatiques. Dans un second temps, les distances entre tous les VST les modèles sont calculés pour former l'ontologie médicale présentant les tumeurs détectées. L'évaluation de ce travail a été conduite sur 74 annotations de lésions hépatiques.

Dans le même contexte, [112] présentent une approche basée sur l'architecture CNN et sur des connaissances du domaine. Le but est de classifier les images de la base constituée de 1500 images en plusieurs catégories. Les auteurs visent à étiqueter ces images avec les nœuds parents 'Fort', 'Tomb' et 'Mosque'. Ce travail applique une méthode basée sur l'apprentissage par transfert (« *transfer learning* ») qui vise à transférer les connaissances d'une ou plusieurs tâches sources vers une ou plusieurs tâches cibles. Les résultats ont montré que l'approche proposée est plus performante que les autres approches d'apprentissage classiques. Le taux de précision trouvée était de 75.59% contre 71.42% pour une méthode de classification classique (sans intégration de connaissances). Ces résultats ont prouvé également que le couplage entre les ontologies et le CNN donne de meilleures classifications sur des données complexes et volumineuses par rapport à un CNN standard.

[113] propose un modèle ontologique basé sur l'architecture HMAX (« *Hierarchical Max-pooling* ») pour extraire des caractéristiques visuelles des images. Il sert également à identifier les concepts et les relations à partir de ces caractéristiques afin de créer une modélisation sémantique. La première étape de cette approche consiste à extraire des fonctionnalités visuelles en appliquant le modèle HMAX. La deuxième étape consiste à identifier les concepts et les relations à partir des catégories d'images afin de créer l'ontologie. Ensuite, les classificateurs d'entités sont formés. Enfin, la dernière phase sert à tester les images classifiées et détecter la relation entre les classificateurs. Des ontologies et des structures sémantiques ont été intégrées dans ce travail pour améliorer la

classification des images. Les résultats obtenus ont montré une meilleure classification par rapport au réseau d'apprentissage profond Inception-V3 pour certaines classes d'images. Le taux de précision de la méthode proposée était de 63,8% contre 61,1% en appliquant la méthode basée sur Inception-V3.

Enfin, [114] proposent un *framework* permettant l'extraction des *labels* issus des rapports de radiologie et de l'ontologie des étiquettes. Ce *framework* vise à prédire une liste complète d'étiquettes pertinentes (par exemple le type de la lésion) qui pourraient être appliquées plus tard dans le processus de diagnostic. Un modèle de réseau d'annotation de lésion nommé LesaNet (« *Lesion Annotation Network* ») a été créé pour apprendre toutes les étiquettes de manière holistique. Ce modèle d'apprentissage profond est basé sur un CNN *multi-label* formé à partir des rapports médicaux associés aux images médicales des lésions. Pour améliorer la précision des prévisions, des relations entre les étiquettes ont été créées. Les expériences montrent que LesaNet peut annoter précisément les lésions en utilisant une ontologie composée de 171 labels avec un taux moyen d'AUC (« *Area Under The Curve* » : c'est une mesure de l'aire de la surface située sous le tracé d'une fonction) de 93,44 %.

III.5 Discussion

Dans ce chapitre, nous avons décrit en premier lieu les notions relatives à la détection des tumeurs de type CHC à partir des images médicales. Nous avons présenté également une vue sur les systèmes de stadification des tumeurs hépatiques. En deuxième lieu, nous avons présenté certaines approches qui traitent l'intégration des ontologies pour la classification des données médicales.

L'étude de ces approches a montré l'importance des ontologies médicales dans l'amélioration de plusieurs tâches cliniques telles que le diagnostic, la prise de décision clinique, et l'automatisation des dossiers médicaux. De plus, nous avons remarqué la richesse des modélisations sémantiques et leurs capacités à spécifier différentes situations complexes dans le domaine médical.

La majorité de ces travaux abordent le sujet du diagnostic médical du foie en se basant sur des modélisations sémantiques. Ils suggèrent des solutions informatiques performantes. Une attention considérable a été accordée à la détection des maladies hépatiques d'une manière générale. Cependant, la plupart de ces études ne discutent pas le processus de diagnostic complet des tumeurs hépatiques de type CHC. En outre, certains travaux traitent uniquement de la détection des tumeurs hépatiques en se basant sur les techniques de l'apprentissage automatique sans donner des informations supplémentaires sur la tumeur détectée (par exemple la taille, la localisation, le type, le stade, *etc.*). De plus, nous avons constaté qu'il n'y a pas beaucoup de travaux qui traitent et modélisent les tumeurs hépatiques de type CHC d'une manière très détaillée, en incluant plusieurs systèmes de stadification par exemple. Aussi, nous avons remarqué que tous les travaux ignorent l'intégration et la modélisation du système de classification Li-RADS dans le processus de détection des lésions hépatiques. Il s'agit d'un système faisant la standardisation du compte rendu radiologique en imagerie hépatique chez les patients à risque de CHC. Il est très utile pour définir, pour toute anomalie observée, la probabilité d'être un CHC. Enfin, la notion du couplage entre les techniques d'ingénierie de connaissances et le *deep learning* est presque absente dans le contexte des tumeurs CHC.

Comme notre objectif principal s’articule sur la classification des données médicales par le biais des ontologies, nous avons travaillé dans cette thèse sur deux volets interdépendants ; (1) étudier et répertorier les ontologies médicales dans le contexte du cancer du foie et (2), proposer une nouvelle approche, basée sur les ontologies créées, pour l’identification et le diagnostic des tumeurs hépatiques. Nous présentons dans cette thèse, une approche qui combine les deux techniques d’intelligence artificielle : les ontologies et le *deep learning*. Notre méthode repose sur l’architecture CNN pour développer un algorithme parallèle permettant la détection des lésions hépatiques de type CHC à partir des séquences IRM avec prise de contraste. Les ontologies que nous proposons sont également appliquées pour modéliser le processus de diagnostic et de stadification des tumeurs hépatiques en se basant sur des systèmes standards (TNM, BCLC et Li-RADS), mais jamais combinés ensemble jusqu’à présent. La méthode proposée permet également de suggérer le traitement à appliquer. Ainsi les contributions majeures de ce travail sont :

- Développer des approches ontologiques permettant la modélisation du cancer du foie, plus précisément pour le CHC. Trois modules principaux seront présentés (1) détecter les lésions à partir de l’image médicale, (2) extraire le stade de la tumeur détectée, et (3) proposer le type de traitement à appliquer.
- Construire les relations et les propriétés entre les concepts et intégrer le processus de raisonnement à travers les règles SWRL pour interroger sémantiquement des données grâce aux ontologies.
- Peupler l’ontologie avec des instances (des cas réels de patients infectés par le CHC).
- Intégrer les ontologies développées dans un *framework* de classification des lésions hépatiques.
- Tester et évaluer l’approche proposée par rapport aux travaux connexes.

Dans la suite de ce manuscrit, nous présentons nos contributions qui s’intéressent à la proposition d’ontologies médicales permettant le diagnostic, la modélisation et la stadification des tumeurs hépatiques. Nous proposons ensuite une approche basée sur le *deep learning* permettant la détection des lésions CHC. Enfin, nous exposons une approche permettant de fusionner les deux approches citées précédemment pour améliorer la détection des tumeurs CHC dans les séquences IRM avec prise de contraste.

REPRÉSENTATION DES CONNAISSANCES MÉDICALES À TRAVERS LES ONTOLOGIES

Sommaire

IV.1 Introduction	75
IV.2 Approche de détection et de classification du CHC <i>via</i> une ontologie : OntHCC 75	
IV.2.1 Phase A : Extraction des informations médicales à partir de l'image médicale du foie	76
IV.2.2 Phase B : La représentation sémantique des connaissances médicales	77
IV.2.3 Présentation de l'ontologie OntHCC	78
IV.2.4 Phase C : Diagnostic du CHC	84
IV.2.5 Raisonnement dans notre modèle ontologique	85
IV.3 Développement d'une ontologie pour la modélisation des rapports radiologiques du CHC (MROnt)	86
IV.3.1 Phase 1 : Population de la base de connaissances	87
IV.3.2 Phase 2 : Analyse des rapports radiologiques	88
IV.3.3 Présentation de l'ontologie MROnt	88
IV.3.4 Les règles SWRL	91
IV.4 Conclusion	94

IV.1 Introduction

La présentation des connaissances *via* les ontologies est devenue un sujet de recherche très populaire dans plusieurs domaines de recherche notamment l'intelligence artificielle, la recherche d'information, le traitement du langage naturel, l'ingénierie des connaissances, *etc.* Les ontologies offrent une connaissance formelle sur un domaine qui peut être partagée entre les personnes et les systèmes hétérogènes. En intelligence artificielle, elles visent à faciliter le partage des connaissances et leur réutilisation dans diverses applications. Dans ce chapitre, nous cherchons à intégrer les techniques d'ingénierie de connaissance pour la détection du CHC à partir d'images IRM. Cette maladie est découverte durant une phase de diagnostic primordiale à travers l'analyse et l'observation des résultats issus de l'imagerie médicale. La décision médicale est prise en se basant sur cette observation radiologique.

Nous proposons une approche ontologique dont les rôles principaux sont : (1) la représentation sémantique de la tumeur hépatique CHC, (2) la détection de cette tumeur à partir de l'imagerie médicale du foie, et (3) l'extraction du stade de la tumeur en se basant sur des systèmes standards de stadification tels que BCLC, TNM et Li-RADS. Nous proposons également une deuxième approche ontologique permettant l'analyse et la représentation des rapports radiologiques reliées aux patients infectés par le CHC.

IV.2 Approche de détection et de classification du CHC *via* une ontologie : OntHCC

Dans le contexte de la classification des lésions hépatiques, nous avons développé une première approche pour l'aide au diagnostic, à la classification et au choix de traitement des tumeurs CHC. L'approche proposée est présentée dans la Figure [IV.1](#), elle est fondée sur l'analyse d'images IRM de foies infectés et sur des rapports radiologiques. Elle permet de rassembler les informations nécessaires et relatives à la maladie concernée (CHC) sur la base de cas réels. Les informations tirées à partir de ces images seront introduites par la suite dans l'ontologie. Pour détecter le CHC à partir d'une IRM, nous prenons en compte plusieurs informations (par exemple, la taille du nodule observé, son profil vasculaire). Ainsi, notre contribution consiste à proposer un modèle ontologique pour la caractérisation des tumeurs hépatiques de type CHC. Notre objectif était de concevoir un système informatique permettant de suivre le processus de détection, de stadification et de traitement de ces lésions. De plus, nous souhaitons enrichir le système proposé en profitant de la richesse sémantique du raisonnement afin d'avoir un modèle compréhensible.

Notre approche prend en entrée des informations sur des nodules hépatiques et génère comme sortie des détails sur la tumeur détectée, à savoir le type, le stade et le type de traitement à appliquer. De plus, l'approche proposée est dotée de mécanismes d'inférences et de raisonnement qui rendent les connaissances compréhensibles par la machine et facilitent ainsi leur analyse par l'utilisateur. Notre approche est constituée essentiellement de trois phases : (a) la collection des données médicales (b) la création et l'enrichissement de l'ontologie et (c) l'implantation et l'interrogation de l'ontologie créée. Nous expliquons dans les sous-sections suivantes chacune de ces phases.

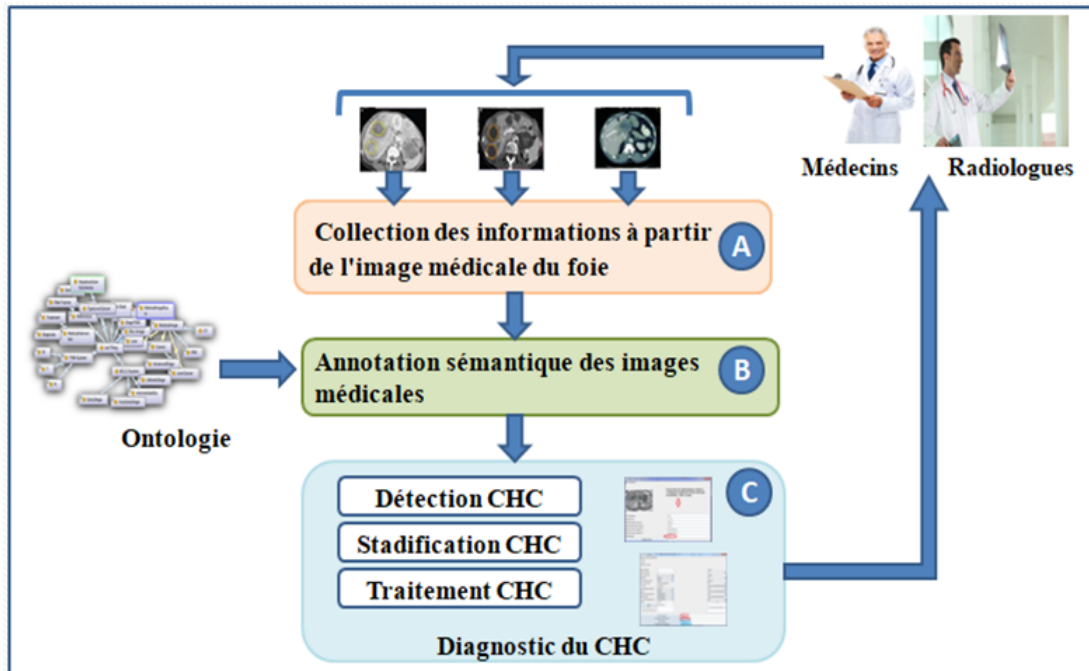


FIGURE IV.1 – Approche ontologique pour le diagnostic, la stadification et le traitement du CHC

IV.2.1 Phase A : Extraction des informations médicales à partir de l’image médicale du foie

Pour réaliser la représentation sémantique des images médicales, il est nécessaire tout d’abord de préciser les concepts, les propriétés et les relations qui existent entre les différents acteurs permettant l’achèvement de cette représentation. Dans cette première phase, nous cherchons à rassembler et collecter les informations médicales et relatives à la maladie concernée (CHC) en revenant à des cas réels des patients infectés. Les informations sont tirées également à partir des rapports radiologiques réels. Ces images médicales seront introduites à l’ontologie sous forme de concepts, de propriétés et de relations entre ces concepts. En outre, l’ontologie possède un rôle important pour annoter ces images médicales et pour assurer ultérieurement l’aspect de raisonnement sémantique.

La représentation ontologique des connaissances médicales est utilisée également pour incrémenter et élargir la base de connaissances médicale liée aux maladies hépatiques et surtout aux lésions de type CHC. Plusieurs informations servent à la détection de ces nodules à partir des images IRM. Parmi ces informations, nous citons, la taille du nodule observé, son profil vasculaire, les caractéristiques des signaux T1 et T2 de l’IRM, le taux de l’alpha-ftoprotéine (AFP), la pression veine portale, le score Child Pugh, la fonction du foie, *etc.* Pour enrichir notre approche, nous avons utilisé l’outil *Weasis*⁶⁵ qui permet la visualisation des images DICOM. Il permet également de lire l’entête de ce format et d’extraire les informations relatives au patient. En effet, les informations extraites seront considérées comme l’entrée de l’ontologie. Les figures [III.2](#) et [III.3](#)

65. <https://nroduit.github.io/en/>

exposent respectivement un exemple d'une image hépatique de type IRM et sa visualisation (avec les données relatives au patient) à l'aide du logiciel *Weasis*.

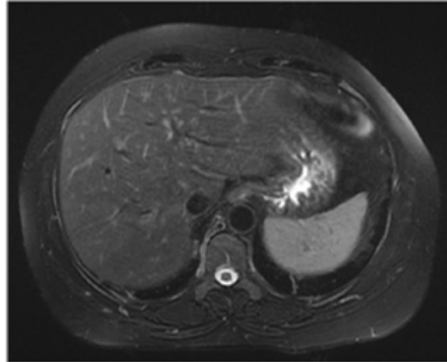


FIGURE IV.2 – Un exemple d'une IRM hépatique

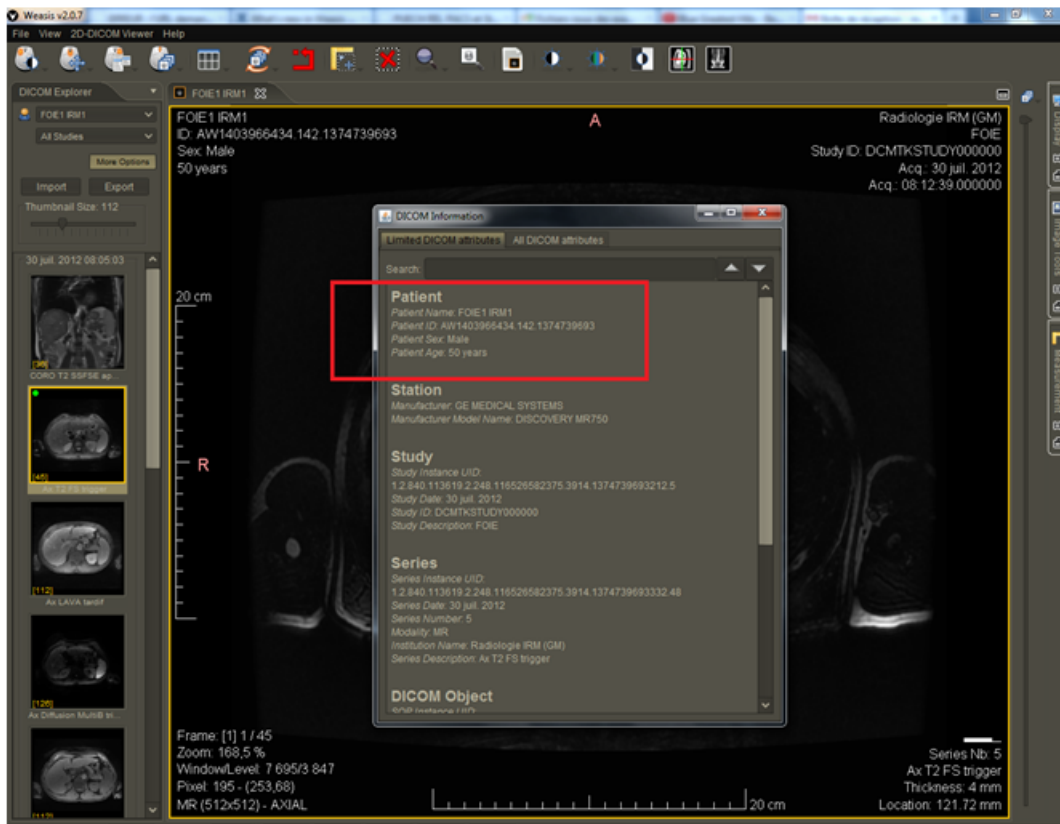


FIGURE IV.3 – Visualisation de l'IRM hépatique *via* Weasis

IV.2.2 Phase B : La représentation sémantique des connaissances médicales

Cette phase offre une description détaillée du cancer de foie en se basant sur une représentation sémantique formelle. Elle génère à la fin notre ontologie médicale (OntHCC) qui sert à la

reconnaissance et la stadification des tumeurs détectées dans les images IRM. Durant cette phase, l'objectif est de mettre l'accent sur le processus de diagnostic du CHC. Nous avons étudié les indications médicales permettant le diagnostic de cette maladie à savoir les facteurs de risques, les symptômes, les interventions médicales possibles, *etc.* Nous avons inclus également des connaissances reliées aux systèmes de stadification permettant d'extraire le stade de chaque tumeur hépatique détectée, en se basant sur les directives de représentation standards et fondamentales des systèmes : BCLC, TNM et Li-RADS.

Nous avons pris ainsi toutes les informations collectées pour construire les concepts, les propriétés et les relations de notre ontologie médicale. Cette ontologie a été ensuite enrichie par des règles d'inférence (SWRL).

Dans la section suivante, nous exposons la construction de notre ontologie en appliquant la méthodologie fondamentale de construction des ontologies « *MethOntology* » [?]. Cette méthodologie assure une représentation ontologique structurée et bien formalisée. Elle est basée sur trois phases principales : i) la conceptualisation qui consiste à définir les concepts du domaine étudié, les relations entre ces concepts, ainsi que des contraintes sur ces concepts ; ii) la formalisation qui consiste à faire l'expression de l'ontologie avec un langage spécifique (les règles SWRL) ; et enfin iii) la phase de validation qui vise à peupler l'ontologie avec les instances. Nous présenterons ainsi la première phase qui consiste à identifier les concepts de l'ontologie. Par la suite, nous identifions tout au long ce chapitre les règles SWRL pour assurer la phase de la formalisation. Pour la validation, elle consiste à comparer les ontologies développées avec les ontologies existantes et à prendre l'avis des experts du domaine. Cette dernière phase est présentée dans le Chapitre VI de ce manuscrit.

IV.2.3 Présentation de l'ontologie OntHCC

Pour la construction de notre modèle ontologique, nous avons consulté diverses ontologies médicales issues de la littérature (par exemple l'ontologie développée par [92]). Nous avons contacté également des experts du domaine à savoir des praticiens hospitaliers (médecins et radiologues). En premier lieu, nous avons réalisé une description médicale globale du CHC en termes de symptômes, facteurs de risques, et type du traitement à appliquer. En effet, nous avons commencé par présenter les concepts génériques et communs entre les maladies hépatiques. En deuxième lieu, nous avons intégré des termes liés à la phase de diagnostic montrant le processus de détection du CHC chez un patient cirrhotique. En troisième lieu, nous avons introduit les concepts modélisant la stadification des nodules hépatiques de type CHC *via* des systèmes de classifications standards qui sont : BCLC, TNM et Li-RADS. Ce prototype contient 49 classes, 78 propriétés, 166 instances, 70 règles SWRL et 1390 axiomes. Des règles SWRL ont été intégrées également pour détecter les lésions CHC à partir des images IRM, pour extraire le stade et pour extraire le traitement à appliquer.

Dans la suite de cette sous-section, nous exposons notre modèle ontologique OntHCC. Nous commençons par présenter les concepts, les propriétés d'objets et les propriétés de type de données et nous finirons par exposer les règles d'inférence SWRL.

IV.2.3.1 Les concepts de l'ontologie OntHCC

La conceptualisation, la première phase de la méthodologie appliquée, caractérise une phase de définition de l'ontologie en termes de concepts ou classes. Dans cette étape, nous présentons les concepts de l'ontologie OntHCC en utilisant Protégé (logiciel développé en Java et dédié pour la création d'ontologies). Notre modèle ontologique développé contient 49 concepts, connectés à travers une dizaine de propriétés d'objets et de données. Le Tableau IV.1 montre une description détaillée de ces concepts.

TABLE IV.1 – Les concepts définis dans l'ontologie OntCHC et leurs significations

Concept	Définition
Liver	Représente l'élément de base du projet qui est le foie, caractérisé par la pression veine portale, le taux de bilirubine, <i>etc.</i>
LiverCancer	Caractérise le type du cancer traité qui est le cancer du foie.
Disease	Présente la maladie.
Cancer	Présente le type de la maladie traitée qui est le cancer.
Hepatocellular-Carcinoma	Précise exactement le type du cancer du foie qui est le carcinome hépatocellulaire CHC.
Nodule	Caractérise la tumeur trouvée dans le foie.
MedicalIntervention	Présente les interventions médicales que suit le médecin pour faire le diagnostic de la maladie.
Symptoms	Présente les symptômes du CHC, par exemple : la jaunisse, la fatigue, <i>etc.</i>
Diagnosis	Rassemble les outils permettant de réaliser le diagnostic du CHC tel que l'IRM, les examens physiques, <i>etc.</i>
Treatment	Présente le type du traitement à appliquer notamment la chirurgie, l'ablation, la chimioembolisation, <i>etc.</i>
References	Présente l'outil utilisé par le médecin ou le radiologue pour détecter le type de la maladie.
RiskFactor	Présente les facteurs de risques qui ont pour rôle d'augmenter le degré de danger du CHC comme l'alcool, le tabac, l'obésité, <i>etc.</i>
Patient	Présente le patient sur lequel se font les opérations médicales du diagnostic et du traitement.
BCLC-System	Présente le système standard de classification des lésions hépatiques. Il prend en entrée plusieurs paramètres comme la taille du nodule, le taux de pression de la veine portale, le score Child-Pugh, <i>etc.</i> , et sa sortie sera automatiquement le stade de la tumeur. Il contient globalement 5 stades.
VeryEarlyStage	Caractérise le stade primaire engendré par le système BCLC, il est noté par A0.
EarlyStage	Caractérise les stades précoces du système BCLC, il peut prendre l'une de ces valeurs : A1, A2, A3, et A4.

IntermediateStage	Présente le stade intermédiaire du système BCLC, il est noté par B
AdvancedStage	Présente le stade avancé du système BCLC. C'est un stade noté par C et montre bien l'augmentation de la maladie dans l'organisme humain.
UltimateStage	C'est le dernier stade du système BCLC, noté par D et montre que la maladie atteint son maximum de propagation dans le corps du patient.
TNM-System	Présente un autre système utilisé pour la classification des tumeurs cancéreuses. Il prend en entrée plusieurs paramètres tels que la taille de la tumeur, <i>etc.</i> , et en sortie il donne le stade de la tumeur. Il peut donner sept stades possibles.
T	Présente le paramètre d'entrée pour le système TNM et s'intéresse à la taille de la tumeur.
N	Présente un autre paramètre du système TNM et s'intéresse à la présence ou l'absence des métastases dans les ganglions lymphatiques régionaux.
M	Présente aussi un paramètre d'entrée pour le système TNM et s'intéresse sur la présence ou l'absence des métastases dans une autre partie du corps différente du foie.
Treatment-Stadification	Présente un autre système de stadification des lésions cancéreuses et classe les stades fournis par les deux autres systèmes de classification en trois stades uniquement qui sont : localisé, régional et avancé.
MedicalImage	Présente l'image médicale du foie, caractérisé par : un identifiant, une taille, une surface et un nombre de blocs. Cette image peut être de type IRM ou de type Scanner
IRM	Présente l'image médicale de type IRM
CT	Présente l'image médicale de type Scanner
Bloc-Image	Présente chaque bloc de l'image médicale, caractérisée par un identifiant, une taille : hauteur et largeur, une position : abscisse et ordonnée, et la nature du tissu s'il est cancéreux ou non.
MedicalImageExam	Présente l'examen médical et se fait à travers une succession d'images médicales.
Li-RADS-Category	Représente la stadification des tumeurs en se basant sur le système Li-RADS.
LR-NC	Représente une catégorie du système Li-RADS, elle caractérise une tumeur non spécifiée à cause d'une imagerie altérée ou incomplète.
LR-1	Représente une catégorie du système Li-RADS, elle caractérise une tumeur certainement bénigne.

LR-2	Représente une catégorie du système Li-RADS, elle caractérise une tumeur probablement bénigne.
LR-3	Représente une catégorie du système Li-RADS, elle caractérise une probabilité intermédiaire de malignité.
LR-M	Représente deux catégories du système Li-RADS, soit LR-4 pour désigner un CHC probable soit LR-5 pour désigner un CHC certain.
LR-TIV	Représente une catégorie du système Li-RADS, elle désigne un envahissement tumoral endoveineux.
Segment	Représente les segments du foie, nous distinguons 8 segments selon Couinaud.

IV.2.3.2 Les propriétés de l'ontologie OntHCC

Dans cette sous-section, nous présentons les propriétés permettant de faire la liaison entre les concepts déjà créés. Ces propriétés servent à la description et l'enrichissement de notre modèle ontologique. Nous distinguons deux catégories de propriétés : les propriétés d'objet et les propriétés de type de données. Les propriétés d'objet permettent de lier les concepts entre eux. Par exemple les concepts LiverCancer et BCLC-System sont liés par la relation « *classifiedInto* ». De même, la relation « *treatedBy* » est utilisée pour lier les concepts LiverCancer et Treatment, « *hasCategory* » est appliqué pour lier les concepts Nodule et LIRADS-Category. Le Tableau IV.2 présente quelques propriétés d'objets utilisées dans notre ontologie.

TABLE IV.2 – Les propriétés d'objets

Relation	Domain	Range
associatedTo	LiverCancer	MedicalImage
composedOf	LiverCancer	Nodule
classified-Into	LiverCancer	BCLC-System, TNM-System
Contains	Liver	LiverCancer
diagnosedBy	LiverCancer	Diagnosis
has-T	LiverCancer	T
has-N	LiverCancer	N
has-M	LiverCancer	M
hasBloc	MedicalImage	Bloc-Image
hasCauses	LiverCancer	RiskFactors
hasStage	LiverCancer	TreatmentStadification
hasSymptoms	LiverCancer	Symptoms
hasTumor	Patient	LiverCancer

relatedTo	MedicalImage	Patient
treatedBy	LiverCancer	Treatment
foundIn	Nodule	MedicalImageExam
IsSequenceOf	MedicalImage	MedicalImageExam

Les propriétés du type de données permettent de décrire les concepts de notre modèle ontologique. Le Tableau IV.3 expose les propriétés du type de données ajoutées dans notre modèle ontologique.

TABLE IV.3 – Les propriétés du type de données

Concept	Propriétés
MedicalImage	ID-Image Image-height Image-widht Image-surface
Liver	hasBiluribine hasFunctionalIndex hasPortalVeinPressure hasScoreChildPugh LiverSize LiverShape LiverEdges
BlocImage	Bloc-ID Bloc-abscissa Bloc-ordinate Bloc-surface Bloc-tissue Bloc-width

Patient	Patient-ID Patient-Age Patient-Name Patient-Sex Patient-Weight isDiabete isObese isAlcoholic hasHereditaryDiseases hasOtherDisease
LiverCancer	hasSpreadBloodVessels hasSpreadLymphNodes hasSpreadOtherPart
Nodule	noduleNumber noduleSize hasVascularProfile hasResultClass hasAFP
MedicalImageExam	hasSignalT1 hasSignalT2

IV.2.3.3 Visualisation de l'ontologie OntHCC

La formalisation de notre modèle ontologique OntHCC a été réalisée en utilisant l'outil Protégé. La définition des concepts, des propriétés et des relations nous ont permis de créer notre modèle ontologique OntHCC. Cette modélisation peut être visualisée à travers des plugins liés à l'outil Protégé. Dans ce cas, nous appliquons le langage OWL pour la construction de notre prototype. Par la suite, nous visualisons cette ontologie *via* le plugin OntoGraf⁶⁶ de Protégé qui permet de visualiser et de naviguer dans des hiérarchies de classes de l'ontologie. Ce plugin donne une

66. <https://protegewiki.stanford.edu/wiki/OntoGraf>

visualisation générale d'un modèle de connaissance particulier. Il offre aussi une lisibilité considérable afin de faciliter la compréhension des concepts créés. La Figure IV.4 montre une partie de notre ontologie *via* le plug-in *OntoGraf*.

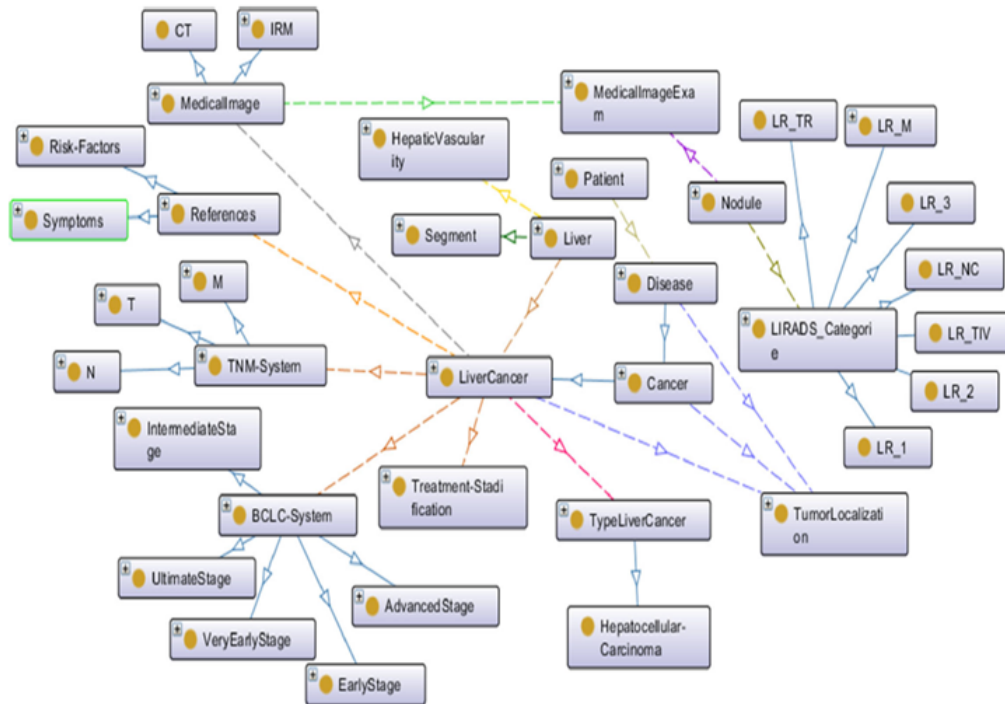


FIGURE IV.4 – Visualisation de l'ontologie OntHCC *via* OntoGraf

IV.2.4 Phase C : Diagnostic du CHC

Cette phase est une validation de l'ontologie créée. Elle sert à assurer l'aspect de raisonnement *via* les moteurs d'inférence. Notre modèle ontologique contient des concepts, des propriétés d'objet et des propriétés de type de donnée pouvant améliorer la définition, la représentation et la stadification des maladies hépatiques de type CHC. Nous distinguons dans la littérature diverses techniques et méthodologies de validation d'une ontologie. Le raisonnement sur les concepts est considéré comme étant la solution la plus performante. Il permet de déduire de nouvelles connaissances et termes aidant à mieux caractériser le diagnostic et la détection des tumeurs CHC. Nous avons utilisé les techniques de raisonnement et d'inférence de l'ontologie pour définir des règles d'inférence.

Dans cette phase, nous incluons les systèmes de stadification comme paramètres d'entrée pour aider à dégager le stade de la tumeur hépatique. Ainsi, tous les paramètres aidant au bon fonctionnement de ces systèmes seront inclus aussi dans l'ontologie. En effet, le système BCLC nécessite de raisonner sur : l'indice fonctionnel, le score Child-Pugh, le taux de Bilirubine, la pression veine portale, le nombre de nodules cancéreux, et la(es) taille(s) de(s) nodule(s). Le système TNM nécessite la connaissance des paramètres suivants : la taille du nodule, et la propagation ou non de

la tumeur dans les autres parties du corps. Pour la classification du traitement, nous avons besoin uniquement de savoir le stade de la tumeur selon BCLC ou TNM.

IV.2.5 Raisonnement dans notre modèle ontologique

La définition des règles d'inférence SWRL nécessite le choix d'un langage de règles, d'un outil d'édition des règles et d'un raisonneur. Dans notre travail, nous appliquons le langage SWRL. L'utilité de ces règles SWRL est expliquée par le fait d'exprimer sous forme de langage en premier ordre des relations ou des liens entre les concepts déjà définis dans l'ontologie. Dans ce travail, les règles SWRL sont utilisées pour la détection du CHC à partir d'images IRM, l'extraction du stade de cette tumeur et l'aide à la prise de décision pour le type de traitement à appliquer. Dans la suite de cette section, nous allons montrer les différents types de règles utilisées dans notre modèle ontologique et le rôle de chacune de ces règles :

Règle 1 : Exemple d'une règle SWRL pour la détection du CHC à partir de l'image médicale

```
Nodule(?x) ^ MedicalImage(?y) ^ MedicalImageExam(?z) ^ IsSequenceOf(?y, ?z) ^  
foundIn(?x, ?z) ^ noduleSize(?x, ?s) ^ swrlb:greaterThan(?s, "2") ^ hasVascularProfile(?x,  
"typical") ^ usesPDC(?z, "Oui") -> hasResultClass(?x, "HCC")
```

La Règle 1 s'explique comme suit : Après l'injection d'un produit de contraste, l'observation d'une masse hépatique qui a la taille supérieure à 2 cm dans l'image médicale et qui contient un profil vasculaire typique est classée comme CHC.

Règle 2 : Exemple d'une règle SWRL relative au système BCLC

```
Liver(?x) ^ LiverCancer(?y) ^ Nodule(?n) ^ composedOf(?y, ?n) ^ noduleSize(?n, ?s) ^  
swrlb:lessThan(?s, 2) ^ noduleNumber(?n, 1) ^ contains(?x, ?y) ^ hasBilirubine(?x, "Normal") ^  
hasPortalVeinPressure(?x, "Normal") ^ hasFunctionalIndex(?x, 0) ^ hasScoreChildPugh(?x,  
"NonApplicable") -> classifiedInto(?y, A0)
```

La Règle 2 vise à préciser le stade d'un nodule hépatique détecté dans une image médicale. Dans ce cas, nous traitons le stade primaire A0 qui satisfait les conditions suivantes : une seule masse hépatique de taille inférieure à 2 cm. Ici le foie est caractérisé par un taux normal de Bilirubine et un taux normal de la pression veine portale.

Règle 3 : Exemple d'une règle relative au système TNM

```
LiverCancer(?x) ^ has-T(?x, T1) ^ has-N(?x, n0) ^ has-M(?x, M0) ^ hasSpreadBloodVessels(?x, "Non") ^ hasSpreadLymphNodes(?x, "Non") ^ hasSpreadOtherPart(?x, "Non") -> classifiedInto(?x, Stage-I)
```

La Règle 3 est appliquée afin de présenter le stade primaire du système TNM, c'est le stade I. Il est caractérisé par la présence d'un nodule de petite taille qui ne s'est pas encore propagé dans le corps.

Règle 4 : Exemple d'une règle relative à localiser une tumeur hépatique

```
LiverCancer(?x) ^ classifiedInto(?x, A0) -> hasStage(?x, Localized)
```

La Règle 4 permet de dire que si une tumeur classifiée comme stade primaire (A0) selon le système BCLC, alors elle sera classifiée comme 'Localized' selon la stadification du traitement.

Règle 5 : Exemple d'une règle SWRL relative au type de traitement à appliquer

```
LiverCancer(?x) ^ classifiedInto(?x, A0) -> treatedBy(?x, Surgery)
```

La règle 5 montre que si la tumeur hépatique observée en IRM est en stade primaire A0, alors le type du traitement qui pourrait être appliqué est la chirurgie. Toutes les autres règles SWRL sont présentées dans l'annexe A.

IV.3 Développement d'une ontologie pour la modélisation des rapports radiologiques du CHC (MROnt)

Dans le cadre de ces travaux de thèse, nous avons proposé une deuxième approche ontologique (voir Figure [IV.5](#)) pour la modélisation de l'information médicale contenue dans les rapports radiologiques, dans le cadre du diagnostic et de suivi de tumeurs du foie. L'interprétation radiologique décrit l'état du foie (*e.g.*, dysmorphie, contours irréguliers) et de la lésion hépatique (taille, localisation, nombre de nodules). De plus, notre système vérifie s'il existe une anomalie liée au système hépatique et autres organes, y compris les observations de pancréas ou des reins par exemple. Il contient également d'autres données liées aux signaux IRM tels que T1, T2, *etc.*, qui peuvent être visualisés après l'injection du produit de contraste. Les rapports médicaux et leurs formats textuels non structurés peuvent entraîner des données médicales ambiguës et complexes. Afin de minimiser cette complexité, nous proposons l'ontologie MROnt (« *Medical Reports Ontology* »). Cette ontologie contient les concepts principaux modélisant ces rapports médicaux, également les propriétés et les relations. MROnt prend en considération les rapports radiologiques, issus des acquisitions IRM notamment, des patients atteints d'une maladie hépatique (essentiellement le CHC).

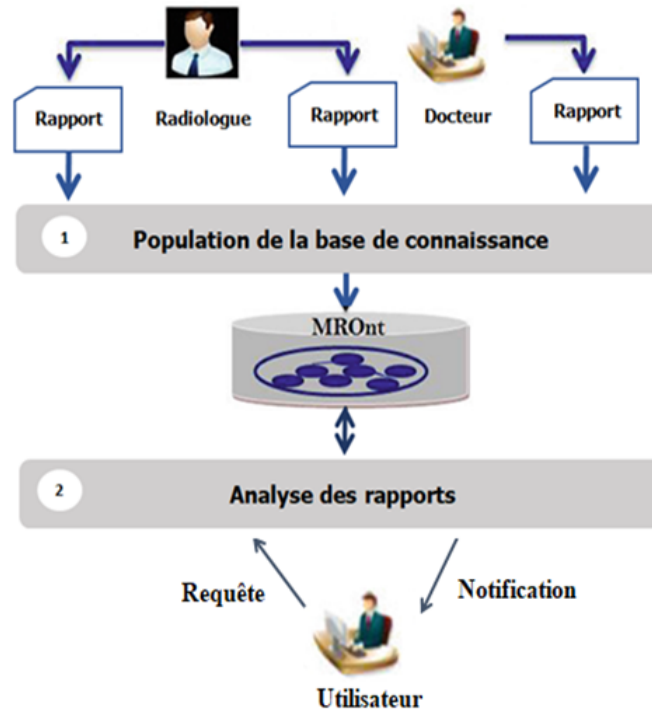


FIGURE IV.5 – Ontologie pour la modélisation des rapports radiologiques du CHC (MROnt)

Notre approche permet un meilleur échange de connaissances et de données cliniques. Nous avons introduit des aspects sémantiques pour analyser les informations contenues dans les rapports radiologiques. En outre, nous avons inclus des règles de raisonnement pour extraire des observations cliniques à partir de rapports de cas réels de tumeurs du foie. Pour ce faire, nous avons utilisé le *framework* Protégé et le langage OWL. Nous avons commencé par extraire les observations radiologiques pour peupler notre ontologie. Puis, nous avons appliqué les règles SWRL afin d’extraire le type de la lésion détectée. Le système fournit à la fin une décision concernant le type du nodule (CHC, kyste ou autres types de maladies hépatiques). Nous avons implémenté une interface graphique permettant d’interroger le prototype MROnt à travers des règles SWRL et SPARQL. Le système donne la liste des observations contenues dans un rapport radiologique en se basant sur l’identifiant du patient. Il fournit ensuite une proposition de décision clinique : la tumeur détectée est de type CHC ou non. MROnt est composée de deux phases :

IV.3.1 Phase 1 : Population de la base de connaissances

La première phase permet la modélisation des rapports radiologiques et vise à rendre ces rapports compréhensibles par la machine. L’objectif de cette phase est de modéliser les rapports radiologiques à partir de l’ontologie MROnt. Elle vise à rendre les rapports médicaux des patients interprétables par la machine. Nous avons extrait les concepts à partir des rapports radiologiques réels.

IV.3.2 Phase 2 : Analyse des rapports radiologiques

La deuxième phase est celle de l'analyse des rapports basée sur un système d'analyse permettant aux utilisateurs (médecins, radiologues ou techniciens de laboratoire) d'être informés de l'évolution de l'état du patient. Cette phase propose un système d'analyse permettant aux utilisateurs (médecins, radiologues ou techniciens de laboratoire) d'être informés sur l'évolution de l'état des patients. Une description de chacune de ces deux phases est mentionnée dans les sous-sections suivantes. L'analyse des rapports radiologiques est une étape essentielle pour soutenir les systèmes de décision médicale. L'objectif est de faciliter le traitement de divers rapports médicaux et de fournir des informations structurées aux utilisateurs. Les règles d'inférence sont introduites pour assurer une bonne interrogation de l'ontologie MROnt. Ainsi, les requêtes SQWRL sont intégrées comme des opérations SQL utilisées pour exploiter les connaissances induites par les règles SWRL.

Dans la section suivante, nous présentons l'ontologie MROnt tout en exposant les concepts, les propriétés et les règles d'inférence.

IV.3.3 Présentation de l'ontologie MROnt

L'ontologie proposée vise à structurer et standardiser la forme des rapports radiologiques réels. Durant le processus de développement de notre modèle ontologique, nous avons utilisé le *framework* Protégé et le langage OWL. MROnt contient 42 concepts, 29 propriétés (18 propriétés d'objets et 11 propriétés de données), 10 règles SWRL, 12 instances et 172 axiomes. Pour la partie évaluation, nous avons utilisé 28 rapports radiologiques réels anonymisés, pour des patients atteints de CHC. Ces rapports sont issus d'examens réalisés dans le CHU de Clermont-Ferrand. Dans cette section, nous présentons une vue globale sur l'ontologie MROnt en terme de concepts, relations et règles de raisonnement.

IV.3.3.1 Les concepts de l'ontologie MROnt

Le modèle MROnt est composé de concepts requis pour aider à la détection de la maladie pendant le processus de diagnostic. Par exemple le concept de « *Radiologist* » désigne la personne qui réalise l'examen IRM et note ses descriptions sous forme d'observations « *Observation* ». Le concept « *Report* » désigne l'examen médical qui contient toutes les informations sur l'état du patient. Ce concept est lié au concept « *Radiologist* » par la relation « *elaboratedBy* ». « *Patient* » désigne la personne qui subit un examen médical et qui suit un type de traitement ou une chirurgie. Ce concept fait référence à tous les patients et leurs informations (par exemple le nom, le sexe, l'âge, *etc.*). Le concept « *MedicalImage* » décrit l'image médicale qu'étudie le radiologue pour extraire et noter ses observations médicales. Cette image fournit une description visuelle complète du foie. Dans ce travail, nous nous sommes intéressés aux images médicales hépatiques de type IRM. « *MedicalImage* » est connecté au « *Patient* » à travers la relation « *hasMedicalImage* ». Le concept « *HepaticLesion* » présente la tumeur hépatique étudiée. Il caractérise les anomalies à l'intérieur du tissu hépatique. Le concept « *Nodule* » décrit la masse hépatique détectée dans une image IRM. Il se caractérise essentiellement par sa forme anormale. La relation « *ComposedOf* »

est utilisée pour relier ces deux derniers concepts. Sans oublier qu'une lésion hépatique peut être composée d'un ou plusieurs nodules. D'autres propriétés d'objet sont définies pour effectuer des liens entre les concepts et les individus. Dans notre modèle, la relation « *observedIn* » est utilisée pour lier les concepts « *Nodule* » et « *Rapport* ». La relation « *hasReport* » connecte le concept « *Patient* » au concept « *Signal* ». Le Tableau IV.4 montre une description des différents concepts utilisés.

TABLE IV.4 – Les concepts définis dans l'ontologie MROnt et leurs significations

Concept	Définition
Patient	Désigne la personne atteinte d'une maladie hépatique (CHC ou autre) ou qui reçoit un traitement particulier. Il est identifié par un identifiant (ID), un sexe, un âge, une date de naissance, <i>etc.</i>
Radiologist	Représente le radiologue qui réalise l'examen des images médicales et note les observations après chaque examen. Il est identifié par un nom.
Doctor	Il analyse les examens IRM et les rapports radiologiques.
Report	Représente un texte noté par le radiologue pour sauvegarder des informations sur l'état du patient, la maladie détectée et le traitement appliqué.
Protocol	Identifie les techniques utilisées par le radiologue pour réaliser l'examen médical, comme le signal T1, T2, T2 FS, <i>etc.</i>
HepaticLesion	Représente la masse hépatique détectée dans les séquences IRM. Elle peut être un CHC, un kyste ou un autre type de lésion. Il se caractérise par son stade, sa taille et sa localisation.
Type	Représente le type de cancer du foie détecté lors de l'observation médicale. Il peut être un CHC, un kyste ou un autre type (<i>e.g.</i> , angiome).
HCC	Désigne une lésion hépatique maligne. Un CHC s'est développé dans le foie et attaque rapidement ses cellules (les hépatocytes).
Cyst	Définit un type de cancer du foie bénin.
Other	Désigne les maladies hépatiques autres que le CHC et le kyste.
Nodule	Caractérise la lésion hépatique détectée dans l'image médicale. Elle est identifiée par une taille. La lésion hépatique peut être constituée par un ou plusieurs nodules
Segment	Représente les segments du foie selon Couinaud. On distingue 8 segments au total.
Observation	Représente les interprétations radiologiques notées par le radiologue ou le médecin durant le processus de diagnostic.
Stage	Fournit des informations sur le stade de la lésion hépatique détectée.
Size	Représente la taille de la lésion hépatique détectée.

Treatment	Représente le traitement à appliquer.
InjectionAC	Fait référence au produit de contraste qu'on injecte au patient avant l'examen IRM.
Date	Désigne la date à laquelle le patient a passé l'examen médical.
Material	Représente le matériel utilisé pour réaliser l'examen IRM
Quality	Représente la qualité de l'image IRM
Phase	Représente la phase de l'examen IRM (par exemple Portal, <i>etc.</i>).

IV.3.3.2 Les propriétés de l'ontologie MRont

Les propriétés sont intégrées pour assurer des liaisons entre les concepts déjà définies dans notre modèle ontologique. Le Tableau IV.5 visualise les propriétés créées dans l'ontologie MRont.

TABLE IV.5 – Les propriétés définies dans l'ontologie MRont

Relation	Domain	Range
composedOf	HepaticLesion	Nodule
contains	MedicalImage	Nodule
describes	Observation	MedicalImage
elaboratedBy	Report	Radiologist
hasDate	Report	Date
hasFeatures	MedicalImage	Features
hasInjection	Patient	InjectionCA
hasLesion	Patient	HepaticLesion
hasMaterial	MRI	Material
hasMedicalImage	Patient	MedicalImage
hasObservation	Report	Observation
hasPhase	MedicalImage	Phase
hasProtocol	Report	Protocol
hasQuality	MedicalImage	Quality
hasReport	Patient	Report
hasSize	Nodule	Size
hasStage	Nodule	Stage
hasTreatment	HepaticLesion	Treatment
hasType	HepaticLesion	Type
LocatedIn	Nodule	Segment
prescribe	Doctor	Report

IV.3.3.3 Visualisation de l'ontologie MROnt

La Figure IV.6 expose une visualisation de l'ontologie MROnt *via* le plugin Ontograp de l'outil Protégé.

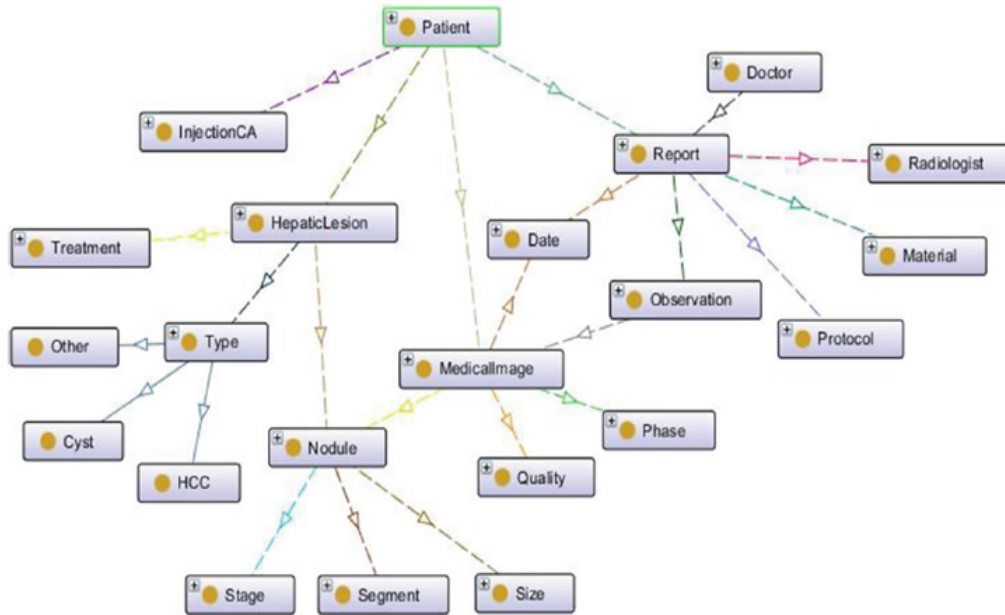


FIGURE IV.6 – Visualisation de l'ontologie MROnt *via* le plugin OntoGraf

Dans la section suivante, nous étudions des exemples des règles utilisées dans notre modèle ontologique.

IV.3.4 Les règles SWRL

Le module de raisonnement est appliqué également dans cette approche ontologique pour assurer l'aspect d'inférence et la déduction de nouveaux termes. Nous avons fait quelques simulations sur Protégé pour nous assurer du bon fonctionnement des règles d'inférence. Ainsi, nous avons peuplé notre ontologie par des instances pour enrichir le prototype développé. Nous avons défini également différentes règles d'inférence (règles qui fondent le processus de déduction de nouvelles connaissances) permettant l'analyse des rapports réels. Par exemple, dans la Règle 1, nous avons utilisé le langage SQWRL pour extraire les informations sur le nodule détecté dans l'image médicale. Plus précisément, nous nous sommes concentrés sur la taille, la localisation dans un segment du foie, et le stade du nodule.

Règle 1 :

$ \begin{aligned} & Patient(?p) \wedge HepaticLesion(?h) \wedge hasLesion(?p,?h) \wedge Nodule(?n) \wedge \\ & composedOf(?h,?n) \wedge Size(?size) \wedge hasSize(?n,?size) \wedge Segment(?segment) \wedge \\ & locatedIn(?n,?segment) \wedge Stage(?stage) \wedge hasStage(?n,?stage) \rightarrow sqwrl : \\ & select(?size) \wedge sqwrl : select(?segment) \wedge sqwrl : select(?stage) \end{aligned} $

La Règle 2 est utilisée pour évaluer l'état du nodule détecté. En particulier, nous vérifions la liste d'observations en utilisant « *swrlb : equal* » (c'est un *built-in* intégré dans le langage SWRL pour ajouter plus de sens sur ce langage). Si toutes les observations requises sont vérifiées, nous pouvons déduire un cancer du foie de type CHC et dans ce cas l'instance de nodule est classée comme « CHC ».

Règle 2 :

$ \begin{aligned} & Patient(?p) \wedge HepaticLesion(?h) \wedge hasLesion(?p,?h) \wedge \\ & Nodule(?n) \wedge composedOf(?h,?n) \wedge MedicalImage(?i) \wedge \\ & contains(?i,?n) \wedge describes(?a,?i) \wedge Observation(?a) \wedge swrlb : \\ & equal(?a,"dysmorphic liver") \wedge Observation(?b) \wedge swrlb : \\ & equal(?b,"hypertrophy segment I") \wedge describes(?b,?i) \wedge Observation(?c) \wedge \\ & swrlb : equal(?c,"hypertrophy segment II") \wedge describes(?c,?i) \wedge \\ & Observation(?d) \wedge swrlb : equal(?d,"micronodular architecture") \wedge \\ & describes(?d,?i) \wedge Observation(?e) \wedge swrlb : equal(?e,"no portal dilation") \wedge \\ & describes(?e,?i) \wedge Observation(?f) \wedge swrlb : equal(?f,"no portal thrombose") \wedge \\ & describes(?f,?i) \wedge Observation(?g) \wedge swrlb : equal(?g,"portal washout") \wedge \\ & describes(?g,?i) \wedge Observation(?j) \wedge swrlb : equal(?j,"hypersignal T1") \wedge \\ & describes(?j,?i) \rightarrow HCC(?h) \wedge ClassifiedAs(?n,?h) \end{aligned} $

Afin de montrer le résultat d'exécution de la Règle 2, nous exposons dans la Figure [IV.7](#) l'évolution de notre modèle ontologique MRont avant et après exécution de la Règle 2. Nous pouvons déterminer le type de la lésion hépatique (CHC dans notre cas) et créer une nouvelle propriété d'objet « *ClassifiedAs* » entre le nodule étudié et l'instance « CHC ».

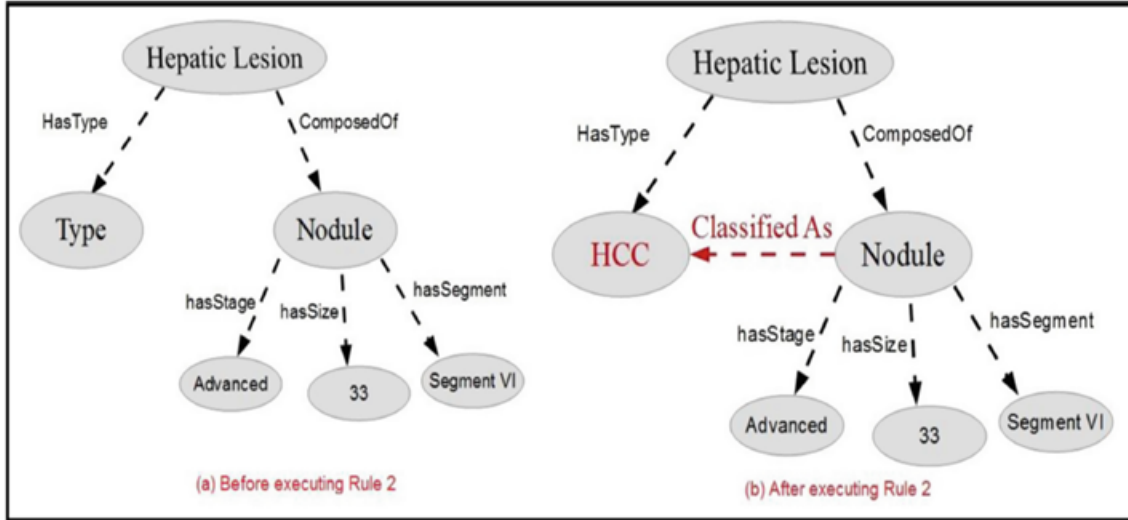


FIGURE IV.7 – Processus de changement d'état et classification du nodule détecté

La Règle 3 permet de déduire que le type du nodule du patient n'est pas un CHC. Ceci s'explique par le fait que les observations extraites ne justifient pas un cancer du foie.

Règle 3 :

$$\begin{aligned}
 & Patient(?p) \wedge HepaticLesion(?h) \wedge hasLesion(?p, ?h) \wedge Nodule(?n) \wedge \\
 & composedOf(?h, ?n) \wedge MedicalImage(?i) \wedge contains(?i, ?n) \wedge describes(?a, ?i) \wedge \\
 & Observation(?a) \wedge swrlb : equal(?a, "dysmorphic liver") \wedge Observation(?b) \wedge \\
 & swrlb : equal(?b, "periferic rehaussement") \wedge describes(?b, ?i) \wedge \\
 & Observation(?c) \wedge swrlb : equal(?c, "no arterialization") \wedge describes(?c, ?i) \wedge \\
 & Observation(?d) \wedge swrlb : equal(?d, "micronodular architecture") \wedge \\
 & describes(?d, ?i) \wedge Observation(?e) \wedge swrlb : equal(?e, "no hepatic sus") \wedge \\
 & describes(?e, ?i) \wedge Observation(?f) \wedge swrlb : \\
 & equal(?f, "no portal thrombose") \wedge describes(?f, ?i) \wedge Observation(?g) \wedge swrlb : \\
 & equal(?g, "no cliac abnormalities") \wedge describes(?g, ?i) \wedge Observation(?h) \wedge \\
 & swrlb : equal(?h, "hypersignal T2") \wedge describes(?h, ?i) \wedge Observation(?j) \wedge \\
 & swrlb : equal(?j, "hypersignal FAT SAT") \wedge describes(?j, ?i) \wedge \\
 & Observation(?k) \wedge swrlb : equal(?k, "hypersignal diffusion") \wedge \\
 & describes(?k, ?i) \rightarrow Other(?h) \wedge ClassifiedAs(?n, ?h)
 \end{aligned}$$

La Règle 4 fournit des informations sur l'état du nodule par rapport à l'examen IRM précédent. Elle prend en compte les observations des examens précédents pour le même patient. Par la suite, elle génère des détails sur la progression des lésions (par exemple, changement de taille, apparition d'une nouvelle tumeur, détection de stade et extraction de type de lésion). Dans l'exemple ci-dessous, nous avons créé la propriété «value» montrant que le nodule CHC détecté a subi un changement de taille de 20 mm à 22 mm.

Règle 4 :

$\begin{aligned} & Patient(?p) \wedge HepaticLesion(?h) \wedge hasLesion(?p, ?h) \wedge Nodule(?n) \wedge \\ & composedOf(?h, ?n) \wedge MedicalImage(?i) \wedge contains(?i, ?n) \wedge describes(?a, ?i) \wedge \\ & Observation(?a) \wedge swrlb : equal(?a, "dysmorphic liver") \wedge Observation(?b) \wedge \\ & swrlb : equal(?b, "periferic rehaussement") \wedge describes(?b, ?i) \wedge \\ & Observation(?c) \wedge swrlb : equal(?c, "no arterialization") \wedge describes(?c, ?i) \wedge \\ & Observation(?d) \wedge swrlb : equal(?d, "micronodular architecture") \wedge \\ & describes(?d, ?i) \wedge Observation(?e) \wedge swrlb : equal(?e, "no hepatic sus") \wedge \\ & describes(?e, ?i) \wedge Observation(?f) \wedge swrlb : \\ & equal(?f, "no portal thrombose") \wedge describes(?f, ?i) \wedge Observation(?g) \wedge swrlb : \\ & equal(?g, "no cliac abnormalities") \wedge describes(?g, ?i) \wedge Observation(?h) \wedge \\ & swrlb : equal(?h, "hypersignal T2") \wedge describes(?h, ?i) \wedge Observation(?j) \wedge \\ & swrlb : equal(?j, "hypersignal FAT SAT") \wedge describes(?j, ?i) \wedge \\ & Observation(?k) \wedge swrlb : equal(?k, "hypersignal diffusion") \wedge \\ & describes(?k, ?i) \wedge hasSize(?n, ?v) \wedge value(?v, 20) \wedge ClassifiedAs(?n, HCC) \rightarrow \\ & hasSize(?n, ?v) \wedge value(?v, 22) \end{aligned}$

IV.4 Conclusion

Dans ce chapitre, nous avons présenté en premier lieu le modèle ontologique permettant l'assistance de détection des nodules CHC à partir des images IRM ainsi la stadification de ces nodules en exploitant les systèmes standards de classification, à savoir BCLC, TNM et Li-RADS. En deuxième lieu, nous avons présenté un modèle ontologique permettant l'analyse et la représentation des connaissances médicales incluses dans des rapports radiologiques en se basant sur des rapports réels des patients infectés par le CHC. Une formalisation détaillée de ces deux modèles OntHCC et MROnt *via* le langage OWL a été expliquée tout en exposant leur richesse sémantique. La consistance des ontologies développées a été approuvée par les règles d'inférence implémentées par le langage SWRL. Nous avons montré également l'intérêt de leur utilisation avec des simulations effectuées sur Protégé.

Dans le chapitre suivant, nous étudions le couplage entre les ontologies et les techniques d'apprentissage pour la classification des images médicales. Nous allons présenter notre approche basée sur le *deep learning* pour la détection des tumeurs de type CHC à partir de séquences IRM. De plus, en exploitant les ontologies développées, nous décrivons leur efficacité pour la détection des nodules CHC.

APPROCHE DE CLASSIFICATION D'IMAGES IRM BASÉE SUR LE COUPLAGE ONTOLOGIES-DEEP LEARNING

Sommaire

V.1 Introduction	96
V.2 Approche de détection et de localisation des tumeurs hépatiques	96
V.2.1 Phase de prétraitement et découpage des images IRM en petits fragments	97
V.2.2 Segmentation du foie	98
V.2.3 Récupération des images d'un fichier format RAW	100
V.2.4 Extraction et ajustement des images IRM/CT	101
V.2.5 Création de la base de données	101
V.3 Description de l'algorithme deep learning	102
V.3.1 Les paramètres de l'algorithme <i>deep learning</i>	102
V.3.2 Parallélisation de prétraitement	103
V.3.3 Phase de prédiction	104
V.3.4 Validation des résultats	104
V.4 Intégration d'ontologies dans la classification des tumeurs hépatiques	104
V.4.1 Les ontologies appliquées	105
V.4.2 Collection des labels	106
V.4.3 Création d'un CNN multi-label	107
V.5 Conclusion	110

V.1 Introduction

L'apprentissage automatique est une méthode d'analyse de données qui sert à l'automatisation et à la création de modèles analytiques tout en permettant la génération de données prédictives. Il est intégré actuellement dans diverses applications de différents domaines à savoir la reconnaissance des formes, la biologie, l'optimisation, le traitement de langage naturel ou la reconnaissance faciale. Dans le contexte médical, il est devenu un outil pour l'analyse et le traitement de grande masse de données médicales vu sa performance pour la résolution des problèmes complexes. Dans ce chapitre, nous présentons une première approche basée sur le *deep learning* permettant la détection et localisation des tumeurs hépatiques. Nous commençons en premier lieu par présenter notre algorithme parallèle de prétraitement permettant de fournir les informations nécessaires pour le processus d'apprentissage. Il prend comme entrée des images médicales IRM et fournit en sortie des fragments de tailles identiques. Par la suite, nous appliquons cet algorithme pour la détection des lésions hépatique de type CHC. En deuxième lieu, nous intégrons les ontologies dans le processus de classification des images IRM pour élargir la base de connaissances et améliorer la phase d'apprentissage.

V.2 Approche de détection et de localisation des tumeurs hépatiques

La détection des tumeurs du foie nécessite un processus de diagnostic primaire en utilisant obligatoirement les images médicales (par exemple IRM ou TDM).

Dans cette section, nous proposons un algorithme d'apprentissage automatique basé sur l'architecture CNN pour détecter les lésions CHC à partir des séquences DCE-MRI (« *Dynamic contrast-enhanced magnetic resonance imaging* »). L'approche proposée (voir Figure [V.1](#)) offre un algorithme basé sur le *deep learning* permettant la détection et la localisation du CHC visualisé dans une image IRM. Notre approche se compose principalement de quatre phases : 1) une phase de prétraitement 2) une phase de préparation des fragments 3) le développement de l'algorithme *deep learning*, et 4) une phase de validation. Dans les sous-sections suivantes, nous détaillons chacune de ces phases.

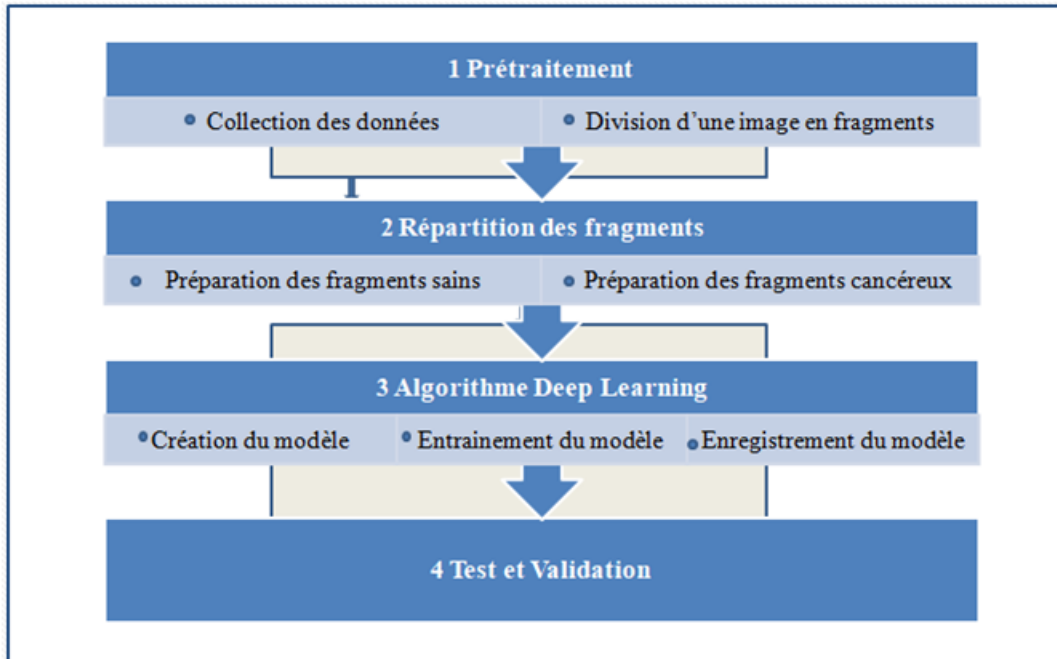


FIGURE V.1 – Approche de détection et de localisation de la tumeur de foie

V.2.1 Phase de prétraitement et découpage des images IRM en petits fragments

Le prétraitement est une phase de préparation des données médicales. Cette première phase est constituée à son tour de plusieurs étapes assurant la manipulation des images médicales et la préparation des bases de données. En premier lieu, nous commençons par réaliser une segmentation du foie, puis nous découpons chaque image en petits fragments de tailles identiques pour pouvoir localiser la tumeur. Chaque fragment est classé en fragment sain ou cancéreux à l'aide d'un algorithme de *deep learning*. L'approche proposée étudie l'évolution des images du foie dans trois phases différentes liées à l'injection du produit de contraste lors de l'examen IRM (voir Figure [V.2](#)) :

- Phase 1 : avant l'injection d'un produit de contraste
- Phase 2 : entre 30 s et 1 min après injection
- Phase 3 : phase portale ou tardive (> 1 min) en fonction des séquences

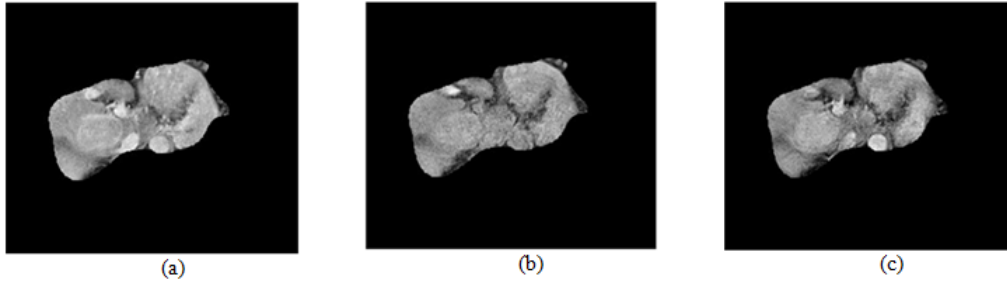


FIGURE V.2 – Les trois phases IRM appliquées (a) phase 1 : avant injection (b) entre 30 s et 1 min après injection (c) phase 3 : phase portale ou tardive (> 1 min) en fonction des séquences

La Figure [V.4](#) visualise des exemples d’images reliées aux différents patients.

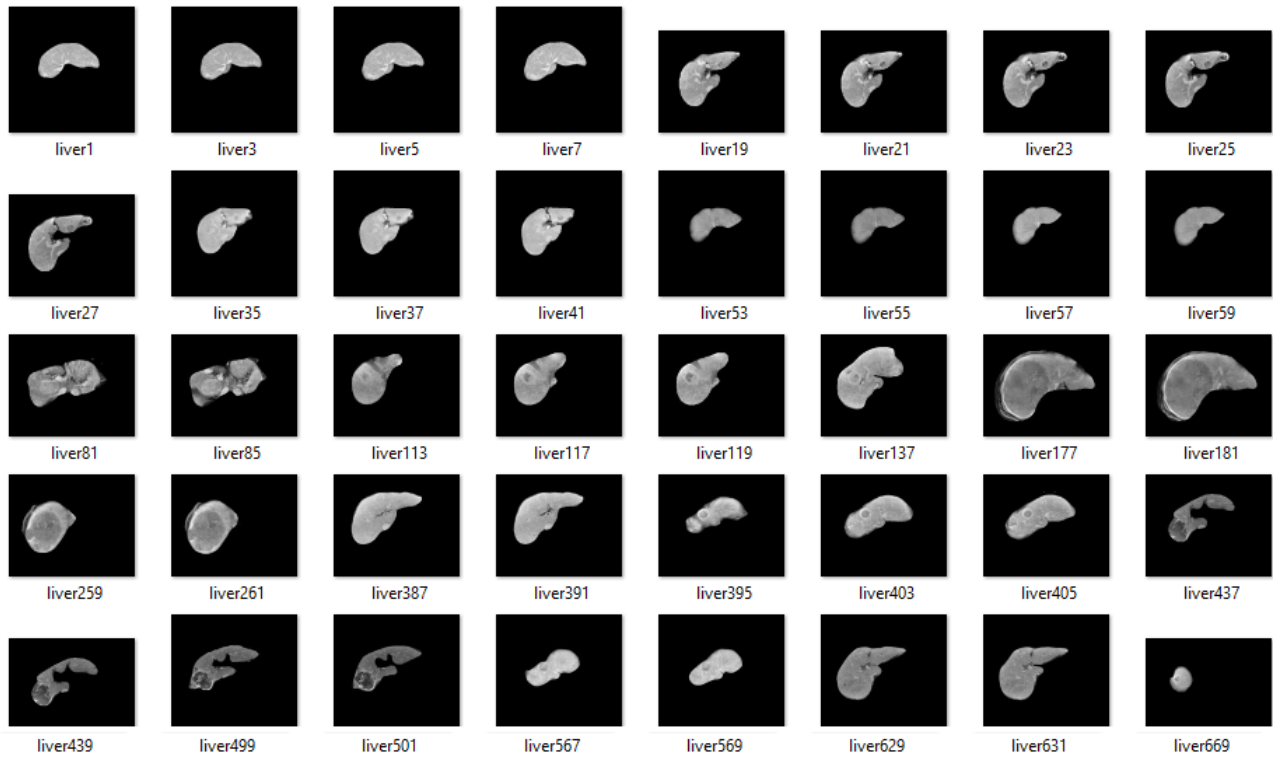


FIGURE V.3 – Les images IRM des différents patients

V.2.2 Segmentation du foie

La segmentation du foie sur des IRM abdominales est une tâche importante pour faciliter la détection des lésions hépatiques. Elle permet également d’aider au diagnostic de pathologies hépatiques telles que la cirrhose. Cette étape consiste à faire l’extraction du foie à partir de l’examen IRM/TDM. Cette partie de projet est implémentée en collaboration avec un autre membre de

l'équipe, Marie-Ange Lebre, qui a développé l'algorithme de segmentation [146]. La Figure V.5 visualise les étapes de segmentation du foie dans une image IRM.

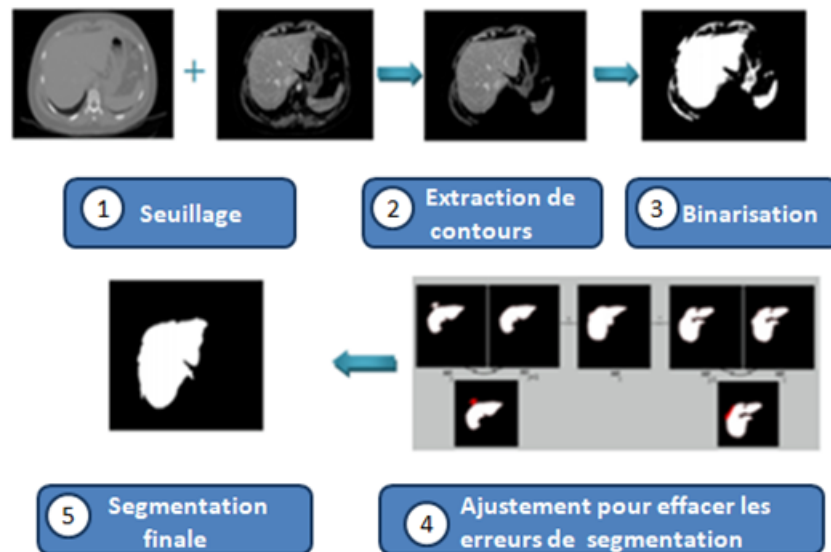


FIGURE V.4 – Les étapes de segmentation du foie dans une image IRM

En résumé, les étapes de la méthode de segmentation suivie sont :

1. Seuillage (ou « *Thresholding* ») : C'est une technique utilisée pour remplacer l'intensité d'un ou plusieurs pixels dans l'image par un seuillage pour conserver les pixels ayant une intensité supérieure à un seuil donné.
2. Extraction de contours : Cette technique consiste à repérer les points de l'image qui correspondent à un changement local significatif de l'intensité.
3. Binarisation : Cette étape caractérise une segmentation binaire (0/1) obtenue à partir des étapes précédentes par une approche *level-set*.
4. Ajustement : Cette étape est appliquée pour effacer les erreurs de segmentation entre différentes coupes successives de l'image traitée.

La Figure V.5 présente deux exemples de segmentations du foie en première phase des images IRM après l'injection du produit de contraste.

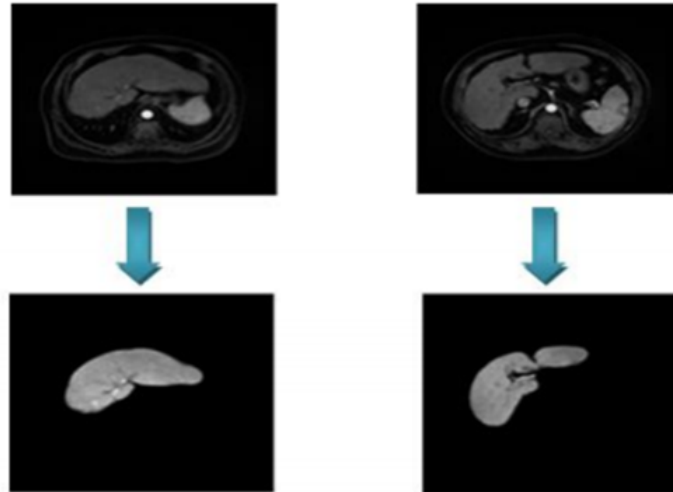


FIGURE V.5 – Exemples d’examens IRM avant et après la segmentation du foie

V.2.3 Récupération des images d’un fichier format RAW

Dans cette partie, nous nous intéressons aux différentes méthodes permettant de découper les images IRM et construire les bases de données nécessaires pour l’entraînement de l’algorithme de *deep learning*. Dans notre travail, nous avons travaillé plus précisément des images de profondeurs 16 bits. La visualisation de ces images nécessite l’intervention d’un logiciel externe. Nous avons choisi ImageJ⁶⁷ (« *Image Processing and Analysis in Java*») pour ouvrir les images codées sur 16 bits. La Figure V.6 montre une visualisation d’une image IRM à partir du logiciel ImageJ.

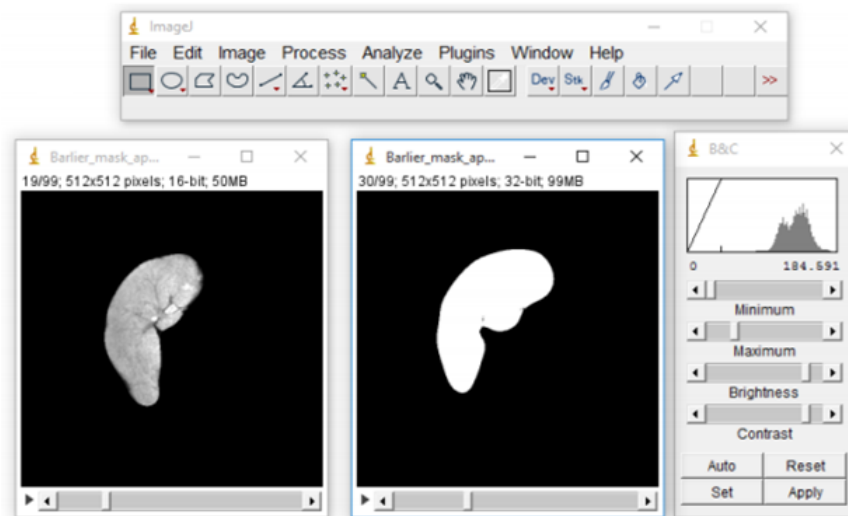


FIGURE V.6 – Visualisation d’une image IRM à partir du logiciel ImageJ

67. <https://imagej.nih.gov/ij/index.html>

V.2.4 Extraction et ajustement des images IRM/CT

Après la récupération des images en format « RAW » et en trois phases (phase 1, phase 2 et phase 3), nous convertissons chaque image récupérée sous la forme d'une matrice contenant les différentes valeurs de pixels. Cette matrice sera parcourue par la suite afin de fixer la première et la dernière ligne qui contiennent des pixels n'appartenant pas au fond (valeur 0). Nous visons dans cette étape à extraire le foie en réduisant le fond, afin de concentrer le traitement qui suit sur cette région d'intérêt. Aussi, elle est nécessaire pour minimiser les erreurs de classification liées à la présence du fond. Pour garder la position des lésions hépatiques dans l'image médicale, nous découpons chaque image ajustée en un ensemble de fragments de tailles identiques en prenant en considération le recouvrement des fragments voisins (voir Figure V.7). Nous distinguons ainsi quatre cas en se basant sur les dimensions de l'image et des fragments.

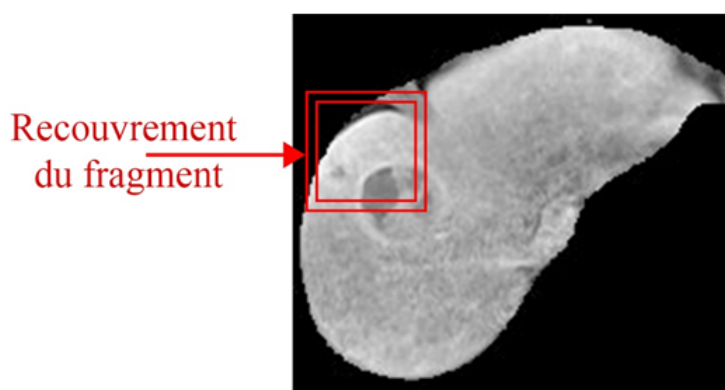


FIGURE V.7 – Visualisation d'un fragment avec recouvrement

Les images utilisées dans notre travail possèdent des tumeurs de différentes tailles. Nous distinguons deux types de fragmentation d'images : (a) un petit pas pour les tumeurs de petites tailles : 2 ou 3 pixels, et un pas autour de 8 pixels ou plus pour les tumeurs de tailles plus grandes. Nous distinguons ainsi deux types de fragments : les fragments cancéreux et les fragments sains. Ces fragments sont utilisés pour former la base de données appliquée pour entraîner l'algorithme *deep learning*. Cette base est formée de cette manière :

- Pour la phase d'entraînement : nous avons consacré environ 75% de fragments.
- Pour la validation, elle comprend les 25% de fragments restants.

V.2.5 Création de la base de données

Afin de regrouper les images en deux classes, nous avons créé plusieurs bases de données pour tester l'efficacité de l'algorithme proposé. Nous avons constaté que la phase 2 après l'injection de contraste fournit plus d'information sur les tumeurs du foie que les autres phases. La base de données utilisée pour l'entraînement du modèle de *deep learning* est divisée en deux dossiers ; le premier pour l'entraînement et le deuxième pour la validation. Chaque dossier est divisé en deux sous-dossiers contenant les fragments cancéreux et les fragments sains.

V.3 Description de l’algorithme deep learning

Dans cette section, nous présentons notre algorithme *deep learning* permettant la détection des lésions CHC à partir des images IRM. Les étapes de l’algorithme proposé sont exposées dans la Figure V.8.

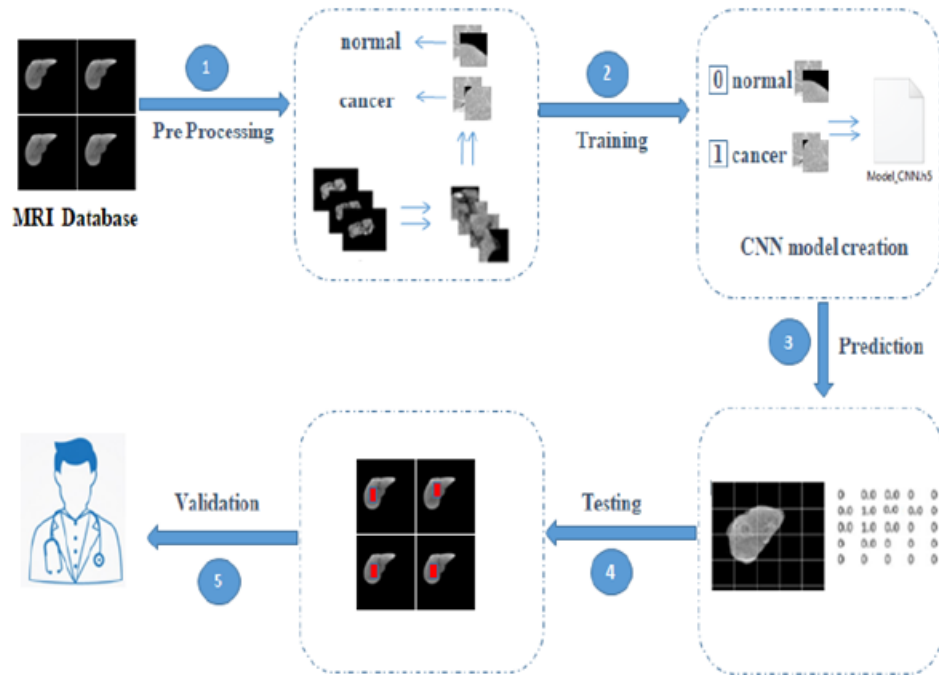


FIGURE V.8 – Les étapes de l’algorithme *deep learning*

Après avoir découpé les images IRM en des fragments (avec recouvrement), et les classifier en des fragments cancéreux et fragments sains pour créer les différentes bases de données, nous passons à l’entraînement du modèle. Notre modèle est basé sur l’architecture CNN qui nécessite tout d’abord une couche d’entrée constituée par des fragments de tailles 64x64, trois couches de convolutions de tailles 64x64, 32x32 et 16x16 respectivement et une couche de classification qui contient deux couches cachées de tailles 128 et 32 et finalement une couche de sortie qui présente un seul neurone. La valeur à la sortie se présente comme une probabilité allant de 0 à 1 où 0 présente une probabilité d’un fragment sain et 1 présente une probabilité d’un fragment cancéreux. En fait, durant la phase de prédiction, l’image est divisée en des fragments ensuite nous prédisons toutes les valeurs des fragments avec le modèle résultant à la fin de l’entraînement puis construire une matrice à la sortie.

V.3.1 Les paramètres de l’algorithme *deep learning*

Les paramètres appliqués dans notre modèle *deep learning* visualisé dans la Figure V.9 sont ;

- Les entrées et sorties : caractérisent les différents fragments de taille 64x64 divisés en deux classes : fragments cancéreux et fragments sains.
- La fonction d'activation : nous avons utilisé la fonction ReLu comme fonction d'activation pour les différents blocs de l'algorithme, sauf la dernière couche dans laquelle la fonction utilisée est sigmoïde, pour obtenir une probabilité.
- *Batch size* : ce paramètre présente le nombre d'échantillons qui vont être propagés à travers le réseau. Nous avons commencé par la valeur 10 en changeant à chaque fois cette valeur pour faire plus de tests.

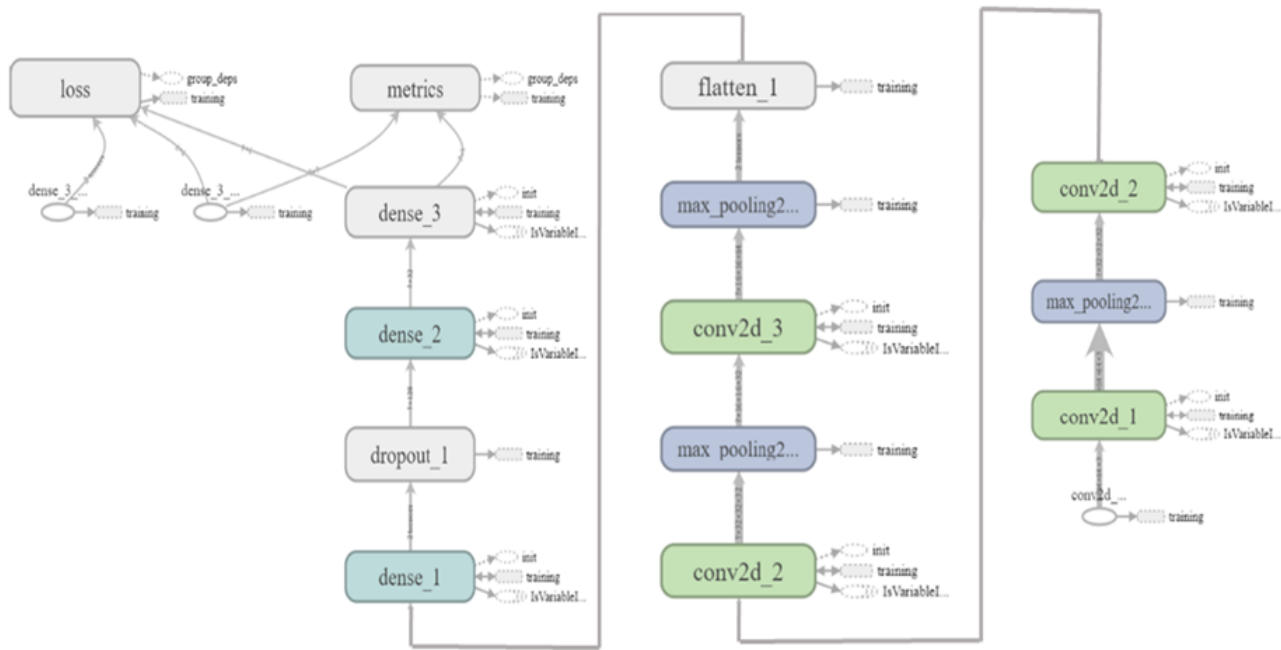


FIGURE V.9 – Visualisation du modèle CNN développé

V.3.2 Parallélisation de prétraitement

La parallélisation de prétraitement consiste à découper d'une manière simultanée les images IRM en des fragments. Ce traitement simultané nécessite un temps de calcul assez important. En effet, nous avons appliqué des mécanismes de répartition de charge pour garantir une meilleure parallélisation de traitement de données. Cette technique est réalisée à travers le découpage en blocs des images médicales IRM. Le découpage en blocs prend en considération le recouvrement des fragments voisins (les pixels de voisinage). Notre algorithme permet en premier lieu de paralléliser le traitement des données d'un seul examen IRM. Il consiste à découper l'image médicale en un ensemble de fragments de taille égale. Tout d'abord, il faut télécharger le volume IRM à découper et définir le pourcentage de recouvrement et le nombre de processeurs à utiliser ainsi que la taille de chaque coupe (Taille = largeur de l'image x longueur de l'image) en utilisant différentes techniques de répartition de charge. Puis, il faut calculer les dimensions d'un bloc (fragment d'image). Par la

suite, nous parcourons l'ensemble de l'image et extrayons les blocs tout en sauvegardant leurs dimensions et leurs positions. Finalement, nous enregistrons les blocs dans un dossier. Le traitement parallèle de ces images s'effectue en utilisant les données calculées aux étapes précédentes.

V.3.3 Phase de prédiction

Après la phase d'apprentissage et l'enregistrement du modèle, nous avons préparé de nouvelles images pour la prédiction. Une telle image est divisée en fragments de 64x64 non superposés et nous avons prédit chaque fragment indépendamment des autres. Comme le montre la Figure 25, la valeur 0 est attribuée à un fragment normal et la valeur 1 indique que le fragment contient potentiellement une tumeur, puisque chaque fragment a son abscisse et son ordonné dans l'image source, nous pouvons construire cette matrice où chaque valeur représente la prédiction du fragment correspondant (voir Figure [V.10](#)).

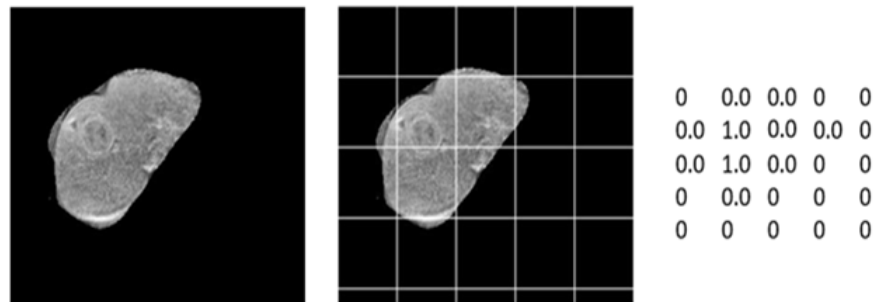


FIGURE V.10 – Description de la phase de prédiction

V.3.4 Validation des résultats

L'objectif de cette étape est de voir l'efficacité et la performance de l'algorithme proposé, en utilisant différentes images de différents patients. Plus de détails sur les tests et les résultats obtenus seront mentionnés dans le chapitre suivant.

Dans la section suivante, nous exposons l'intégration des ontologies dans le processus de classification des images 2D *via* un CNN *multi-label*. L'objectif est de montrer l'adaptation des ontologies avec les algorithmes *deep learning* et de de comparer cette approche avec le CNN classique.

V.4 Intégration d'ontologies dans la classification des tumeurs hépatiques

Afin d'accroître la performance du processus de classification des images, nous avons cherché à intégrer les ontologies dans l'approche présentée dans les sections précédentes. L'idée est de profiter de la base de connaissances offertes par les ontologies pour décrire les images médicales, pour fournir des informations sur les tumeurs détectées (par exemple, le type, la taille et le stade). L'avantage d'utiliser les techniques d'ingénierie de connaissance dans le domaine médical est de

faciliter le diagnostic des tumeurs. En outre, notre approche consiste à développer un CNN *multi-label* afin d'intégrer les informations sémantiques extraites à partir des ontologies développées OntHCC et MROnt. Ainsi, nous pouvons étendre l'algorithme CNN développé par les membres de l'équipe pour supporter plus d'informations. Ces informations décrivent les tumeurs détectées dans les images IRM du foie. La première phase de l'approche consiste à préparer les données images sous la forme des fragments. Par la suite, nous allons introduire les concepts et les propriétés des ontologies tels que « *Nodule* », « *Stage* », « *Tumor* », « *Size* », *etc.*, sous la forme des labels (étiquettes). Par la suite, il y aura l'entraînement du CNN et la génération de la classe ainsi que la catégorie décrivant chaque tumeur détectée sous la forme d'un vecteur de valeurs.

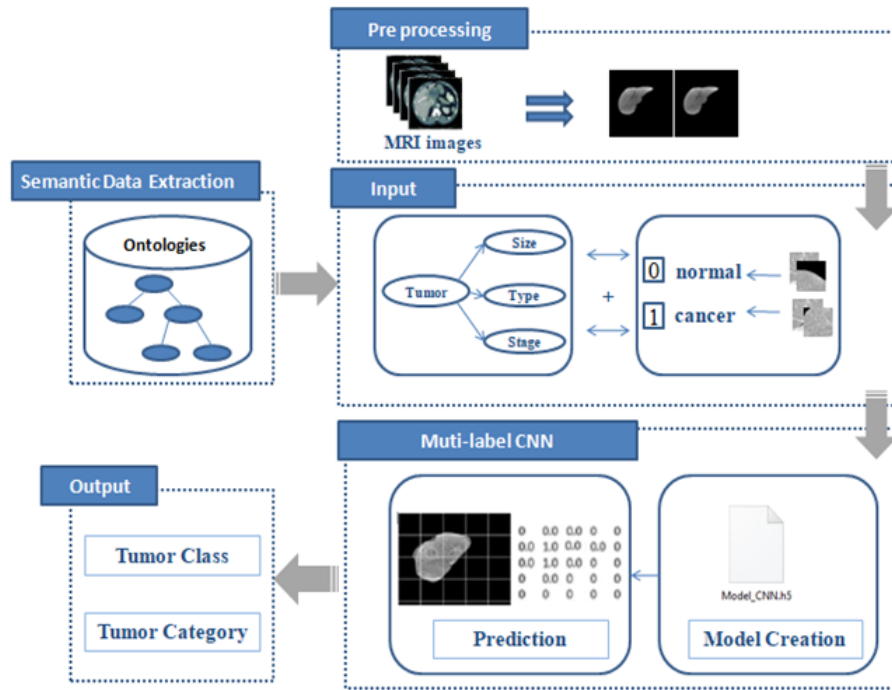


FIGURE V.11 – Approche de couplage ontologies/*deep learning*

Dans les sections suivantes, nous présentons une vue globale sur l'approche proposée. Nous exposons tout d'abord les ontologies appliquées, puis nous expliquons le processus fonctionnel de cette contribution. Il est composé essentiellement de cinq étapes (voir Figure [V.11](#)) ; (1) une phase de prétraitement, (2) une phase d'extraction de données sémantiques, (3) une phase de préparation des données et de la base de données pour l'entraînement du deep learning, (4) la création du CNN, (5) l'extraction de la classe et de la catégorie de la tumeur.

V.4.1 Les ontologies appliquées

Pour enrichir la classification des tumeurs hépatiques détectées dans les images IRM, nous avons intégré l'aspect sémantique au processus d'apprentissage. Ces informations sont extraites à partir des ontologies OntHCC [9] et MROnt [10]. L'OntHCC, présentée dans le chapitre précédent, (voir Figure [IV.4](#), page 84 de ce manuscrit) est un prototype qui traite le diagnostic du cancer du

foie. Il est dédié à l'analyse de l'image médicale appliquée pour la détection de différents types de tumeurs hépatiques y compris le CHC. Les informations collectées de cette ontologie concernent essentiellement la taille du nodule observé, son profil vasculaire, le taux d'AFP, *etc.* De plus, elles sont appliquées pour extraire le stade de la lésion détectée en appliquant différents systèmes de classification tels que BCLC [35], TNM [36] et Li-RADS [37]. Parmi les concepts de ce modèle Ontologique, nous trouvons par exemple : « *Disease* », « *Patient* », « *Diagnosis* », « *Treatment* », « *Liver* », « *Nodule* », « *TNM-System* », « *BCLC-System* » et « *LIRADS-Category* ». Cette ontologie a été peuplée et évaluée avec de rapports médicaux réels liés aux patients infectés par le CHC.

L'ontologie MROnt (voir Figure IV.6, page 91 de ce manuscrit) est aussi intégrée dans notre travail. C'est une représentation sémantique caractérisant les rapports médicaux des patients infectés par le cancer du foie. Ces rapports radiologiques comprennent des informations sur l'examen clinique telles que les techniques utilisées, les résultats, les notes radiologiques et les conclusions [38]. L'objectif du prototype MROnt est de modéliser et d'analyser les rapports médicaux des examens IRM. Ce modèle ontologique représente les données des rapports radiologiques afin de les rendre sous une forme standard et lisible par la machine. Des règles de raisonnement ont été appliquées pour analyser l'ensemble de données cliniques. Ceci permet à l'utilisateur d'être informé sur l'état du patient. Le modèle proposé a été appliqué spécifiquement à un ensemble de rapports médicaux sur le CHC. Dans le cas des rapports radiologiques, nous avons pris les observations notées par les radiologues comme instances. Elles sont appliquées par la suite pour prédire le type de la tumeur détectée dans l'image IRM grâce aux techniques de raisonnement (les règles SWRL). Plus de détails sur l'évaluation des ontologies OntHCC et MROnt seront présentés dans le chapitre suivant.

V.4.2 Collection des labels

La classification des images médicales en deux classes «avec tumeur» et «sans tumeur» n'est pas suffisante. Nous avons besoin d'un sens plus précis pour pouvoir visualiser plus d'informations sur ces images IRM. La classification multi-label est une sorte de classification dans laquelle un échantillon de données est classé en plusieurs classes. En effet, les labels servent à prédire des descriptions des tumeurs détectées. Ils sont introduits au CNN à travers un fichier d'extension (.tsv) décrivant les images de la base *deep learning* et prenant les noms des images suivies d'une description de chaque image pour former de différentes catégories. L'objectif est de trouver un modèle qui met en correspondance des entrées aux vecteurs binaires tout en attribuant une valeur de 0 ou 1 pour chaque élément (label). Le résultat est un vecteur binaire où chaque valeur correspond à l'existence ou non de ce label pour chaque image de la base. Dans notre travail, nous avons extrait des données sémantiques à partir des ontologies citées précédemment (voir Figure V.12). En effet, nous avons utilisé des concepts présentant la taille de la tumeur détectée (par exemple 2 mm, 1,9 mm, 1 cm, *etc.*), le stade de cette tumeur (par exemple précoce, intermédiaire, avancé, *etc.*), le type (CHC, autre tumeur, *etc.*) et l'examen médical comme l'IRM ou le scanner. Par la suite, nous allons faire une classification *multi-label* pour chaque image. De cette manière, chaque image de la base *deep learning* est liée à plusieurs étiquettes.

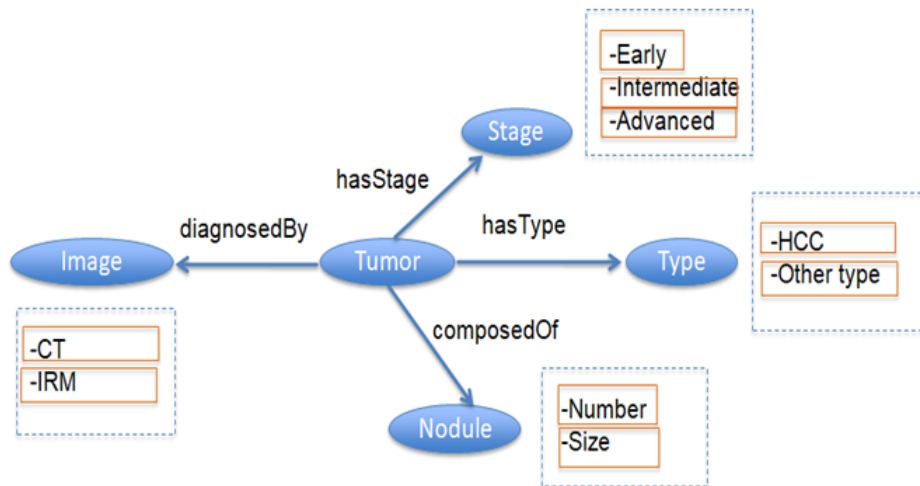


FIGURE V.12 – Les concepts appliqués dans l’algorithme de classification

La base de données appliquée pour l’entraînement de l’algorithme de *deep learning* est connectée à une liste de labels contenant chacun une description de la tumeur détectée. De cette façon, le réseau créé sera plus vaste et plus performant pour mieux détecter les parties de l’image et la lésion dedans. Chaque lésion détectée est associée ainsi à un vecteur de valeurs présentant une description de cette lésion. La Figure [V.13](#) expose un exemple de connexion entre une lésion et des étiquettes associées.

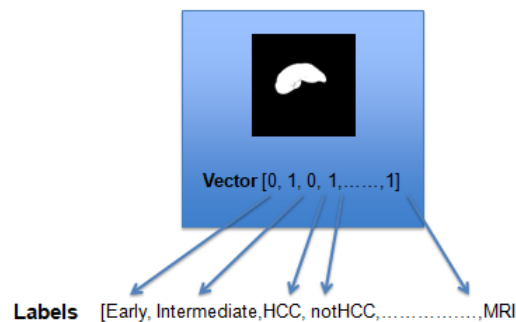


FIGURE V.13 – Vecteur des labels appliqués

V.4.3 Création d’un CNN multi-label

La première phase de l’approche proposée est la préparation des données. Cette étape consiste à faire le prétraitement des images appliquées pour former la base de données de *deep learning*. Nous avons utilisé les mêmes données de l’algorithme précédent en exploitant les trois phases des images IRM prises pour chaque patient. Ceci dépend du temps après l’injection de contraste, nous avons : Phase 1 : sans injection, Phase 2 : entre (30 s-1 min) après l’injection de contraste, et Phase 3 > 1 min) après l’injection de contraste. Ces phases sont importantes pour vérifier l’évolution des intensités d’imagerie tissulaire dans ces séquences. Pour former la base de données, nous avons utilisé des fragments d’images de tailles identiques. Dans notre cas, les fragments sont de tailles

64x64 pixels, définis avec une taille de bloc de 58x58 pixels et un recouvrement de 3 pixels dans toutes les directions. Nous avons choisi spécifiquement cette taille de fragments après avoir fait un ensemble d'expériences en modifiant à chaque fois les tailles. Les résultats ont montré que la taille 64x64 est la plus adéquate pour faire les tests. De plus, nos anciens travaux de l'équipe tels que [48] [49] et [50] ont approuvé la même constatation. Par la suite, nous avons fait une division des images IRM associées à un ensemble de patients. Nous avons utilisé des mécanismes de redimensionnement (c'est la technique de remplissage dans notre cas) dans le cas où la taille de l'image est inférieure à la taille du fragment. Cet algorithme élimine chaque patch qui avait zéro comme moyenne d'intensité globale dans le cas d'un fragment avec un arrière-plan uniquement. La Figure V.14 montre un exemple d'une image IRM après formation de fragments avec recouvrement; les fragments contenant une tumeur sont présentés avec une couleur rouge.

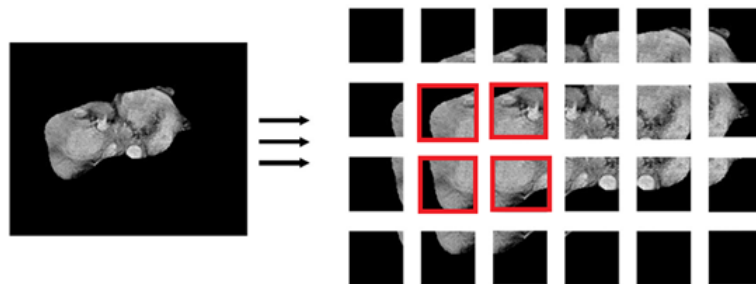


FIGURE V.14 – Visualisation des fragments d'une image IRM contenant une tumeur

Après la phase de prétraitement, nous passons à la création d'une liste d'étiquettes en prenant les données sémantiques à partir des ontologies OntHCC et MROnt. Ces données caractérisent des informations sur la lésion hépatique observée dans l'image telles que la taille, le stade et le type. La méthode proposée (voir Figure V.15) caractérise une séquence de blocs de convolution permettant l'extraction et l'identification des caractéristiques des images. Les images utilisées sont interconnectées avec un bloc entièrement connecté qui fait la classification des fragments

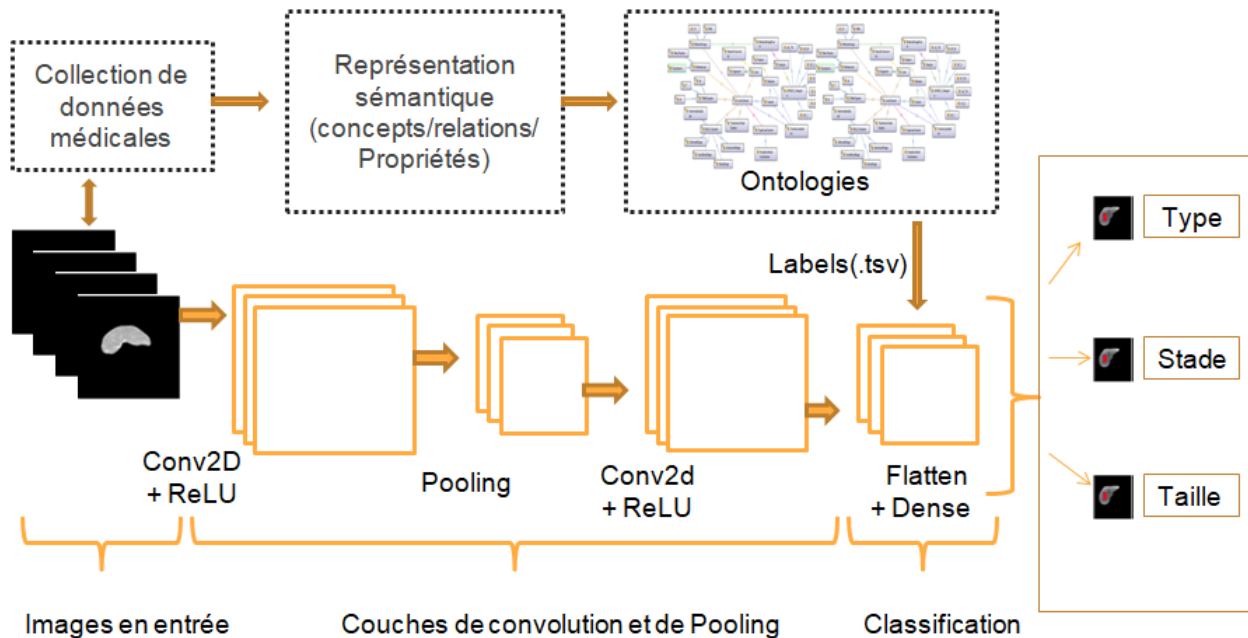


FIGURE V.15 – Visualisation des fragments d’une image IRM contenant une tumeur

Afin de créer le modèle d’apprentissage et la base de données, nous avons initialisé les différentes variables utiles pour l’exécution de notre algorithme : Le « nombre d’epochs », le « batch size », et les valeurs de poids et des biais. Pour le nombre d’epochs, nous avons fixé cette variable à 100. Nous avons utilisé une autre variable « *Callbacks* » qui permet l’interruption de la phase d’entraînement lorsque la valeur d’erreur n’a pas subi N (variable à fixer dès le début de l’algorithme) changements successivement. La valeur d’erreur consiste à dégager la différence entre les résultats obtenus après la phase de « *Forward propagation* » et les résultats introduits comme des entrées à l’algorithme. Pour le « *Batch size* » : cette variable dépend des ressources graphiques disponibles pour assurer l’exécution de l’algorithme. Nous avons utilisé le notebook « *Google Collaboratory* », qui offre un accès gratuit aux ressources de calcul (GPU de google, etc.)GPU un partage facile. Les valeurs de poids et des biais sont initialisées d’une façon aléatoire.

La phase de prédiction nécessite un ensemble de nouvelles images pour visualiser l’algorithme. Chaque nouvelle image utile pour la prédiction est divisée en des fragments de taille 64x64 non superposés et nous faisons la prédiction de chaque fragment. Le résultat d’exécution de cette étape est une matrice de valeurs où la valeur 0 est attribuée à un fragment normal et la valeur 1 est attribuée à un fragment cancéreux. Un vecteur de labels est également généré durant cette phase. Par conséquent, chaque tumeur sera définie par une classe (0 ou 1) et également avec une liste de labels présentant une description de son état. Les tests d’évaluation de cette approche seront présentés dans le chapitre suivant.

V.5 Conclusion

Dans ce chapitre, nous avons présenté notre contribution pour la détection et la localisation des lésions hépatiques de type CHC. Nous avons présenté un *framework* pour le traitement et la classification parallèle des données médicales. Nous avons appliqué l'architecture CNN pour le développement de ce *framework* de détection de lésions hépatiques. La première partie de ce chapitre a été consacrée à l'algorithme *deep learning* permettant le traitement parallèle des images médicales en prenant en compte des fragments. Cette approche assure le traitement parallèle d'une grande masse d'images médicales. Dans la deuxième partie, nous avons intégré les ontologies dans le processus de classification des tumeurs CHC afin d'améliorer la classification de ces images. Nous avons constaté que la classification multi-label permet non seulement de vérifier l'existence ou non d'une tumeur hépatique mais offre également plus d'informations sur la tumeur détectée.

Dans le chapitre suivant nous aborderons les prototypes qui mettent en œuvre les contributions théoriques étudiées dans ce chapitre ainsi que leurs évaluations. Nous présenterons en premier lieu les outils techniques appliquées pour le développement de ces approches et nous passerons en deuxième lieu à la phase de validation de ces prototypes.

PROTOTYPES ET EXPÉRIMENTATIONS

Sommaire

VI.1 Introduction	112
VI.2 Outils de développement	112
VI.2.1 Développement des ontologies	112
VI.2.2 Développement des modèles deep learning	114
VI.3 Prototypes développés	116
VI.3.1 Prototypes ontologiques	116
VI.3.2 Module A : Création de l'ontologie	117
VI.3.3 Module B : Construction des règles SWRL	117
VI.3.4 Module C : Affichage du résultat	118
VI.3.5 Prototype MROnt	120
VI.3.6 Prototype de détection et de localisation de la tumeur de foie	123
VI.4 Evaluation et validation des résultats	127
VI.4.1 Evaluation du Prototype OntHCC	127
VI.4.2 Evaluation des algorithmes de deep learning	130
VI.5 Conclusion	135

VI.1 Introduction

Dans les chapitres précédents, nous avons présenté une vue sur les approches proposées permettant de faire la détection et la stadification des tumeurs hépatiques de type CHC. De même, nous avons exposé la problématique de détection de ces tumeurs à partir des images médicales de type IRM. Afin de montrer la faisabilité des approches théoriques, nous avons développé des prototypes basés sur divers outils et langages informatiques conduisant à la mise en pratique de ces méthodes. Par la suite, nous présentons les architectures logicielles tout en détaillant leurs fonctionnalités et nous terminons par évaluer les différentes approches proposées.

VI.2 Outils de développement

Dans cette section, nous présentons les outils utilisés pour le développement des approches proposées dans les chapitres précédents, nous commençons par énumérer les développements des ontologies OntHCC et MRont. Par la suite, nous présentons les outils et les *frameworks* utilisés pour la réalisation des algorithmes basés sur le *deep learning*.

VI.2.1 Développement des ontologies

VI.2.1.1 Protégé

Protégé est un éditeur libre et *open source* fondé par l'Université de Stanford, destiné essentiellement pour la création et la sauvegarde des ontologies sous plusieurs formats RDF, RDFS, XML Schema, OWL, *etc.* C'est un outil informatique basé sur Java, appliqué pour la construction et la représentation des systèmes intelligents. Protégé offre un environnement de modélisation des connaissances en se basant sur les éditeurs Protégé-Frames et Protégé-OWL. Il supporte également plusieurs langages d'interprétation et d'interrogation des requêtes à savoir SQWRL et SPARQL. L'architecture de cet éditeur peut être adaptée pour développer des applications ontologiques avec un niveau de complexité différent. Des plug-ins ont été développés pour ajouter des fonctions dédiées à Protégé; OntoViz ⁶⁸ et OWLViz ⁶⁹ sont considérés parmi les plug-ins de Protégé les plus appliqués pour la visualisation des ontologies. SwrlTab ⁷⁰ est utilisé également pour la création et l'édition des règles SWRL.

VI.2.1.2 Eclipse

Eclipse ⁷¹ est un environnement de développement intégré (IDE) polyvalent créé par IBM (« *International Business Machines Corporation* ») en novembre 2001. Il s'appuie sur une conception modulaire visant à assurer un environnement de production de logiciels extensible, libre, et universel. Eclipse est développé en Java tout en fournissant les outils et les *frameworks* pour la

68. <https://protegewiki.stanford.edu/wiki/OntoViz>

69. <https://protegewiki.stanford.edu/wiki/OWLViz>

70. <https://prfowlotegewiki.stanford.edu/wiki/SWRLTab>

71. <https://eclipse.org/>

réalisation de logiciels, incluant les activités de programmation. Il supporte plusieurs langages de programmation tels que Python, C / C ++, PHP, *etc.*. Dans ce travail, nous avons utilisé ce logiciel pour la mise en œuvre des prototypes ontologiques. Il est appliqué aussi pour l'implémentation de quelques interfaces graphiques qui concernent la détection et la stadification des nodules hépatiques de type CHC.

VI.2.1.3 Jena

Apache Jena⁷² caractérise un ensemble d'outils destinés pour la construction d'applications du Web sémantique. L'API (« *Application Programming Interface* ») Java est l'un de ces outils, elle permet la manipulation de plusieurs langages (par exemple OWL, RDF, RDFS, et SPARQL). Jena peut aussi créer et manipuler des graphes RDF puisqu'elle contient des classes d'objets présentant les graphes, les propriétés les ressources, et les littéraux. Un graphe ou un modèle peut ainsi être représenté par l'interface « *Model* ». Elle utilise divers moteurs d'inférence pour assurer le raisonnement des modèles ontologiques à travers une diversité de raisonneurs (par exemple Pellet ou HermiT). Dans notre travail, nous avons utilisé cette API pour assurer le lien entre Java et Protégé ce qui permet de lire l'ontologie médicale construite.

VI.2.1.4 SWRL

SWRL est défini comme étant un langage de règles destiné pour le Web sémantique. Ce langage est une combinaison des langages OWL-DL (« *Ontology Web Language Description Logics* ») et RuleML (« *Rule Markup Language* »). Il permet l'intégration des règles d'inférences, écrites dans Protégé, dans les ontologies. Ces règles seront par la suite exécutées *via* les raisonneurs intégrés dans Protégé (comme Pellet ou Hermit). Les règles SWRL sont formées à partir des concepts, des propriétés et des instances. Elles sont sauvegardées dans l'éditeur SwrlTab de *Protégé*. Dans notre travail, nous avons utilisé SWRL pour la construction des règles de raisonnement permettant en premier lieu la détection du CHC dans l'image médicale du foie. Par la suite, ces règles décrivent la stadification de chaque nodule détecté en se basant sur les paramètres standards reliés à chaque système utilisé. Aussi, les règles SWRL sont appliquées pour donner une décision primaire concernant le type du traitement à appliquer.

VI.2.1.5 SPARQL

SPARQL (« *Query Language for RDF* ») est un langage standard destiné pour la création et l'édition des requêtes en retournant des informations sous la forme d'un ensemble de liaisons ou d'un graphe RDF (« *Resource Description Framework* »). Il interroge surtout les données détenues dans les modèles sémantiques. SPARQL est considéré comme l'une des technologies clés du Web sémantique. SPARQL est utile pour rechercher des motifs de graphe (obligatoires ou optionnels). Les résultats des requêtes SPARQL peuvent être des ensembles de résultats ou des graphes RDF. Dans notre travail, ce protocole de requêtes est utilisé pour extraire des informations portant sur le type des nodules hépatiques et leurs stades selon des conditions spécifiques.

72. <https://jena.apache.org/>

VI.2.1.6 SQWRL

SQWRL ⁷³ (« *Semantic Query-Enhanced Web Rule Language* ») est un langage de requête basé sur SWRL. Il peut être utilisé pour interroger les modèles ontologiques OWL. Ce langage fournit une diversité d'opérations de type SQL (« *Structured Query Language* ») permettant la mise en forme des connaissances ou des données sémantiques, connaissances qui sont déjà préparées dans l'ontologie.

VI.2.1.7 Weasis

Weasis est un logiciel multifonctionnel permettant la visualisation des fichiers DICOM (« *Digital Imaging and Communications in Medicine* »). Il est employé dans certains services cliniques, les hôpitaux, et les réseaux de santé. Weasis est un logiciel multiplateforme, libre et *open source* qui assure une intégration flexible de plusieurs systèmes tels que PACS (« *Picture Archiving and Communication System* »), RIS (« *Radiological Information System* »), *etc.* Une image médicale portant le format DICOM est constituée par l'entête et le contenu de l'image elle-même. Pour récupérer les informations relatives au patient, nous avons recours à un logiciel qui lit cet entête. Weasis permet d'importer tout un examen médical en fournissant toutes les informations relatives au patient concerné.

VI.2.2 Développement des modèles deep learning

VI.2.2.1 Python

Python ⁷⁴ est un langage de programmation interprété et de haut niveau créé en 1991. Il permet d'assurer une programmation structurée, orientée objet et fonctionnelle. Python est un langage *open source* qui s'est situé en tête de l'analyse de données la gestion d'infrastructure, ou dans le domaine du développement de logiciels. Python est aussi un langage multiplateforme qui fonctionne sur la majorité des plateformes, des smartphones aux ordinateurs centraux, sur Windows à Unix avec notamment GNU/Linux en passant par macOS, ou encore Android, iOS, et il peut être aussi traduit en Java ou .NET. Il est destiné également à optimiser la productivité des programmeurs en offrant des outils de haut niveau avec une syntaxe simple à appliquer. Ce langage dispose d'un système dynamique qui prend en charge plusieurs paradigmes de programmation orientée objet, impérative, et fonctionnelle.

VI.2.2.2 Microsoft Visual Studio

Visual Studio ⁷⁵ code est un éditeur fondé par Microsoft, composé de plusieurs logiciels de développement et conçu pour plusieurs systèmes (Windows, Linux et macOS). C'est un ensemble d'outils de développement permettant de créer une diversité d'applications (par exemple des applications Web ASP.NET, des applications mobiles, des services Web XML, *etc.*). Dans ce travail,

73. <http://protege.cim3.net/cgi-bin/wiki.pl?SQWRL>

74. <https://www.python.org/>

75. <https://visualstudio.microsoft.com/fr/>

nous avons utilisé Visual Studio (version 2019) pour écrire les modèles d'apprentissage *deep learning*.

VI.2.2.3 Google Colaboratory Notebook

Google Colaboratory ⁷⁸ (ou « *Google Colab* ») est un outil d'analyse de données offert par Google. C'est un service cloud basé sur le Notebook Jupyter ⁷⁹ et conçu surtout pour les applications *deep learning*. Cette plateforme est capable d'entraîner plusieurs modèles de Machine Learning implantés directement dans le *cloud*. Google Colab est une application open-source capable de créer et partager des documents contenant du code, des équations, des visualisations et du texte. Elle est utile aussi pour faire la transformation des données, la simulation numérique, et la modélisation statistique. Elle fournit également un accès GPU totalement gratuit ce qui la distingue des autres services *cloud*.

VI.2.2.4 TensorFlow

TensorFlow ⁸⁰ est une plateforme *open source* développée par Google et destinée pour entraîner des modèles de l'apprentissage automatique. Cet outil offre un environnement complet et flexible d'outils et de bibliothèques permettant aux chercheurs d'avancer dans le domaine du *machine learning*, et aux utilisateurs de déployer d'une manière facile des applications qui utilisent cette technologie. L'architecture flexible de TensorFlow permet de faire le calcul sur plusieurs processeurs dans une machine, un serveur ou un périphérique mobile.

VI.2.2.5 Keras

Keras ⁸¹ est une bibliothèque *open source* écrite par François Chollet permettant d'interagir avec des algorithmes de réseaux neuronaux. Elle a été fondée dans le cadre du projet ONEIROS (« *Open-ended Neuro-Electronic Intelligent Robot Operating System* »). Elle a la capacité de fonctionner avec plusieurs autres outils (par exemple TensorFlow, Microsoft Cognitive Toolkit, Theano, PlaidML, *etc.*). Keras est utilisée pour assurer une expérimentation rapide avec les réseaux de neurones profonds, elle est caractérisée essentiellement par son ergonomie et sa modularité. L'utilisation de Keras s'avère nécessaire si nous avons besoin d'une bibliothèque de *deep learning*. Cet API assure un prototypage rapide grâce à sa convivialité, sa modularité et son extensibilité. Elle prend en charge également les réseaux convolutifs et les réseaux récurrents tout en fonctionnant d'une façon transparente sur le processeur et/ou le GPU.

78. <https://colab.research.google.com/notebooks/>

79. <https://jupyter.org/>

80. <https://www.tensorflow.org/>

81. <https://keras.io/>

VI.2.2.6 Numpy

NumPy⁸² est une bibliothèque libre et *open source* pour le langage de programmation Python qui permet plus de stockage de données avec moins de mémoire. Elle est appliquée pour faire la manipulation des matrices ou des tableaux multidimensionnels ainsi que des fonctions mathématiques opérant sur ces tableaux. Cette bibliothèque offre de multiples fonctions permettant par exemple de créer directement un tableau depuis un fichier ou au contraire, de sauvegarder un tableau dans un fichier, et de manipuler des vecteurs ou des matrices. Numpy contient également des fonctions sophistiquées des outils pour intégrer C, C++ et Fortran. NumPy peut également être utilisé comme un conteneur multidimensionnel efficace de données génériques. Plusieurs types de données peuvent être définis. Cela permet à NumPy de s'intégrer facilement dans une grande variété de bases de données.

VI.2.2.7 OpenCV

Open CV (« *Open Computer Vision* ») est une bibliothèque graphique libre, développée par Intel sous licence BSD (« *Berkeley Software Distribution Licence* »). Elle est conçue essentiellement pour le traitement d'images y compris des vidéos en temps réel. La première version a été lancée en 2000. Elle fonctionne sur la plupart des systèmes d'exploitation et pour plusieurs langages de programmation (par exemple Python, C++ et Java). Elle est réputée pour une meilleure efficacité de calcul surtout pour les applications en temps réel. Initialement codée en C et C++, cette bibliothèque regroupe une diversité de fonctions permettant de faciliter le traitement ou la reconnaissance d'images (par exemple : le seuillage, les filtres de détection de contours, les filtres de lissage et les filtres morphologiques, *etc.*).

VI.3 Prototypes développés

Dans cette section, nous présentons en premier lieu les prototypes ontologiques développés, nous exposons l'architecture fonctionnelle du prototype OntHCC puis celle du prototype MRont. En deuxième lieu, nous exposons l'architecture fonctionnelle du prototype de détection des tumeurs du foie basé sur le *deep learning*.

VI.3.1 Prototypes ontologiques

VI.3.1.1 Prototype OntHCC

La création des prototypes ontologiques vise à modéliser les concepts d'un domaine sous une forme structurée et de faciliter son intégration dans un système. Dans cette section, nous détaillons l'architecture fonctionnelle de notre modèle ontologique OntHCC. Ce modèle s'appuie sur trois modules principaux (voir Figure VI.1) : (A) La construction de l'ontologie médicale en suivant la méthodologie « *MethOntology* », (B) La définition des règles SWRL qui présentent la détection

82. <https://numpy.org/>

des lésions CHC à partir les images IRM. (C) L'implantation et l'interrogation de l'ontologie en utilisant le langage Java. La lecture de l'ontologie est faite en faisant appel à l'API Jena qui fait la liaison avec le fichier (.owl) généré par Protégé. Une description plus détaillée de chacun de ces modules est présentée dans les sous-sections suivantes.

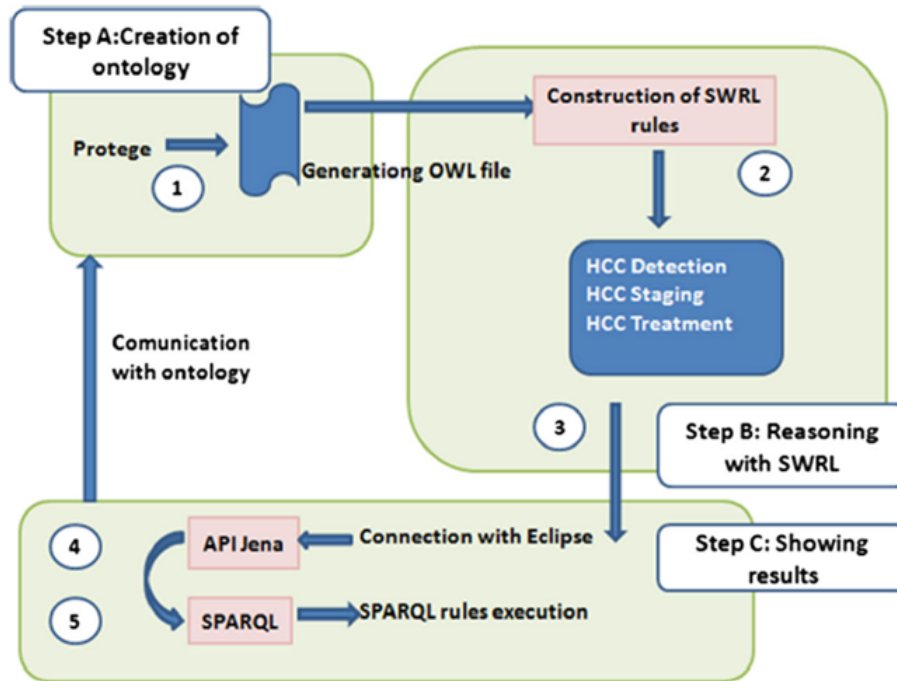


FIGURE VI.1 – Architecture fonctionnelle du prototype OntHCC

VI.3.2 Module A : Création de l'ontologie

La création de cette ontologie médicale vise principalement à la modélisation et la stadification des tumeurs de type CHC. L'objectif de ce module est de créer un fichier OWL contenant l'ontologie OntHCC composée de concepts, relations et propriétés. Parmi les notions évoquées, nous trouvons les symptômes, les facteurs de risques, les types de diagnostic, le type du traitement et aussi les systèmes de stadification des lésions hépatiques. Nous avons intégré les systèmes BCLC, TNM, et Li-RADS. Nous avons enrichi l'ontologie à travers les instances nécessaires pour chaque classe créée. Un fichier OWL est généré automatiquement en exploitant Protégé portant le format RDF/XML.

VI.3.3 Module B : Construction des règles SWRL

Ce module s'intéresse la construction des règles de raisonnement SWRL avec 3 volets principaux : (1) le premier volet consiste à la représentation de l'image médicale du foie en se basant sur l'examen radiologique. Dans ce cas, la présence d'une tumeur au sein de l'image revient à trouver le type de cette tumeur. Pour assurer cette détection, nous avons suivi les critères de BCLC pour

vérifier s'il s'agit d'une lésion CHC ou non. (2) Le deuxième volet correspond à la stadification de cette tumeur en appliquant les systèmes de stadification BCLC, TNM et Li-RADS. (3) Le troisième volet consiste à la précision du type de traitement à appliquer par la suite. D'une manière générale, ce module traite principalement le diagnostic, la stadification et le traitement du CHC.

VI.3.4 Module C : Affichage du résultat

Ce module se concentre sur l'interrogation de l'ontologie OntHCC à partir des règles SPARQL sous la forme des requêtes. L'exécution de ces règles se fait à travers Java par l'intermédiaire de l'API Jena. Les résultats d'exécution de ces requêtes sont affichées sous la forme d'interfaces graphiques implémentées en Java. La détection du CHC *via* l'IRM se fait en appliquant les critères BCLC. Dans cette approche, nous utilisons les notions fondamentales appliquées dans les routines cliniques pour détecter le nodule CHC. La Figure VI.2 montre l'interface graphique affichant les informations relatives à l'image médicale qui visualise le nodule hépatique. Elle montre la décision clinique s'il s'agit d'un CHC.

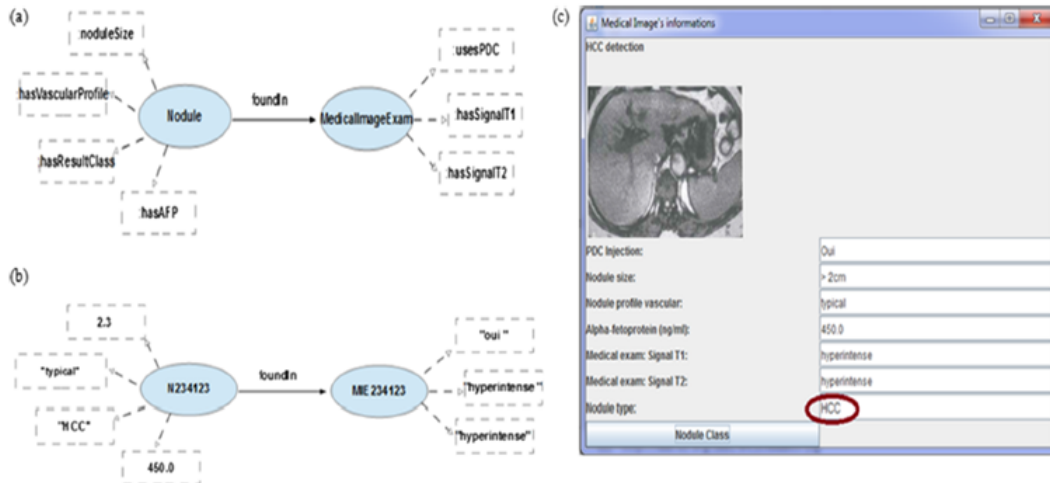


FIGURE VI.2 – Interface de détection des tumeurs CHC

L'intégration des systèmes de stadification des lésions hépatiques est un élément essentiel parce qu'il nous renseigne sur le stade des tumeurs CHC détectées. La Figure VI.3 expose l'interface graphique présentant les systèmes de stadification utilisés dans notre approche. La première étape consiste à choisir soit d'utiliser BCLC, soit TNM, soit une combinaison des deux systèmes.

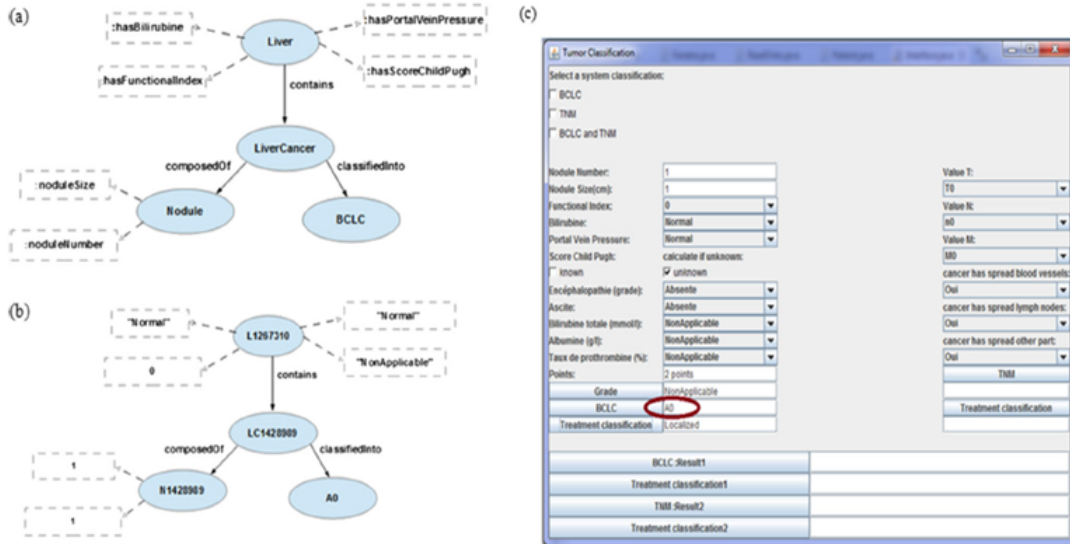


FIGURE VI.3 – Interface de stadification des tumeurs hépatiques

Pour aborder le système de stadification BCLC, il faut remplir en premier lieu tous les champs qui présentent réellement les paramètres standards de ce système (par exemple la taille du nodule, le nombre des nodules composant la tumeur, la pression veine portale, *etc*). Parmi les paramètres fonctionnels de ce système, nous citons le score Child-Pugh qui doit être calculé nécessairement en fonction de cinq mesures. Dans la Figure VI.4, nous affichons le calcul de ce score. Puis, nous intégrons ce calcul dans l'interface principale de stadification pour qu'il soit compatible avec les autres paramètres. Par la suite, le score sera affiché après un clic sur le bouton « Grade ».

Calcul Score ChildPugh

Encéphalopathie (grade):	Gradell
Ascite:	Minime
Bilirubine totale (mmol/l):	35-50
Albumine (g/l):	>35
Taux de prothrombine (%):	<40
ScoreChildPugh	GradeC

FIGURE VI.4 – Calcul du score Child-Pugh

Le fonctionnement des systèmes de stadification des tumeurs hépatiques se fait par l'intermédiaire des requêtes SPARQL. La Figure VI.5 montre la requête SPARQL relative à la mise en marche du système BCLC. Après le remplissage de tous les paramètres, l'action du bouton « BCLC » affiche automatiquement le stade de la tumeur. Nous avons choisi comme exemple d'application le stade primaire A0 du système BCLC.

```

(a) SPARQL query:
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX ex: <http://www.semanticweb.org/stagiaire1/ontologies/2016/5/untitled-ontology-49#>
SELECT ?im ?medicalimage ?medicalageexam ?moduleSize ?vascularProfile ?PDC ?Class ?AFP ?SignalT1 ?SignalT2
WHERE [
  ?medicalimage ex:isSequenceOf ?z.
  ?im ex:foundIn ?medicalageexam.
  ?im ex:moduleSize ?moduleSize.
  ?im ex:vascularProfile ?vascularProfile.
  ?z ex:usePDC ?PDC.
  ?im ex:hasResultClass ?Class.
  ?im ex:hasAFP ?AFP.
  ?medicalageexam ex:hasSignalT1 ?SignalT1.
  ?medicalageexam ex:hasSignalT2 ?SignalT2]

(b) SPARQL query:
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX ex: <http://www.semanticweb.org/stagiaire1/ontologies/2016/5/untitled-ontology-49#>
SELECT ?stage
WHERE (?liver ex:contains ?cancer
  ?liver ex:hasScoreChildPugh ?score.
  ?liver ex:hasFunctionalIndex ?index.
  ?liver ex:hasBilirubine ?bilirubin.
  ?liver ex:hasPortalVeinPressure ?pressure
  ?cancer ex:composedOf ?module.
  ?module ex:moduleNumber ?number.
  ?im ex:moduleSize ?size.
  ?cancer ex:classifiedInto ?stage)

(c) SPARQL query:
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX ex: <http://www.semanticweb.org/stagiaire1/ontologies/2016/5/untitled-ontology-49#>
SELECT ?treatment ?cancer
WHERE (?cancer ex:has-T ?t.
  ?cancer ex:has-N ?n.
  ?cancer ex:has-M ?m.
  ?cancer ex:hasSpreadBloodVessels ?x.
  ?cancer ex:hasSpreadLymphNodes ?y.
  ?cancer ex:hasSpreadOtherPart ?z.
  ?cancer ex:classifiedInto ?stage)

(d) SPARQL query:
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX ex: <http://www.semanticweb.org/stagiaire1/ontologies/2016/5/untitled-ontology-49#>
SELECT ?treatment ?cancer
WHERE (?cancer ex:classifiedInto ?stage.
  ?cancer ex:treatedBy ?treatment)

```

FIGURE VI.5 – Exemples de requêtes SPARQL : (a) Règle SPARQL pour la détection de la tumeur hépatique, (b) Règle SPARQL pour la stadification BCLC, (c) Règle SPARQL pour la stadification TNM, (d) Règle SPARQL pour le type de traitement à appliquer

Notre approche inclut aussi le type de traitement à appliquer en se basant sur les critères BCLC. Cette fonctionnalité peut aider les médecins ou les radiologues à prendre les décisions cliniques. La Figure VI.6 montre que le stade B selon le système BCLC peut être traité par la TACE (chémioembolisation transartérielle).

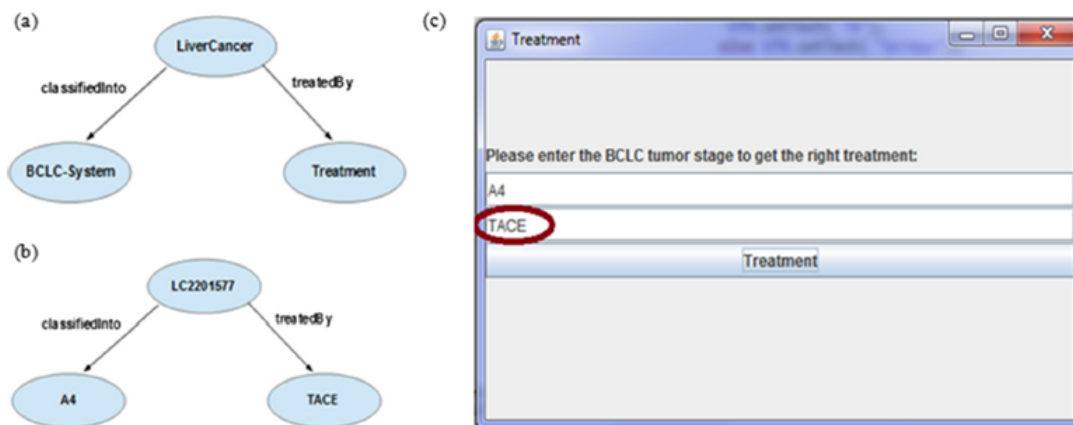


FIGURE VI.6 – Interface de traitement du CHC

VI.3.5 Prototype MRONT

Le prototype MRONT a pour objectif de modéliser les informations contenues dans les rapports radiologiques à partir d'un modèle sémantique. Ces rapports sont reliés à des patients infectés par le CHC. Le modèle créé permet de présenter et partager des connaissances médicales. Il caractérise

une représentation formelle basée sur des concepts, des relations, des propriétés et des instances utilisés pour réaliser un vocabulaire référentiel. Dans ce travail, nous introduisons cette représentation sémantique pour analyser les informations cliniques. Des règles SWRL sont construites aussi pour extraire les observations cliniques à partir de rapports radiologiques réels. L'architecture fonctionnelle du prototype proposé (voir Figure VI.7) est composée de trois modules : (1) Module de collection d'information, (2) Module de modélisation sémantique et (3) Module de l'implantation et de l'interrogation de l'ontologie.

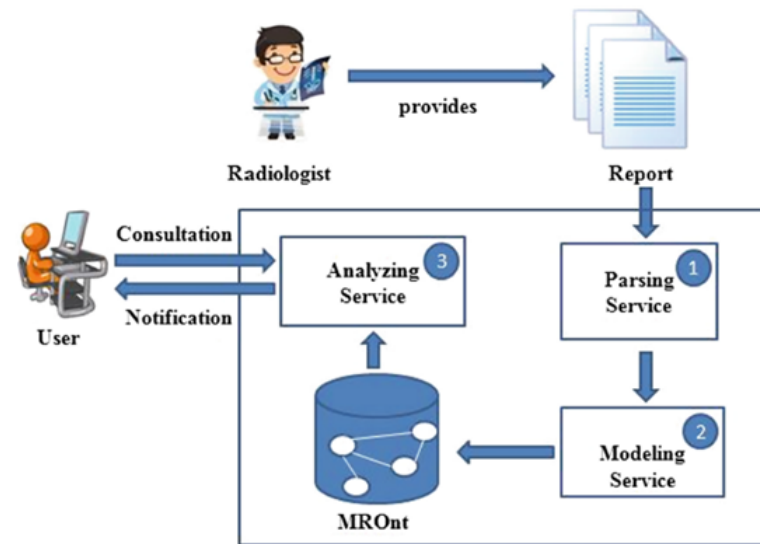


FIGURE VI.7 – Architecture fonctionnelle du prototype MRont

La première étape consiste à analyser les rapports médicaux à l'aide de techniques d'analyse notamment EMF⁸³ (« *Eclipse Modeling Framework* »). Par la suite, nous enrichissons l'ontologie MRont avec des informations médicales qui sont des examens IRM des patients en utilisant l'API Jena. Ce package offre la possibilité d'interroger des données médicales tirées des rapports médicaux. Puis, nous testons les règles SWRL et SQWRL construites pour montrer l'efficacité de notre système d'analyse pour détecter le type de la lésion hépatique. Les données utilisées dans ce travail proviennent du CHU de Clermont-Ferrand, France. Les rapports médicaux, acquis par les médecins et les radiologues, ont été anonymisés et stockés d'une manière sécurisée dans des serveurs situés au CHU. Dans ce travail, nous avons utilisé 28 rapports médicaux pour les patients qui ont été diagnostiqués avec un cancer du foie. Ces rapports caractérisent différents états pathologiques de patients. Ils contiennent tous les étapes de l'examen médical en commençant par des informations générales jusqu'à donner des détails sur les interprétations cliniques notés par les radiologues.

Ce système d'analyse de rapport médical permet de peupler l'ontologie MRont avec des examens IRM. Il rassemble toutes les informations et les inclut dans l'ontologie MRont. Tout d'abord, il prend en compte des informations générales sur le patient (par exemple, l'identifiant, sexe, âge,

83. <https://www.eclipse.org/modeling/emf/>

date de naissance, *etc.*). Ensuite, nous nous intéressons à l'interprétation radiologique qui commence par présenter les techniques appliquées pour la réalisation de l'examen IRM. Cette interprétation décrit l'état du foie (par exemple dysmorphie, contours irréguliers, *etc.*) et la lésion hépatique (par exemple taille, localisation, nombre de nodules, *etc.*). L'approche proposée vérifie également s'il existe une anomalie liée au système hépatique, y compris les observations des organes près du foie (par exemple le pancréas et les reins). Aussi, elle introduit d'autres données liées aux signaux IRM (par exemple T1, T2, et T2 FAT SAT) qui peuvent être visualisés après l'injection du produit de contraste au patient. Enfin, ce système d'analyse de rapport médical génère le résultat sous la forme d'une conclusion indiquant le type du nodule (CHC, kyste ou autre type de maladies hépatiques). Toutes les étapes mentionnées sont effectuées suite à l'exécution des règles SWRL et SQWRL. Ces règles permettent la réalisation d'un processus de raisonnement de haut niveau. Les règles SQWRL sont utilisées pour interroger les données cliniques tirées à partir des rapports radiologiques. L'exemple présenté dans la Figure [?] expose un cas d'exécution d'une règle SQWRL. Il s'agit d'une requête de sélection qui affiche toutes les observations liées à l'examen médical identifiée par « MI124589 ». Parmi les observations sélectionnées, nous trouvons par exemple une dysmorphie du foie, une architecture micronodulaire, pas de thrombose portale, et un hypersignal T1.

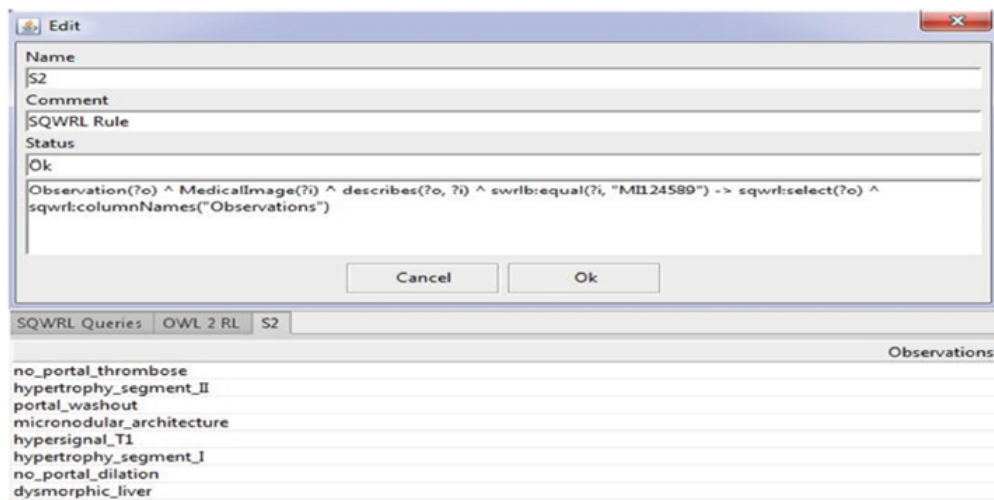


FIGURE VI.8 – Exécution d'une requête SPARQL

La Figure VI.9 représente la liste des observations déduites pour un patient identifié par «P3534533». L'examen du patient traité dans cet exemple est caractérisé par les observations suivantes : une thrombose portale, une architecture micronodulaire, un hypersignal T1, et une lésion d'une taille de 21 mm située dans le segment III du foie, *etc.* En appliquant notre système d'analyse, nous pouvons déduire que ce patient est affecté par le CHC.

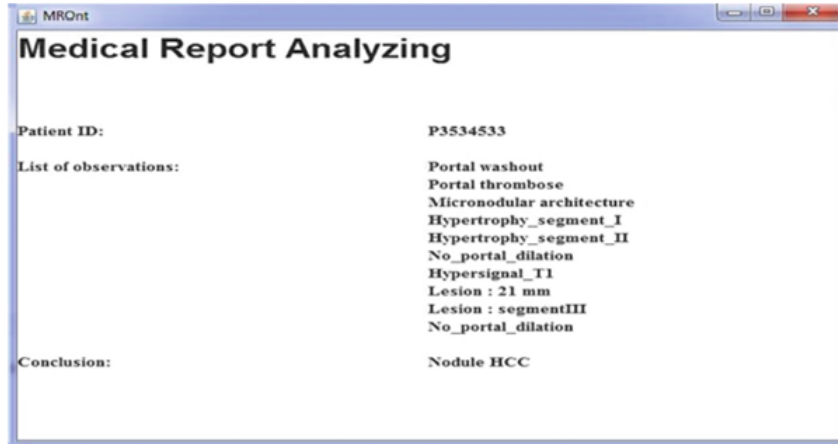


FIGURE VI.9 – Extraction des observations radiologiques d'un patient

VI.3.6 Prototype de détection et de localisation de la tumeur de foie

Dans cette section, nous étudions la réalisation du prototype de détection et de localisation de la tumeur du foie ainsi que le processus de son développement. Cette section donne une visualisation des différentes étapes permettant la mise en pratique de cette approche. En premier lieu, nous présentons en détail l'architecture fonctionnelle du prototype. En deuxième lieu, nous abordons la réalisation de notre application en mettant l'accent sur les différentes étapes de développement et les résultats obtenus. L'architecture fonctionnelle du prototype proposé est présentée dans la Figure VI.10.

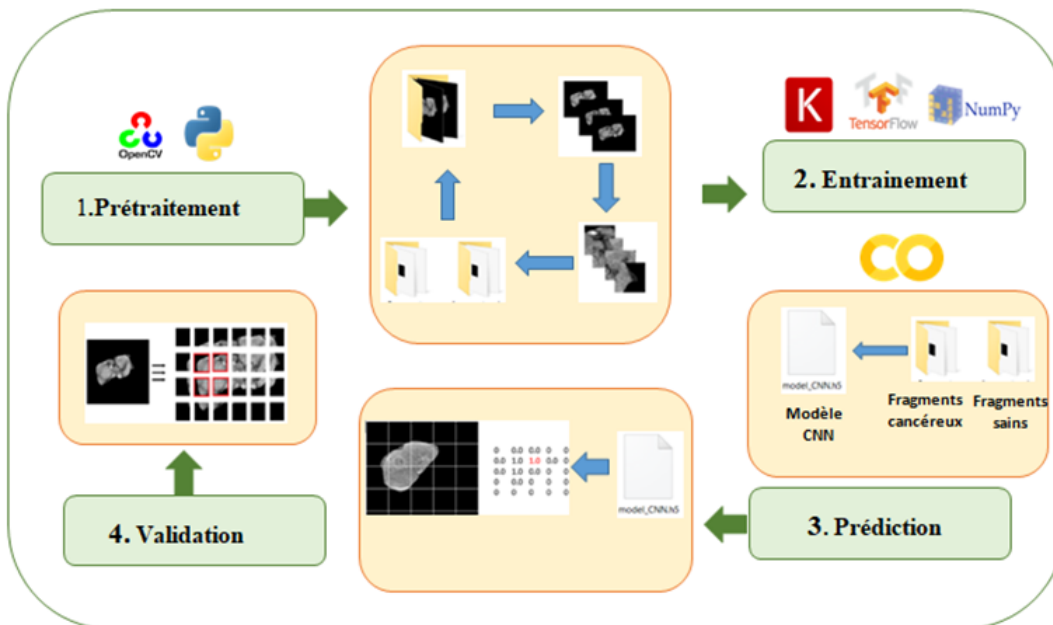


FIGURE VI.10 – Architecture fonctionnelle de détection et localisation de la tumeur de foie

La première phase de cette approche consiste à préparer les images IRM à utiliser pour former la base *deep learning*. Nous récupérons ainsi trois phases IRM à partir des séquences DCE-MRI. La Figure V.2 (page 98) apporte un aperçu sur les images IRM durant les différentes phases utilisées après l'injection de contraste au patient.

Les images utilisées sont de taille 64x64 pixels pour notre cas (avec 58x58 comme taille de chaque bloc avec un recouvrement de 3 pixels dans les 4 directions). Une division de ces images est faite tout en considérant le recouvrement des fragments voisins (voir Figure V.7) (page 101). Au cas où la taille d'une image est inférieure à celle de la taille du fragment, nous utilisons les mécanismes de redimensionnement sans perte. Nous ajoutons ainsi des valeurs « 0 » à la fin de la matrice jusqu'à atteindre la taille du fragment. Le programme développé élimine aussi chaque fragment qui admet un moyen total d'intensité égale à 0 : c'est le cas d'avoir un fragment avec un fond noir. Les résultats de cette phase sont présentés dans les Figures VI.11, VI.12 et VI.13.



FIGURE VI.11 – Exécution de l’algorithme pour l’extraction des fragments

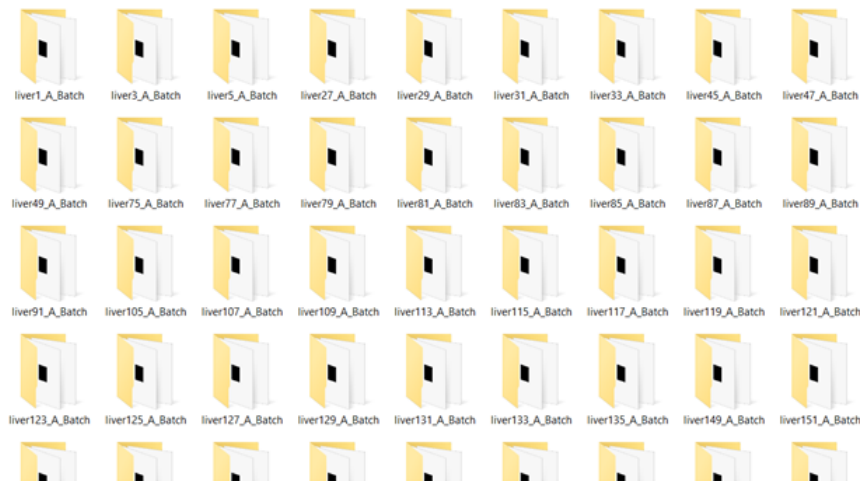


FIGURE VI.12 – Résultat de l’algorithme d’extraction exécuté sur tous les images

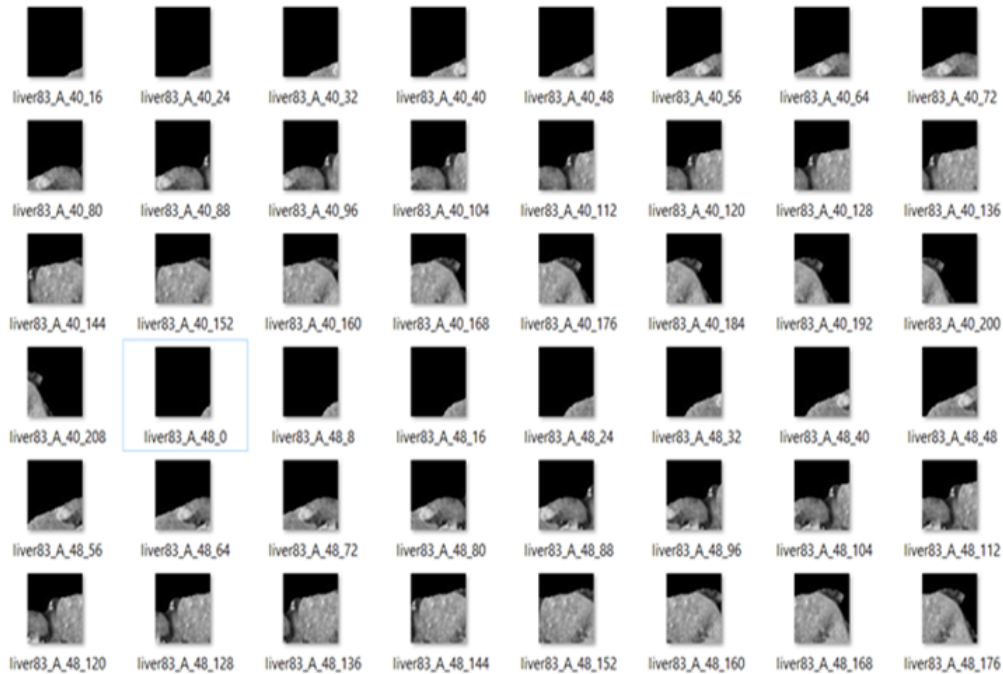


FIGURE VI.13 – Exemple d’extraction et d’enregistrement des fragments d’une image IRM

L’algorithme prend comme entrée le chemin qui englobe les images à traiter, il cherche alors toutes les images disponibles dans ce chemin et divise ensuite chaque image en fragments en se basant sur la taille du fragment désirée et le choix du pas. Ensuite il crée un dossier avec le même nom de l’image et enregistre chaque fragment dans ce dossier avec un nom composé du nom de l’image concaténé avec l’abscisse et l’ordonnée de ce dernière dans l’image. Durant cette partie, nous utilisons les différents fragments résultants de l’étape précédente pour les trois phases de l’injection de contraste afin de créer les bases de données. Chaque base est constituée de deux dossiers : un pour la partie d’entraînement (TrainingSet) et l’autre pour la phase de validation (TestSet), chaque dossier est divisé à son tour en deux autres dossiers. Les deux derniers dossiers identifient les deux classes de notre algorithme de *deep learning* (fragments cancéreux et fragments non cancéreux).

Le modèle créé est une suite de blocs de convolution interconnectés destinés à l’extraction et l’identification des motifs images. Afin de créer le modèle et implémenter la base des fragments, nous initialisons les différentes variables pertinentes dans notre algorithme : le nombre d’epochs, la valeur d’erreur et le *batch size*. Les valeurs de poids et des biais sont initialisées d’une façon aléatoire. La dernière étape présente l’entraînement et l’enregistrement du modèle (.h5). Les Figures [VI.14](#) et [VI.15](#) illustrent le déroulement de la phase d’entraînement sur une machine (en utilisant Visual code ainsi que sur le Google Colab.

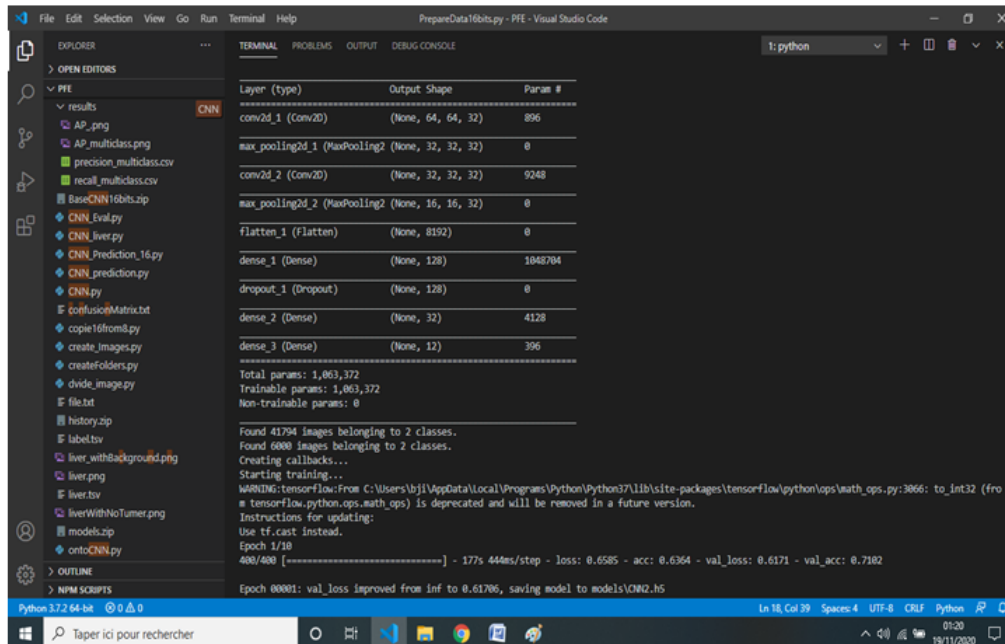


FIGURE VI.14 – Exécution de l’algorithme de *deep learning* sur une machine

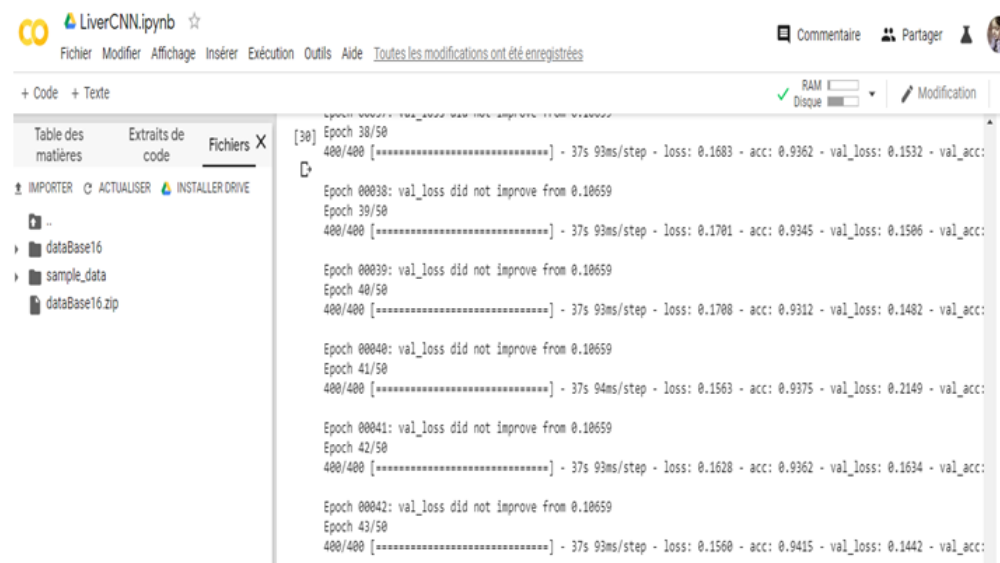


FIGURE VI.15 – Exécution de l’algorithme de *deep learning* sur Google Colab

Après la phase d’entraînement et l’enregistrement du modèle, nous avons préparé une nouvelle image pour la prédiction. L’image est divisée en des fragments de 64x64 sans recouvrement puis nous avons prédit chaque fragment indépendamment des autres. Comme montre la Figure [V.12](#), la valeur 0 est attribuée à un fragment sain et la valeur 1 indique que le fragment contient potentiellement une tumeur, puisque chaque nom de fragment comporte son abscisse et son ordonné dans

l'image source, nous pouvons construire cette matrice où chaque valeur représente la prédiction du fragment correspondant. La Figure VI.16 montre un exemple d'exécution du CNN multi-label qui donne comme sortie le type du fragment ainsi qu'un vecteur de valeur mentionnant des descriptions sur la tumeur détectée.

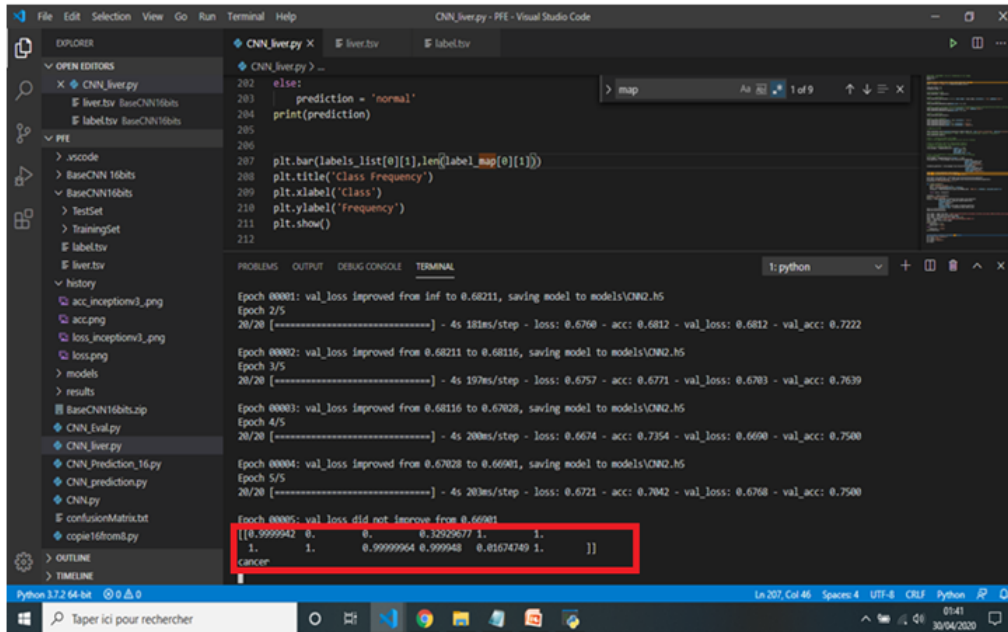


FIGURE VI.16 – Exemple d'exécution d'un modèle CNN multi-label

VI.4 Evaluation et validation des résultats

Dans cette section, nous exposons les expérimentations réalisées pour l'évaluation des approches développées. Nous présentons en premier lieu l'évaluation de l'ontologie OntHCC et en deuxième lieu l'évaluation des algorithmes de *deep learning*.

VI.4.1 Evaluation du Prototype OntHCC

Dans cette sous-section, nous expliquons les étapes suivies pour réaliser la troisième phase de la méthodologie « *MethOntology* » qui est la validation de l'ontologie. Pour évaluer l'ontologie OntHCC, nous avons appliqué l'une des techniques les plus courantes [115]. Elle consiste à comparer les concepts de OntHCC aux autres concepts créés dans d'autres ontologies et langues. Nous utilisons cette technique pour indiquer la couverture globale de notre ontologie OntHCC par rapport aux autres modèles ONLIRA [72], LiCO [29] et Liver Cancer Ontology [92]. Nous tenons en compte tous les concepts présentant surtout les tumeurs hépatiques, les symptômes, les facteurs de risques, les observations radiologiques, les descriptions des tumeurs CHC, *etc.* Les résultats de cette évaluation sont indiqués dans les Tableaux VI.1 et VI.2 présentant respectivement les évaluations quantitative et qualitative.

TABLE VI.1 – Evaluation Quantitative

Modèle ontologique	Concepts	Concepts couverts par OntHCC	Taux de couverture
ONLIRA	11	8	72%
Liver Cancer Ontology	23	23	100%
LiCO	20	15	75%

TABLE VI.2 – Evaluation Qualitative

Modèle Ontologique	Instances des concepts couverts par OntHCC	Instances des concepts non couverts par OntHCC	Taux de couverture
ONLIRA	311	59	84%
LiCO	246	65	80%

Pour chacune de ces ontologies [72], LiCO [29] et Liver Cancer Ontology [92], nous avons choisi les concepts les plus adaptés avec les modules de détection et de traitement des lésions hépatiques. Nous avons éliminé le module de la stadification car la majorité des rapports médicaux que nous avons utilisés ne fournissent pas des informations sur les stades des tumeurs. Donc, nous n'avons pas pris ces concepts reliés à notre modèle dans l'étude comparative. Nous avons choisi respectivement 23 concepts de [92], 20 de LiCO et 11 d'ONLIRA. Notre étude consiste à remplir les quatre ontologies avec des instances prises des rapports radiologiques. Les instances obtenues ont été collectées pour faire l'évaluation et l'étude comparative. En ce qui concerne l'évaluation quantitative (voir Tableau VI.1), nous constatons que l'OntHCC couvre 100% du modèle proposé par [92]. Pour les autres ontologies, nous trouvons qu'OntHCC couvre ONLIRA et LiCO respectivement de 72% et 75% de nos concepts. Pour évaluer quantitativement notre modèle, nous avons utilisé seulement les modèles qui contiennent des concepts non couverts par notre ontologie (ONLIRA et LiCO). Nous avons pris des exemples de rapports médicaux pour réaliser cette évaluation qualitative. Le Tableau VI.2 montre les mesures de cette évaluation. Les résultats obtenus montrent que les mesures quantitatives sont supérieures aux mesures qualitatives. Nous constatons ainsi que la majorité des instances sont couvertes par OntHCC. Par la suite, notre modèle est plus efficace quantitativement puisqu'il contient le plus grand nombre de concepts caractérisant la détection et le traitement de CHC.

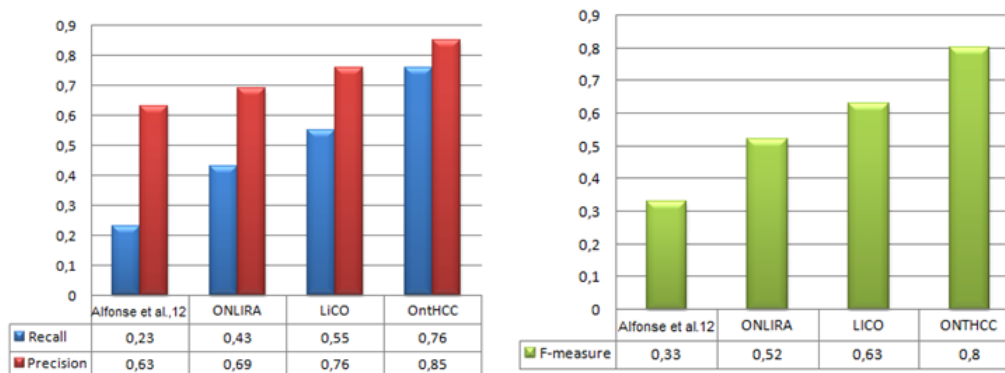


FIGURE VI.17 – Evaluation du prototype OntHCC

La deuxième partie de l'évaluation consiste à faire une étude de cas pour montrer l'efficacité de notre modèle. Le rappel, la précision et le F-mesure sont les mesures appliquées pour identifier le taux de pertinence pertinente des instances au modèles OntHCC. Nous avons utilisé ces instances pour évaluer OntHCC par rapport à d'autres modèles. Le but de ces tests est de trouver le taux de validité des concepts instanciés. Les résultats obtenus (voir Figure VI.17) montrent qu'OntHCC contient les valeurs de rappel 76%, précision 85% et F-mesure 80% les plus élevées.

Pour élargir notre évaluation, nous avons cherché à comparer notre approche aux autres études. Nous avons constaté que toutes les méthodes proposées discutent le sujet de la représentation sémantique des données cliniques en particulier pour les maladies du foie. La plupart des travaux proposés traite le processus de diagnostic des cancers du foie en se basant sur les examens radiologiques (IRM, ou TDM, *etc*). Ces travaux fournissent des modèles ontologiques en utilisant le langage OWL. Ils visent à élargir la base de connaissances efficaces à des fins cliniques. Ils fournissent des ontologies pour l'interprétation des images médicales [33], [39], [99], [82], et [128]. D'une part, quelques travaux mettent l'accent sur l'extraction des caractéristiques de l'image médicale en proposant des prototypes par exemple OntoVIP [82], Reportviewer [39], et VHOSWS [128]. D'autre part, les autres travaux se concentrent sur le processus de stadification en intégrant des systèmes de stadification comme BCLC [35] et TNM [92], [98], et [94]. Le module de traitement est également traité dans ces travaux. De plus, plusieurs techniques ont été appliquées pour améliorer l'efficacité de ces travaux par exemple tester avec des données réelles, interroger les modèles ontologiques avec des requêtes de recherche sémantique (par exemple, règles SPARQL, règles SWRL) [39], [98], et [82] tandis que d'autres travaux tentent d'appliquer la norme mesures lors de l'étape de validation telles que le rappel et la précision [35].

Pour récapituler, malgré l'efficacité de ces approches, peu de ces travaux ont proposé un modèle sémantique complet pour le CHC. Cela confirme l'originalité de notre approche. En effet, il réalise trois missions principales : la détection de CHC à partir de l'image médicale, la stadification de CHC en utilisant les systèmes fréquemment utilisés par les radiologues tels que BCLC et TNM. De plus, notre approche a le potentiel d'intégrer le système Li-RADS en particulier dans la phase de détection. Il pourrait éliminer les erreurs de détection en offrant des informations complémentaires sur la lésion hépatique. En plus, les interfaces développées en Java pourraient être utilisées facilement par les radiologues.

TABLE VI.3 – Description de la base d’apprentissage

Base de données (nombre de fragments)	Normal	cancer	Total
Training Data	26,162	15,632	41,794
Testing Data	3,588	2,412	6,000
Total	29,750	18,044	47,794

TABLE VI.4 – Métrique de performance

TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
SEN	$TP/TP+FN$
ACC	$TP/TP+FP$
SPC	$TN/(TN+FP)$
F1 Score	$(2*Recall * Precision) / (Recall + Precision)$

VI.4.2 Evaluation des algorithmes de deep learning

Pour faire l’évaluation des algorithmes de *deep learning* proposés, nous avons utilisé des images IRM pour différents cas de patients. La base de données appliquée contient un ensemble de fragments répartis comme suit (voir Tableau VI.3); pour la phase d’entraînement, nous avons utilisé 26 162 patches normaux et 15 632 patches cancéreux. Pour la phase de test, nous avons utilisé 3 588 patches normaux et 2 412 patches cancéreux. De cette manière, nous avons au total 47 794 fragments répartis en 29 750 patches normaux et 18 044 patches cancéreux. Chaque patch est de taille 64×64 pixels.

VI.4.2.1 Métriques de performance

Parmi les mesures de performance qui permettent l’évaluation des systèmes d’apprentissage profond, nous trouvons les valeurs TN, TP, FN, FP; la sensibilité (SEN), le taux de bonne classification ou accuracy (ACC), la spécificité (SPC) et le score F1. Le Tableau VI.4 présente la signification de ces mesures. Nous avons calculé ces valeurs pour les deux algorithmes proposés : l’algorithme standard CNN et le nouvel algorithme CNN multi-label. Nous avons fait également une autre étude de comparaison avec les travaux précédents de l’équipe en utilisant respectivement les techniques SVM [48] et U-Net [49]. Il n’était pas facile de comparer avec les autres travaux connexes car nous n’avons pas travaillé avec le même type de données. Plus de détails sur les résultats obtenus sont mentionnés dans les sous-sections suivantes.

VI.4.2.2 Résultats sans couplage

Le modèle de la détection des lésions CHC proposé a été implémenté avec Python en utilisant une machine avec les configurations système suivantes : un Processeur i7 avec une RAM de 8 Go. L'exécution de l'algorithme *deep learning* nécessite un temps de calcul très important, nous avons utilisé le *framework* TensorFlow pour accélérer le temps de calcul. L'exécution de l'algorithme avec Google Colaboratory Notebook a atteint 98% de précision (voir Figure VI.18(a)) et le taux de perte est de 0,03 (voir Figure VI.18(b)).

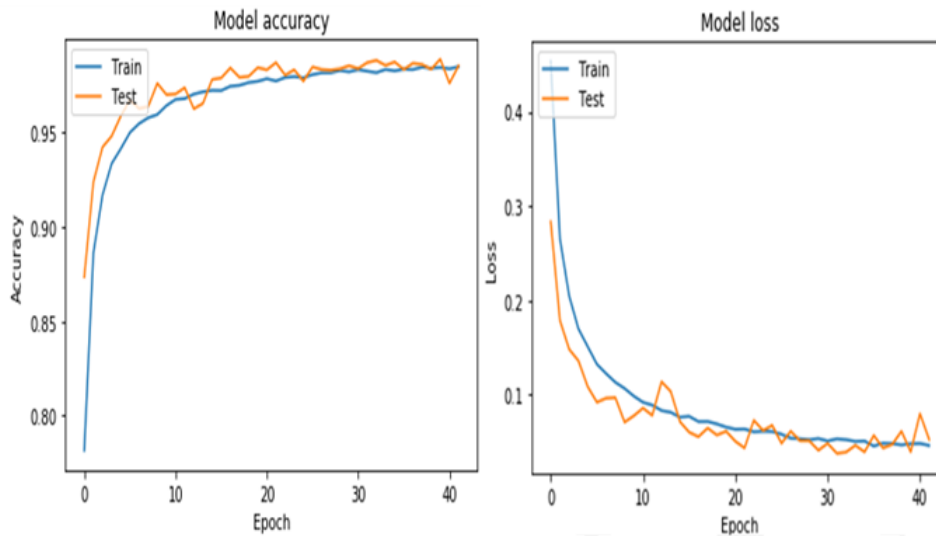


FIGURE VI.18 – Le modèle CNN : (a) taux de précision, (b) taux d'erreur

Pour faire l'évaluation de ce modèle, nous avons utilisé des données issues de 9 patients. Les résultats obtenus sont illustrés respectivement dans les tableaux VI.5 et VI.6 pour les deux phases 2 (avec contraste) et 3 (après contraste). Pour tester l'efficacité de notre modèle, nous avons utilisé les mêmes données que celles utilisés dans les travaux précédents de l'équipe [48] et [49]. La performance du modèle CNN proposé est déterminée par sa capacité à détecter les fragments cancéreux et normaux dans chaque image IRM. Nous avons utilisé les mêmes critères d'évaluation que nos travaux précédents. Le modèle proposé a fourni des meilleurs résultats de classification par rapport aux anciens travaux de l'équipe. En effet, la précision a clairement indiqué que l'algorithme proposé est profondément efficace dans la détection des fragments CHC. Pour la phase 2, les résultats atteints sont : 80% de SEN, 71% d'ACC, 92% de SPC et 74% en score F1. Pour la phase 3, les résultats atteints sont 81% de SEN, 68% d'ACC, 90% de SPC et 73% de score F1. La précision totale a atteint 90% durant les deux phases. En comparant ces résultats avec les travaux mentionnés dans [48] et [49], nous constatons que le score F1 a été augmenté de 66,5% dans [49] à 74% dans notre travail. De plus, la valeur de la sensibilité a été augmentée en phase 3 de 76% dans [48] à 81%. De plus, la valeur de la spécificité a été augmentée respectivement dans les deux phases 2 et 3 de 87% et 82% à 92% et 90%.

TABLE VI.5 – Evaluation durant la phase 2 (avec contraste)

	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5	Patient 6	Patient 7	Patient 8	Patient 9	Total
TP	6	3	6	30	116	5	98	27	163	454
TN	65	17	38	169	454	40	503	43	628	1957
FP	14	0	5	11	22	5	70	2	31	160
FN	4	0	6	0	10	0	65	5	15	105
SEN	0.5	1	0.5	1	0.92	1	0.6	0.84	0.91	0.80
ACC	0.5	1	0.5	0.73	0.84	0.5	0.58	0.93	0.84	0.71
SPC	0.87	1	0.88	0.93	0.95	0.88	0.87	0.95	0.95	0.92
F1 Score	0.5	1	0.5	0.84	0.87	0.66	0.59	0.88	0.87	0.74

TABLE VI.6 – Evaluation durant la phase 3 (après contraste)

	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5	Patient 6	Patient 7	Patient 8	Patient 9	Total
TP	6	3	6	17	111	5	109	24	159	440
TN	67	17	38	62	454	40	537	44	618	1877
FP	12	0	5	9	21	5	51	6	32	141
FN	4	0	6	1	14	0	48	3	27	103
SEN	0.6	1	0.5	0.94	0.88	1	0.69	0.88	0.85	0.81
ACC	0.33	1	0.54	0.65	0.84	0.5	0.68	0.8	0.83	0.68
SPC	0.84	1	0.88	0.87	0.95	0.88	0.91	0.88	0.95	0.90
F1 Score	0.42	1	0.52	0.77	0.86	0.66	0.68	0.84	0.84	0.73

VI.4.2.3 Résultats avec couplage

Pour l'évaluation du CNN multi-label, nous avons calculé les mesures suivantes : la sensibilité, la spécificité et la précision totale. Nous avons testé également avec des patients IRM infectés par le CHC. Nous avons exécuté l'algorithme plusieurs fois pour vérifier son efficacité par rapport au CNN standard précédent. Pour chaque exécution, nous avons modifié les paramètres *deep learning* tels que la taille du batch, le nombre d'époques, *etc.* et nous enregistrons la précision et la valeur de perte pour voir l'efficacité de l'approche proposée. La Figure VI.19 expose la précision et le taux de perte obtenus pour le CNN multi-label. Le modèle proposé atteint 98% en précision et 0,003 comme taux d'erreur.

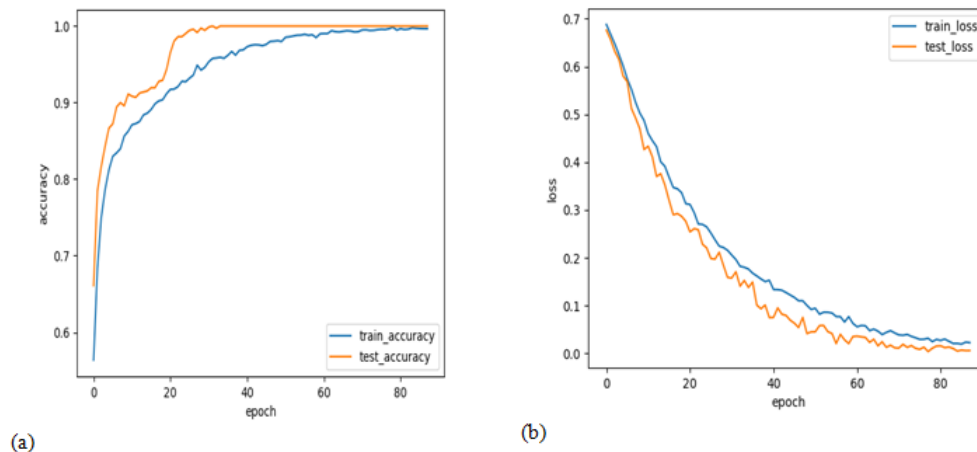


FIGURE VI.19 – Le modèle CNN multi-label : (a) taux de précision, (b) taux d'erreur

Une comparaison des deux algorithmes est présentée dans les tableaux VI.7, VI.9 et VI.11. En ce qui concerne les valeurs présentées dans le Tableau VI.7, le CNN multi-label est plus performant que le CNN standard. Le modèle CNN atteint 80% de SEN, 74% de SPC, 93% d'ACC et 77% de F1-score. Tandis que le CNN multi-label atteint 82% de SEN, 95% d'ACC, 84% de SPC et 85% de F1-score. De plus, le nombre des faux positifs (FP) et des faux négatifs (FN) a été diminué pour le CNN multi-label comme mentionné dans le Tableau VI.9. En ce qui concerne la précision totale (voir Tableau VI.11), elle augmente de 90% à 93% dans le cas CNN multi-label (dernière itération). Cette méthode est également plus performante que les algorithmes développés par notre équipe qui utilisent les techniques SVM et U-Net [48] et [49]. En effet, le score F1 augmente de 66,5% et 74% en considérant respectivement [48] et [49] à 84 % grâce au CNN multi-label. De plus, en ce qui concerne le taux de sensibilité, nous constatons qu'il est passé de 82% et 79% en considérant respectivement [48] and [49] à 85%. Enfin, les taux de spécificité sont passés de 87,6% et 83% en considérant respectivement [48] et [49] à 95% dans notre cas.

TABLE VI.7 – Comparaison des modèles CNN et CNN multi-label

CNN				Multi-label CNN			
SEN	ACC	SPC	F1-score	SEN	ACC	SPC	F1-score
0.8	0.74	0.93	0.77	0,85	0,82	0,95	0,84

TABLE VI.9 – Taux des TP, TN, FP et FN des modèles CNN et CNN multi-label

Iteration	CNN				Multi-label CNN			
	TP	TN	FP	FN	TP	TN	FP	FN
1	299	1802	360	340	281	1869	306	294
2	310	1840	310	340	290	1860	380	320
3	305	1895	350	250	295	1805	295	255
4	298	1802	402	298	301	1799	297	293
5	302	1798	300	280	305	1795	290	210
6	310	1890	300	250	305	1940	229	216
7	300	2000	270	230	302	1938	231	219
8	300	1900	270	200	313	1937	180	190
9	302	2119	199	150	320	1880	170	130
10	452	2051	154	113	501	2082	105	82

TABLE VI.11 – Taux de précision des modèles CNN et CNN multi-label

Iteration	CNN	Multi-label CNN
	Total Acc	Total Acc
1	0.77	0.76
2	0.76	0.75
3	0.78	0.79
4	0.77	0.78
5	0.78	0.8
6	0.80	0.83
7	0.82	0.83
8	0.83	0.85
9	0.87	0.88
10	90.36	93,24

VI.5 Conclusion

Dans ce chapitre, nous avons présenté nos contributions pratiques à travers les prototypes développés. En premier lieu, nous avons proposé les architectures fonctionnelles de chaque prototype en mentionnant pour chacune les démarches pratiques suivies pour leur développement. Afin d'automatiser le plus possible ces étapes, nous nous sommes appuyés sur les mécanismes de raisonnement et d'inférence à travers les règles SWRL.

OntHCC et MROnt sont conçus pour détecter et analyser les tumeurs CHC. Elles sont alimentées par des données médicales réelles. La mise en œuvre nécessite l'utilisation de plusieurs outils informatiques. Dans la première section, nous avons énuméré les outils techniques appliqués pour l'implémentation des prototypes. Dans deuxième section, nous avons détaillé les étapes menées dans l'implémentation. La troisième section présente les évaluations effectuées pour montrer la validation et la faisabilité des modèles proposés.

Ce chapitre décrit également une méthode de couplage entre les ontologies et le *deep learning* en proposant le modèle CNN multi-label. Ce travail est une approche nouvelle par rapport à de nos travaux précédents qui ont proposé des algorithmes d'apprentissage automatique pour la détection du CHC à partir de séquences IRM. De plus, ce travail repose sur l'utilisation de techniques d'ingénierie des connaissances mises à profit pour soutenir le processus de classification. Nous avons intégré des prototypes OntHCC et MROnt pour enrichir le processus de classification. Nous avons testé cet algorithme avec des données réelles issues d'une cohorte de 9 patients. Les résultats ont montré l'efficacité de la stratégie de classification proposée par rapport à nos travaux précédents.

CONCLUSION GÉNÉRALE

Sommaire

VII.1 Bilan des contributions	137
VII.2 Perspectives	138

Le cancer du foie est considéré parmi les maladies les plus mortelles dans le monde. Le développement de ce cancer survient souvent au cours de l'évolution d'une maladie chronique du foie telles que la cirrhose ou l'hépatite B ou C. Son importance en fait de cette tumeur un sujet d'étude intéressant pour plusieurs travaux de recherche. Le CHC (Carcinome Hépatocellulaire) est le cancer hépatique primitif le plus courant. Dans cette thèse, nous nous sommes intéressés à la détection des tumeurs CHC à partir des images médicales IRM plus précisément à partir des séquences DCE-MRI. En outre, nous avons présenté en premier lieu des approches qui servent à l'optimisation du diagnostic des tumeurs du foie à l'aide des modèles sémantiques. En deuxième lieu, nous avons cherché à intégrer ces connaissances dans le processus d'apprentissage pour enrichir un modèle *deep learning* de type CNN.

VII.1 Bilan des contributions

En raison de sa performance, l'ontologie a été largement appliquée pour la réduction de la complexité de plusieurs problèmes médicaux. Dans cette thèse, nous avons créé des approches ontologiques afin de mettre en œuvre la représentation des lésions hépatiques. Ces ontologies servent à caractériser les tumeurs CHC à travers des modélisations sémantiques. En fait, nous nous sommes appuyés sur les règles SWRL pour ajouter des modules de raisonnement et d'inférences et pour améliorer les résultats des requêtes et donc pour mieux modéliser les tumeurs CHC. Nous avons proposé des modèles ontologiques pour la caractérisation et la stadification du CHC *via* l'imagerie médicale (IRM dans notre cas). En outre, elles sont alimentées par des données cliniques pour enrichir la base de connaissances des lésions hépatiques. La deuxième partie de cette thèse porte sur la détection et la classification des tumeurs CHC à partir des images IRM. Ces algorithmes sont fondés sur l'apprentissage profond. Par la suite, nous avons intégré les connaissances sémantiques dans les algorithmes de détection et classification des images médicales. Comme synthèse de nos travaux, nous avons réalisé les contributions suivantes :

- Approche ontologique pour la stadification des tumeurs hépatiques : Proposer un prototype ontologique « OntHCC » permettant la stadification des lésions CHC en appliquant des systèmes connus dans les routines cliniques telles que BCLC, TNM et Li-RADS. Ce modèle vise à détecter les nodules hépatiques caractérisant le CHC en se basant sur des cas réels de patients infectés. Il offre également des outils informatiques d'aide au diagnostic des tumeurs hépatiques. Nous avons créé également les règles SWRL pour enrichir les ontologies créées afin d'assurer les notions de raisonnement et d'inférence.
- Approche ontologique pour l'analyse des rapports radiologiques : Proposer un prototype ontologique nommé « MROnt » qui offre une modélisation sémantique des observations cliniques issues des rapports radiologiques des patients infectés par le CHC. Nous avons intégré les notes écrites par les radiologues pour enrichir l'ontologie développée.
- Approche basée sur le *deep learning* pour la détection des tumeurs CHC à partir des images médicales : Proposer une approche de classification des images médicales en appliquant

l'architecture CNN. Nous avons étendu l'algorithme proposé en intégrant les connaissances médicales offertes par les ontologies développées dans le processus de classification. Nous avons cherché à coupler les techniques d'ingénierie de connaissances pour améliorer les performances de l'algorithme CNN proposé.

- Développement et évaluation des approches proposées : Pour la mise en pratique, nous avons exposé les architectures fonctionnelles tout en présentant l'environnement logiciel appliqué pour chacune de ces approches. Par la suite, nous avons passé à l'évaluation et la validation de ces contributions.

VII.2 Perspectives

Dans nos futurs travaux :

- Nous visons à étendre notre base de données avec plus de séquences IRM provenant de différentes sources médicales. En plus, nous souhaitons travailler sur d'autres maladies hépatiques. Cela permet d'assurer la flexibilité des ontologies OntHCC et MROnt et d'améliorer l'extraction des informations cliniques intégrées pour bien détecter ces lésions hépatiques.
- Nous avons également l'intention de formaliser davantage notre ontologie en proposant d'autres règles SWRL pour que la recherche soit plus pertinente. De plus, intégrer plus de règles de raisonnement caractérise une piste de recherche très importante à mieux détecter et stadifier les nodules de type CHC. La représentation des connaissances *via* les ontologies et le *deep learning* peuvent caractériser des modèles utilisés par des systèmes de raisonnement logique pour produire de nouveaux vocabulaires sémantiques utilisés comme référence dans le domaine médical.
- Nous souhaitons approfondir le couplage entre le *deep learning* et les ontologies en utilisant d'autres méthodes. Nous prévoyons également d'appliquer la stratégie proposée dans différents systèmes intelligents permettant de soutenir les décisions cliniques. Le fait de combiner les deux approches avec plus de données médicales pourrait être une piste de recherche très importante pour améliorer la détection des lésions hépatiques.
- Nous visons aussi à évaluer notre contribution pour d'autres domaines. Nous comptons aussi tester nos algorithmes sur d'autres images médicales autres que l'IRM (TDM par exemple) pour plusieurs examens successifs.

Liste des publications

Dans des revues internationales à comité de lecture

- Rim Messaoudi, Faouzi Jaziri, Achraf Mtibaa, Manuel Grand-Brochier, Hawa Mohamed Ali, Ali Amouri, Hela Fourati, Pascal Chabrot, Faïez Gargouri, Antoine Vacavant : Ontology-Based Approach for Liver Cancer Diagnosis and Treatment. J. Digit. Imaging 32(1) : 116-130 (2019) (**Quartile=Q1, SJR=0.97**)
- Rim Messaoudi, Achraf Mtibaa, Antoine Vacavant, Faïez Gargouri, Faouzi Jaziri : Ontologies for Liver Diseases Representation : A Systematic Literature Review. J. Digit. Imaging 33(3) : 563-573 (2020) (**Quartile=Q1, SJR=0.97**)
- Rim Messaoudi, Faouzi Jaziri, Achraf Mtibaa, Faïez Gargouri, Antoine Vacavant : “ Ontology-Driven Approach for Liver MRI classification and HCC detection ”, International Journal of Pattern Recognition and Artificial Intelligence, 2020 (**Quartile=Q3, SJR=0.34**)

Dans les conférences internationales à comité de lecture et avec actes

- Rim Messaoudi, Taher Labidi, Antoine Vacavant, Faïez Gargouri, Manuel Grand-Brochier, Ali Amouri, Hela Fourati, Achraf Mtibaa, Faouzi Jaziri : An Ontological Model for Analyzing Liver Cancer Medical Reports. EMCIS 2018 : 369-382 (**Quartile=Q1, SJR=0.97**)
- Rim Messaoudi, Faouzi Jaziri, Antoine Vacavant, Achraf Mtibaa, Faïez Gargouri : A Novel Deep Learning Approach for Liver MRI Classification and HCC Detection. ICPRAI 2020 : 635-645 (**Conférence Internationale à comité de lecture**)

Bibliographie

- [1] Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge acquisition*, 5(2), 199-220.
- [2] Guarino, N., & Welty, C. (2000, October). A formal ontology of properties. In *International Conference on Knowledge Engineering and Knowledge Management* (pp. 97-112). Springer, Berlin, Heidelberg.
- [3] Neches, R., Fikes, R. E., Finin, T., Gruber, T., Patil, R., Senator, T., & Swartout, W. R. (1991). Enabling technology for knowledge sharing. *AI magazine*, 12(3), 36-36.
- [4] Guarino, N., & Giarretta, P. (1995). Ontologies and knowledge bases. *Towards very large knowledge bases*, 1-2.
- [5] Guarino, N. (Ed.). (1998). *Formal ontology in information systems : Proceedings of the first international conference (FOIS'98)*, June 6-8, Trento, Italy (Vol. 46). IOS press.
- [6] Guarino, N. (1997). Understanding, building and using ontologies. *International journal of human-computer studies*, 46(2-3), 293-310.
- [7] Sowa, J. F. (1995). Top-level ontological categories. *International journal of human-computer studies*, 43(5-6), 669-685.
- [8] Jin, L., Chen, W., Hayashi, Y., Ikeda, M., Mizoguchi, R., Takaoka, Y., & Ohta, M. (1999). An ontology-aware authoring tool-Functional structure and guidance generation. In *Proc. of AI-ED 99, Le Mans*.
- [9] Inaba, A., Supnithi, T., Ikeda, M., Mizoguchi, R., & Toyoda, J. I. (2000, June). How can we form effective collaborative learning groups?. In *International Conference on Intelligent Tutoring Systems* (pp. 282-291). Springer, Berlin, Heidelberg.
- [10] Maedche, A., & Staab, S. (2001). Ontology learning for the semantic web. *IEEE Intelligent systems*, 16(2), 72-79.
- [11] Baneyx, A. (2007). *Construire une ontologie de la Pneumologie Aspects théoriques, modèles et expérimentations* (Doctoral dissertation, Université Pierre et Marie Curie-Paris VI).

- [12] Tartir, S., Arpinar, I. B., & Sheth, A. P. (2010). Ontological evaluation and validation. In *Theory and applications of ontology : Computer applications* (pp. 115-130). Springer, Dordrecht.
- [13] Brewster, C., Alani, H., Dasmahapatra, S., & Wilks, Y. (2004). Data driven ontology evaluation.
- [14] Gruninger, M., & Fox, M. S. (1994). The design and evaluation of ontologies for enterprise engineering. In *Workshop on Implemented Ontologies, European Conference on Artificial Intelligence (ECAI)*.
- [15] Uschold, M., & King, M. (1995). Towards a methodology for building ontologies (pp. 19-1). Edinburgh : Artificial Intelligence Applications Institute, University of Edinburgh.
- [16] Schreiber, G., Wielinga, B., & Jansweijer, W. (1995, August). The KACTUS view on the 'O'word. In *IJCAI workshop on basic ontological issues in knowledge sharing* (pp. 159-168).
- [17] Fernandez, M., Gomez-Perez, A., & Juristo, N. (1997, March). Methontology : from ontological art towards ontological engineering. In *Proceedings of the AAAI97 spring symposium series on ontological engineering* (pp. 33-40).
- [18] Horrocks, I., Parsia, B., Patel-Schneider, P., & Hendler, J. (2005, September). Semantic web architecture : Stack or two towers ?. In *International Workshop on Principles and Practice of Semantic Web Reasoning* (pp. 37-41). Springer, Berlin, Heidelberg.
- [19] Bak, J., Jedrzejek, C., & Falkowski, M. (2009, November). Usage of the Jess engine, rules and ontology to query a relational database. In *International Workshop on Rules and Rule Markup Languages for the Semantic Web* (pp. 216-230). Springer, Berlin, Heidelberg.
- [20] Clouard, R., Elmoataz, A., Porquet, C., & Revenu, M. (1995). BORG : a knowledge-based system for the automation of image segmentation tasks.
- [21] Smola, A., & Vishwanathan, S. V. N. (2008). *Introduction to machine learning*. Cambridge University, UK, 32(34), 2008.
- [22] Deo, R. C. (2015). Machine learning in medicine. *Circulation*, 132(20), 1920-1930.
- [23] Kononenko, I. (2001). Machine learning for medical diagnosis : history, state of the art and perspective. *Artificial Intelligence in medicine*, 23(1), 89-109.
- [24] Svozil, D., Kvasnicka, V., & Pospichal, J. (1997). Introduction to multi-layer feed-forward neural networks. *Chemometrics and intelligent laboratory systems*, 39(1), 43-62.
- [25] Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., et al. (2007). The OBO Foundry : coordinated evolution of ontologies to support biomedical data integration. *Nature biotechnology*, 25(11), 1251-1255.
- [26] Bertaud-Gounot, V., Duvauferrier, R., & Burgun, A. (2012). Ontology and medical diagnosis. *Informatics for Health and Social Care*, 37(2), 51-61.
- [27] Van Soest, J., Lustberg, T., Grittner, D., Marshall, M. S., Persoon, L., Nijsten, B., et al. (2014). Towards a semantic PACS : Using Semantic Web technology to represent imaging data. *Studies in health technology and informatics*, 205, 166.

- [28] Caputo, B., Müller, H., Martinez-Gomez, J., Villegas, M., Acar, B., Patricia, N., et al. (2014, September). ImageCLEF 2014 : Overview and analysis of the results. In International Conference of the Cross-Language Evaluation Forum for European Languages (pp. 192-211). Springer, Cham.
- [29] del Mar Roldán-García, M., Uskudarli, S., Marvasti, N. B., Acar, B., & Aldana-Montes, J. F. (2018). Towards an ontology-driven clinical experience sharing ecosystem : Demonstration with liver cases. *Expert Systems With Applications*, 101, 176-195.
- [30] Wang, Y., Lin, Z., Liu, Z., Harris, S., Kelly, R., Zhang, J., et al. (2013). A unifying ontology to integrate histological and clinical observations for drug-induced liver injury. *The American journal of pathology*, 182(4), 1180-1187.
- [31] Fung, K. W., & Bodenreider, O. (2019). Knowledge representation and ontologies. In *Clinical research informatics* (pp. 313-339). Springer, Cham.
- [32] Ramadier, L., & Lafourcade, M. (2015, June). Augmentation d'index par propagation sur un réseau lexical Application aux comptes rendus de radiologie. In *TALN : Traitement Automatique des Langues Naturelles*.
- [33] Chan, L. W., Wong, S. C., Chiau, C. C., Chan, T. M., Tao, L., Feng, J., & Chiu, K. W. (2017). Association patterns of ontological features signify electronic health records in liver cancer. *Journal of healthcare engineering*, 2017.
- [34] Chan, L. W., Liu, Y., Chan, T., Law, H. K., Wong, S. C., Yeung, A. P., et al. (2015). PubMed-supported clinical term weighting approach for improving inter-patient similarity measure in diagnosis prediction. *BMC medical informatics and decision making*, 15(1), 1-8.
- [35] Yim, W. W., Kwan, S. W., & Yetisgen, M. (2016). Tumor reference resolution and characteristic extraction in radiology reports for liver cancer stage prediction. *Journal of biomedical informatics*, 64, 179-191.
- [36] Grishman, R., & Sundheim, B. M. (1996). Message understanding conference-6 : A brief history. In *COLING 1996 Volume 1 : The 16th International Conference on Computational Linguistics*.
- [37] Bagga, A., & Baldwin, B. (1998, May). Algorithms for scoring coreference chains. In *The first international conference on language resources and evaluation workshop on linguistics coreference* (Vol. 1, pp. 563-566).
- [38] Luo, X. (2005, October). On coreference resolution performance metrics. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing* (pp. 25-32).
- [39] Oberkampff, H., Zillner, S., Overton, J. A., Bauer, B., Cavallaro, A., Uder, M., & Hammon, M. (2015). Semantic representation of reported measurements in radiology. *BMC medical informatics and decision making*, 16(1), 1-12.
- [40] Marwede, D., Fielding, M., & Kahn, T. (2007). RadiO : a prototype application ontology for radiology reporting tasks. In *AMIA Annual Symposium Proceedings* (Vol. 2007, p. 513). American Medical Informatics Association.

- [41] Harkema, H., Roberts, I., Gaizauskas, R., & Hepple, M. (2005). Information extraction from clinical records. In Proceedings of the 4th UK e-science all hands meeting (pp. 19-22).
- [42] Tommasi, T., Caputo, B., Welter, P., Güld, M. O., & Deserno, T. M. (2009, September). Overview of the CLEF 2009 medical image annotation track. In Workshop of the Cross-Language Evaluation Forum for European Languages (pp. 85-93). Springer, Berlin, Heidelberg.
- [43] Stevens, R., Baker, P., Bechhofer, S., Ng, G., Jacoby, A., Paton, N. W., et al.(2000). TAM-BIS : transparent access to multiple bioinformatics information sources. *Bioinformatics*, 16(2), 184-186.
- [44] Zeiler, M. D. (2012). Adadelta : an adaptive learning rate method. arXiv preprint arXiv :1212.5701.
- [45] Chaouachi, A., Kamel, R. M., & Nagasaka, K. (2010). A novel multi-model neuro-fuzzy-based MPPT for three-phase grid-connected photovoltaic system. *Solar energy*, 84(12), 2219-2229.
- [46] Babic, B. R., Nestic, N., & Miljkovic, Z. (2011). Automatic feature recognition using artificial neural networks to integrate design and manufacturing : Review of automatic feature recognition systems. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing : AI EDAM*, 25(3), 289.
- [47] Le Cun, Y., & Fogelman-Soulié, F. (1987). Modèles connexionnistes de l'apprentissage. *Intellectica*, 2(1), 114-143.
- [48] Pavan, A. L., Benabdallah, M., Lebre, M. A., de Pina, D. R., Jaziri, F., Vacavant, A., et al.(2018, April). A parallel framework for HCC detection in DCE-MRI sequences with wavelet-based description and SVM classification. In Proceedings of the 33rd Annual ACM Symposium on Applied Computing (pp. 14-21).
- [49] Fabijańska, A., Vacavant, A., Lebre, M. A., Pavan, A. L., de Pina, D. R., Abergel, A., et al. (2018, September). U-CatchHCC : an accurate HCC detector in hepatic DCE-MRI sequences based on an U-Net framework. In International Conference on Computer Vision and Graphics (pp. 319-328). Springer, Cham.
- [50] Messaoudi, R., Jaziri, F., Vacavant, A., Mtibaa, A., & Gargouri, F. (2020, October). A Novel Deep Learning Approach for Liver MRI Classification and HCC Detection. In International Conference on Pattern Recognition and Artificial Intelligence (pp. 635-645). Springer, Cham.
- [51] Lee, J. G., Jun, S., Cho, Y. W., Lee, H., Kim, G. B., Seo, J. B., & Kim, N. (2017). Deep learning in medical imaging : general overview. *Korean journal of radiology*, 18(4), 570.
- [52] Li, Y. (2017). Deep reinforcement learning : An overview. arXiv preprint arXiv :1701.07274.
- [53] Suzuki, K. (2017). Overview of deep learning in medical imaging. *Radiological physics and technology*, 10(3), 257-273.

- [54] Barney, B. (2010). Introduction to parallel computing. Lawrence Livermore National Laboratory, 6(13), 10.
- [55] Flynn, M. J. (1972). Some computer organizations and their effectiveness. *IEEE transactions on computers*, 100(9), 948-960.
- [56] Renedo, F. P., de la Revilla Negro, J., & Panero, J. C. (2008). Carcinoma hepatocelular. *Medicine-Programa de Formación Médica Continuada Acreditado*, 10(12), 770-776.
- [57] El-Serag, H. B., & Rudolph, K. L. (2007). Hepatocellular carcinoma : epidemiology and molecular carcinogenesis. *Gastroenterology*, 132(7), 2557-2576.
- [58] Torre, L. A., Siegel, R. L., Ward, E. M., & Jemal, A. (2016). Global cancer incidence and mortality rates and trends—an update. *Cancer Epidemiology and Prevention Biomarkers*, 25(1), 16-27.
- [59] Petrick, J. L., Kelly, S. P., Altekruse, S. F., McGlynn, K. A., & Rosenberg, P. S. (2016). Future of hepatocellular carcinoma incidence in the United States forecast through 2030. *Journal of Clinical Oncology*, 34(15), 1787.
- [60] Costentin, C. E., Sogni, P., Falissard, B., Barbare, J. C., Bendersky, N., Farges, O., & Goutte, N. (2020). Geographical disparities of outcomes of hepatocellular carcinoma in France : the heavier burden of alcohol compared to hepatitis C. *Digestive diseases and sciences*, 65(1), 301-311.
- [61] Goutté, N., Sogni, P., Bendersky, N., Barbare, J. C., Falissard, B., & Farges, O. (2017). Geographical variations in incidence, management and survival of hepatocellular carcinoma in a Western country. *Journal of hepatology*, 66(3), 537-544.
- [62] Tsapis, N. (2017). Agents de contraste pour l'imagerie médicale-Les exemples de l'IRM et de l'ultrasonographie. *médecine/sciences*, 33(1), 18-24.
- [63] Blanc, J. F., Debaillon-Vesque, A., Roth, G., Barbare, J. C., Baumann, A. S., Boige, V., et al. (2021). Hepatocellular carcinoma : French Intergroup Clinical Practice Guidelines for diagnosis, treatment and follow-up (SNFGE, FFCD, GERCOR, UNICANCER, SFCD, SFED, SFRO, AFEF, SIAD, SFR/FRI). *Clinics and Research in Hepatology and Gastroenterology*, 45(2), 101590.]
- [64] Ba-Ssalamah, A., Zacherl, J., Noebauer-Huhmann, I. M., Uffmann, M., Matzek, W. K., Pinker, K., et al. (2009). Dedicated multi-detector CT of the esophagus : spectrum of diseases. *Abdominal imaging*, 34(1), 3-18.
- [65] Llovet, J. M., Di Bisceglie, A. M., Bruix, J., Kramer, B. S., Lencioni, R., Zhu, A. X., et al. (2008). Design and endpoints of clinical trials in hepatocellular carcinoma. *Journal of the National Cancer Institute*, 100(10), 698-711.
- [66] Forner, A., Reig, M. E., de Lope, C. R., & Bruix, J. (2010, February). Current strategy for staging and treatment : the BCLC update and future prospects. In *Seminars in liver disease* (Vol. 30, No. 01, pp. 061-074). Thieme Medical Publishers.
- [67] Di Muzio B, Morgan MA et al. Liver cancer (BCLC staging). *Radiopaedia.org*. Available at <https://radiopaedia.org/articles/barcelona-clinic-liver-cancer-bclc-staging-classification>. Accessed 2005

- [68] Dilts, D. M., Sandler, A., Cheng, S., Crites, J., Ferranti, L., Wu, A., et al. (2008). Development of clinical trials in a cooperative group setting : the eastern cooperative oncology group. *Clinical Cancer Research*, 14(11), 3427-3433.
- [69] Schima, W., & Heiken, J. (2018). LI-RADS v2017 for liver nodules : how we read and report. *Cancer Imaging*, 18(1), 1-11.
- [70] Cerny, M., Olivie, D., Billiard, J. S., Wagner, M., Aubé, C., Lucidarme, O., et al. (2018). LI-RADS pour le diagnostic de carcinome hépatocellulaire en TDM et IRM. *Journal d'imagerie diagnostique et interventionnelle*, 1(3), 195-206.
- [71] Kumar, A., Dyer, S., Kim, J., Li, C., Leong, P. H., Fulham, M., & Feng, D. (2016). Adapting content-based image retrieval techniques for the semantic annotation of medical images. *Computerized Medical Imaging and Graphics*, 49, 37-45.
- [72] Kökciyan, N., Türkay, R., Üsküdarlı, S., Yolum, P., Bakır, B., & Acar, B. (2014). Semantic description of liver CT images : an ontological approach. *IEEE journal of biomedical and health informatics*, 18(4), 1363-1369.
- [73] Marvasti, N. B., Kökciyan, N., Türkay, R., Yazici, A., Yolum, P., Üsküdarlı, S., & Acar, B. (2014). ImageCLEF Liver CT Image Annotation Task 2014. In *CLEF (working notes)* (pp. 329-340).
- [74] Möller, M., Regel, S., & Sintek, M. (2009, May). Radsem : Semantic annotation and retrieval for medical images. In *European Semantic Web Conference* (pp. 21-35). Springer, Berlin, Heidelberg.
- [75] Depeursinge, A., Kurtz, C., Beaulieu, C., Napel, S., & Rubin, D. (2014). Predicting visual semantic descriptive terms from radiological image data : preliminary results with liver lesions in CT. *IEEE transactions on medical imaging*, 33(8), 1669-1676.
- [76] Noble, W. S. (2006). What is a support vector machine ?. *Nature biotechnology*, 24(12), 1565-1567.
- [77] Bownik, M. (2003). Riesz wavelets and generalized multiresolution analyses. *Applied and Computational Harmonic Analysis*, 14(3), 181-194.
- [78] Liu, S., Cai, W., Song, Y., Pujol, S., Kikinis, R., & Feng, D. (2013). A bag of semantic words model for medical content-based retrieval.
- [79] Langlotz, C. P. (2006). RadLex : a new method for indexing online educational materials.
- [80] Seifert, S., Thoma, M., Stegmaier, F., Hammon, M., Kramer, M., Huber, M., et al. (2011, February). Combined semantic and similarity search in medical image databases. In *Medical Imaging 2011 : Advanced PACS-based Imaging Informatics and Therapeutic Applications* (Vol. 7967, p. 796703). International Society for Optics and Photonics.
- [81] Hudelot, C., Atif, J., & Bloch, I. (2008). Fuzzy spatial relation ontology for image interpretation. *Fuzzy Sets and Systems*, 159(15), 1929-1951.
- [82] Gibaud, B., Forestier, G., Benoit-Cattin, H., Cervenansky, F., Clarysse, P., Friboulet, D. et al. (2014). OntoVIP : An ontology for the annotation of object models used for medical image simulation. *Journal of biomedical informatics*, 52, 279-292.

- [83] Gkoutos, G. V., Green, E. C., Mallon, A. M., Hancock, J. M., & Davidson, D. (2005). Using ontologies to describe mouse phenotypes. *Genome biology*, 6(1), 1-10.
- [84] Schofield, P. N., Gkoutos, G. V., Gruenberger, M., Sundberg, J. P., & Hancock, J. M. (2010). Phenotype ontologies for mouse and man : bridging the semantic gap. *Disease models & mechanisms*, 3(5-6), 281-289.
- [85] De Matos, P., Alcántara, R., Dekker, A., Ennis, M., Hastings, J., Haug, K., et al. (2010). Chemical entities of biological interest : an update. *Nucleic acids research*, 38(suppl-1), D249-D254.
- [86] Gurcan, M. N., Tomaszewski, J., Overton, J. A., Doyle, S., Ruttenberg, A., & Smith, B. (2017). Developing the Quantitative Histopathology Image Ontology (QHIO) : A case study using the hot spot detection problem. *Journal of biomedical informatics*, 66, 129-135.
- [87] Othmani, A., Meziat, C., & Loménie, N. (2010, November). Ontology-driven image analysis for histopathological images. In *International Symposium on Visual Computing* (pp. 1-12). Springer, Berlin, Heidelberg.
- [88] Hwang, K. H., Lee, H., Koh, G., Willrett, D., & Rubin, D. L. (2017). Building and querying RDF/OWL database of semantically annotated nuclear medicine images. *Journal of digital imaging*, 30(1), 4-10.
- [89] Rubin, D. L., Mongkolwat, P., Kleper, V., Supekar, K., & Channin, D. S. (2008, March). Medical Imaging on the Semantic Web : Annotation and Image Markup. In *AAAI Spring Symposium : semantic scientific knowledge integration* (pp. 93-98).
- [90] Quinn, J. (1999). An HL7 (health level seven) overview. *Journal of AHIMA*, 70(7), 32-4.
- [91] Abdel-Badeeh, M. S., Mostafa, M. A., Moawed, I. F., Alfonse, M., & AlMarzoki, G. (2016). Ontological engineering in medical informatics. *International Journal of Genomics, Proteomics, Metabolomics & Bioinformatics (IJGPMB)*, 1, 9-13.
- [92] Alfonse, M., Aref, M. M., & Salem, A. B. M. (2014). An ontology-based system for cancer diseases knowledge management. *International Journal of Information Engineering and Electronic Business*, 6(6), 55-63.
- [93] Moawad, I. F., Marzoqi, G. A., & Salem, A. B. M. (2012). Building OBR-based OWL ontology for viral hepatitis. *Egyptian Computer Science Journal, ECS*, 36(1), 89-98.
- [94] Al-Marzoqi, G., Alfonse, M., Moawad, F. I., & Roushdy, M. (2015). A survey on applying ontological engineering : approach for hepatobiliary system diseases. In *The 7th International Conference on Information Technology* (pp. 370-375).
- [95] Masci, A. M., Roach, J., De Bono, B., Grenon, P., & Cowell, L. G. (2011). Bridging Multiple Ontologies : Representation of the Liver Immune Response. In *ICBO*.
- [96] Bruc, N., Magariu, G., & Verlan, T. (2011, November). Elaborating of ultrasound images ontology in ultrasound diagnostics. In *2011 E-Health and Bioengineering Conference (EHB)* (pp. 1-4). IEEE.

- [97] Kafalı, Ö., Sindlar, M., van der Weide, T., & Stathis, K. (2013, August). ORC : an Ontology Reasoning Component for diabetes. In 2nd International Workshop on Artificial Intelligence and Netmedicine (NetMed'13).
- [98] Galal, A. M., Moawad, I. F., & Salem, A. B. M. (2012, December). Web service based approach for viral hepatitis Ontology sharing and diagnosing. In International Conference on Advanced Machine Learning Technologies and Applications (pp. 257-266). Springer, Berlin, Heidelberg.
- [99] Alfonse, M., Aref, M. M., & Salem, A. B. M. (2014). An ontology-based system for cancer diseases knowledge management. *International Journal of Information Engineering and Electronic Business*, 6(6), 55-63.
- [100] Luque, E. F., Rubin, D. L., & Moreira, D. A. (2015, June). Automatic classification of cancer tumors using image annotations and ontologies. In 2015 IEEE 28th International Symposium on Computer-Based Medical Systems (pp. 368-369). IEEE.
- [101] Franca, F., Schulz, S., Bronsert, P., Novais, P., & Boeker, M. (2015, September). Feasibility of an ontology driven tumor-node-metastasis classifier application : A study on colorectal cancer. In 2015 International Symposium on Innovations in Intelligent Systems and Applications (INISTA) (pp. 1-7). IEEE.
- [102] Schulz, S., & Boeker, M. (2013). BioTopLite : An upper level ontology for the life sciences. Evolution, design and application. *INFORMATIK 2013–Informatik angepasst an Mensch, Organisation und Umwelt*.
- [103] Glimm, B., Horrocks, I., Motik, B., Stoilos, G., & Wang, Z. (2014). HermiT : an OWL 2 reasoner. *Journal of Automated Reasoning*, 53(3), 245-269.
- [104] Kaur, P., & Khamparia, A. (2015). Diagnosis of liver cancer ontology using SPARQL. *international journal of applied engineering research*, 10(69), 15-18.
- [105] Starlinger, J., Kittner, M., Blankenstein, O., & Leser, U. (2017). How to improve information extraction from German medical records. *It-Information Technology*, 59(4), 171-179.
- [106] Huang, W. C., Chung, P. C., Tsai, H. W., Chow, N. H., Juang, Y. Z., Tsai, H. H., et al. (2019, March). Automatic HCC detection using convolutional network with multi-magnification input images. In 2019 IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS) (pp. 194-198). IEEE.
- [107] Hamm, C. A., Wang, C. J., Savic, L. J., Ferrante, M., Schobert, I., Schlachter, T., et al. (2019). Deep learning for liver tumor diagnosis part I : development of a convolutional neural network classifier for multi-phasic MRI. *European radiology*, 29(7), 3338-3347.
- [108] Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. (2017, February). Inception-v4, inception-resnet and the impact of residual connections on learning. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 31, No. 1).
- [109] Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net : Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (pp. 234-241). Springer, Cham.

- [110] Donadello, I., & Dragoni, M. (2019, September). Ontology-driven food category classification in images. In International Conference on Image Analysis and Processing (pp. 607-617). Springer, Cham.
- [111] Lazzez, O., Ouarda, W., & Alimi, A. M. (2018). DeepVisInterests : CNN-Ontology Prediction of Users Interests from Social Images. arXiv preprint arXiv :1811.10920.
- [112] Gupta, U., & Chaudhury, S. (2015, December). Deep transfer learning with ontology for image classification. In 2015 Fifth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG) (pp. 1-4). IEEE.
- [113] Filali, J., Zghal, H., & Martinet, J. (2019, February). Ontology and hmax features-based image classification using merged classifiers. In International Conference on Computer Vision Theory and Applications 2019 (VISAPP'19).
- [114] Yan, K., Peng, Y., Sandfort, V., Bagheri, M., Lu, Z., & Summers, R. M. (2019). Holistic and comprehensive annotation of clinically significant findings on diverse CT images : learning from radiology reports and label ontology. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 8523-8532).
- [115] Sabou, M., & Fernandez, M. (2012). Ontology (network) evaluation. In Ontology engineering in a networked world (pp. 193-212). Springer, Berlin, Heidelberg.
- [116] Subramaniam, S., Kelley, R. K., & Venook, A. P. (2013). A review of hepatocellular carcinoma (HCC) staging systems. Chinese clinical oncology, 2(4), 33-33.
- [117] Bachimont, B. (2004). Arts et sciences du numérique : ingénierie des connaissances et critique de la raison computationnelle. Mémoire de HDR.
- [118] Baader, F., Calvanese, D., McGuinness, D., Patel-Schneider, P., & Nardi, D. (Eds.). (2003). The description logic handbook : Theory, implementation and applications. Cambridge university press.
- [119] Bernard E., Introduction aux Ontologies. Professeur à l'Université d'Aix-Marseille, 2010
- [120] Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. Scientific american, 284(5), 34-43.
- [121] Biebow, B., Szulman, S., & Clément, A. J. (1999, May). TERMINAE : A linguistics-based tool for the building of a domain ontology. In International Conference on Knowledge Engineering and Knowledge Management (pp. 49-66). Springer, Berlin, Heidelberg.
- [122] Rubin, D. L., Mongkolwat, P., Kleper, V., Supekar, K., & Channin, D. S. (2008, March). Medical Imaging on the Semantic Web : Annotation and Image Markup. In AAAI Spring Symposium : semantic scientific knowledge integration (pp. 93-98).
- [123] Elbyed, A. (2016). Resources-Based Ontology Matching to Access Learning Resources. International Journal of Computer and Information Engineering, 9(10), 2261-2268.
- [124] Furst, F. (2002). L'ingénierie ontologique, rapport technique. Institut de recherche en Informatique de Nantes.
- [125] Gómez-Pérez, A., Fernández, M., & De Vicente, A. (1996, August). Towards a method to conceptualize domain ontologies. In ECAI96 Workshop on Ontological Engineering, Budapest (pp. 41-51).

- [126] De Ganay, C., & Gillot, D. (2017). Pour une intelligence artificielle maîtrisée, utile et démystifiée. Report, Office parlementaire d'évaluation des choix scientifiques et technologiques.
- [127] Gómez-Pérez, A., Fernández-López, M., & Corcho, O. (2006). *Ontological Engineering : with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web*. Springer Science & Business Media.
- [128] Levy, M. A., O'Connor, M. J., & Rubin, D. L. (2009). Semantic reasoning with image annotations for tumor assessment. In *AMIA Annual Symposium Proceedings (Vol. 2009, p. 359)*. American Medical Informatics Association.
- [129] Mezaris, V., Kompatsiaris, I., & Strintzis, M. G. (2003, September). An ontology approach to object-based image retrieval. In *Proceedings 2003 International Conference on Image Processing (Cat. No. 03CH37429) (Vol. 2, pp. II-511)*. IEEE.
- [130] Mazuel, L. (2007). Utilisation des ontologies pour la modélisation logique d'une commande en langue naturel. *Rencontre des étudiants chercheurs en informatique pour le traitement automatique des langues (RECITAL 2007)*, 427-436.
- [131] Pierra, G. (2008). Context representation in domain ontologies and its use for semantic integration of data. In *Journal on data semantics X (pp. 174-211)*. Springer, Berlin, Heidelberg.
- [132] Porzel, R., & Malaka, R. (2004, August). A task-based approach for ontology evaluation. In *ECAI Workshop on Ontology Learning and Population, Valencia, Spain (pp. 1-6)*. Citeseer.
- [133] Suárez-Figueroa, M. C., Gómez-Pérez, A., Motta, E., & Gangemi, A. (2012). Introduction : Ontology engineering in a networked world. In *Ontology engineering in a networked world (pp. 1-6)*. Springer, Berlin, Heidelberg.
- [134] Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. (2018). Convolutional neural networks : an overview and application in radiology. *Insights into imaging*, 9(4), 611-629.
- [135] Mitchell, D. G., Bruix, J., Sherman, M., & Sirlin, C. B. (2015). LI-RADS (Liver Imaging Reporting and Data System) : Summary, discussion, and consensus of the LI-RADS Management Working Group and future directions. *Hepatology*, 61(3), 1056-1065.
- [136] Elsayes, K. M., Hooker, J. C., Agrons, M. M., Kielar, A. Z., Tang, A., Fowler, K. J., et al. (2017). 2017 version of LI-RADS for CT and MR imaging : an update. *Radiographics*, 37(7), 1994-2017.
- [137] Chabert-Ranwez, S. (2000). *Composition automatique de documents hypermédia adaptatifs à partir d'ontologies et de requêtes intentionnelles de l'utilisateur (Doctoral dissertation, Université Montpellier II-Sciences et Techniques du Languedoc)*.
- [138] Uschold M.(1996) *Converting an Informal Ontology into Ontolingua : Some Experiences*. A slightly abridged version of this paper appears in the *Proceedings of the Workshop on Ontological Engineering held in conjunction with ECAI 96, Budapest*.

- [139] Noy, N. F., & McGuinness, D. L. (2000). Développement d'une ontologie 101 : Guide pour la création de votre première ontologie. Université de Stanford, Stanford, Traduit de l'anglais par Anila Angjeli. <http://www.bnf.fr/pages/infopro/normes/pdf/no-DevOnto.pdf>.
- [140] Sánchez, D., Batet, M., Isern, D., & Valls, A. (2012). Ontology-based semantic similarity : A new feature-based approach. *Expert systems with applications*, 39(9), 7718-7728.
- [141] Smeulders, A. W., Worring, M., Santini, S., Gupta, A., & Jain, R. (2000). Content-based image retrieval at the end of the early years. *IEEE Transactions on pattern analysis and machine intelligence*, 22(12), 1349-1380.
- [142] Smith, L. N. (2017, March). Cyclical learning rates for training neural networks. In 2017 IEEE winter conference on applications of computer vision (WACV) (pp. 464-472). IEEE.
- [143] Makantasis, K., Protopapadakis, E., Doulamis, A., Doulamis, N., & Loupos, C. (2015, September). Deep convolutional neural networks for efficient vision based tunnel inspection. In 2015 IEEE international conference on intelligent computer communication and processing (ICCP) (pp. 335-342). IEEE.
- [144] Rumelhart, D. E., Hinton, G. E., & McClelland, J. L. (1986). A general framework for parallel distributed processing. *Parallel distributed processing : Explorations in the microstructure of cognition*, 1(45-76), 26.
- [145] Edge, S. B., & Compton, C. C. (2010). The American Joint Committee on Cancer : the 7th edition of the AJCC cancer staging manual and the future of TNM. *Annals of surgical oncology*, 17(6), 1471-1474.
- [146] Lebre, M. A., Vacavant, A., Grand-Brochier, M., Rositi, H., Strand, R., Rosier, H., et al. (2019). A robust multi-variability model based liver segmentation algorithm for CT-scan and MRI modalities. *Computerized Medical Imaging and Graphics*, 76, 101635.
- [147] Machine-learning Activation Functions, lien : <https://riptutorial.com/machine-learning/example/31624/activation-functions>
- [148] Anatomie du foie, lien : <https://www.e-cancer.fr/Patients-et-proches/Les-cancers/Cancer-du-foie/Anatomie-du-foie>
- [149] Elsinger, F., Lunt, C., Harris A., and Chang S.(2020) LI-RADS Liver Imaging Reporting And Data System, Luzerner Kantonsspital and Vancouver General Hospital, lien : <https://radiologyassistant.nl/abdomen/liver/li-rads>

Annexe

Annexe A

Exemples de règles SWRL relatives au système BCLC

```
Liver(?x) ^ LiverCancer(?y) ^ Nodule(?n) ^ composedOf(?y, ?n) ^  
noduleSize(?n, ?s) ^ swrlb:lessThan(?s, "2"^^xsd:int) ^ noduleNumber(?n,  
"1"^^xsd:int) ^ contains(?x, ?y) ^ hasBiliribine(?x, "Normal") ^  
hasPortalVeinPressure(?x, "Normal") ^ hasFunctionalIndex(?x, "0"^^xsd:int) ^  
hasScoreChildPugh(?x, "NonApplicable") -> classifiedInto(?y, A0)
```

```
Liver(?x) ^ LiverCancer(?y) ^ Nodule(?n) ^ composedOf(?y, ?n) ^  
noduleSize(?n, ?s) ^ swrlb:greaterThan(?s, "5"^^xsd:int) ^ noduleNumber(?n,  
"1"^^xsd:int) ^ contains(?x, ?y) ^ hasBiliribine(?x, "NonApplicable") ^  
hasPortalVeinPressure(?x, "NonApplicable") ^ hasFunctionalIndex(?x,  
"2"^^xsd:int) ^ hasScoreChildPugh(?x, "GradeA") -> classifiedInto(?y, C)
```

```
Liver(?x) ^ LiverCancer(?y) ^ Nodule(?n) ^ composedOf(?y, ?n) ^  
noduleSize(?n, ?s) ^ swrlb:greaterThan(?s, "5"^^xsd:int) ^ noduleNumber(?n,  
"1"^^xsd:int) ^ contains(?x, ?y) ^ hasBiliribine(?x, "NonApplicable") ^  
hasPortalVeinPressure(?x, "NonApplicable") ^ hasFunctionalIndex(?x,  
"2"^^xsd:int) ^ hasScoreChildPugh(?x, "GradeB") -> classifiedInto(?y, C)
```

```
Liver(?x) ^ LiverCancer(?y) ^ Nodule(?n) ^ composedOf(?y, ?n) ^  
noduleSize(?n, ?s) ^ contains(?x, ?y) ^ hasBiliribine(?x, "NonApplicable") ^  
hasPortalVeinPressure(?x, "NonApplicable") ^ hasFunctionalIndex(?x,  
"3"^^xsd:int) ^ hasScoreChildPugh(?x, "GradeC") -> classifiedInto(?y, D)
```

```
Liver(?x) ^ LiverCancer(?y) ^ Nodule(?n) ^ composedOf(?y, ?n) ^  
noduleSize(?n, ?s) ^ contains(?x, ?y) ^ hasBiliribine(?x, "NonApplicable") ^  
hasPortalVeinPressure(?x, "NonApplicable") ^ hasFunctionalIndex(?x,  
"4"^^xsd:int) ^ hasScoreChildPugh(?x, "GradeC") -> classifiedInto(?y, D)
```

```
Liver(?x) ^ LiverCancer(?t) ^ contains(?x, ?t) ^ classifiedInto(?t, ?i) ^  
hasFunctionalIndex(?x, ?f) ^ hasBiliribine(?x, ?y) ^ hasPortalVeinPressure(?x,  
?p) ^ hasScoreChildPugh(?x, ?s) ^ contains(?x, ?c) -> sqwrl:select(?x, ?f, ?y,  
?p, ?s) ^ sqwrl:orderBy(?x) ^ sqwrl:columnNames("Liver", "Functional  
Index", "Bilirubine", "PortalVeinPressure", "Score Child Pugh")
```

Liver(?x) ^ LiverCancer(?y) ^ Nodule(?n) ^ composedOf(?y, ?n) ^
noduleSize(?n, ?s) ^ swrlb:greaterThan(?s, "5"^^xsd:int) ^ noduleNumber(?n,
"1"^^xsd:int) ^ contains(?x, ?y) ^ hasBiliribine(?x, "NonApplicable") ^
hasPortalVeinPressure(?x, "NonApplicable") ^ hasFunctionalIndex(?x,
"1"^^xsd:int) ^ hasScoreChildPugh(?x, "GradeB") -> classifiedInto(?y, C)

Liver(?x) ^ LiverCancer(?y) ^ Nodule(?n) ^ composedOf(?y, ?n) ^
noduleSize(?n, ?s) ^ swrlb:greaterThan(?s, "5"^^xsd:int) ^ noduleNumber(?n,
"1"^^xsd:int) ^ contains(?x, ?y) ^ hasBiliribine(?x, "NonApplicable") ^
hasPortalVeinPressure(?x, "NonApplicable") ^ hasFunctionalIndex(?x,
"1"^^xsd:int) ^ hasScoreChildPugh(?x, "GradeA") -> classifiedInto(?y, C)

Liver(?x) ^ LiverCancer(?y) ^ Nodule(?n) ^ composedOf(?y, ?n) ^
noduleSize(?n, ?s) ^ swrlb:greaterThan(?s, "5"^^xsd:int) ^ noduleNumber(?n,
"1"^^xsd:int) ^ contains(?x, ?y) ^ hasBiliribine(?x, "NonApplicable") ^
hasPortalVeinPressure(?x, "NonApplicable") ^ hasFunctionalIndex(?x,
"0"^^xsd:int) ^ hasScoreChildPugh(?x, "GradeA") -> classifiedInto(?y, B)

Liver(?x) ^ LiverCancer(?y) ^ Nodule(?n) ^ composedOf(?y, ?n) ^
noduleSize(?n, ?s) ^ swrlb:lessThan(?s, "3"^^xsd:int) ^ noduleNumber(?n,
"3"^^xsd:int) ^ contains(?x, ?y) ^ hasBiliribine(?x, "NonApplicable") ^
hasPortalVeinPressure(?x, "NonApplicable") ^ hasFunctionalIndex(?x,
"0"^^xsd:int) ^ hasScoreChildPugh(?x, "NonApplicable") -> classifiedInto(?y,
A4)

Exemples de règles SWRL relatives au système TNM

LiverCancer(?x) ^ has-T(?x, T2) ^ has-N(?x, n0) ^ has-M(?x, M0) ^
hasSpreadBloodVessels(?x, "Oui") ^ hasSpreadLymphNodes(?x, "Non") ^
hasSpreadOtherPart(?x, "Non") -> classifiedInto(?x, Stage-II)

LiverCancer(?x) ^ has-T(?x, T3a) ^ has-N(?x, n0) ^ has-M(?x, M0) ^
hasSpreadBloodVessels(?x, "Non") ^ hasSpreadLymphNodes(?x, "Non") ^
hasSpreadOtherPart(?x, "Non") -> classifiedInto(?x, Stage-IIIa)

```
LiverCancer(?x) ^ has-T(?x, T3b) ^ has-N(?x, n0) ^ has-M(?x, M0) ^  
hasSpreadBloodVessels(?x, "Oui") ^ hasSpreadLymphNodes(?x, "Non") ^  
hasSpreadOtherPart(?x, "Non") -> classifiedInto(?x, Stage-IIIB)
```

```
LiverCancer(?x) ^ has-T(?x, T1) ^ has-N(?x, n0) ^ has-M(?x, M0) ^  
hasSpreadBloodVessels(?x, "Non") ^ hasSpreadLymphNodes(?x, "Non") ^  
hasSpreadOtherPart(?x, "Non") -> classifiedInto(?x, Stage-I)
```

Exemples de règles SWRL relatives au type du traitement à appliquer

```
LiverCancer(?x) ^ classifiedInto(?x, A2) -> treatedBy(?x,  
Surgery)
```

```
LiverCancer(?x) ^ classifiedInto(?x, A4) -> treatedBy(?x, TACE)
```

```
LiverCancer(?x) ^ classifiedInto(?x, A1) -> treatedBy(?x, ARF)
```

```
LiverCancer(?x) ^ classifiedInto(?x, A3) -> treatedBy(?x, TACE)
```

```
LiverCancer(?x) ^ classifiedInto(?x, A4) -> treatedBy(?x,  
Liver_transplant)
```

```
LiverCancer(?x) ^ classifiedInto(?x, C) -> treatedBy(?x,  
Targeted-chemotherapy(
```