

**UNIVERSITE SAAD DAHLEB DE BLIDA**

**Faculté des sciences**

Département d'informatique



**MEMOIRE DE MASTER**

**En Informatique**

Option : Ingénierie Des Logiciels

**THÈME :**

**Systeme d'aide au Diagnostic Médical fondé sur un  
Modèle de Réseau Bayésien Flou et une Ontologie  
Floue**

Réalisé par

IBNAISSA MOHAMED CHAKIB

BELLIL MOUNIR

Encadré par

Mr. Riali Ishak

Mme. Fareh Messaouda

14 Juillet 2021



## Remerciements

Nous remercions ALLAH de nous avoir donné la santé et le courage afin de pouvoir réussir ce travail. Ce travail est l'aboutissement d'un long cheminement au cours duquel nous avons bénéficié d'encadrement, des encouragements et du soutien de plusieurs personnes, à qui nous tenons à dire profondément et sincèrement merci.

Nous exprimons notre grande gratitude à notre promoteur et co-promotrice Mr. RIALI ISHAK et Mme. FAREH MESSAOUDA, d'avoir accepté de nous encadrer durant notre travail et pour leurs précieux conseils et leurs orientations.

Nous tenons à remercier également Dr BRAHIMI et Dr BENDALI, qui ont voulu nous apporter leur aide en nous faisant part de leur propre expérience.

Nous remercions également tous les membres du jury pour nous avoir honorés par leur présence et pour avoir accepté d'évaluer notre travail.

## **Dédicace**

*Je dédie ce travail*

*A mes chers parents,*

*Qui ont sacrifié leur vie pour ma réussite et m'ont éclairé le chemin par leur conseils judicieux, j'espère un jour, je pourrai leurs rendre un peu de ce qu'ils ont fait pour moi, que dieu leur prête bonheur et longue vie,*

*mon cher frère Aymen et chère amie Maroua Belkhir,*

*A tout ma famille,*

*A tous ceux qui sont chers,*

*A mes chers amis,*

*Une dédicace spéciale à mon binôme Mounir,*

*Pour tous les moments inoubliables que nous avons vécus ensemble,*

***IBNAISSA MOHAMED CHAKIB.***

## **Dédicace**

*Je dédie ce travail*

*A mes très chers parents, source de vie, d'amour et d'affection,*

*Qui n'ont jamais cessé, de formuler des prières à mon égard, de me soutenir et de m'épauler pour que je puisse atteindre mes objectifs.*

*A mes cher frères,*

*A ma chère sœur,*

*Pour ses soutiens moral et leurs conseils précieux tout au long de mes études.*

*A mes chères amies, spécialement Maroua, Pour leurs aides et supports dans les moments difficiles.*

*A Chakib, cher ami avant d'être binôme, Pour sa entente et sa sympathie.*

*A vous cher lecteur,*

**BELLIL MOUNIR.**



## Résumé

L'incertitude en médecine a des conséquences néfastes pour le patient. Cependant, les décisions médicales sont difficiles à prendre car elles sont souvent fondées sur des informations incertaines. Plusieurs propositions ont émergé pour faire face à l'incertitude dans les ontologies classiques. Néanmoins, les ontologies classiques ont des limites dans l'apport du raisonnement sous incertitude.

Dans ce contexte, les ontologies floues ont été proposées pour remédier les défauts des ontologies classiques en utilisant la logique floue. Ainsi, les réseaux bayésiens flous ont été reconnus comme des techniques très performantes en matière d'aide à la décision.

L'objectif de ce travail est de réaliser un système automatique basé sur les réseaux bayésiens flous et les ontologies floues, afin de prédire le diagnostic médical, à partir des données incertaines et floues. Nous allons en premier lieu concevoir et implémenter une ontologie floue, en suivant la méthodologie FODM (fuzzy ontology development methodology). Ensuite, nous procédons à la construction du réseau bayésien flou, par l'apprentissage de la structure puis, l'apprentissage de paramètres. Enfin, le système prédire si le patient est atteint par la maladie de l'hépatite C ou non, tout en sachant ses informations biologiques et ses données pour faire l'inférence.

**Mots clés :** Connaissance incertaine, logique floue, réseaux bayésiens flous, ontologies floues.

## Abstract

Uncertainty in medicine has negative consequences for the patient. However, medical decisions are difficult to make because they are often based on uncertain information. Several proposals have emerged to deal with the uncertainty in classical ontologies. Nevertheless, classical ontologies have limits in the contribution of reasoning under uncertainty.

In this context, fuzzy ontologies have been proposed to remedy the shortcomings of classical ontologies by using fuzzy logic. Thus, fuzzy Bayesian networks have been recognized as very powerful techniques for decision support.

The objective of this work is to realize an automatic system based on fuzzy Bayesian networks and fuzzy ontologies, in order to predict the medical diagnosis, from uncertain and fuzzy data. We will first design and implement a fuzzy ontology, following the FODM (fuzzy ontology development methodology). Then, we proceed to the construction of the fuzzy Bayesian network, by learning the structure and then learning the parameters. Finally, the system predicts whether the patient has hepatitis C disease or not, knowing their biological information and data to make the inference.

**Keywords :** uncertain knowledge, fuzzy logic, fuzzy bayesian networks, fuzzy ontology.

## ملخص

عدم اليقين في الطب له عواقب سلبية على المريض. ومع ذلك، يصعب اتخاذ القرارات الطبية لأنها غالبًا ما تستند إلى معلومات غير مؤكدة. ظهرت عدة مقترحات للتعامل مع عدم اليقين في الأنطولوجيا الكلاسيكية. ومع ذلك، فإن الأنطولوجيا الكلاسيكية لها حدود في تقديم تفكير في ظل عدم اليقين.

في هذا السياق، تم اقتراح أنطولوجيا ضبابية لمعالجة أوجه القصور في الأنطولوجيا الكلاسيكية باستخدام المنطق الضبابي. وكذلك، تم التعرف على شبكات بايز الغامضة على أنها تقنيات قوية جدًا لدعم القرار.

الهدف من هذا العمل هو تحقيق نظام آلي قائم على شبكات بايزية ضبابية وأنطولوجيا ضبابية، من أجل التنبؤ بالتشخيص الطبي، من بيانات غير مؤكدة و غامضة. سنقوم أولاً بتصميم وتنفيذ منهجية تطوير علم الوجود الغامض، باتباع منهجية FODM (منهجية تطوير الأنطولوجيا الضبابية). بعد ذلك، ننتقل إلى بناء شبكة بايز الغامضة، من خلال تعلم الهيكل ثم تعلم الإعدادات. أخيراً، يتنبأ النظام بما إذا كان المريض مصاباً بمرض التهاب الكبد الوبائي سي، ومعرفة المعلومات والبيانات البيولوجية الخاصة به للاستدلال.

**الكلمات المفتاحية:** معرفة غير مؤكدة، منطق غامض، شبكات بايزية ضبابية، أنطولوجيا غامضة.

---

# Table des matières

---

<b>Table des figures</b>	<b>13</b>
<b>Liste des tableaux</b>	<b>15</b>
<b>Liste des abréviations</b>	<b>17</b>
<b>Introduction Générale</b>	<b>19</b>
Contexte de travail . . . . .	19
Problématique . . . . .	20
Objectifs du travail . . . . .	20
Organisation du mémoire . . . . .	21
<b>1 Les ontologies classiques et les ontologies floues</b>	<b>23</b>
1.1 Introduction . . . . .	23
1.2 Définition des Ontologies . . . . .	23
1.3 Composants d'une ontologie . . . . .	25
1.4 Cycle de vie d'une Ontologie . . . . .	25
1.5 Processus de développement d'une ontologie . . . . .	26
1.6 Méthodologies pour la construction d'ontologies . . . . .	27
1.6.1 La méthode SENSUS . . . . .	27
1.6.2 Methontology . . . . .	27
1.7 Langage de représentation d'ontologies . . . . .	27
1.8 Connaissances incertaines . . . . .	29
1.8.1 Définition de l'Incertitude . . . . .	29
1.8.2 Classification de l'incertitude . . . . .	29
1.8.3 Logique floue . . . . .	31
1.8.4 Exemples sur des connaissances floues . . . . .	31
1.9 Ontologies Floues . . . . .	32
1.9.1 Définition des ontologies floues . . . . .	32

1.9.2	Composants des ontologies floues . . . . .	33
1.9.3	La méthodologie de développement de l'ontologie floue . . . . .	34
1.10	Conclusion . . . . .	36
<b>2</b>	<b>Réseau Bayésien et réseau Bayésien flou</b>	<b>37</b>
2.1	Introduction . . . . .	37
2.2	Réseau bayésien . . . . .	37
2.2.1	Exemple de réseau bayésien . . . . .	38
2.2.2	Apprentissage dans les réseaux bayésiens . . . . .	38
2.2.3	Apprentissage de structure . . . . .	39
2.2.4	Apprentissage de paramètres . . . . .	39
2.2.5	Inférence dans les réseaux bayésiens . . . . .	40
2.3	Limites des réseaux bayésiens . . . . .	41
2.4	Réseaux bayésiens flous . . . . .	41
2.5	Exemple . . . . .	41
2.5.1	Méthode de pondération . . . . .	42
2.5.2	Distribution probabiliste floue . . . . .	44
2.5.3	Méthode de l'évidence virtuelle . . . . .	45
2.6	Travaux connexes . . . . .	45
2.6.1	Medical Decision Support Tool from a Fuzzy-Rules Driven Bayesian Network . . . . .	45
2.6.2	AlzFuzzyOnto : Une ontologie floue pour l'aide à la décision dans le domaine de la maladie d'Alzheimer . . . . .	46
2.6.3	HEPO : The Hepatitis Ontology For Abductive Medical Diagnostic Systems . . . . .	46
2.6.4	Bayesian Network Model for Diagnosis of Psychiatric Diseases . . . . .	47
2.6.5	A novel clinical decision support system for liver fibrosis using evolutionary multi-objective method based numerical association analysis . . . . .	47
2.6.6	Applying Machine Learning to Evaluate for Fibrosis in Chronic Hepatitis C . . . . .	47
2.6.7	Discussion . . . . .	48
2.7	Conclusion . . . . .	49
<b>3</b>	<b>Conception de système</b>	<b>51</b>
3.1	Introduction . . . . .	51
3.2	Motivation . . . . .	51
3.3	Schéma global de notre système . . . . .	53
3.4	Prétraitement des données . . . . .	54
3.4.1	Nettoyage des données . . . . .	58

3.5	Construction d'une ontologie floue pour l'hépatite C . . . . .	59
3.5.1	Phase 1 : Objectifs et le champ d'application de l'ontologie . . . . .	59
3.5.2	Phase 2 : Identifier le besoin de représenter les connaissances floues . . . . .	60
3.5.3	Phase 3 : Déterminer les connaissances floues et précises . . . . .	60
3.5.4	Phase 4 : Considérer la réutilisation des ontologies existantes . . . . .	61
3.5.5	Phase 7 : Définir les éléments d'ontologie floue . . . . .	62
3.5.6	Algorithme du Fuzzy C-means clustering . . . . .	63
3.5.7	Phase 8 : Définir les éléments d'ontologie précis . . . . .	65
3.5.8	Phase 9 : Formalisation . . . . .	65
3.5.9	Instanciation de l'ontologie floue . . . . .	66
3.6	Construction de réseau Bayésien Flou . . . . .	67
3.6.1	Apprentissage de la structure . . . . .	67
3.6.2	Apprentissage des paramètres . . . . .	69
3.7	Diagnostic de l'état du patient en se basant sur le modèle RBF . . . . .	70
3.8	Conclusion . . . . .	71
<b>4</b>	<b>Implémentation et test</b>	<b>73</b>
4.1	Introduction . . . . .	73
4.2	Implémentation . . . . .	73
4.2.1	Outils de programmation . . . . .	73
4.3	Les interfaces de notre application . . . . .	77
4.3.1	Interface d'accueil . . . . .	77
4.3.2	Interface : Ajouter patient . . . . .	78
4.3.3	Interface : Modifier patient . . . . .	79
4.3.4	Interface Inférence patient . . . . .	80
4.4	Test et validation . . . . .	81
4.4.1	Mesures d'évaluation . . . . .	81
4.4.2	Résultats obtenus . . . . .	82
4.5	Conclusion . . . . .	83
	<b>Conclusion générale</b>	<b>85</b>
	<b>Bibliographie</b>	<b>87</b>

---

# Table des figures

---

1.1	Ontologie avec les concepts en nœud et les relations en arc[1] . . . . .	24
1.2	Cycle de vie d'une ontologie [2] . . . . .	25
1.3	Processus général de construction d'une ontologie . . . . .	26
1.4	Description de l'ontologie de l'incertitude développée par URW3-XG [3]. . . . .	29
1.5	Les types de l'incertitude [3]. . . . .	30
1.6	Classification des températures d'une pièce en deux sous-ensembles [4]. . . . .	31
1.7	Exemple d'ensembles considérés en logique booléenne [5] . . . . .	32
1.8	Exemple d'ensembles considérés en logique floue [6]. . . . .	32
1.9	Concepts Floue [7] . . . . .	33
1.10	Relation floue [7] . . . . .	34
1.11	Flux de travail du FODM [8] . . . . .	35
2.1	Exemple de réseau bayésien simple . . . . .	38
2.2	Exemple d'un RBF . . . . .	42
3.1	Schéma global du système . . . . .	53
3.2	Une portion des lignes du dataset . . . . .	57
3.3	Informations sur le dataset . . . . .	57
3.4	Nombre de 'null' pour chaque colonne . . . . .	58
3.5	Une portion des lignes du dataset après nettoyage . . . . .	59
3.6	Diagramme de classes . . . . .	61
3.7	Certaines fonctions d'appartenance floues construites à partir des Centre de masse . . . . .	62
3.8	La structure globale visualisée de l'ontologie floue . . . . .	66
3.9	Exemple du type de donnée . . . . .	66
3.10	Structure de réseau bayésien . . . . .	69
3.11	Réseau Bayésien . . . . .	70
3.12	Le processus d'inférence adapté pour le diagnostic . . . . .	71
4.1	Interface graphique de Protégé . . . . .	74
4.2	Hiérarchie de concept . . . . .	74

4.3	Relations entre concepts . . . . .	75
4.4	Les instances de concept . . . . .	75
4.5	Création d'un type de données flou avec le plug-in. . . . .	76
4.6	L'interface d'accueil. . . . .	78
4.7	Ajouter Patient. . . . .	79
4.8	Modifier Patient. . . . .	79
4.9	Diagnostiquer Patient. . . . .	80
4.10	Résultat de la prédiction . . . . .	80
4.11	Résultat graphique de la prédiction . . . . .	81
4.12	Résultats obtenus par chaque classifieur . . . . .	83

---

# Liste des tableaux

---

2.1	Tableau de comparaison entre les travaux existants . . . . .	48
3.1	Descriptions du dataset . . . . .	54
3.2	Attributs et abréviations du dataset . . . . .	55
3.3	Concepts précis et concepts flous . . . . .	61
3.4	Concepts flous . . . . .	65
3.5	Concepts précis . . . . .	65
3.6	Catégorisation des attributs numériques . . . . .	68
4.1	Matrice de confusion . . . . .	82
4.2	Illustration de (TP, FP, TN, FN) a pour chaque classe [9] . . . . .	82
4.3	Comparaison des 'accuracy' (%) obtenu par notre système et les classifieurs de Machine Learning . . . . .	83



---

# Liste des abréviations

---

**DTD** Document Type Définition. 28

**FODM** fuzzy ontology development methodology. 7, 8, 34, 53, 59

**OMS** L'Organisation mondiale de la santé. 51, 52

**OWL** Ontology Web Language. 28

**RB** Réseau Bayésien. 39, 40, 45, 46, 49

**RBF** réseau bayésien flou. 20, 42, 43, 51, 52, 54

**RDF** Resource Description Framework. 28

**SADM** systèmes d'aide à la décision médicale. 20

**URW3-XG** Uncertainty Reasoning for the World Wide Web Incubator Group. 13, 29

**VHC** Virus de l'Hépatite C. 19, 47, 51, 52

**W3C** World Wide Web Consortium. 28

**XML** Extensible Markup Language. 27, 28



---

# Introduction générale

---

## Contexte de travail

Les ontologies ont reçu une importance très remarquable durant les dernières années, elles sont capables de faciliter la gestion et le partage des connaissances et de promouvoir l'interopérabilité sémantique entre les systèmes d'information du domaine médical.

Cependant, les décisions médicales sont difficiles à prendre car elles doivent être fondées souvent sur des informations incertaines d'une part, et d'autre part, les technologies du Web sémantique classiques ne permettent pas de manipuler des informations médicales incertaines.

Récemment, plusieurs extensions et propositions ont émergé pour faire face à l'incertitude dans les ontologies classiques. Dans ce contexte, les ontologies floues ont été proposées pour remédier les défauts des ontologies classiques en utilisant la logique floue ; ils permettent de traiter des connaissances vagues et imprécises qui apparaissent de manière inhérente dans la plupart des domaines et des applications du monde réel.

Par ailleurs, les réseaux bayésiens flous ont été reconnus comme des techniques très performantes en matière d'aide à la décision. Ils sont issus d'une hybridation entre la théorie bayésienne des probabilités conditionnelles et la théorie des graphes avec la logique floue. Ces techniques ont prouvé leur efficacité dans plusieurs domaines, tels que le domaine d'aide à la décision et le domaine de modélisation de systèmes complexes.

L'hépatite C est une maladie du foie causée par un virus VHC (Virus de l'Hépatite C). Ce virus peut entraîner des hépatites aiguës comme des hépatites chroniques. La gravité des hépatites C est variable et peut aller d'une forme bénigne, d'une durée limitée à quelques semaines, à une maladie grave qui s'installe à vie [10].

En Algérie, l'hépatite C touche environ 320.000 personnes [11], mais ce chiffre reste « sous-estimé » selon l'Association SOS Hépatite. Une prévalence de 3,43% est enregistrée au niveau de 6 wilayas de l'Est pour l'hépatite C durant l'enquête régionale réalisée par le ministère de la Santé. Des milliers de personnes recensées dans les deux zones endémiques (Khenchela et Barika) et des dizaines de décès sont comptabilisés chaque année.

# Problématique

Les médecins prennent des décisions très difficiles en présence des informations incertaines. Il est donc intéressant de concevoir et de mettre en œuvre des systèmes automatisés dans le domaine de la santé qui facilitent la tâche de diagnostic médical. Les médecins sont alors face à des situations très critiques et doivent prendre des décisions qui ne sont pas évidentes[12].

De nos jours, plusieurs travaux de réalisation de systèmes d'aide à la décision médicale ont vu le jour. Ces SADM (systèmes d'aide à la décision médicale) tentent de remédier à diverse problématique dans le domaine médical afin d'assister les praticiens par des outils informatiques dans leur processus de prise de décision.

Le but ultime donc est d'améliorer les pratiques médicales et le développement des systèmes d'aides à la décision efficaces. Ceci dépend fortement de la disponibilité des outils capables de partager et de représenter la sémantique des données médicales collectées en tenant compte de leur nature incertaine.

Dans notre cas, le système d'aide au diagnostic médical que nous proposons traite la maladie de l'hépatite C car dans le cadre de nos recherches on a trouvé que les travaux existants dans ce domaine sont peu et limités.

Les situations d'incertitude sont donc de plus en plus fréquentes dans le domaine de la médecine. Donc gérer ces incertitudes est un besoin nécessaire.

La représentation des connaissances dans le domaine du diagnostic médical n'est pas une tâche facile. Parce que les informations manipulées sont le plus souvent incertaines.

Ce problème nous amène à poser quelques questions :

Comment représenter les connaissances incertaines dans le domaine de diagnostic médicale ?

Comment améliorer les pratiques médicales et le développement des systèmes d'aides à la décision médical pour le traitement de la connaissance médical de l'hépatite C vu sa nature incertaine et surmonter les limites des systèmes existants et de proposer une meilleur technique ?

## Objectifs du travail

L'objectif de ce travail est de concevoir et mettre en place un système d'aide au diagnostic médical sur l'hépatite C pour une prise de décision relative aux observations sur les patients.

Nous gérerons l'incertitude inhérente à la pratique médicale. Nous avons choisi d'utiliser les ontologies floues pour représenter les informations biologiques et les données sur les patients et desréseau bayésien flou (RBF) pour faire l'inférence et prédire si le patient est malade ou non.

Nous avons utilisé ces techniques pour leurs efficacités et leurs utilités. Ces techniques ont prouvé leur efficacité dans plusieurs domaines.

L'objectif est de concevoir un système d'aide au diagnostic médical guidé par le réseau bayésien flou à partir d'un ensemble de données concernant la maladie d'hépatite C qui sont stockées dans une ontologie floue. Pour cette raison, nos objectifs se sont concentrés sur ;

- La construction d'une ontologie floue pour représenter les informations biologiques et les données des patients.
- La construction d'un réseau bayésien flou pour faire l'inférence.
- L'implémentation d'un système d'aide au diagnostic médical basé sur l'ontologie floue et le réseau bayésien flou.

## Organisation du mémoire

Pour mener à bien notre recherche, nous avons organisé notre travail comme suit :

- Chapitre 1 : Des notions sur les ontologies classiques et les ontologies flous.
- Chapitre 2 : Présentera le Réseau bayésien et le réseau bayésien flou ainsi que les travaux connexes .
- Chapitre 3 : Conception du système d'aide au diagnostic médical, en présentant la construction d'une ontologie floue pour l'hépatite C, ainsi que la construction du réseau Bayésien Flou.
- Chapitre 4 : Implémentation et l'évaluation des performances, ainsi que les différents résultats obtenus.

Ce mémoire sera finalisé par une conclusion générale reprenant les principaux points abordés, et en exposant certaines perspectives d'approfondissement, et d'amélioration de notre travail.



## *Chapitre 1*

---

# **Les ontologies classiques et les ontologies floues**

---

## **1.1 Introduction**

Les ontologies ont été reconnues comme une composante essentielle pour le partage et la réutilisation des connaissances. Elles doivent aussi pouvoir représenter non seulement des concepts précis, mais aussi des concepts imparfaits (flous) en rajoutent quelques notions que nous-allons broche dans ce chapitre.

La logique floue peut être incorporée à l'ontologie pour représenter les informations d'incertitude. En règle générale, l'ontologie floue est générée à partir d'une hiérarchie de concepts prédéfinies. Cependant, construire une hiérarchie de concepts pour un certain domaine peut être une tâche difficile et fastidieuse.

Dans ce chapitre, nous décrirons tout d'abord la définition générale d'une ontologie classique, ses composants et le cycle de vie, et quelques définitions sur la logique floue et les connaissances incertaines, Ensuite nous présentons les définitions sur les ontologies floues, les composants et les approches de la construction des composants d'ontologies floues.

## **1.2 Définition des Ontologies**

La définition d'une ontologie diffère d'un sens à un autre ; En Informatique les ontologies sont nées dans le besoin de représenter les connaissances. Elles sont une formalisation logique, structurée des connaissances pour un domaine donné et sont réalisées en fonction d'un usage spécifique [13].

En 1993, GRUBER [14] propose la définition suivante : «spécification explicite d'une conceptualisation».

En 1995, GUARINO [15] a modifié légèrement la définition de GRUBER, et la définit par : «spécification partielle et formelle d'une conceptualisation».

En 1997, ces deux dernières définitions sont regroupées dans celle de BORST [16] comme «spécification formelle d'une conceptualisation partagée».

- Formelle : l'ontologie doit être lisible par une machine, ce qui exclut le langage naturel.
- Explicite : la définition explicite des concepts utilisés et des contraintes de leurs utilisations.
- Conceptualisation : le modèle abstrait d'un phénomène du monde réel par identification des concepts clefs de ce phénomène.
- Partagée : l'ontologie n'est pas la propriété d'un individu, mais représente un consensus accepté par une communauté d'utilisateur.

Une ontologie est un ensemble structuré sous la forme d'un graphe orienté qui contient [17] :

- Des noeuds représentant le vocabulaire d'un domaine particulier.
- Des arcs représentant les relations nommées entre les concepts.

A partir de cette structure, la sémantique de chaque mot est déduite par les relations que ce mot possède dans l'ontologie, ce qui permet de restreindre les interprétations possibles (voir **la figure 1.1**).

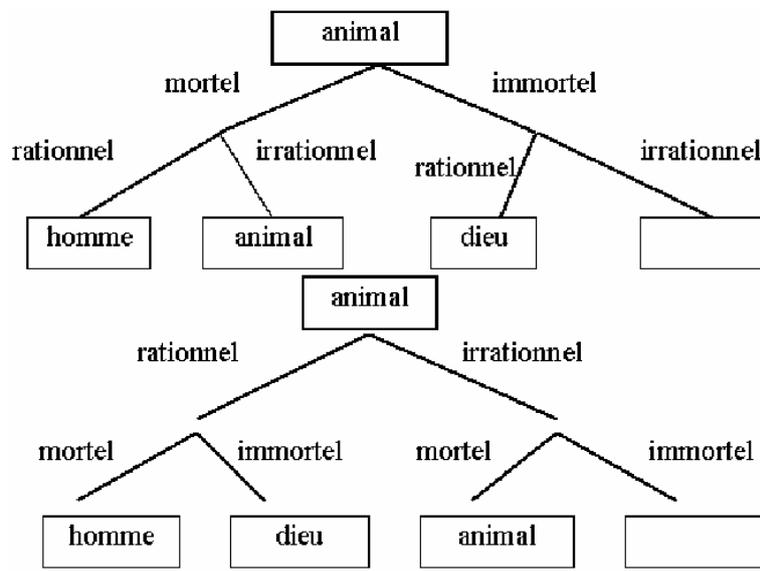


FIGURE 1.1 – Ontologie avec les concepts en nœud et les relations en arc[1]

Cependant pour être exploitable par une machine, une ontologie doit respecter certaines règles :

- Être défini par une syntaxe formelle et une sémantique non ambiguë.
- Permettre la déduction de nouvelles connaissances présentes implicitement dans l'ontologie.

## 1.3 Composants d'une ontologie

Les ontologies fournissent un vocabulaire commun d'un domaine. Elle est composée de cinq types de composants [18] :

- (a) **Concept** : Les concepts peuvent être une pensée, un principe ou une notion profonde. Un concept peut être divisé en trois parties :
  - Un terme.
  - Une notion (intension) : comprend la sémantique du concept formulé sous la forme propriétés, attributs, contraintes et règles.
  - Les objets (extension) : ce sont les instances du concept.
- (b) **Relation** : La relation désigne les différentes interactions et corrélations entre les concepts d'une ontologie. Une relation peut être : sous classe de, partie de, associé a, instance de, est un etc...
- (c) **Fonction** : Une fonction constitue des cas particuliers des relations, dans laquelle un élément de la relation, le nième est définie en fonction de n-élément précédents.
- (d) **Axiome** : Les axiomes sont utilisés pour décrire les assertions de l'ontologie qui seront considérés après comme vrais. Cette détermination a pour but de définir les significations des composant, les contraintes sur les valeurs des attribues et les arguments de relation.
- (e) **Instance** : Les instances constituent la définition extensionnelle de l'ontologie. Ces objets véhiculent les connaissances (statiques, factuelles) à propos du domaine.

## 1.4 Cycle de vie d'une Ontologie

Les ontologies doivent être considérées comme des objets techniques évolutifs et possédants un cycle de vie qui nécessite d'être précisé. **La figure 1.2** montre le cycle de vie d'une ontologie.

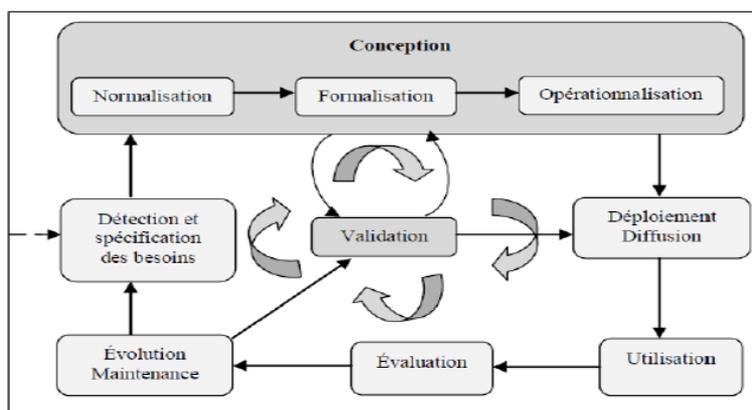


FIGURE 1.2 – Cycle de vie d'une ontologie [2]

Le cycle de vie d'une ontologie comporte plusieurs étapes nous détaillons chaque étape par une explication :

- Etape initiale : détection et spécification des besoins qui permet notamment de circonscrire précisément le domaine de connaissances.
- Deuxième étape : conception qui se subdivise en trois phases (normalisation, formalisation, opérationnalisation).
- Etape de déploiement et de diffusion.
- Etape d'utilisation et une étape d'évaluation. Après chaque utilisation significative, l'ontologie et les besoins doivent être réévalués.
- Enfin, une sixième étape consacrée à l'évolution et à la maintenance du modèle.
- La validation du modèle de connaissances est au centre du processus et se fait de manière itérative.

## 1.5 Processus de développement d'une ontologie

Les ontologies nous ont permis de dégager un consensus sur le processus de conception d'une ontologie exploitable dans un système à base de connaissance [19]. **La figure 1.3** illustre ce consensus qui repose sur un enchaînement de trois étapes, permettant de passer des données brutes à l'ontologie opérationnelle.

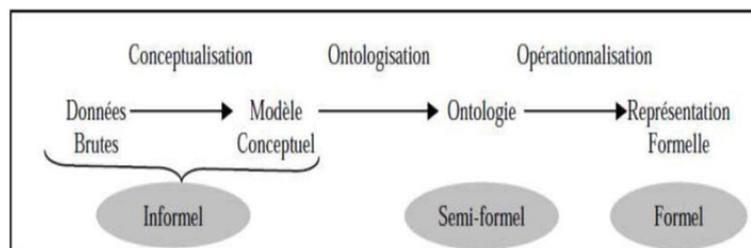


FIGURE 1.3 – Processus général de construction d'une ontologie

- (a) **Conceptualisation** : La première étape, appelée conceptualisation, permet d'aboutir à un modèle informel. Cette étape consiste à partir des données brutes, à dégager les concepts et les relations entre ces concepts permettant de décrire de manière informelle les entités cognitives du domaine. Elle est réalisée par un expert du domaine assisté de l'ingénieur de la connaissance et aboutit à un modèle conceptuel. Le modèle obtenu consiste en un ensemble de termes désignant les entités du domaine de connaissances [13].
- (b) **L'ontologisation** : L'ontologisation consiste un formalisme partiel sans perte d'information du modèle conceptuel obtenu dans l'étape précédente. Ce formalisme partiel facilite sa représentation ultérieure dans un langage complètement formel et opérationnel . Cependant, le modèle obtenu est souvent qualifié de semi-formel pour justifier le fait que certaines

connaissances ne peuvent pas être totalement formalisées.

En fait, cette étape produit un résultat en deux parties :

- Une partie formelle : Disposant d'une sémantique précise ou du moins consensuelle.
- Une partie informelle : ne disposant pas d'une sémantique claire ou consensuelle, ou tout ou moins d'une sémantique fixée à priori et donc exprimée dans un langage naturel ou semi-structuré.

(c) **L'opérationnalisation** : C'est la dernière étape de développement d'une ontologie appelée parfois représentation. Elle consiste à formaliser complètement l'ontologie obtenue précédemment dans le cadre d'un langage de représentation de connaissances, formel et opérationnel. Cette étape est menée par un spécialiste du langage de représentation et par l'ingénieur de la connaissance [20].

## 1.6 Méthodologies pour la construction d'ontologies

Les principales méthodologies utilisées pour construire les ontologies sont présentés ci-dessous :

### 1.6.1 La méthode SENSUS

C'est une méthode qui consiste à construire une ontologie de domaine à partir d'une plus grande ontologie qui est l'ontologie SENSUS. La méthode propose de relier les termes spécifiques du domaine à cette ontologie et détailler dans SENSUS. Le résultat de ce processus est le squelette de cette nouvelle ontologie, qui est générée automatiquement en utilisant ce processus et l'outil OntoSaurus [21].

### 1.6.2 Methontology

La méthode METHONTOLOGY est née dans les travaux de software development process et dans knowledge engineering méthodologies. Cette méthodologie inclut : l'identification du processus de développement de l'ontologie, le cycle de vie basé sur l'évolution de prototypes et les techniques de gestion de projet de développement et des activités de support. Les outils ODE et WeODE ont été construit pour donner un support technique à cette méthode [22].

## 1.7 Langage de représentation d'ontologies

Les langages de construction d'ontologie sont nombreux. Parmi eux on a :

(a) **XML et XML Schéma** : XML (Extensible Markup Language) constitue le premier langage de descriptions d'ontologies sur le web. XML est destiné pour l'échange de données sur le web ; rend possible la description des concepts et leurs hiérarchisations sous

forme de taxonomies à l'aide des balises. De même XML reste très léger et limité pour exprimer des contraintes et des relations entre des concepts ou des axiomes, ces lacunes ont motivé le développement d'autres langages comme DTD (Document Type Définition) et XML Schéma. En effet, DTD et XML Schéma offrent des grammaires qui permettent de spécifier des restrictions sur la structure des documents (taxonomies) et préciser les noms et le typage des attributs. Cependant, ils n'offrent aucun mécanisme pour exprimer des contraintes ou des relations sémantiques entre les concepts [18].

(b) **RDF et RDF Schéma**

a) **RDF (Resource Description Framework)** : C'est le langage de base du Web sémantique, publié par le consortium W3C (World Wide Web Consortium) le 10 février 2004. C'est un modèle de graphe destiné à décrire de façon formelle les ressources Web et leurs métadonnées, de façon à permettre le traitement automatique de telles descriptions. RDF a pour rôle de relier chaque concept de base à d'autres définitions, afin de lui donner un sens [23].

— **Syntaxe de RDF**

Il existe 3 différentes syntaxes de RDF :

- RDF : description de ressources Web (métadonnées).
- RDF Schéma (RDFS) : vocabulaires de description des ontologies.
- Syntaxe XML : échange des métadonnées et des schémas.

b) **RDF Schéma** : « RDF Schéma est un langage permettant de définir des propriétés sémantiques pour les ressources par un schéma. Avec RDFS, de nouvelles notions sémantiques apparaissent. Le principal est la distinction entre une classe (concept d'une ontologie) et une instance (individu d'une ontologie) » [24].

(c) **OWL (Ontology Web Language)** : OWL est une recommandation de W3C depuis février 2004. Il s'agit d'une extension du langage RDF et RDFS. OWL est un langage basé sur la syntaxe XML, très expressif avec un vocabulaire très riche. Il permet de définir des ontologies de domaines complexes. De plus il permet de faire du raisonnement pour déduire des informations qui ne sont pas explicitement présents dans l'ontologie [24]. OWL se compose de trois sous langages d'expressivité croissante : OWL Lite, OWL DL et OWL Full.

## 1.8 Connaissances incertaines

Dans cette section nous présentons la notion de l'incertitude et la logique floue.

### 1.8.1 Définition de l'Incertitude

L'incertitude est relative à la vérité d'une information, et caractérise son degré de conformité à la réalité. Elle fait référence à la nature de l'objet ou du fait concerné, à sa qualité, à son essence ou à son occurrence [25].

### 1.8.2 Classification de l'incertitude

Une ontologie simple a été développée par URW3-XG (Uncertainty Reasoning for the World Wide Web Incubator Group) [3] la mission du groupe est de mieux définir le défi de raisonner et de représenter des informations incertaines disponibles via le World Wide Web pour démontrer certaines fonctionnalités de base d'échange d'informations incertaines. De plus, cette ontologie a été utilisée pour classer les cas d'utilisation. Dans le but d'obtenir une couverture relativement complète des fonctionnalités liées au raisonnement d'incertitude sur les informations disponibles sur le World Wide Web. Considérer l'incertitude comme une connaissance permettra, par la suite, d'intégrer l'information dans les processus de raisonnement afin de créer de nouvelles connaissances. Pour ce faire, nous nous sommes basés sur l'ontologie décrite dans le rapport du URW3-XG [3], comme il est illustré dans la figure suivante (voir **la figure 1.4**).

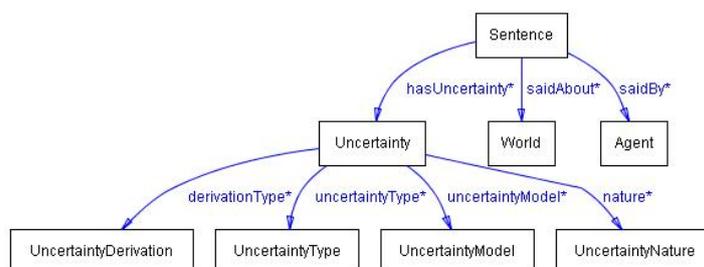


FIGURE 1.4 – Description de l'ontologie de l'incertitude développée par URW3-XG [3].

Une brève description de l'ontologie est présentée ci-dessous :

- **La phrase (sentence)**, c'est l'expression dans un langage logique qui s'évalue à une valeur de vérité (formule, axiome, assertion). L'incertitude sera donc associée aux phrases.
- **Le monde (world)**, représente le monde sur lequel la phrase est dite.
- **L'agent**, représentant celui qui fait la déclaration de phrase. Il peut être un humain ou une machine.
- **L'incertitude (Uncertainty)**, une déclaration sur l'incertitude associée à la phrase.

Elle sera décrite par les informations suivantes :

- **La nature de l'incertitude (Uncertainty Nature)**, cette classe capture les informations sur la nature de l'incertitude, c'est-à-dire si l'incertitude est inhérente au phénomène exprimé par la phrase, ou si elle est le résultat d'un manque de connaissance de l'agent.
  - **Aléatoire**, variabilité qui est une propriété inhérent du monde.
  - **Epistémique**, L'incertitude épistémique est due à l'agent dont la connaissance est limitée.
- **La dérivation de l'incertitude (Uncertainty Derivation)**, La dérivation de l'incertitude est une information objective ou subjective.
  - **Objective**, dérivé d'une manière formelle, processus de dérivation répétable.
  - **Subjective**, jugement subjectif, peut-être une supposition.
- **Les types de l'incertitude (Uncertainty Type)** (voir la figure 1.4),
  - **L'ambiguïté**, les référents des termes dans une phrase sur le monde ne sont pas clairement spécifiés et il ne peut donc pas être déterminé si la phrase est satisfaite.
  - **L'aléatoire**, une classe pour laquelle il n'existe pas une loi statistique régissant si les instances sont satisfaites.
  - **L'imprécision**, il n'y a pas de correspondance précise entre les termes de la phrase et les référents dans le monde,
  - **L'inconsistance**, on parle de l'inconsistance si une phrase contient des connaissances contradictoires.
  - **L'incomplétude**, on parle de l'incomplétude si les informations sur le monde sont incomplètes, certaines informations sont manquantes.

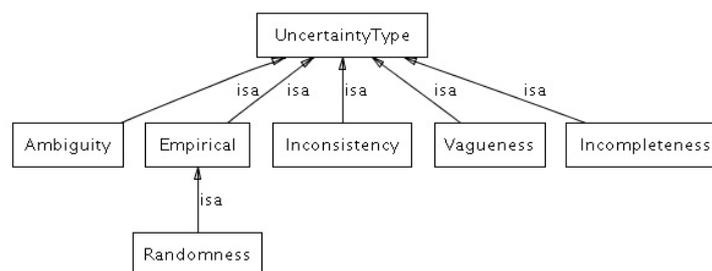


FIGURE 1.5 – Les types de l'incertitude [3].

- **Le modèle d'incertitude (UncertaintyModel)**, cette classe contient des informations sur les théories mathématiques qui peuvent être utilisées pour faire face à l'incertitude à savoir, **les théories des probabilités, la logique floue, la théorie des ensembles aléatoires, fonctions de croyance, Combinaison de plusieurs modèles (hybride)**

### 1.8.3 Logique floue

La logique floue est considérée Parmi les méthodes de représentation des connaissances, la logique floue (fuzzy logic) est utilisée pour le traitement des problèmes de réglage et de prise de décision établit par le professeur Lotfi Zadeh en 1965 [6].

### 1.8.4 Exemples sur des connaissances floues

Dans le cadre de la logique classique, une proposition est soit vraie, soit fausse (1 ou 0). Par exemple, la logique classique peut facilement partitionner la température d'une pièce en deux sous-ensembles, <froid> si la température moins de 15 degrés et <chaud> si la température est 15 degrés ou plus. Toutes les températures de moins de 15 degrés sont alors considérées comme appartenant à l'ensemble « froid ». On leur affecte une valeur de 1. Toutes les températures atteignant 15 degrés ou plus ne sont pas considérées comme appartenant à l'ensemble « froid ». On leur attribue une valeur de 0 [4].

Le raisonnement humain s'appuie fréquemment sur des connaissances ou des données inexactes, incertaines ou imprécises. Une personne placée dans une pièce dont la température est soit de 14.95 degrés soit de 15.05 degrés, ne fera certainement pas de distinction entre ces deux valeurs. **La figure 1.6** a montré la différence entre un ensemble classique et l'ensemble flou. Cette personne sera pourtant capable de dire si la pièce est « froide » ou « chaude », sans pour cela utiliser de température limite ni de mesure précise. La logique floue permet de définir des sous-ensembles, comme « froid » ou « chaud », en introduisant la possibilité pour une valeur d'appartenir plus ou moins à chacun de ces sous-ensembles. La logique floue permet de faire intervenir les notions d'imprécision et d'incertitude dans un système. Cela permet par exemple de faire intervenir une température « d'environ 15 degrés » dans un contrôleur flou [4].

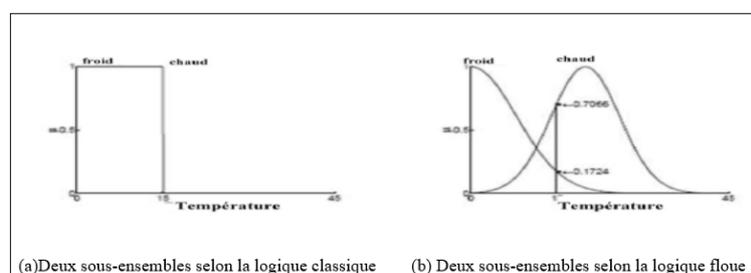


FIGURE 1.6 – Classification des températures d'une pièce en deux sous-ensembles [4].

Exemple :

(a) **En logique booléenne (classique)** : Le degré d'appartenance ( $\mu$ ) ne peut prendre que deux valeurs (0 ou 1). La température peut être (voir **la figure 1.7**) :

— Faible : (T) faible = 1, (T) moyenne = 0, (T) élevée = 0

— Moyenne : (T) faible = 0, (T) moyenne = 1, (T) élevée = 0

— Elevée : (T) faible = 0, (T) moyenne = 0, (T) élevée = 1

Elle ne peut pas prendre deux qualificatifs à la fois [4].

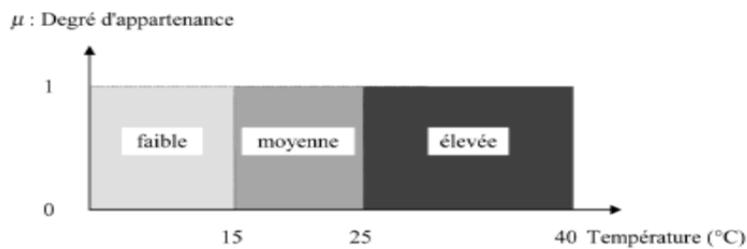


FIGURE 1.7 – Exemple d'ensembles considérés en logique booléenne [5]

- (b) **En logique floue :** Le degré d'appartenance devient une fonction qui peut prendre une valeur réelle comprise entre 0 et 1 inclus.  $\mu$  moyenne (T), par exemple, permet de quantifier le fait que la température puisse être considérée comme moyenne. Dans ce cas, la température peut être considérée, à la fois, comme faible avec un degré d'appartenance de 0,2 et comme moyenne avec un degré d'appartenance de 0,8.  $\mu$  faible(T) = 0,2,  $\mu$  moyenne(T) = 0,8,  $\mu$  élevée(T) = 0 [5] (voir la figure 1.8).

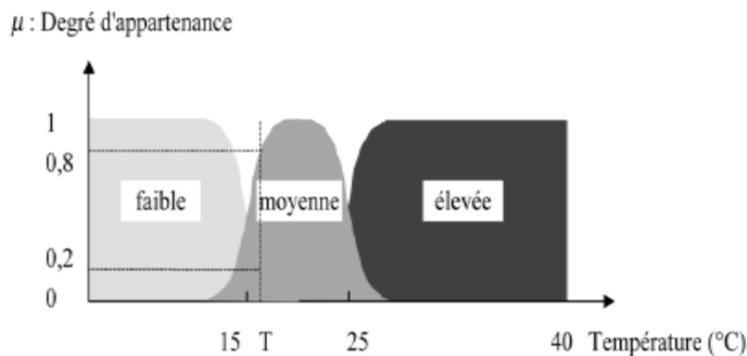


FIGURE 1.8 – Exemple d'ensembles considérés en logique floue [6].

## 1.9 Ontologies Floues

### 1.9.1 Définition des ontologies floues

Les ontologies floues se basent sur l'intégration de la logique floue à la définition des ontologies précises pour représenter les imprécisions dans le cadre du web sémantique.

L'ontologie floue peut être définie comme une ontologie qui possède, en plus des composants de base (concepts, relations, axiomes et instances), de nouveaux composants conformes à la logique floue : les concepts flous et les relations floues [20].

## 1.9.2 Composants des ontologies floues

Une ontologie floue comprend les concepts et les relations précises gardent les mêmes définitions et jouent les mêmes rôles que dans une ontologie classique. Les concepts et les relations flous, définis selon la logique floue et servent à la représentation des éléments flous du domaine modélisé, ils ont des répercussions sur l'expression des axiomes et des instances [7].

- Des concepts précis (exacts) et des concepts flous
- Des relations précises (exactes) et des relations floues
- Axiomes
- Instance

Les concepts et les relations flous, définis selon la logique floue, servent à la représentation des éléments flous du domaine modélisé. Ils ont des répercussions sur l'expression des axiomes et des instances [26].

- (a) **Concepts flous** : Un concept flou est considéré comme un ensemble flou. Ses instances possèdent ainsi des degrés d'appartenance prenant la valeur dans l'intervalle  $[0,1]$ . Ces degrés sont calculés conformément aux formules de calcul relatives à la fonction d'appartenance définie pour le concept. **La figure 1.9** montre deux concepts flous : Employé Jeune et Employé Âgé.

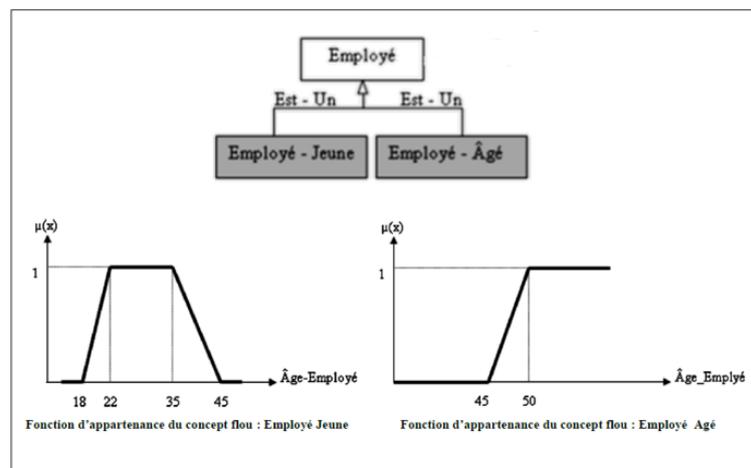


FIGURE 1.9 – Concepts Floue [7]

- (b) **Relations Floues** : Une relation floue est aussi définie comme un ensemble flou. Ses instances possèdent ainsi des degrés d'appartenance prenant leurs valeurs dans l'intervalle  $[0,1]$ . Ces degrés sont calculés conformément aux formules de calcul relatives à la fonction d'appartenance définie pour la relation. La relation floue « Acquiert », relie les deux concepts précis « Employé » et « Compétence-Acquise ». Elle dépend de la valeur de la propriété « Niveau-Acquisition ». Elle a deux ensembles flous : Compétence-Maîtrisée et Compétence-Parfaitement-Maîtrisée, **la figure 1.10** montré la relation floue [26].

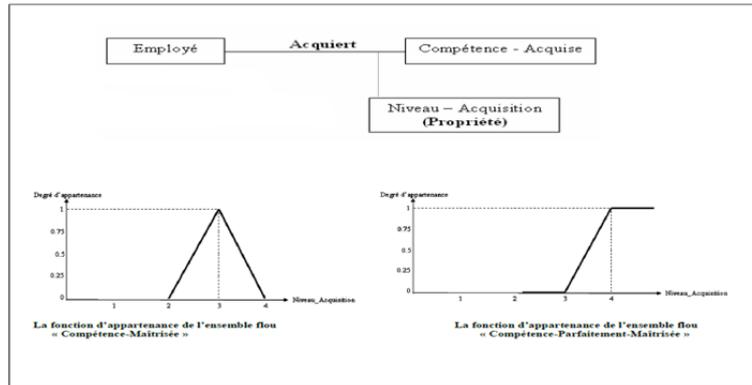


FIGURE 1.10 – Relation floue [7]

(c) **Axiomes** : A part leurs utilisations classiques, les axiomes servent à exprimer les formules de calcul des fonctions d'appartenance.

Par exemple : Axiomes pour le concept flou « Employé-Jeune » de la figure 1.8 :

- $\text{Emp-Jeune}(\text{Age}) = (\text{Age} - 18) / (22 - 18)$  si Age appartient à  $[18; 22]$ ;
- $\text{Emp-Jeune}(\text{Age}) = 1$  si Age appartient à  $[22; 35]$ ;
- $\text{Emp-Jeune}(\text{Age}) = (45 - \text{Age}) / (45 - 35)$  si Age appartient à  $[35; 45]$ ;
- $\text{Emp-Jeune}(\text{Age}) = 0$  ailleurs

(d) **Instances** : Dans les ontologies floues l'appartenance d'une instance à un composant flou n'est pas Sure. Elle est déterminée par une variable probabiliste prenant ses valeurs entre 0 et 1, et Déterminée par une fonction d'appartenance [7].

### 1.9.3 La méthodologie de développement de l'ontologie floue

FODM (fuzzy ontology development methodology) [8] Un paradigme formel de développement d'ontologie floue est présenté sur la base des méthodes de développement d'ontologies existantes. Il met l'accent sur l'introduction de nouveaux changements apportés par des ontologies floues dans le processus de développement.

FODM suppose une connaissance préalable des principes de l'ontologie classique et floue des lecteurs potentiels. Elle ne vise pas à réformer complètement les méthodes actuelles de développement d'ontologies classiques. Au lieu de cela, elle est construite sur la base de méthodes de développement d'ontologies existantes avec des considérations supplémentaires floues.

Le but du FODM est de fournir une abstraction formelle des activités qui doivent être réalisées tout au long du processus de développement. Elle vise à présenter la première approche méthodologique pour construire des ontologies floues à partir de zéro, plutôt que de convertir des ontologies existantes en ontologies floues. En général, toutes les activités ou tâches sont regroupées en onze phases (voir la figure 1.11) pour former l'ensemble du cycle de vie de la construction d'une ontologie floue.

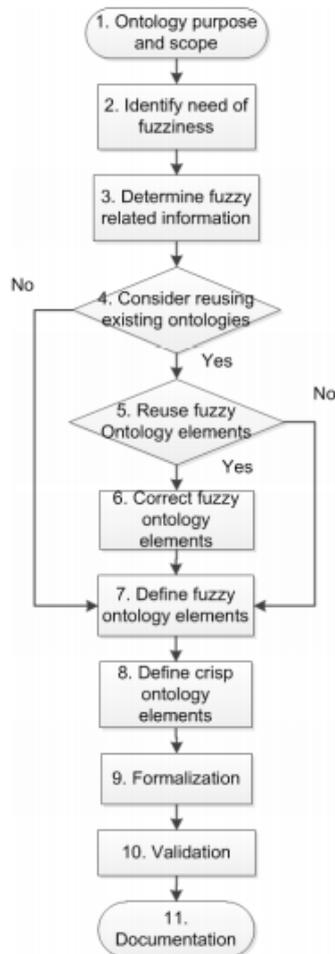


FIGURE 1.11 – Flux de travail du FODM [8]

**PHASE 1 :** Des questions de base doivent être soulevées et traitées explicitement afin de clarifier le but et la portée de l’ontologie.

- 1) Quel est le domaine ou la portée des informations à modéliser ?
- 2) L’ontologie est-elle la meilleure technique de modélisation par rapport aux autres solutions ?
- 3) Quel est le type d’ontologie en fonction de la détermination du domaine ou de la portée ?
- 4) Qui sera impliqué dans le développement de l’ontologie et quels rôles ils vont jouer ?

**PHASE 2 :** La deuxième phase vise à identifier si le flou doit être introduit dans la conception de l’ontologie.

**PHASE 3 :** L’objectif principal de l’étape 3 est de déterminer les informations floues liées. Suite à l’étape 2, une meilleure compréhension des informations vagues présentes dans le domaine pourraient être obtenue.

**PHASE 4 :** La vérification des ontologies existantes pertinentes pour le domaine ou la portée d’intérêt et la détermination de leur réutilisabilité sont les principales tâches définies dans cette phase.

**PHASE 5 :** La réponse à la question de savoir si les ontologies existantes peuvent être réutilisées pourrait devenir claire après l'étape 4. Si une ou plusieurs ontologies existantes sont analysées pour être utiles dans la conception de l'ontologie, une vérification fine doit être effectuée sur ces ontologies potentielles. Si seuls les éléments d'ontologie croustillants des ontologies existantes sont identifiés comme utiles, cela conduit à l'étape 7 qui sera spécifiée dans la sous-section Phase 7.

**PHASE 6 :** Dans cette phase, l'implication d'experts du domaine est nécessaire pour corriger les éléments d'ontologie floue hérités des ontologies floues existantes. Les spécifications et la modélisation de l'imprécision fournies par les éléments d'ontologie floue existants peuvent ne pas garantir un ajustement parfait pour capturer les informations identifiées comme vagues dans le domaine d'intérêt.

**PHASE 7 :** Le but de cette phase est de définir différents éléments d'ontologie floue pour fournir des approximations correctes de la nature des informations vagues et imprécises dans le domaine.

**PHASE 8 :** Cette phase se concentre sur le traitement de certaines connaissances dans le domaine. Outre les informations floues liées, le reste de la base de connaissances du domaine est défini comme des éléments d'ontologie précis en fonction de leurs attributs spécifiques.

**PHASE 9 :** Un certain langage doit être sélectionné pour formaliser l'ontologie conçue dans un format lisible par machine.

**PHASE 10 :** L'ontologie conçue doit passer par une vérification approfondie pour s'assurer qu'elle a représenté le modèle prévu du monde.

**PHASE 11 :** Documentation pour présenter les principes d'ingénierie de l'ontologie conçue, y compris les descriptions des différents éléments de l'ontologie, les détails de conception, la méthode d'utilisation et la maintenance.

## 1.10 Conclusion

Dans ce chapitre, nous nous sommes intéressés aux ontologies classiques et aux connaissances incertaines, logique floue et la façon de surmonter ce problème avec les ontologies floues que nous jugeons capables d'engendrer d'appréciables apports pour divers domaines, dont celui du web sémantique. Nous allons par la suite abordé le réseau bayésien et le réseau bayésien flou, Nous allons également présenté les différents travaux connexes.

## Chapitre 2

---

# Réseau Bayésien et réseau Bayésien flou

---

## 2.1 Introduction

Le réseau bayésien est un modèle graphique puissant pour traiter les connaissances incertaines. Cependant, l'une des principales limites auxquelles sont confrontés les utilisateurs du réseau bayésien est l'expression de connaissances floues. Face à ce problème, dans [27] ont proposé l'intégration de l'évidence floue dans les réseaux bayésiens qui combine la logique floue avec les réseaux bayésiens. Cela permet d'obtenir les avantages communs des deux méthodes et de surmonter le problème de l'ambiguïté des données.

Dans ce chapitre nous allons définir les réseaux bayésiens et les réseaux bayésiens flous. Nous présentons par la suite notre étude sur les travaux existants des systèmes d'aide à la décision médicale.

## 2.2 Réseau bayésien

Les réseaux bayésiens sont un couplage entre la théorie des graphes et la théorie probabiliste [28]. Un modèle graphique probabiliste représentant des variables aléatoires sous la forme d'un graphique orienté acyclique. Il est composé d'un ensemble de variables et d'un ensemble d'arcs entre les variables.

Un graphe orienté acyclique  $G = (V, E)$  avec  $V$  l'ensemble des nœuds du réseau et  $E$  l'ensemble des arcs. L'ensemble  $V$  est donc un ensemble de variables aléatoires discrètes chaque nœud de  $V$  est conditionnellement indépendant de ses non-descendants, compte tenu de ses parents immédiats [29].

Les réseaux bayésiens sont principalement utilisés pour le diagnostic (médical et industriel),

l'analyses des risques, la détection du spam et le data mining

$$P(V) = \prod_{(x \in v)} p(x|pa(x))$$

Avec  $x$  est le noeud, et  $Pa(x)$  est le parent de noeud  $x$ .

Les réseaux bayésiens sont des modèles graphiques probabilistes représentant des connaissances et permettant de calculer des probabilités conditionnelles à l'aide du théorème de Bayes et d'apporter des solutions à différents types de problèmes. Ils décrivent la distribution des probabilités associées à un ensemble de variables, dont certaines sont directement dépendantes et d'autres sont conditionnellement indépendantes [30].

### 2.2.1 Exemple de réseau bayésien

Ce matin-là, alors que le temps est clair et sec, M. Holmes sort de sa maison. Il s'aperçoit que la pelouse de son jardin est humide. Il se demande alors s'il a plu pendant la nuit, ou s'il a simplement oublié de débrancher son arroseur automatique. Il jette un coup d'œil sur la pelouse de son voisin, M. Watson, et s'aperçoit qu'elle est également humide.

Il en déduit qu'il a probablement plu et décide de partir au travail sans vérifier son arroseur automatique. La représentation graphique du modèle causal utilisé par M. Holmes est composée de 4 noeuds : Arroseur (A), Pluie (P), Jardin humide (J), Jardin du voisin humide (W)[31] (voir la figure 2.1).

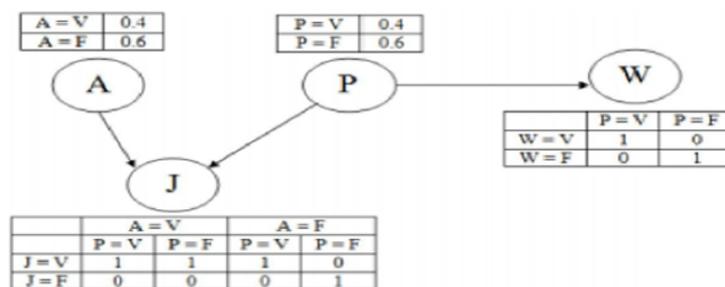


FIGURE 2.1 – Exemple de réseau bayésien simple

### 2.2.2 Apprentissage dans les réseaux bayésiens

Nous avons montré précédemment qu'un réseau bayésien est composé du graphe et d'un ensemble de probabilités conditionnelles. Le problème d'apprentissage est divisé en deux parties :

- Apprentissage des paramètres, nous supposons que la structure du réseau est fixe et où nous devons estimer la probabilité conditionnelle de chaque noeud du réseau bayésien.
- Apprentissage des structures, pour trouver le meilleur graphique représentant la tâche à résoudre.

### 2.2.3 Apprentissage de structure

Pour construire la structure d'un RB, on a deux sources d'informations. Des informations provenant des experts du domaine, et des données statiques. Par conséquent on a deux méthodes de construction d'un modèle de réseau bayésien, le choix de la méthode à utiliser est imposé par la source des informations disponibles.

(a) **La construction d'un réseau bayésien en collaboration avec les experts du domaine :**

La construction d'un RB en utilisant les connaissances des experts du domaine est une tâche difficile et consommatrice du temps. Elle est menée par un ensemble de spécialistes. Un expert en RB dirige le processus de construction, il pose des questions pertinentes, et explique les suppositions encodées dans le modèle au reste du groupe des spécialistes.

(b) **La construction d'un réseau bayésien à partir des données statique (modélisation basé sur les données) :**

La modélisation basée sur les données consiste à identifier la structure d'un RB à partir d'une source de données. La base de données peut être complète, comme elle peut être incomplète. Par conséquent, les méthodes d'apprentissage de structure des réseaux bayésiens se divise en deux classes :

La classe des algorithmes d'apprentissage de structure à partir des données complètes : il existe différents méthodes d'apprentissages de structures à partir de données complètes parmi la méthode basées sur la recherche d'indépendances conditionnelles, sa procédure est de faire des tests sur les indépendances conditionnelles permettant au final de retrouver la structure recherchée comme l'algorithme PC et IC [32], algorithme BNPC [32], l'algorithme QFCI [33], et des méthodes de recherche employant un score comme l'algorithme K2[33].

Et la classe des algorithmes d'apprentissage de structure à partir des données incomplètes parmi ces algorithmes : l'algorithme EL et l'algorithme SEM [33].

### 2.2.4 Apprentissage de paramètres

L'apprentissage des paramètres d'un RB se fait à partir de données relatives au problème à modéliser. Elle consiste à calculer des probabilités a posteriori (après observation d'évènements), dans ce cas on distinct deux type de donnée (données complètes et données incomplètes).

(a) **A partir des données complètes :** Nous cherchons ici à estimer les distributions de probabilités (ou les paramètres des lois correspondantes) à partir de données disponibles. L'estimation de distributions de probabilités, est un sujet très vaste et complexe.

— **Apprentissage statistique :** Dans le cas où toutes les variables sont observées, la méthode la plus simple et la plus utilisée est l'estimation statistique qui consiste à estimer la probabilité d'un évènement par la fréquence d'apparition de l'évènement dans la base de données. Cette approche appelée maximum de vraisemblance (MV).

— **Apprentissage bayésien** : L'estimation bayésienne suit un principe quelque peu différent. Il consiste à trouver les paramètres les plus probables sachant que les données ont été observées, en utilisant des a priori sur les paramètres.

(b) **A partir de données incomplètes** : Les réseaux Bayésiens (RBS) sont des structures fondamentales pour l'inférence probabiliste et modélisation. Les tables de probabilités conditionnelles, peuvent avoir besoin d'être appris.

Lorsque la structure de RB est connue, mais les données sont incomplètes, plusieurs algorithmes ont été utilisés : Espérance Maximisation (EM), par chaînes de Markov, méthode de Monte Carlo [34], comme l'échantillonnage de Gibbs, et les méthodes de descente de gradient. La méthode la plus couramment utilisée est fondée sur l'algorithme itératif EM, et d'autres algorithmes comme MMHC [35] et MDL [36]. Ces méthodes choisissent la structure de RB de candidat basé sur la fonction de score quantitative.

## 2.2.5 Inférence dans les réseaux bayésiens

Un réseau bayésien est donc un cadre probabiliste formalisant l'indépendance conditionnelle d'un système. La loi des probabilités conditionnelles permet de calculer des probabilités a priori pour toute variable du réseau. Mais ce qu'on appelle "inférence" dans un réseau bayésien, c'est l'estimation (ou la mise à jour) des distributions de certaines variables lorsque l'on fixe (avec une observation par exemple) des états pour d'autres variables.

On peut déduire certaines variables en utilisant la loi des probabilités conditionnelles si la variable instanciée est en amont de la variable à inférer, ou théorème de Bayes dans le cas contraire. Dans le cas où plusieurs variables sont instanciées, les auteurs [6] ont développé des algorithmes dits de « propagation de message » en appliquant les théorèmes de probabilités conditionnelles et de Bayes au voisinage des variables instanciées, et en propageant à leur tour les résultats de leur implémentation à leur voisin et ainsi de suite.

— **Inférence exacte** : On vise maintenant à calculer la distribution de probabilité a posteriori d'un ensemble de variables de requêtes, compte tenu d'un événement observé, c'est-à-dire de certaines affectations de valeurs à des variables de preuve.

—  $X$  : variable de question/requête

—  $E$  : l'ensemble des variables d'évidence

—  $e$  : un événement particulier

—  $Y$  : l'ensemble des variables cachées

— L'ensemble complet des variables est :  $\mathbf{X} = X \cup E \cup Y$

Dans l'inférence exacte il y a deux types : par énumération, par élimination de variable.

— **Inférence approximative** : Les méthodes d'inférence exactes qui viennent d'être discutées ne peuvent pas être utilisées pour les grands réseaux. Pour cela, nous considérons des

approches approximatives. Nous verrons des algorithmes basés sur un échantillonnage aléatoire (Monte Carlo) dont la précision dépendra d'un certain nombre d'échantillons.

— **Méthodes d'échantillonnage directes**

- La forme la plus simple d'échantillonnage aléatoire est de générer des événements sans variable d'évidence.
- La distribution de probabilité à partir de laquelle un échantillon pour une variable est choisi et basée sur les valeurs attribuées aux parents.

## 2.3 Limites des réseaux bayésiens

Aujourd'hui, les réseaux bayésiens constituent l'un des formalismes cohérents pour l'acquisition, la représentation et l'utilisation des connaissances par ordinateur.

L'évidence classique est une observation d'une variable aléatoire ayant une valeur particulière. Cependant, il n'est pas toujours possible d'observer la valeur exacte d'une variable ou d'avoir une confiance complète sur une observation, d'où l'apparition de l'évidence incertaine.

Dans quelque circonstance, l'utilisation des Réseaux Bayésiens se trouve limitée par l'impossibilité d'utiliser des variables de natures flous. Face à ce problème, elle existe l'évidence floue dans les réseaux bayésiens. Avec l'utilisation des ensembles flous qui nous emmènerons à la notion de réseaux bayésiens flous.

## 2.4 Réseaux bayésiens flous

Diverses recherches ont récemment permis d'utiliser la théorie des ensembles flous dans le contexte des réseaux bayésiens. Différentes approches ont été proposées, mais à ce jour aucune n'est parvenue à un consensus.

Le réseau bayésien flou est une extension du réseau bayésien standard car il se combine avec la théorie des ensembles flous. L'inférence dans les réseaux bayésiens flous est basée sur un raisonnement probabiliste flou. Lorsqu'un nœud est flou, plusieurs états peuvent être observés à la fois avec des degrés d'appartenances graduels. Il existe plusieurs méthodes pour faire une inférence probabiliste floue, nous citons les principales méthodes qui sont : méthode de pondération, méthode de l'évidence virtuelle et la méthode de distribution probabiliste floue[12].

## 2.5 Exemple

Pour mettre en évidence les trois approches, nous présentons dans ce qui suit un exemple qui sera utilisé modélisant la relation de causalité entre les plaquettes(PLAT) et la fatigue(F). La fatigue est une condition dans laquelle il y a une baisse des plaquettes dans le corps pour la coagulation sanguine[12].

Le graphe RBF qui correspond à cette situation est illustré dans la figure 2.2.

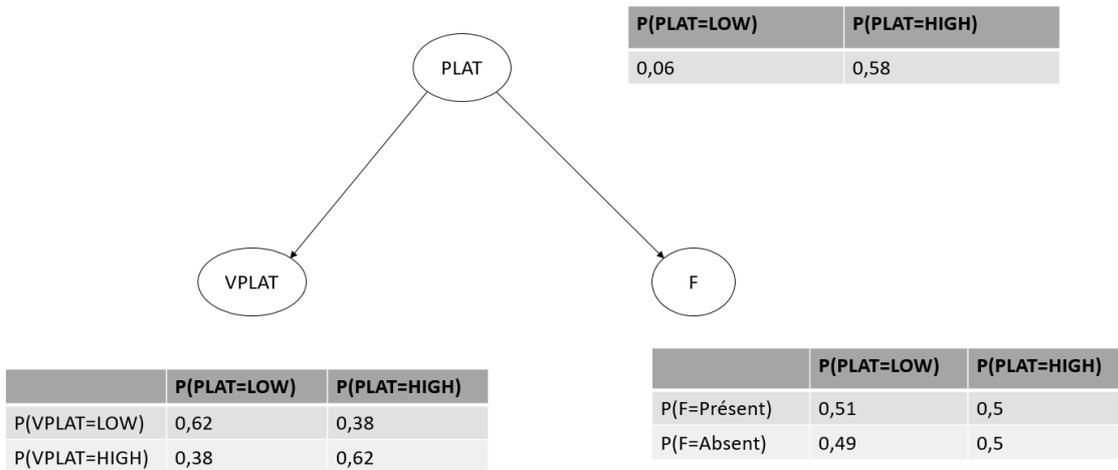


FIGURE 2.2 – Exemple d’un RBF

Dans cet exemple, nous allons considérer la variable plaquettes(Plat) comme floue, c’est pour cette raison, nous allons fuzzifier ses états, i.e., attribuer une fonction d’appartenance à chacun de ces états. Le tableau suivant illustre cette étape :

Etat	Interval	Type et paramètre de la fonction
LowPlat	[0, 255000[	Trapezoid("LowPlat",72459.19,72459.19,123198.34,194120.7)
HighPlat	[255000, 226464]	Trapezoid("HighPlat",123198.34,194120.7,247017.81,247017.81)

On suppose que la valeur de plaquettes(Plat) d’un patient P1 soit mesurée à 150000, les valeurs d’appartenances à chaque sous-ensemble floue. Les valeurs sont comme suit :

$$\mu_{LowPlat}(150000) = 0.62.$$

$$\mu_{HighPlat}(150000) = 0.38.$$

Supposons qu’on veut calculer la requête suivante (Requête 1) :

$$P(F = Absent \mid Plat)$$

Cette requête représente la probabilité que le patient P1 ne souffre pas de la fatigue sachant que le plaquettes est low avec un degré de 0.62 et high avec un degré de 0.38.

### 2.5.1 Méthode de pondération

L’idée de cette méthode est d’étendre les différentes règles utilisées dans les réseaux bayésiens en associant une valeur de degré d’appartenance à chaque loi, cette méthode consiste à étendre la loi Bayes afin d’incorporer les degrés d’appartenances dans les probabilités comme des poids lors de l’inférence. En effet, les règles bayésiennes floues peuvent ensuite être définies pour

prendre en charge l'inférence bayésienne floue dans un modèle RBF [12].

Il existe trois cas possible qui sont :

- Le nœud de requête est flou
- Le nœud d'observation est flou
- Le nœud de requête et le nœud d'observation sont flous

Enfin, chaque règle bayésienne floue utilise la loi de la marginalisation floue.

Soient A et B deux variables aléatoires. Dans le cas où la variable A est floue la loi de Bayes pondérée est définie comme suit :

$$P(A | B_j) = \frac{\sum_{i \in I} \mu A(A_i) P(B_i | A_i) P(A_i)}{P(B_j)} \quad (2.1)$$

Ou,  $\mu A(A_i)$  représente le degré d'appartenance de l'élément à l'état i et B<sub>j</sub> représente la variable B à l'état j.

Dans le cas où la variable B es floue : la loi de Bayes pondérée est définie comme suit :

$$P(A_j | B) = \frac{\sum_{i \in I} \mu A(B_i) P(B_i | A_j) P(A_j)}{P(B)} \quad (2.2)$$

Ou,  $\mu A(B_i)$  représente le degré d'appartenance de l'élément à l'état i et A<sub>j</sub> représente la variable A à l'état j.

En outre, la loi de probabilité marginale de la variable floue A est donnée comme suit :

$$P(A) = \sum_{i \in I} \mu A(A_i) P(A_i) \quad (2.3)$$

Ou,  $\mu A(A_i)$  représente le degré d'appartenance de l'élément à l'état A<sub>i</sub>.

Dans le cas où les deux variables sont floues :

$$P(A | B) = \frac{\sum_{i \in I} \sum_{j \in J} \mu A(A_i) \mu B(B_j) P(B_j | A_i) P(A_i)}{P(B)} \quad (2.4)$$

Ou,  $\mu A(A_i)$  représente le degré d'appartenance de l'élément à l'état i et  $\mu B(B_j)$  représente le degré d'appartenance de l'élément à l'état j.

La Requete 1 sera traitée utilisant la méthode de pondération proposée dans [37] comme suit :

$$\begin{aligned} P(F = absent | Plat) &= \frac{\sum_{Plati \in I} \mu_{Plati}(150k) P(Plat = Plati | F = false) P(F = false)}{P(Plat)} \\ &= \frac{\mu_{low}(150k) P(Plat = low | F = Absent) P(F = Absent)}{P(Plat)} \\ &+ \frac{\mu_{high}(150k) P(Plat = high | F = Absent) P(F = Absent)}{P(Plat)} \\ &= \frac{0.62 * 0.49 * 0.49 + 0.38 * 0.5 * 0.49}{0.49} \\ &= 0.47 \end{aligned}$$

$$\begin{aligned}
P(F = Absent) &= P(F = false | Plat = low)P(Plat = low) \\
&\quad + P(F = Absent | Plat = high)P(Plat = high) \\
&= 0.62 * 0.06 + 0.5 * 0.93 \\
&= 0.49
\end{aligned}$$

$$\begin{aligned}
P(Plat) &= \mu_{low}(150k) * P(Plat = low) + \mu_{high}(150k) * P(Plat = high) \\
&= 0.62 * 0.49 + 0.38 * 0.51 \\
&= 0.49
\end{aligned}$$

## 2.5.2 Distribution probabiliste floue

Cette méthode consiste à construire des distributions probabilistes floues [38, 39], les degrés d'appartenances seront directement intégrés dans les distributions des probabilités pour construire des distributions probabilistes floues .

Ainsi, cette méthode est fondée sur deux étapes. La première étape sert à calculer les degrés d'appartenances et les représenter en notation particulière (par des vecteurs), lors de la deuxième étape, les degrés d'appartenances calculés en première étape seront incorporés directement dans les probabilités conditionnelles pour répondre à des requêtes.

Pour répondre à la Requête 1 en utilisant la méthode des distributions probabilistes floues, nous devons tout d'abord représenter les degrés d'appartenances relatifs au noeud PLAT comme un Vecteur.

$$\begin{aligned}
PLAT &= [Low_{\mu_{low}(150k)}, High_{\mu_{low}(150k)}] \\
&= [Low_{0.62}, High_{0.38}]
\end{aligned}$$

Par la suite, il s'agit d'incorporer les degrés d'appartenances dans La CPT du noeud F.

$$\begin{aligned}
F &= \{P(F = Present|PLAT = High), P(F = Absent|PLAT = High)\} \mu_{High}(150000), \\
&\quad \{P(F = Present|PLAT = Low), P(F = Absent|PLAT = Low)\} \mu_{Low}(150000)] \\
&= [\{Present_{0.5}, Absent_{0.5}\}_{0.38}, \{Present_{0.51}, Absent_{0.49}\}_{0.62}] \\
&= [\{Present_{0.5*0.38+0.5*0.38}, Absent_{0.51 * 0.62 + 0.49 * 0.62}\}] \\
&= [\{Present_{0.38}, Absent_{0.62}\}]
\end{aligned}$$

Nous avons obtenu à la fin une distribution probabiliste floue pour la variable F en se basent sur les observations de la variable PLAT, A partir de cette distribution, nous pouvons constater que  $P(F = Absent|PLAT) = 0.62$  et  $P(F = Present|PLAT) = 0.38$

### 2.5.3 Méthode de l'évidence virtuelle

Elle consiste à ajouter pour chaque nœud flou dans le RB, un nœud fils appelé « nœud virtuel » qui sert à stocker les observations incertaines ou floues. En effet, les degrés d'appartenances seront mis dans la CPT du nœud virtuel et seront incorporées avec les probabilités conditionnelles comme des poids lors de l'inférence bayésienne en appliquant un algorithme d'apprentissage standard.

Pour calculer la requête  $P(F = Absent|PLAT)$ , nous devons remplacer le nœud d'observation PLAT par son nœud virtuel VPLAT, Dans ce Cas la requête devient :

$$\begin{aligned}P(F = Absent|VPLAT = low) &= \frac{P(VPLAT = low|F = Absent)P(F = Absent)}{P(VPLAT = low)} \\ &= \frac{0.62 * 0.06}{0.39} \\ &= 0.10.\end{aligned}$$

$$\begin{aligned}P(VPLAT = low) &= P(VPLAT = low|PLAT = low)P(PLAT = low) \\ &\quad + P(VPLAT = low|PLAT = High)P(PLAT = High) \\ &= 0.62 * 0.06 + 0.38 * 0.93 \\ &= 0.39.\end{aligned}$$

## 2.6 Travaux connexes

Dans cette section nous allons donner un aperçu sur les travaux les plus connexes à notre problématique.

### 2.6.1 Medical Decision Support Tool from a Fuzzy-Rules Driven Bayesian Network

- **Maladie** : La pneumonie est une infection pulmonaire d'origine bactérienne ou virale, potentiellement mortelle. Elle peut être provoquée par des virus, des bactéries ou des champignons. Selon le germe en cause, elles peuvent être bénignes ou au contraire très graves .
- **Objectif** : Construction d'un système intelligent d'aide à la décision (DSS) pour le diagnostic médical et la thérapie [40].
- **Méthodes** : Une nouvelle méthodologie pour la construction de RB en utilisant des règles "if-then" et les principaux aspects de la logique floue est clairement présentée. La modélisation et le raisonnement concerne la mise en œuvre de toutes les règles dans un

réseau avec une topologie spécifique sont donnés. La méthode qu'ils ont présentée dans l'article peut être généralisée à des bases de règles floues similaires.

- **Résultats** : Les résultats ont montré que la méthodologie proposée pour la construction de réseaux bayésiens (RB) à partir de règles floues donne une décision initiale sur la gravité des infections pulmonaires, fournissant des résultats similaires à ceux obtenus avec l'intuition des médecins.

## 2.6.2 AlzFuzzyOnto : Une ontologie floue pour l'aide à la décision dans le domaine de la maladie d'Alzheimer

- **Maladie** : La Maladie d'Alzheimer (MA) est la plus fréquente des maladies neurodégénératives, c'est-à-dire les maladies liées à une détérioration irréversible du fonctionnement des cellules du système nerveux .
- **Objectif** : Construction d'une ontologie floue de domaine, appelée « AlzFuzzyOnto », Cette ontologie permet une représentation sémantique des données médicales relatives au diagnostic et à la prise en charge de la Maladie d'Alzheimer [41].
- **Méthodes** : Etendre l'ontologie noyau de domaine Mind pour représenter plus de connaissances et ajouter à Mind un module capable d'assurer la représentation des connaissances relatives à la prise en charge de la MA tout en prenant en considération les incertitudes et les imprécisions y afférentes.
- **Résultats** : Une ontologie floue de domaine représentant des connaissances spécifiques à l'aide à la décision pour la maladie d'Alzheimer, l'ontologie « AlzFuzzyOnto ».

## 2.6.3 HEPO : The Hepatitis Ontology For Abductive Medical Diagnostic Systems

- **Maladie** : L'hépatite désigne toute inflammation aigüe ou chronique du foie. Dans le travail, les auteurs couvrent trois types d'hépatite, notamment l'hépatite A, l'hépatite B et l'hépatite C .
- **Objectif** : Construction de l'ontologie de l'hépatite «HEPO» , pour la conception de systèmes de diagnostic médical [42].
- **Méthodes** : Pour la conception de l'ontologie, HEPO, ils ont inclus quatre experts du domaine ; deux médecins qui travaillent comme gastro-entérologues tandis que les deux autres étaient des experts en ontologie. L'ontologie développée est basée sur OWL, l'ontologie proposée est construite grâce à un outil d'édition open source, Protégé 4.3.
- **Résultats** : Une ontologie hépatite, HEPO a été développée et validée via des requêtes SPARQL.

## 2.6.4 Bayesian Network Model for Diagnosis of Psychiatric Diseases

- **Maladie** : Les maladies psychiatriques la dépression, les addictions, l'anxiété et les phobies, les troubles de comportement alimentaires, les psychoses (dont la schizophrénie) ,
- **Objectif** : construction du réseau bayésien[43].
- **Méthodes** : les dépendances mathématiques détaillées du diagnostic de schizophrénie en utilisant le réseau bayésien.
- **Résultats** : un modèle RB dans le but d'améliorer le processus de diagnostic effectué par les médecins.

## 2.6.5 A novel clinical decision support system for liver fibrosis using evolutionary multi-objective method based numerical association analysis

- **Maladie** : La cirrhose (les données sur la fibrose hépatique du VHC ont été Utilisé) .
- **Objectif** : un système d'aide à la décision clinique pour la découverte automatique de règles d'association numériques et ne nécessitent pas de modifier ou de changer les données [44].
- **Méthodes** : La méthode MOPNAR basée sur l'intelligence artificielle a été modélisée et adaptée pour la découverte de règles d'association numériques sans utiliser de pré-processus.
- **Résultats** : un système d'aide à la décision clinique qui peut aider les médecins à obtenir des résultats plus pratiques pour le diagnostic. Basée sur la méthode MOPNAR.

## 2.6.6 Applying Machine Learning to Evaluate for Fibrosis in Chronic Hepatitis C

- **Maladie** : Hépatite C chronique.
- **Objectif** : les modèles de prédiction du risque clinique construits à l'aide de ML peuvent détecter efficacement la présence de fibrose hépatique chez les patients atteints d'hépatite C chronique [45].
- **Méthodes** : Les algorithmes de Machine Learning utilisés (Logistic Regression, Naïve Bayes, Decision Tree, Random Forest, Extreme Gradient Boosting, k-Nearest Neighbor, Support Vector Machine, Neural Networks, Ensemble Method)

- **Résultats** : un système d'aide à la décision clinique basée sur les modèles dérivés de la ML aide les médecins à pratiquer la médecine préventive en veillant à ce que moins de patients atteints d'hépatite C chronique présentant une fibrose hépatique.

## 2.6.7 Discussion

Le tableau de comparaison des travaux existants mentionnés auparavant qui représentent les systèmes d'aide à la décision médicale chaque système se diffère d'un autre par la représentation des critères ('maladie', 'ontologie', 'réseau bayésien', 'connaissance floues', 'Machine Learning'), qui sont les critères que nous proposons. Le résultat de notre analyse est présenté dans le tableau 2.1 Les critères liés au notre champ de travail sont :

- **Ontologie** (onto) indique si le système est basé sur une ontologie ou non.
- **Réseau bayésien** (RB) indique si un réseau bayésien est utilisé.
- **Connaissances floues** (CF) indique si les auteurs ont traité la connaissance floue.
- **Machine Learning** (ML) indique si les méthodes de machine learning ont été appliqués.

Nom système /Auteur	Maladie	Onto	RB	CF	ML
Zarikas et al., [40]	La pneumonie	non	oui	oui	non
AlzFuzzyOnto [41]	Alzheimer	oui	non	oui	non
HEPO [42]	Hépatite A/B/C	oui	non	non	non
Curiac et al., [43]	Schizophrénie	non	oui	non	non
Altay et al., [44]	La cirrhose	non	non	non	oui
Akella et al., [45]	Hépatite C chronique	non	non	non	oui

TABLE 2.1 – Tableau de comparaison entre les travaux existants

En faisant une analyse du tableau comparatif, nous pouvons remarquer que :

- Chaque système représente une certaine maladie ; le système de Zarikas et al., décrit la maladie (la pneumonie) par contre AlzFuzzyOnto décrit la maladie (Alzheimer)
- La création de chaque système est à la base de l'utilisation ou la non utilisation d'une ontologie comme il est mentionné dans le tableau les auteurs des systèmes (AlzFuzzyOnto, HEPO) ont utilisé une ontologie pour construire leur système, par contre les créateurs des systèmes (Zarikas et al., Curiac et al., Altay et al., Akella et al.,) n'ont pas utilisé l'ontologie.
- Pour les réseaux bayésiens, nous remarquons selon notre étude que beaucoup de systèmes n'ont pas utilisé le réseau pour leur système d'aide au diagnostic médical. Les systèmes (Zarikas et al., Curiac et al.,) ont bien présenté que leur système est formé selon un réseau bayésien.
- Les connaissance floues sont utilisé seulement dans les systèmes (Zarikas et al., AlzFuzzyOnto).

Comme nous l'avons indiqué, plusieurs systèmes d'aide à la décision médicale ont été développés. Ces systèmes sont similaires dans certaines caractéristiques. Malgré cela, leur étude nous a permis d'identifier un ensemble de limites :

- Ils utilisent des ontologies classiques pour représenter les données médicales, ces ontologies sont basées sur une logique booléenne qui ne permet pas de représenter la connaissance médicale floue.
- Ils utilisent des RB standards pour faire le diagnostic médical, ces derniers ne peuvent pas faire des inférences probabilistes lorsque les observations sont floues.

## **2.7 Conclusion**

Dans ce chapitre nous avons couvert les principaux aspects concernant les réseaux bayésiens et les limites des réseaux bayésiens pour enchaîner sur les réseaux bayésiens flous. Avec l'utilisation des ensembles flous. Après nous allons aborder le chapitre de la conception de ce travail.



## Chapitre 3

---

# Conception de système

---

### 3.1 Introduction

Après la compréhension des deux domaines phares de notre système l'ontologie floue et le réseau bayésien flou, viens, l'étape de conception du système avec la construction ontologie floue, ainsi que, le réseau bayésien flou.

Nous allons présenter dans ce chapitre la motivation du travail, la conception du système avec le schéma global de notre système. Nous commencerons par le prétraitement de données et nous présenterons les différentes phases du processus de développement d'ontologies floues et la construction de réseau bayésien flou (RBF). Enfin, un diagnostic de l'état du patient en se basant sur le modèle RBF.

### 3.2 Motivation

L'hépatite C est une maladie du foie causée par un virus VHC (Virus de l'Hépatite C). Ce virus peut entraîner des hépatites aiguës comme des hépatites chroniques. La gravité des hépatites C est variable et peut aller d'une forme bénigne, d'une durée limitée à quelques semaines, à une maladie grave qui s'installe à vie [10].

L'hépatite C est une maladie relativement fréquente. On estime que 3% de la population mondiale a une infection chronique par le virus de l'hépatite C (VHC). Ce virus est responsable d'environ 70% des cas d'hépatites chroniques, cause majeure de cirrhose et de carcinome hépatocellulaire. L'hépatite C est fréquente partout dans le monde. Les régions de OMS (L'Organisation mondiale de la santé) les plus touchées sont la région de la Méditerranée orientale et la région européenne, avec une prévalence estimée en 2015 respectivement à 2,3% et 1,5% [10].

La période d'incubation pour l'hépatite C va de deux (02) semaines à six (06) mois. Après l'infection initiale, 80% environ des individus sont asymptomatiques. Ceux atteints d'une forme aiguë présentent des symptômes parmi les suivants : fièvre, fatigue, manque d'appétit, nausées,

vomissements, douleurs abdominales, coloration plus sombre des urines, coloration grisâtre des selles, douleurs articulaires et ictère (jaunissement de la peau et du blanc des yeux) [10].

L'Organisation mondiale de la santé (OMS) [46] a fixé pour objectif l'élimination de l'infection par le (VHC), en tant que menace pour la santé publique en 2030 (définie comme une diminution de 90% des nouvelles infections, associée à une réduction de la mortalité liée au (VHC) de 65%). L'élimination de l'infection par le (VHC), en tant que menace pour la santé publique, nécessite que 90% des personnes infectées soient diagnostiquées et que 80% des personnes diagnostiquées soient traitées.

Le diagnostic précoce permet d'empêcher les problèmes de santé qui pourraient découler de l'infection et de prévenir la transmission du virus.

La plupart des systèmes d'aide à la décision médicale qui prennent en considération l'incertitude se reposent sur les RBs classiques ou bien sur les ontologies classiques. Néanmoins, il est possible d'améliorer encore la représentation et le raisonnement avec la connaissance médicale incertaine en combinant les ontologies floues et les RBs flous pour faire face à l'incertitude dans les ontologies médicales du domaine riches en termes d'incertitude.

Dans ce regard et à travers le présent chapitre, nous présentons notre système, qui combine les ontologies floues et les réseaux bayésiens flous afin de traiter à la fois les connaissances probabilistes et les connaissances floues du domaine médical, en particulier l'hépatite C. L'avantage de notre solution réside dans le fait que :

- La représentation de la connaissance médicale floue en utilisant une ontologie floue. Cette ontologie définit des vocabulaires structurés, regroupant des concepts utiles d'un domaine et de leurs relations qui servent à organiser et échanger des connaissances floues de façon non ambiguë.
- L'ontologie floue est utilisée pour construire des bases de connaissances. Une base de connaissance peut être formée par une ontologie et un ensemble d'instances de ces classes.
- La représentation de la connaissance probabiliste et incomplète en exploitant les capacités offertes par les RBF.
- Le réseau bayésien flou est utilisé essentiellement pour calculer les probabilités des événements sous des observations floues liés les uns aux autres par des relations de cause à effet.
- Traitement de la connaissance floue et probabiliste à la fois.

Pour conclure, notre système vise à profiter des avantages des ontologies floues et les RBFs, nous avons combiné entre elles pour bien fournir une sémantique au domaine et une meilleure représentation des données du patient « Ontologie floue » plus une inférence basée sur un modèle RBF pour prédire l'état du patient.

### 3.3 Schéma global de notre système

Notre objectif consiste à développer un système d'aide au diagnostic médical pour prédire le diagnostic sur la maladie de l'hépatite C fondé sur un modèle d'ontologie floue et un moteur d'inférence basé principalement sur l'inférence fournie par le réseau bayésien flou. Notre système se compose de quatre (4) modules différents illustrés dans **la figure 3.1**, nous détaillons chaque module avec une brève explication.

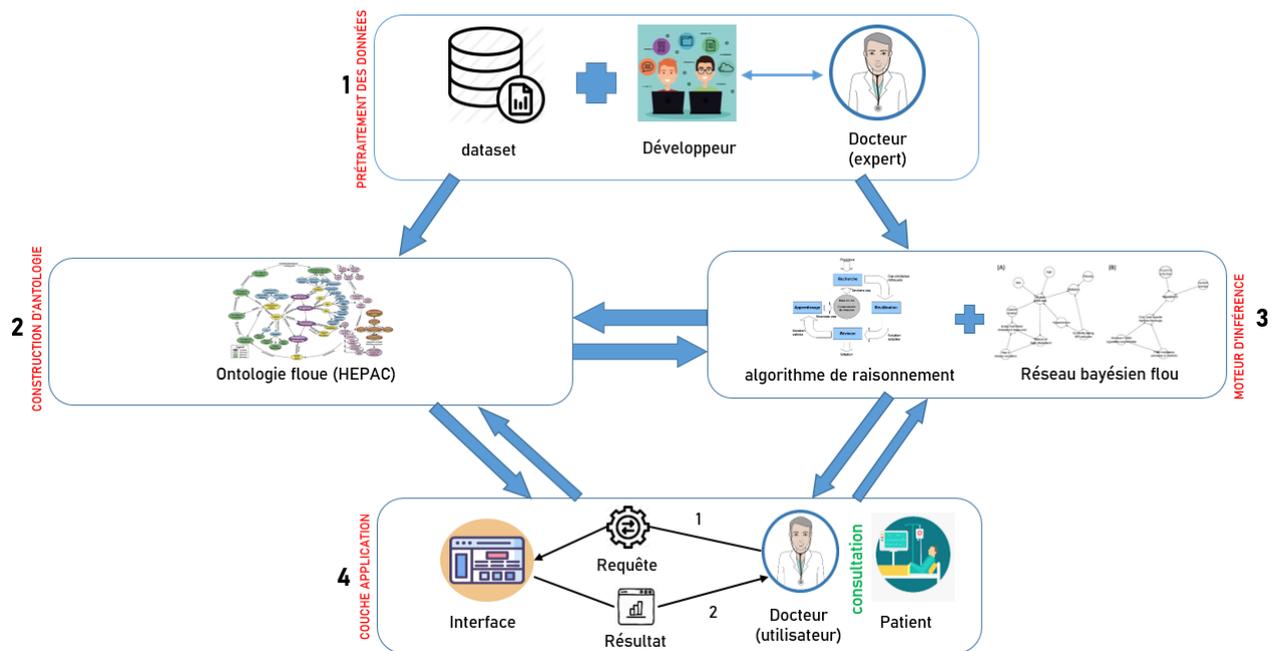


FIGURE 3.1 – Schéma global du système

1. **PRETRAITEMENT DES DONNEES** vise à présenter et analyser les données et faire les opérations de prétraitements nécessaires. Par la suite, nous allons présenter le dataset utilisé dans notre système. Nous allons montrer ainsi les prétraitements effectués pour rendre le dataset utilisable à des fins de décision.
2. **CONSTRUCTION D'ONTOLOGIE** facilite la représentation des données relatives aux patients et l'hépatite C en utilisant une ontologie floue. Cette ontologie forme la base de connaissances floues de notre système. Dans cette base de connaissances floues, une distinction claire a été établie entre les données floues et les données certaines. Par la suite, nous allons montrer les étapes de construction de l'ontologie floue de l'hépatite C en se basant sur la méthodologie (FODM).
3. **MOTEUR D'INFERENCE** contient le moteur d'inférence qui se chargera de déterminer le diagnostic médical pour un patient donné en se basant sur les informations en entrée. Le moteur d'inférence dans notre système tient en compte des données floues et des données probabilistes pour déterminer le diagnostic médical. Effectivement, il est basé sur un

modèle RBF pour calculer la probabilité qu'un patient souffre de l'hépatite C ou non, en utilisant les informations stockées dans la base de connaissances floues de notre système. Par la suite, nous allons montrer les étapes de construction d'un modèle RBF de l'hépatite C et nous présenterons ainsi l'algorithme d'inférence général adapté.

4. **COUCHE APPLICATION** est destiné aux utilisateurs finaux de notre système. A travers des interfaces conviviales, les médecins peuvent faire des requêtes de diagnostic pour un patient donné.

Le détail de chaque module sera représenté dans les sections suivantes.

### 3.4 Prétraitement des données

Dans cette section nous présentons le dataset utilisé et allons analyser le dataset afin d'appliquer les prétraitements nécessaires pour le rendre exploitable dans notre système.

Le dataset utilisé dans notre système contient les valeurs de laboratoire des donneurs de sang et des patients atteints d'hépatite C et des valeurs démographiques comme l'âge [47]. Nous présentons dans le tableau ci-dessous la description du dataset :

Caractéristiques de l'ensemble de données :	Multivarié	Nombre d'instances	615	Surface :	Vie
Caractéristiques des attributs :	Entier, réel	Nombre d'attributs	14	Date du don :	2020-06-10
Taches associées :	Classification, Clustering	Des valeurs manquantes ?	oui	Nombre de visites Web :	42952

TABLE 3.1 – Descriptions du dataset

Source :

Creators : Ralf Lichtinghagen, Frank Klawonn, Georg Hoffmann

Donor : Ralf Lichtinghagen : Institute of Clinical Chemistry ; Medical University Hannover (MHH) ; Hannover, Germany ;

Donor : Frank Klawonn ; Helmholtz Centre for Infection Research ; Braunschweig, Germany ;

Donor : Georg Hoffmann ; Trillium GmbH ; Grafrath, Germany ;

— **Attributs** : Le tableau ci-dessous représente la liste des attributs du dataset avec leurs abréviations et types :

Attributs	Abréviations	Type	intervalles
Category	Category	Nominale (String)	(values : '0=Blood Donor', '0s=suspect Blood Donor', '1=Hepatitis', '2=Fibrosis', '3=Cirrhosis')
Age	Age	Numérique (Integer)	[23 :77]
Sexe	Sex	Nominale (String)	f ou m
Albumine	ALB	Numérique (Real)	[14.9 :82.2]
phosphatase alcaline	ALP	Numérique (Real)	[11.3 :416.6]
alanine aminotransférase	ALT	Numérique (Real)	[0.9 :325.3]
aspartate aminotransférase	AST	Numérique (Real)	[10.6 :324]
bilirubine	BIL	Numérique (Real)	[0.8 :209]
Cholinestérases	CHE	Numérique (Real)	[1.42 :16.41]
cholestérol	CHOL	Numérique (Real)	[1.43 :9.67]
Créatinine	CREA	Numérique (Real)	[8 :1079.1]
gamma-glutamyl transférase	GGT	Numérique (Real)	[4.5 :650.9]
Protéines totales	PROT	Numérique (Real)	[44.8 :86.5]

TABLE 3.2 – Attributs et abréviations du dataset

— **Information sur les attributs** :

**Albumine** : L'albumine est la protéine la plus abondante (60%) dans le sang. Elle est fabriquée par les hépatocytes (les cellules du foie), mais peut aussi provenir de l'alimentation (on la trouve par exemple dans le blanc d'œuf ou encore dans le lait)[48].

**phosphatase alcaline** : La phosphatase alcaline (ALP) est une enzyme produite par plusieurs tissus dont les os, le foie, les canaux biliaires, l'intestin et le placenta. Une élévation des taux de phosphatase alcaline dans le sang peut donc provenir de différentes causes (ou isoenzymes), le plus souvent une atteinte du foie ou des os [49].

**Alanine aminotransférase** : L'alanine aminotransférase (ALAT ou SGPT) est une enzyme présente principalement dans le foie.

Un taux bas d'ALAT dans le sang n'a aucune signification clinique. Un niveau plus élevé que la norme indique généralement que des cellules du foie sont endommagées et laissent échapper leur contenu dans le sang. Un taux élevé d'ALAT (300 unités par litre [U/L] et plus) est généralement causé par une hépatite virale aiguë [50].

**Aspartate aminotransférase :** L'aspartate aminotransférase (ASAT ou SGOT) est une enzyme présente dans plusieurs tissus, en particulier dans le foie et les muscles, incluant le muscle cardiaque. Son dosage est surtout indiqué pour le diagnostic et le suivi des maladies du foie.

Un niveau bas d'ASAT n'a aucune signification clinique. Un niveau plus élevé que la norme indique que des cellules du foie ou des muscles sont endommagées et laissent échapper leur contenu dans le sang. Un taux élevé d'ASAT (350 unités par litre [U/L] et plus) est généralement causé par une hépatite virale aiguë [51].

**Bilirubine :** La bilirubine non conjuguée (indirecte, libre) est un produit de la dégradation de l'hémoglobine des globules rouges. Le foie transforme cette bilirubine en bilirubine directe qui pourra être éliminée dans l'intestin par la vésicule biliaire. Les taux de bilirubine totale sont donc à la fois un reflet de la destruction des globules rouges ainsi que du bon fonctionnement du foie, de la vésicule et des canaux biliaires.

Un taux élevé de bilirubine totale (hyperbilirubinémie), surtout causé par une augmentation de la fraction non conjuguée, peut signaler un problème de destruction des globules rouges (plusieurs causes), une réaction à la suite d'une transfusion sanguine ou une cirrhose [52].

**Cholinestérases :** est présente dans le foie, le pancréas, le cœur, la substance blanche du cerveau et le sérum. Elle ne doit pas être confondue avec l'acétylcholinestérase des globules rouges. Cette enzyme est sécrétée par le foie et son activité est abaissée dans les pathologies hépatiques chroniques [53].

**cholestérol :** Le cholestérol est un type de lipide (gras) fabriqué par le foie et présent dans les aliments que l'on consomme. Pour être en bonne santé, nous avons besoin de cholestérol parce qu'il accomplit plusieurs fonctions essentielles au bon fonctionnement de votre organisme [54].

**créatinine :** La créatinine est un produit de dégradation de la créatine et est éliminée essentiellement par les reins. Sa concentration sérique dépend de la capacité d'élimination rénale et de la masse musculaire. L'élévation de la créatinine plasmatique est un signe de dysfonctionnement de la filtration glomérulaire [55].

**gamma-glutamyl transférase :** La gamma-glutamyl transférase (GGT) est une enzyme présente dans le foie et les canaux biliaires.

Une élévation des taux de GGT dans le sang indique une atteinte du foie, mais sans spécifier de quelle nature (hépatite, cirrhose, atteinte des voies biliaires, etc.) [56].

**protéines totales :** L'albumine produite par le foie et les globulines, principalement les anticorps (gammaglobulines) sont les principales fractions de protéines retrouvées dans le sang. Le taux de protéines totales dans le sang fournit une indication générale concernant l'état d'hydratation, le statut nutritionnel ou encore le bon fonctionnement de certains organes importants comme le foie, les reins ou la moelle osseuse. Un résultat de protéines totales (protéines sériques totales) doit être interprété à l'aide d'autres tests dont le taux

d'albumine, le rapport A/G et parfois l'électrophorèse des protéines[57].

Un taux anormalement bas de protéines dans le sang (hypoprotéinémie) suggère une maladie du foie (faible production d'albumine)[57].

La figure ci-dessous représente le Dataset.

	Unnamed: 0	Category	Age	Sex	ALB	ALP	ALT	AST	BIL	CHE	CHOL	CREA	GGT	PROT
0	1	0=Blood Donor	32	m	38.5	52.5	7.7	22.1	7.5	6.93	3.23	106.0	12.1	69.0
1	2	0=Blood Donor	32	m	38.5	70.3	18.0	24.7	3.9	11.17	4.80	74.0	15.6	76.5
2	3	0=Blood Donor	32	m	46.9	74.7	36.2	52.6	6.1	8.84	5.20	86.0	33.2	79.3
3	4	0=Blood Donor	32	m	43.2	52.0	30.6	22.6	18.9	7.33	4.74	80.0	33.8	75.7
4	5	0=Blood Donor	32	m	39.2	74.1	32.6	24.8	9.6	9.15	4.32	76.0	29.9	68.7
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
610	611	3=Cirrhosis	62	f	32.0	416.6	5.9	110.3	50.0	5.57	6.30	55.7	650.9	68.5
611	612	3=Cirrhosis	64	f	24.0	102.8	2.9	44.4	20.0	1.54	3.02	63.0	35.9	71.3
612	613	3=Cirrhosis	64	f	29.0	87.3	3.5	99.0	48.0	1.66	3.63	66.7	64.2	82.0
613	614	3=Cirrhosis	46	f	33.0	NaN	39.0	62.0	20.0	3.56	4.20	52.0	50.0	71.0
614	615	3=Cirrhosis	59	f	36.0	NaN	100.0	80.0	12.0	9.07	5.30	67.0	34.0	68.0

FIGURE 3.2 – Une portion des lignes du dataset

Pendant cette phase de préparation, les données brutes sont vérifiées avec soin afin de déceler d'éventuelles erreurs. Les informations sont présentés comme suit dans la figure ci-dessous.

```

RangeIndex: 615 entries, 0 to 614
Data columns (total 14 columns):
#   Column      Non-Null Count  Dtype
---  -
0   Unnamed: 0  615 non-null    int64
1   Category    615 non-null    object
2   Age         615 non-null    int64
3   Sex         615 non-null    object
4   ALB         614 non-null    float64
5   ALP         597 non-null    float64
6   ALT         614 non-null    float64
7   AST         615 non-null    float64
8   BIL         615 non-null    float64
9   CHE         615 non-null    float64
10  CHOL        605 non-null    float64
11  CREA        615 non-null    float64
12  GGT         615 non-null    float64
13  PROT        614 non-null    float64
dtypes: float64(10), int64(2), object(2)

```

FIGURE 3.3 – Informations sur le dataset

D'après ces informations, nous constatons des colonnes qui contiennent des 'null'. **La figure 3.4** indique le nombre de 'null' présent dans chaque colonne.

Unnamed:	0
Category	0
Age	0
Sex	0
ALB	1
ALP	18
ALT	1
AST	0
BIL	0
CHE	0
CHOL	10
CREA	0
GGT	0
PROT	1

FIGURE 3.4 – Nombre de 'null' pour chaque colonne

Les causes des données non disponibles :

- Incohérences avec d'autres données et donc supprimées.
- Non saisies car non ou mal comprises.
- Considérées peu importantes au moment de la saisie.

### 3.4.1 Nettoyage des données

Dans cette section nous allons présenter les opérations effectuées pour traiter les valeurs manquantes. En effet, plusieurs techniques peuvent être exploitées pour traiter ces valeurs manquantes, parmi lesquelles :

- **Ignorer le tuple** : Peu efficace quand le pourcentage de valeurs manquantes est élevé.
- **Compléter manuellement les données** : Laborieux ou infaisable
- **Utiliser la moyenne de l'attribut**
- **Utiliser la moyenne de l'attribut pour la même classe** : mieux

La meilleure technique pour notre cas c'est d'utiliser la moyenne de l'attribut pour la même classe. Cette technique consiste à remplacer les 'null' par la moyenne de la classe : mesuré la tendance centrale qui représente le centre d'un groupe de nombres dans une distribution statistique.

Nous avons utilisé la moyenne pour remplacer les valeurs 'null' parce que par la suite on va utiliser l'approche maximum de vraisemblance (MV), cette approche est utilisée dans le cas où toutes les variables sont observés ( pas de données manquantes).

Après le remplacement des null part des moyennes de la classe correspondante, nous avons vérifié les résultats présentés dans la figure 3.5.

	Unnamed: 0	Category	Age	Sex	ALB	ALP	ALT	AST	BIL	CHE	CHOL	CREA	GGT	PROT
0	1	0=Blood Donor	32	m	38.5	52.50000	7.7	22.1	7.5	6.93	3.23	106.0	12.1	69.0
1	2	0=Blood Donor	32	m	38.5	70.30000	18.0	24.7	3.9	11.17	4.80	74.0	15.6	76.5
2	3	0=Blood Donor	32	m	46.9	74.70000	36.2	52.6	6.1	8.84	5.20	86.0	33.2	79.3
3	4	0=Blood Donor	32	m	43.2	52.00000	30.6	22.6	18.9	7.33	4.74	80.0	33.8	75.7
4	5	0=Blood Donor	32	m	39.2	74.10000	32.6	24.8	9.6	9.15	4.32	76.0	29.9	68.7
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
610	611	3=Cirrhosis	62	f	32.0	416.60000	5.9	110.3	50.0	5.57	6.30	55.7	650.9	68.5
611	612	3=Cirrhosis	64	f	24.0	102.80000	2.9	44.4	20.0	1.54	3.02	63.0	35.9	71.3
612	613	3=Cirrhosis	64	f	29.0	87.30000	3.5	99.0	48.0	1.66	3.63	66.7	64.2	82.0
613	614	3=Cirrhosis	46	f	33.0	68.28392	39.0	62.0	20.0	3.56	4.20	52.0	50.0	71.0
614	615	3=Cirrhosis	59	f	36.0	68.28392	100.0	80.0	12.0	9.07	5.30	67.0	34.0	68.0

FIGURE 3.5 – Une portion des lignes du dataset après nettoyage

## 3.5 Construction d'une ontologie floue pour l'hépatite C

La méthodologie FODM a été proposée dans un cadre général pour le développement des ontologies floues. Dans notre étude, nous allons suivre cette méthodologie afin de construire une ontologie floue d'hépatite C formalisée en Fuzzy-OWL 2. La méthodologie FODM organise les tâches en plusieurs phases. Dans ce qui suit, nous allons expliquer chaque phase.

### 3.5.1 Phase 1 : Objectifs et le champ d'application de l'ontologie

**Q1** : Quel est le domaine ou la portée des informations à modéliser ?

**Réponse** : Le domaine médical plus spécialement la maladie de l'hépatite C qui va être inspecté et ses différents symptômes sont l'intérêt de la modélisation, et le champ d'application est de trouver une ressource telle que le sens des concepts médicaux et les relations sémantiques qui existent entre eux puissent être représentés formellement pour pouvoir être exploités sans ambiguïté par un système informatique.

**Q2** : L'ontologie est-elle la meilleure technique de modélisation par rapport aux autres solutions ?

**Réponse** : La communication entre les médecins et les patients devrait être obtenue en utilisant un vocabulaire formalisé pour exprimer les symptômes de la maladie de l'hépatite C, l'ontologie est la technique de modélisation la plus prometteuse pour fournir cette spécification

de conceptualisation. Une ontologie plus compacte et vraisemblablement à des inférences plus efficaces en temps de calcul.

**Q3** : Quel sera le type de l'ontologie envisagée ?

**Réponse** : Le type de l'ontologie est une « ontologie du domaine » est fonctionnelle et orientée objet. Elle est utilisée pour représenter un domaine sous forme de base de connaissances. Elle présente les concepts-clés, les attributs et les instances relatifs au domaine.

Selon la description, nous avons modélisé des informations limitées à un domaine spécifique qui est la maladie de l'hépatite C.

**Q4** : Qui sera impliqué dans le développement de l'ontologie et quels rôles ils vont jouer ?

**Réponse** : Les développeurs (nous) pourraient être les principaux participant, tandis que les experts (Médecin) pourraient apporter des connaissances approfondies à caractérisation.

### **3.5.2 Phase 2 : Identifier le besoin de représenter les connaissances floues**

Selon la description du domaine cible, le diagnostic médical de l'hépatite C et les informations de modélisation prévues sont les symptômes et les analyses biologiques. Il est supposé qu'avec des techniques avancées et des travaux existants on sera capable de déduire si le patient a l'hépatite C ou non. Toutes les informations sont sûres et pourraient être conceptualisées par des éléments d'ontologie précis. Cependant, les experts et le médecin ont mis en avant une exigence particulière pour la modélisation d'ontologie.

Outre de déduire si le patient a l'hépatite C ou non, on aimerait connaître la classification de l'hépatite C. Ainsi, combien le patient est impliqué à ces classes avec des valeurs numériques explicites connues selon une spécification linguistique implique du flou.

Par exemple, le degré d'appartenance est flou pour chacune des classes de l'hépatite C à savoir : Blood Donor, Suspect blood donor, Hepatitis, Fibrosis, Cirrhosis. Sur la base de cette analyse, le flou est nécessaire pour gérer l'imprécision inhérente à la classification de l'hépatite C pour un patient. La conclusion tirée de cette phase est qu'une ontologie floue, au lieu d'une ontologie classique est déterminée pour modéliser le domaine de l'Hépatite C.

### **3.5.3 Phase 3 : Déterminer les connaissances floues et précises**

Puisqu'une ontologie floue est déterminée pour modéliser le domaine de diagnostic médical hépatites C, une coopération étroite entre les médecins et les ingénieurs en ontologie est requise à ce stade. Une distinction claire entre les informations liées floues et certaines informations doit être établie. Après avoir recueilli les connaissances et les suggestions des médecins, les ingénieurs en ontologie établissent un diagnostic précis de la maladie entre les informations floues et certaines informations.

L'ensemble des concepts est divisé en deux parties : des concepts précis et des concepts flous. Ensuite, ils peuvent être modélisés avec différents traitements, respectivement.

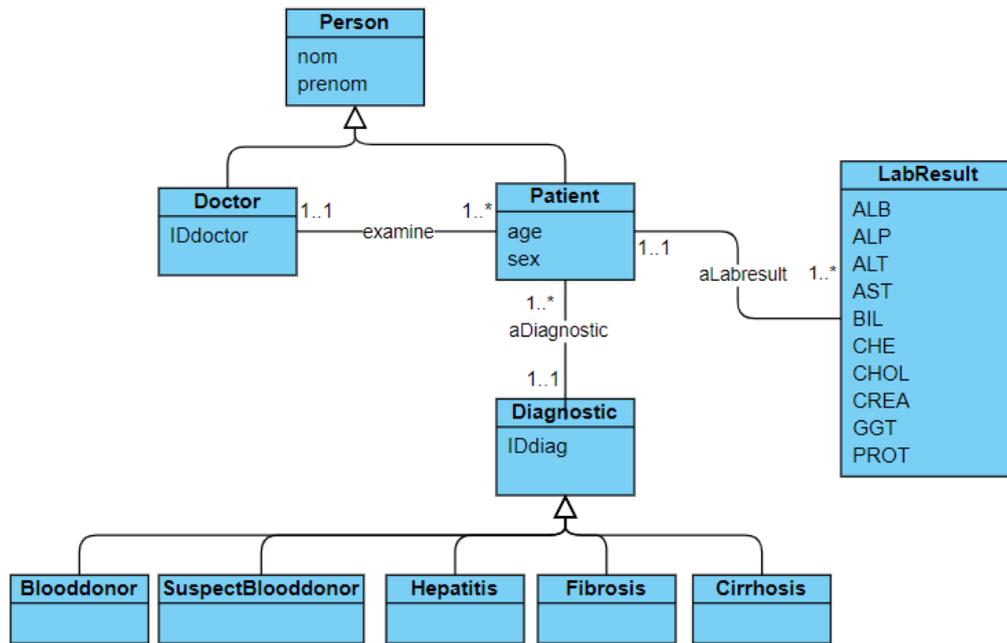


FIGURE 3.6 – Diagramme de classes

Concepts précis	Concepts flous
Personne Patient Doctor	ALB ALP ALT AST BIL CHE CHOL CREA GGT PROT.

TABLE 3.3 – Concepts précis et concepts flous

### 3.5.4 Phase 4 : Considérer la réutilisation des ontologies existantes

Puisqu’aucune ontologie floue existante n’a été réutilisée dans la construction, alors, les phases 5 et 6 pourraient être ignorées.

### 3.5.5 Phase 7 : Définir les éléments d'ontologie floue

Au cours de **la phase 3**, la base de connaissances sur l'hépatites C a été divisée en deux catégories : **les connaissances précises et les connaissances floues**. Dans le but de représenter des informations vagues et imprécises à l'aide d'une ontologie floue, différents éléments d'ontologie floue sont définis dans cette phase.

Les définitions des types de données floues et la fonction d'appartenance obtenue avec DATIL [58] sont présentées dans le tableau ci-dessous. DATIL (DATatypes with Imprecision Learner) est un logiciel qui apprend automatiquement les types de données floues pour les ontologies floues à partir de différents types d'entrées.

Datil implémente plusieurs algorithmes de clustering non supervisés : k-means, fuzzy c-means et mean-shift.

Pour chaque propriété de données dans une ontologie avec une plage numérique Datil collecte un tableau de nombres réels correspondant aux valeurs de la propriété pour différents individus. Un algorithme de clustering fournit un ensemble de centre de masse à partir de ces tableaux de valeurs. Ces centres de masse sont utilisés comme paramètres pour construire des fonctions d'appartenance floues partitionnant le domaine (voir **la figure 3.7**).

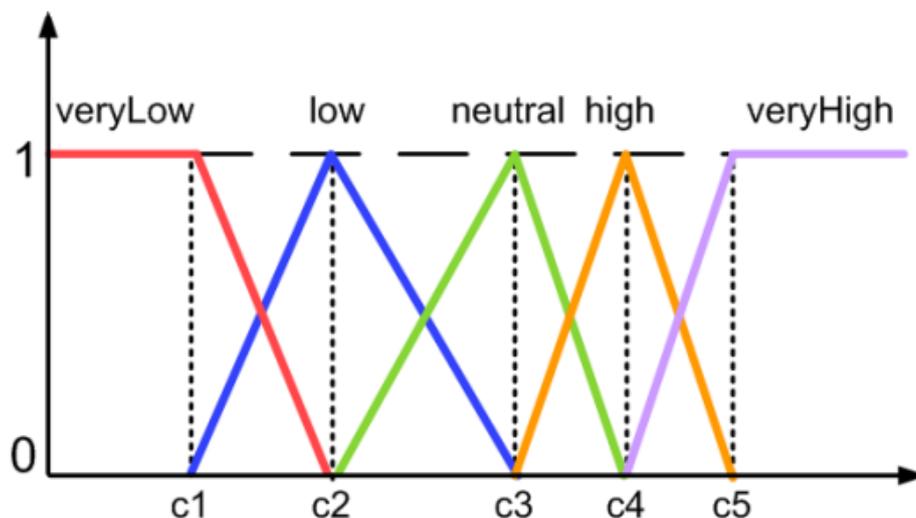


FIGURE 3.7 – Certaines fonctions d'appartenance floues construites à partir des Centre de masse

L'algorithme de clustering utilisé dans note cas est le "Fuzzy C-means"[59]. Cet algorithme fonctionne en attribuant une appartenance à chaque point de données correspondant à chaque centre de cluster sur la base de la distance entre le centre de cluster et le point de données.

Plus les données sont proches du centre de cluster, plus leurs appartenances au centre de cluster particulier sont élevées. Il est clair que la somme des membres de chaque point de données doit être égale à un.

Après chaque itération, les membres et les centres de cluster sont mis à jour selon la formule :

$$\mu_{ij} = 1 / \sum_{k=1}^c (d_{ij} | d_{ik})^{(2/m-1)} \quad (3.1)$$

$$\nu_j = \left( \sum_{i=1}^n (\mu_{ij})^m x_i \right) / \left( \sum_{i=1}^n (\mu_{ij})^m \right), \forall j = 1, 2, \dots, c \quad (3.2)$$

Où :

'n' est le nombre de points de données.

'm' est l'indice de flou  $m \in [1, +\infty[$ .

'c' représente le nombre de centre de cluster.

'vj' représente le j ème centre de cluster.

' $\mu_{ij}$ ' représente l'appartenance des ièmes données au jième centre de cluster.

' $d_{ij}$ ' représente la distance euclidienne entre les ièmes données et le jième centre de cluster.

L'objectif principal de l'algorithme de c-moyenne floue est de minimiser :

$$J(U, V) = \sum_{i=1}^n \sum_{j=1}^c (\mu_{ij})^m \| x_i - \nu_j \|^2 \quad (3.3)$$

Où : ' $\|x_i - \nu_j\|$ ' est la distance euclidienne entre la ième donnée et le jième centre de cluster.

### 3.5.6 Algorithme du Fuzzy C-means clustering

Soit  $\mathbf{X} = x_1, x_2, x_3, \dots, x_n$  l'ensemble des points de données et  $\mathbf{V} = v_1, v_2, v_3, \dots, v_c$  l'ensemble des centres.

1) Sélectionner au hasard les centres de cluster «c».

2) Calculer l'appartenance floue ' $\mu_{ij}$ ' en utilisant :

$$\mu_{ij} = 1 / \sum_{k=1}^c (d_{ij} | d_{ik})^{(2/m-1)} \quad (3.4)$$

3) Calculer les centres flous 'vj' en utilisant :

$$\nu_j = \left( \sum_{i=1}^n (\mu_{ij})^m x_i \right) / \left( \sum_{i=1}^n (\mu_{ij})^m \right), \forall j = 1, 2, \dots, c \quad (3.5)$$

4) Répéter les étapes 2) et 3) jusqu'à ce que la valeur minimale 'J' soit atteinte ou

$\|U(k+1) - U(k)\| < \beta$ . où,

'k' est l'étape d'itération.

' $\beta$ ' est le critère de terminaison entre [0, 1].

' $U = (\mu_{ij})_{n \times c}$ ' est la matrice d'appartenance floue..

'J' est la fonction objectif.

Nous présentons ci-dessous la liste des concepts flous avec ces types de données flous et ses fonctions d'appartenance et ses axiomes :

Concept Flou	Type de donnée	Fonction d'appartenance	Axiom
Age	LowAge	left-shoulder (17.74, 82.26, 39.14, 56.23)	$LOWAge \equiv Age \cap \exists hasAge.LowAge$
	HighAge	right-shoulder (17.74, 82.26, 39.14, 56.23)	$HIGHAge \equiv Age \cap \exists hasAge.HighAge$
ALB	LowALB	left-shoulder (11.85, 85.25, 29.53, 39.75)	$LowALB \equiv ALB \cap \exists hasALB.LowALB$
	NeutralALB	triangular (11.85, 85.25, 29.53, 39.75, 46.28)	$NeutralALB \equiv ALB \cap \exists hasALB.NeutralALB$
	HighALB	right-shoulder (11.85, 85.25, 39.75, 46.28)	$HighALB \equiv ALB \cap \exists hasALB.HighALB$
ALP	LowALP	left-shoulder (-2.43, 430.33, 47.9, 73.83)	$LowALP \equiv ALP \cap \exists hasALP.LowALP$
	NeutralALP	triangular (-2.43, 430.33, 47.9, 73.83, 105.71)	$NeutralALP \equiv ALP \cap \exists hasALP.NeutralALP$
	HighALP	right-shoulder(-2.43, 430.33, 73.83, 105.71)	$HighALP \equiv WBC \cap \exists hasWBC.HighALP$
ALT	LowALT	left-shoulder(-10.15, 336.35, 19.81, 50.19)	$LowALT \equiv ALT \cap \exists hasALT.LowALT$
	HighALT	right-shoulder(-10.15, 336.35, 19.81, 50.19)	$HighALT \equiv ALT \cap \exists hasALT.HighALT$
AST	LowAST	left-shoulder(-6.8, 341.4, 27.49, 166.6)	$LowAST \equiv AST \cap \exists hasAST.LowAST$
	HighAST	right-shoulder(-6.8, 341.4, 27.49, 166.6)	$HighAST \equiv AST \cap \exists hasAST.HighAST$
BIL	LowBIL	left-shoulder(-8.42, 218.22, 9.35, 162.87)	$LowBIL \equiv BIL \cap \exists hasBIL.LowBIL$
	HighBIL	right-shoulder(-8.42, 218.22, 9.35, 162.87)	$HighBIL \equiv BIL \cap \exists hasBIL.HighBIL$
CHE	LowCHE	left-shoulder(0.26, 17.57, 6.54, 9.81)	$LowCHE \equiv CHE \cap \exists hasCHE.LowCHE$
	HighCHE	right-shoulder(0.26, 17.57, 6.54, 9.81)	$HighCHE \equiv CHE \cap \exists hasCHE.HighCHE$
CHOL	LowCHOL	left-shoulder(0.83, 10.27, 4.6, 6.4)	$LowCHOL \equiv CHOL \cap \exists hasCHOL.LowCHOL$
	HighCHOL	right-shoulder(0.83, 10.27, 4.6, 6.4)	$HighCHOL \equiv CHOL \cap \exists hasCHOL.HighCHOL$
CREA	LowCREA	left-shoulder(-18.85, 1105.95, 71.58, 101.21)	$LowCREA \equiv CREA \cap \exists hasCREA.LowCREA$
	HighCREA	right-shoulder(-18.85, 1105.95, 71.58, 101.21)	$HighCREA \equiv CREA \cap \exists hasCREA.HighCREA$

GGT	LowGGT	left-shoulder(-24.26, 679.66, 30.77, 344.94)	$LowGGT \equiv GGT \sqcap \exists hasGGT.LowGGT$
	HighGGT	right-shoulder(-24.26, 679.66, 30.77, 344.94)	$HighGGT \equiv GGT \sqcap \exists hasGGT.HighGGT$
PROT	LowPROT	left-shoulder(41.97, 89.33, 67.69, 75.29)	$LowPROT \equiv PROT \sqcap PROT \exists hasPROT.LowPROT$
	HighPROT	right-shoulder(41.97, 89.33, 67.69, 75.29)	$HighPROT \equiv PROT \sqcap \exists hasPROT.HighPROT$

TABLE 3.4 – Concepts flous

### 3.5.7 Phase 8 : Définir les éléments d'ontologie précis

Dans cette phase, le reste de l'ontologie, à savoir les éléments d'ontologie précis, doit être défini pour modéliser des informations précises dans le domaine cible.

Concept précis	Définition
Personne	Etre humain, sans distinction de sexe.
Patient	Personne soumise à un examen médical, suivant un traitement ou subissant une intervention chirurgicale.
Doctor	Personne habilitée à exercer la médecine après avoir été admise à différents examens sanctionnant plusieurs années d'études médicales (universitaires et hospitalières ) et après avoir soutenu une thèse de doctorat.

TABLE 3.5 – Concepts précis

### 3.5.8 Phase 9 : Formalisation

Dans cette phase, Fuzzy-OWL 2 est sélectionné comme langage de formalisme pour représenter le modèle d'ontologie conçu. Pour effectuer facilement la transformation du modèle conceptuel en expressions formatées OWL 2, l'éditeur d'ontologie protégé et son extension Fuzzy OWL sont utilisés dans cette étape. Protégé permet des implémentations visualisées et faciles de l'ontologie de caractérisation floue (voir **la figure 3.8** et **la figure 3.9** ).

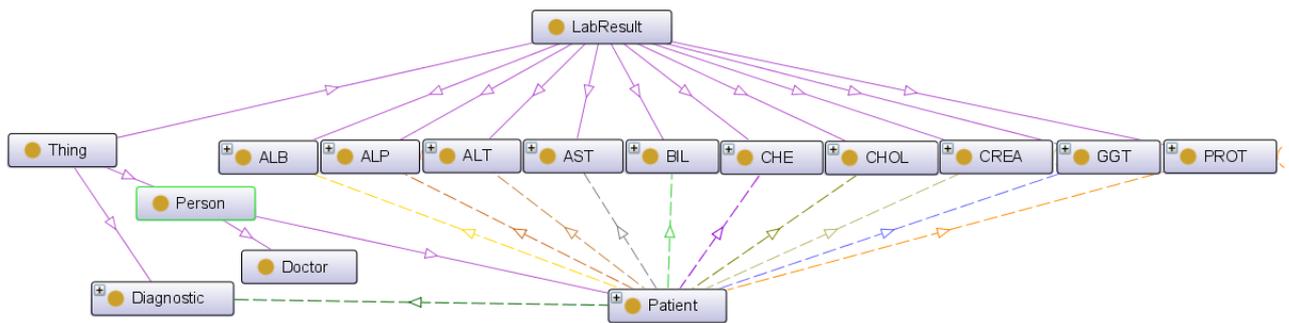


FIGURE 3.8 – La structure globale visualisée de l’ontologie floue

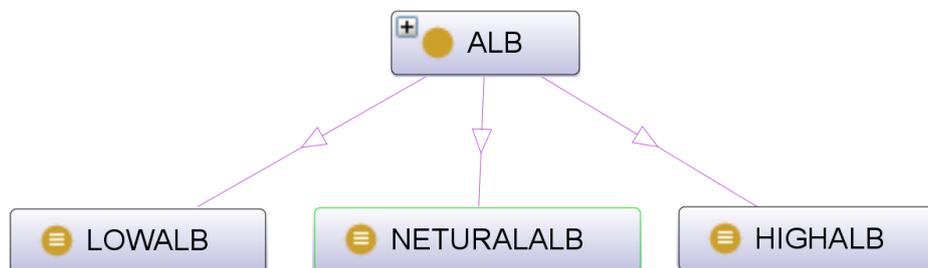


FIGURE 3.9 – Exemple du type de donnée

### 3.5.9 Instanciation de l’ontologie floue

L’instanciation (appelée aussi le peuplement) de l’ontologie floue vise à remplir la partie assertionnelle de l’ontologie. Le pseudo-code suivant « Algorithme 1 » formalise cette opération.

---

#### Algorithm 1 Algorithme de peuplement

---

Input : Ontology O1 + dataset CSV

Output : Ontology O1

$LabResult \leftarrow GetClasses(O1)$

**for all**  $Ligne_i \in CSV$  **do**

$P_i \leftarrow createIndividualP() \text{ patient}$

$Vect_i \leftarrow RetriveInformation(ligne_i)$

$P_i.addDatapropertyAge(Vect_i[age])$

$P_i.addDatapropertyGender(vect_i[gender])$

**for all**  $Subclass_j \in LabResult$  **do**

$C_j \leftarrow createIndividualSubclass_j()$

$C_j.addDataPropertySubclass_j(vect_i[subclass_j])$

$P_i.addObjectProperty("hasLabResult", C_j)$

**end for**

**end for**

---

L'algorithme de peuplement sert à créer des individus pour les concepts de l'ontologie floue . Tout d'abord, on doit récupérer les concepts générale de notre ontologie O1 en utilisant la fonction getClasses qui prend en paramètre l'ontologie O1.

Puis, on doit parcourir les lignes du dataset CSV pour récupérer les valeurs des 13 attributs et les placer dans un vecteur vecti.

Puis on doit créer un individu de concept Patient en utilisant la fonction CreateIndividuP et affecter à ce patient deux dataproperty qui sont l'âge et le sexe en appelant les deux fonctions addDatapropertyAge et addDatapropertyGender qui prend en paramètre les valeurs correspondantes dans le vecteur vecti.

En suite, on doit parcourir les sous concepts des concepts généraux (LabResult) qui se présentent dans notre ontologie O1 pour qu'on puisse créer des individus des sous concepts en utilisant la fonction createIndividualSubclass et affecter les dataproperty en appelant la fonction addDataPropertySubclass qui prend en paramètre la valeur correspondante dans le vecteur vecti.

Puis relier le patient qui a été crée au début avec ces individus des concepts généraux en utilisant la fonction addObjectProperty qui prend en paramètre le concept général et le sous-concept.

## 3.6 Construction de réseau Bayésien Flou

Les algorithmes d'apprentissage dans les réseaux bayésiens reviennent à définir un modèle RB qui s'adapte mieux aux connaissances du problème à représenter. Dans le reste de cette section nous décrivons l'apprentissage de la structure adapté dans notre système. Par la suite nous présentons l'apprentissage des paramètres. Nous concluons par présenter l'algorithme de raisonnement proposé pour le diagnostic.

### 3.6.1 Apprentissage de la structure

Le but de l'apprentissage de la structure est de trouver la structure graphique qui représente le mieux le problème à partir des données disponibles (dataset).

Pour faire l'apprentissage de structure, il nous faut deux étapes :

— **Etape 1** : catégorisation des attributs numériques

Après avoir collecté des données avec des experts du domaine, nous avons pu définir plusieurs classes pour chaque attribut. Cette étape est nécessaire pour construire le réseau. Nous convertissons les valeurs des facteurs existants dans l'ensemble de données en catégories. Ces catégories sont déterminées par intervalles avec l'aide d'experts du domaine en fonction de leur expérience.

Le reste des attributs est généralement sous des formes standardisées.

Le tableau ci-dessous représente le dataset obtenu après la définition des catégories pour chaque facteur à nominaliser.

Attributs	Intervalles	Normalisation
Age	Min = 23 ans Max = 77 ans	Young $\leq 39$ Adulte $> 39$
ALB	Min = 14.9 Max = 82.2	LowALB $< 29.53$ 29.53 $\leq$ NeutralALB $< 37.75$ HighALB $\geq 39.75$
ALP	Min = 11.3 Max = 416.6	LowALP $< 47.9$ 47.9 $\leq$ NeutralALP $< 73.83$ HighALP $\geq 73.83$
ALT	Min = 0.9 Max = 325.3	LowALT $< 19.81$ HighALT $\geq 50.19$
AST	Min = 10.6 Max = 324	LowAST $< 27.49$ HighAST $\geq 27.49$
BIL	Min = 0.8 Max = 209	LowBIL $< 9.35$ HighBIL $\geq 162.87$
CHE	Min = 1.42 Max = 16.41	LowCHE $< 6.54$ HighCHE $\geq 6.54$
CHOL	Min = 1.43 Max = 9.67	LowCHOL $< 4.6$ HighCHOL $\geq 4.6$
CREA	Min = 8 Max = 1079.1	LowCREA $< 71.58$ HighCREA $\geq 71.58$
GGT	Min = 4.5 Max = 650.9	LowGGT $< 30.77$ HighGGT $\geq 30.77$
PROT	Min = 44.8 Max = 86.5	LowPROT $< 67.69$ HighPROT $\geq 67.69$

TABLE 3.6 – Catégorisation des attributs numériques

— **Étape 2** : Construction du réseau bayésien

À l'aide de l'expert du domaine, nous avons pu construire le réseau bayésien de notre système.

La figure ci-dessous représente la structure du réseau Bayésien.

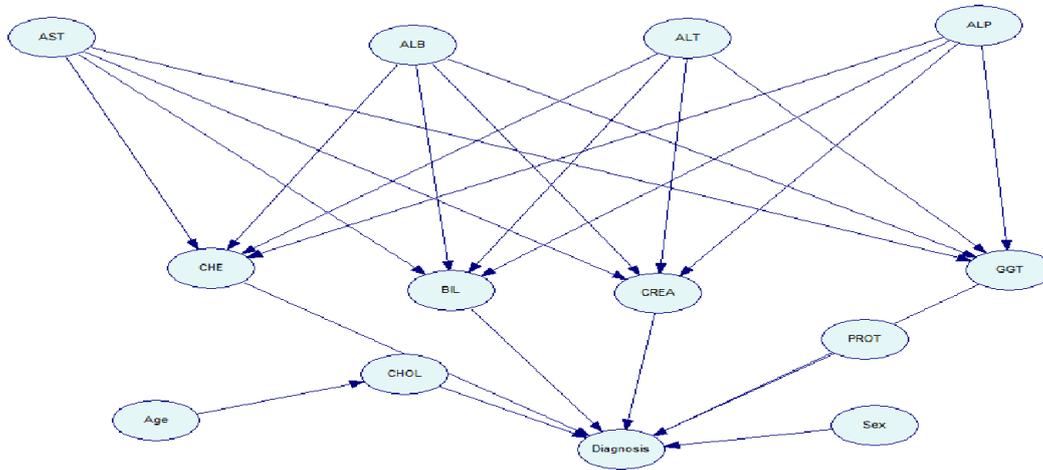


FIGURE 3.10 – Structure de réseau bayésien

### 3.6.2 Apprentissage des paramètres

L'apprentissage des paramètres, où nous avons réalisé la structure du réseau, il faudra estimer les probabilités conditionnelles de chaque nœud du réseau.

Nous avons nettoyé notre base d'apprentissage dans la première étape, qui ne contient plus de données manquantes. Il s'agit maintenant d'estimer la probabilité d'un événement en se basant sur la fréquence d'apparition de l'événement dans la base d'apprentissage. On appelle cette approche maximum de vraisemblance (MV).

Soit  $N_{ijk}$  le nombre d'occurrences simultanées dans la base où de la variable  $X_i$  est dans l'état  $x_k$  et ses parents sont dans la configuration  $x_j$ . Les paramètres du réseau  $\Theta_{i,j,k}^{(MV)}$  sont donnés comme suit :

$$\Theta_{i,j,k}^{(MV)} = N_{ijk}^* = \sum_{i=1}^n P(X_i = x_k \mid P_a(X_i) = x_j) \quad (3.6)$$

La figure ci-dessous montre les résultats obtenus des probabilités pour chaque nœud.

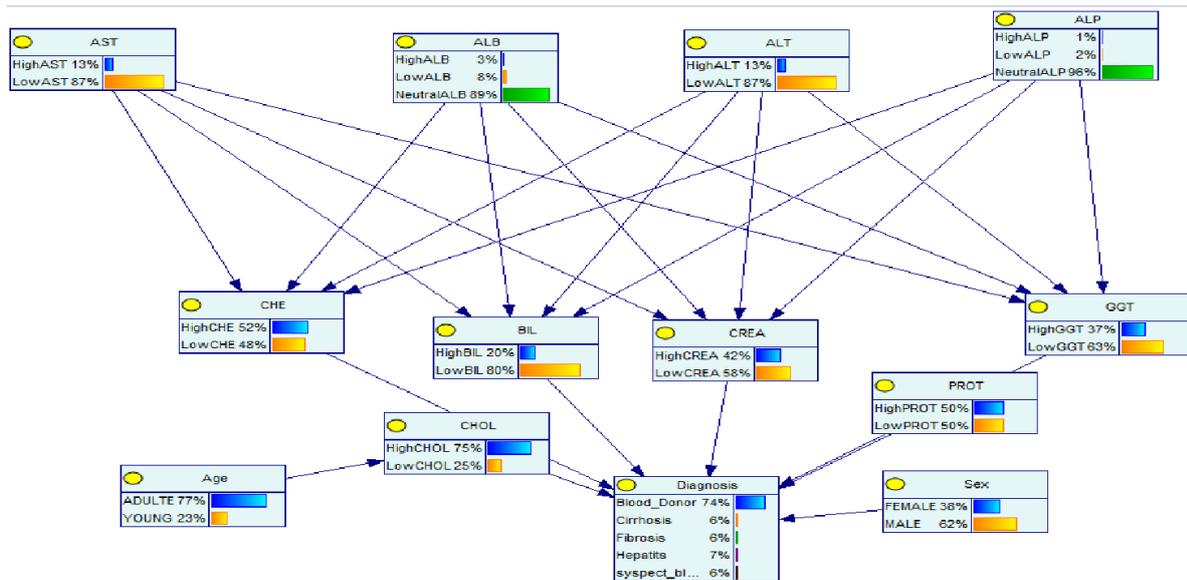


FIGURE 3.11 – Réseau Bayésien

### 3.7 Diagnostic de l'état du patient en se basant sur le modèle RBF

Dans cette section nous décrivons le diagnostic de l'état du patient en se basant sur le modèle RBF avec un diagramme représenté dans la figure 3.12.

le processus d'inférence adapté pour le diagnostic se déroule comme suit :

- Un contrôleur va vérifier la validité des données en répondant à cette question : Est-ce qu'ils sont conformes ou pas ? Deux cas sont possibles :

- **Données non valides** : les données ne sont pas valides. Alors, 'Notifier au docteur de l'invalidité des données'.
- **Données valides** : les données sont valides. Alors, Stocker les données dans l'ontologie.
  - Par la suite, les données enregistrées seront fuzzifier selon les fonctions d'appartenance déjà mentionnées dans le tableau 3.4.
  - Envoyer les données au RBF .
  - L'inférence du RBF s'exécute.
  - L'inférence donne un résultat de la probabilité de chaque classe de nœud 'Diagnostic'.
  - Après le processus va choisir le MAX('Blood Donor', 'Hepatits', 'Fibrosis', 'Cirrhosis' et 'suspect blood donor').
  - Au final notifier le docteur avec un message du diagnostic obtenu.

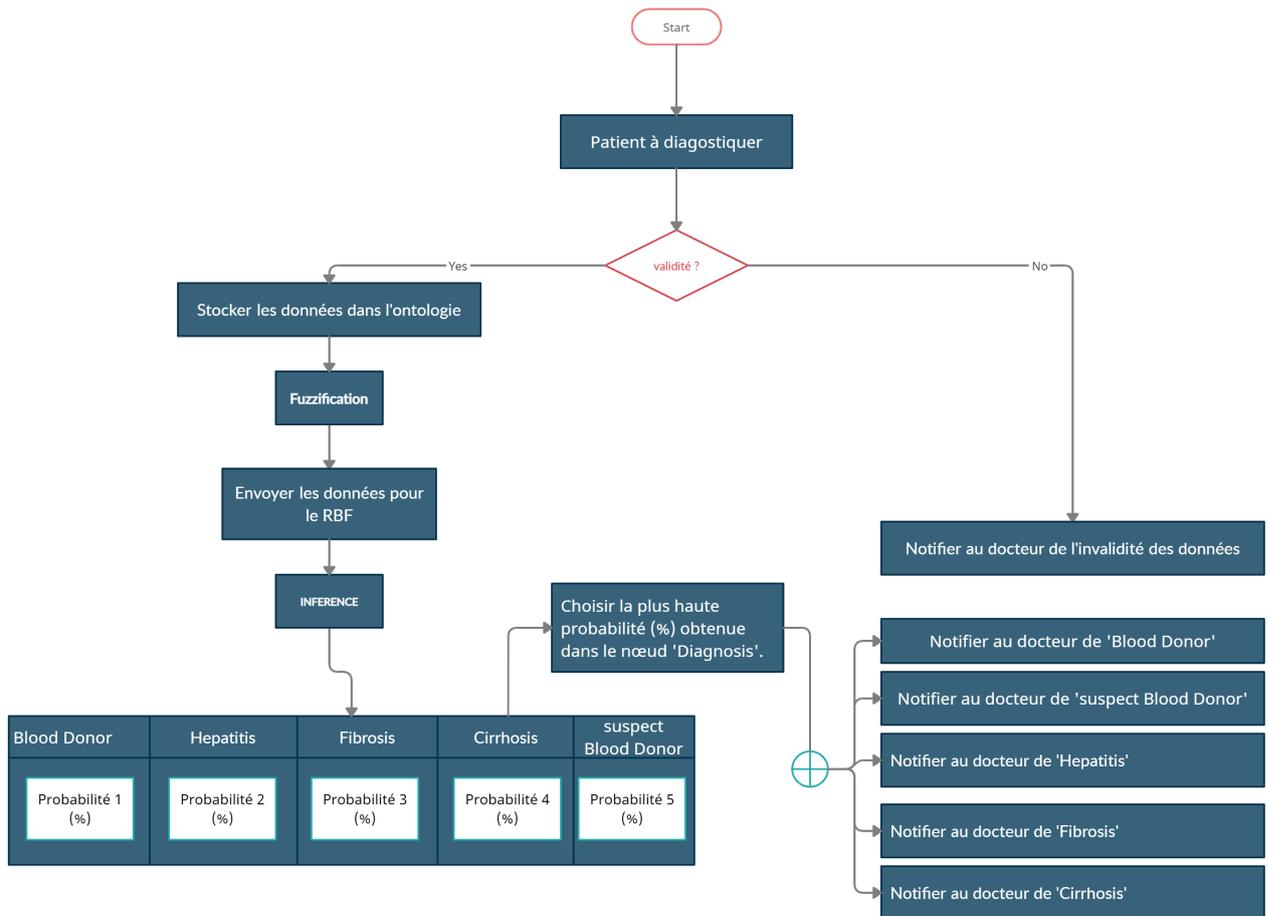


FIGURE 3.12 – Le processus d’inférence adapté pour le diagnostic

### 3.8 Conclusion

Nous avons présenté dans ce chapitre les bases de notre système avec le schéma global, le dataset et le processus de création de l’ontologie floue en utilisant la méthodologie FODM.

Nous avons procédé par la suite à la construction de réseau Bayésien flou ainsi que le diagnostic de l’état du patient en se basant sur le modèle RBF.

Au chapitre suivant nous allons implémenter l’ontologie floue modélisée dans ce chapitre, tester notre système et valider les résultats trouvés.



## Chapitre 4

---

# Implémentation et test

---

## 4.1 Introduction

Après l'étape conception du système, nous allons entamer l'étape finale qui consiste à l'implémentation de notre ontologie floue et le système d'aide au diagnostic.

Dans ce chapitre, nous allons présenter l'implémentation de notre ontologie floue et le système d'aide au diagnostic et les différents outils utilisés, avec le test du système et la validation des résultats.

## 4.2 Implémentation

L'implémentation de l'ontologie floue est la même que l'implémentation de l'ontologie classique. La différence est d'ajouter la partie floue. Ainsi pour implémenter notre ontologie nous utilisons l'éditeur Protégé version 4.3 avec le plug-in Fuzzy OWL 2.

### 4.2.1 Outils de programmation

Dans cette section, nous allons présenter les outils utilisés pour l'implémentation de notre système.

#### — Protégé

Protégé est un éditeur d'ontologies open-source qui supporte le langage OWL (Web Ontology Language). Il est disponible en téléchargement gratuit sur le site protégé (stanford.edu).

L'installation de protégé 4.3 est très simple. Il suffit de double-cliquer sur le programme téléchargé à partir du site. Ensuite, choisissez l'emplacement d'installation et le type de JVM (Java Virtual machine) puis suivez les autres étapes.

La figure suivante montre l'interface principale de l'éditeur protégé 4.3 après l'installation.

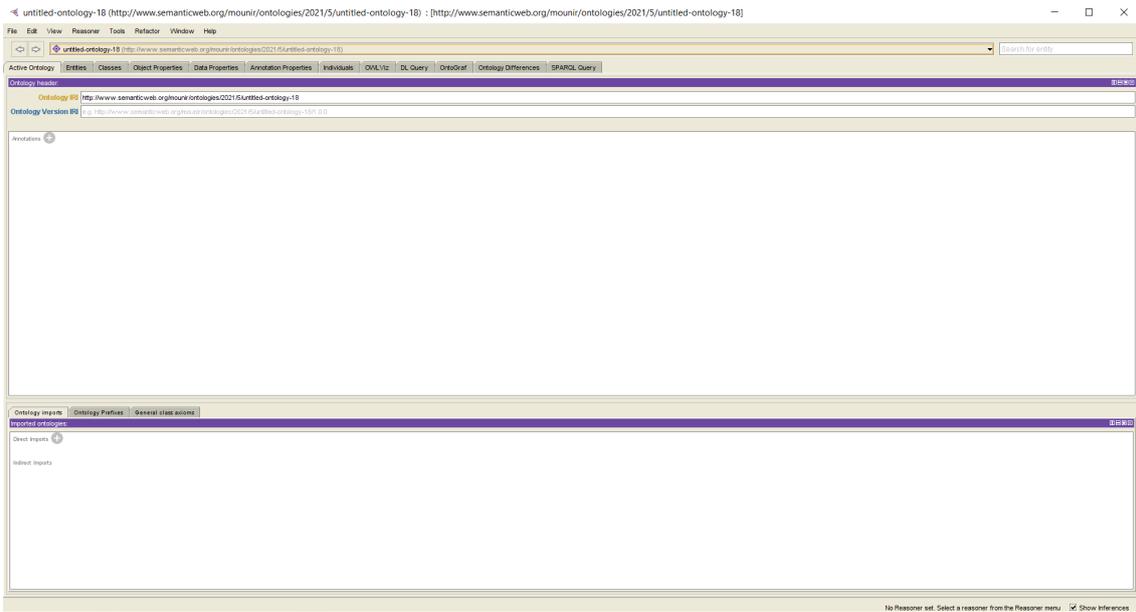


FIGURE 4.1 – Interface graphique de Protégé

## — Définition des concepts

La figure suivante représente l'hierarchie des concepts utilisés dans notre ontologie.

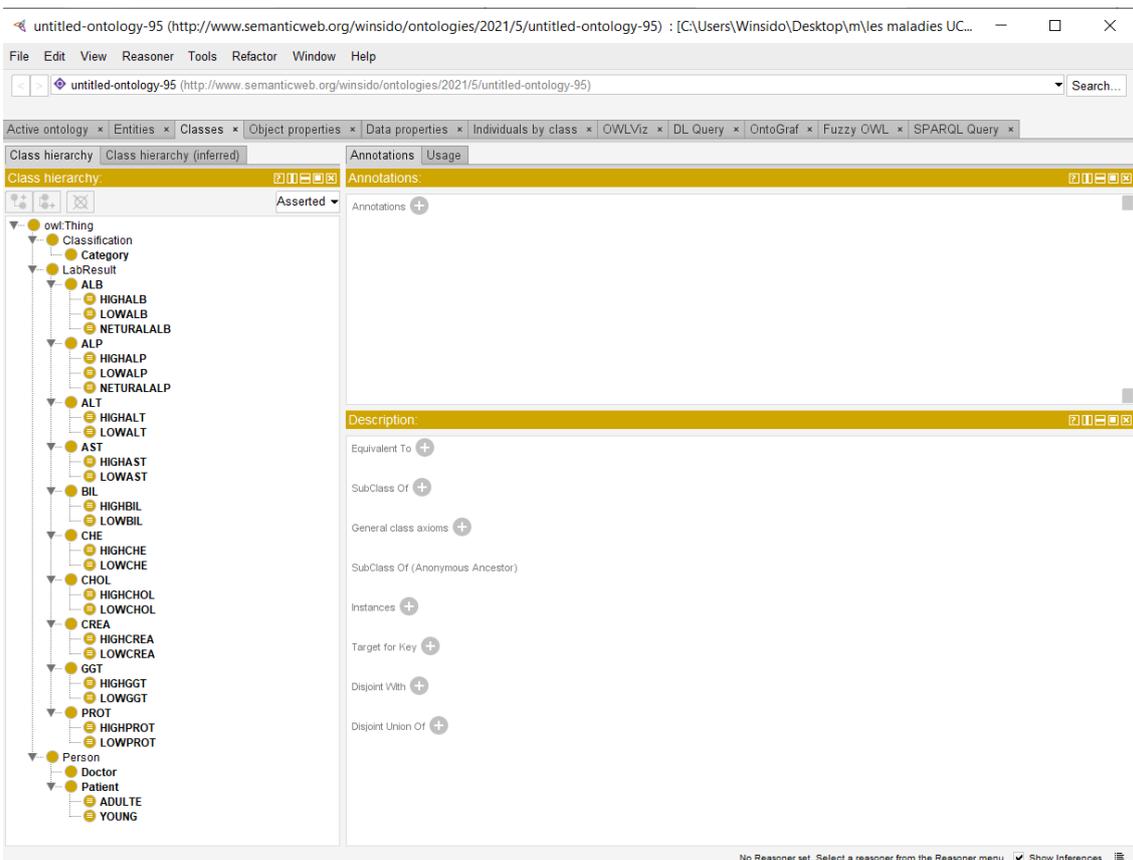


FIGURE 4.2 – Hiérarchie de concept

## — Définition des relations entre les concepts

La figure suivante représente les relations (propriétés de l'objet) entre les concepts

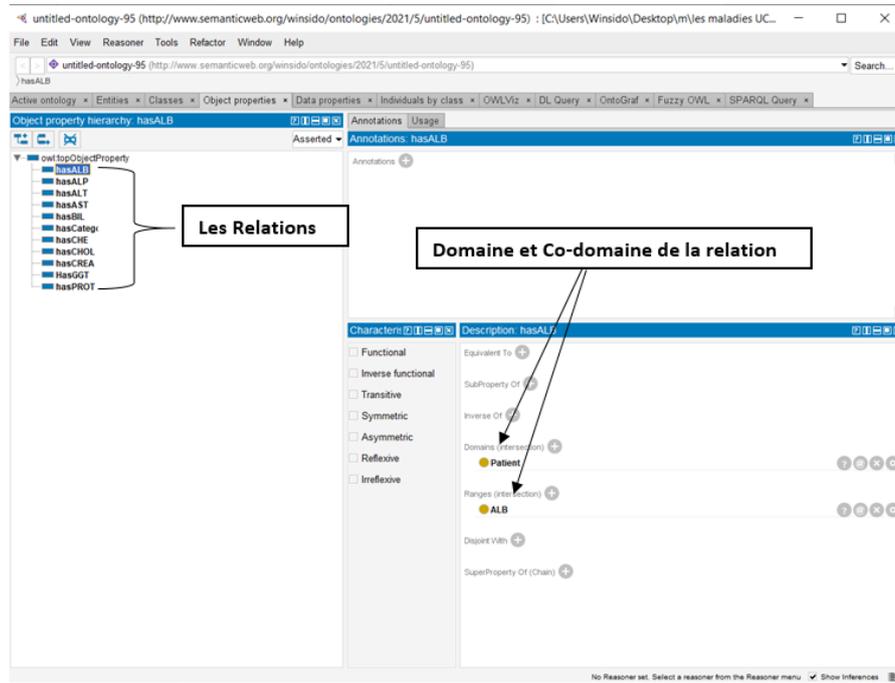


FIGURE 4.3 – Relations entre concepts

## — Instanciation des concepts

La figure suivante illustre un groupe d'instances (individus) pour différents concepts.

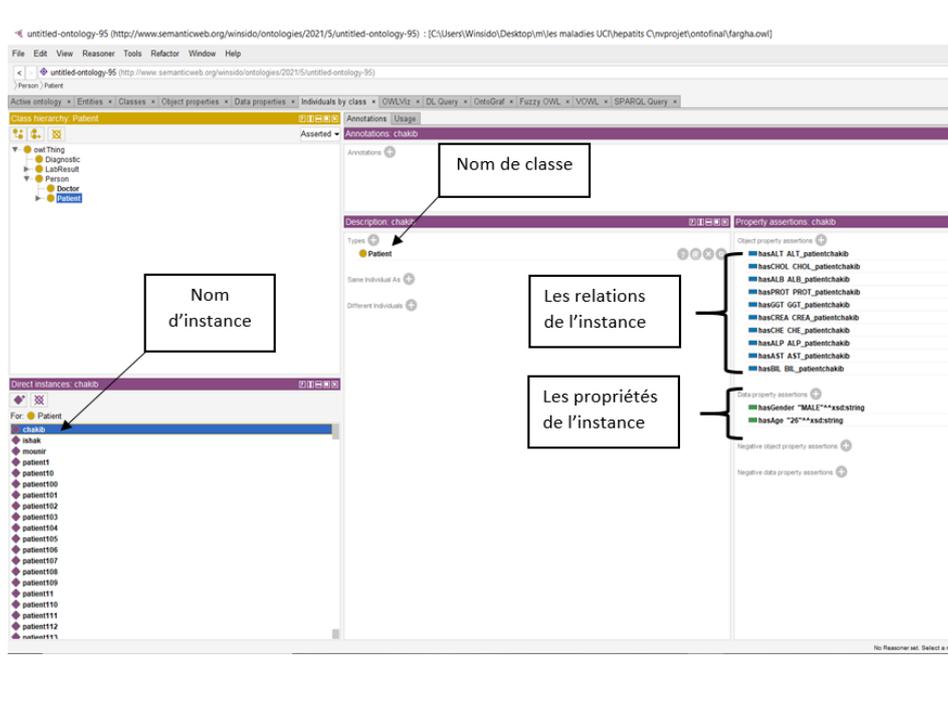


FIGURE 4.4 – Les instances de concept

## — Plugin Fuzzy OWL 2

Fuzzy OWL 2 est un plugin pour l'éditeur Protégé qui permet aux utilisateurs de modifier, enregistrer les ontologies Fuzzy OWL 2 et soumettre des requêtes au moteur d'inférence Fuzzy DL. Il est élaboré par Fernando Bobillo et Umberto Straccia en 2011 [26].

L'installation de plugin Fuzzy OWL 2 est très simple, seulement nous ne copions que le fichier "FuzzyOWL.jar" dans le fichier "plugins" de protégé et nous copions le fichier "dlib" dans le fichier protégé.

Après avoir exécuté le protégé, nous remarquons un nouvel onglet « fuzzy OWL » ce qui indique que le plugin Fuzzy OWL 2 est installé.

La figure 4.5 illustre le fonctionnement du plug-in en montrant comment créer un nouveau type de données flou. Nous spécifions le nom du type de données et le type de la fonction d'appartenance. Ensuite, le plug-in demande les paramètres nécessaires en fonction du type.

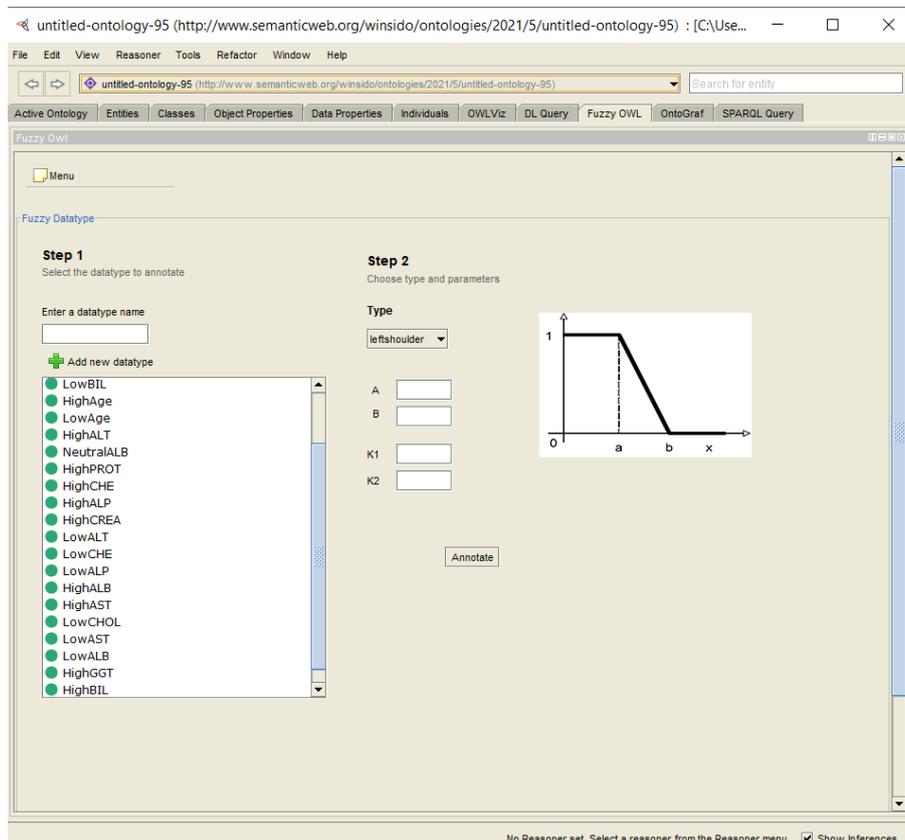


FIGURE 4.5 – Création d'un type de données flou avec le plug-in.

#### — **GeNIe**

GeNIe Modeler est une interface utilisateur graphique (GUI) pour SMILE Engine et permet la construction et l'apprentissage interactifs de modèles. Il est écrit pour l'environnement Windows mais peut également être utilisé sur macOS et Linux sous Wine. Il a été minutieusement testé sur le terrain depuis 1998. Il a reçu une large acceptation dans les universités et l'industrie, et compte des milliers d'utilisateurs dans le monde entier.

#### — **SMILE**

SMILE est un moteur de raisonnement et d'apprentissage/découverte causale pour les modèles graphiques, tels que les réseaux bayésiens, les diagrammes d'influence et les modèles d'équations structurelles. Techniquement, il s'agit d'une bibliothèque de classes C++ qui peuvent être intégrées dans un logiciel utilisateur existant via son API, améliorant les produits utilisateur avec des capacités de modélisation de décision. SMILE est entièrement portable et disponible pour la plupart des plates-formes informatiques - du centre de données à l'embarqué. Nous proposons des wrappers pour SMILE qui permettent de l'utiliser à partir de Java, Python, R, .NET et d'autres environnements de développement.

#### — **JAVA**

C'est un langage de programmation orienté objet, développé par Sun Microsystems. Il permet de créer des logiciels compatibles avec de nombreux systèmes d'exploitation (Windows, Linux, Macintosh, Solaris).

#### — **NetBeans**

NetBeans est un environnement de développement intégré (EDI), placé en open source par Sun en juin 2000 sous licence CDDL (Common Development and Distribution License) et GPLv2. En plus de Java, NetBeans permet la prise en charge native de divers langages tels le C, le C++, le JavaScript, le XML, le Groovy, le PHP et le HTML, ou d'autres (dont Python et Ruby) par l'ajout de greffons. Il offre toutes les facilités d'un IDE moderne (éditeur avec coloration syntaxique, projets multi-langage, refactoring, éditeur graphique d'interfaces et de pages Web).

## **4.3 Les interfaces de notre application**

Dans cette section nous allons présenter les interfaces de notre système.

### **4.3.1 Interface d'accueil**

Interface d'accueil est l'interface principale de notre système. Elle contient les boutons des fonctionnalités (voir **la figure 4.6**).



FIGURE 4.6 – L'interface d'accueil.

Une interface simple, lisible et qui propose les fonctionnalités principales de notre système d'une manière générale. Les boutons affichés ont pour mission d'accéder aux différentes fonctions du système (voir la figure 4.6).

**Bouton "Chercher"** : permet de rechercher un patient en introduisant son nom dans le champ du texte à gauche.

**Bouton "Afficher"** : permet d'afficher la liste des patients sauvegardés.

**Bouton "Ajouter"** : permet d'ouvrir une interface d'ajout pour permettre à l'utilisateur d'introduire un nouveau patient avec tous ses informations nécessaires et de le stocker dans l'ontologie floue.

**Bouton "Modifier"** : permet d'ouvrir une interface de modification pour permettre à l'utilisateur de modifier les informations d'un patient précis.

**Bouton "Supprimer"** : permet d'afficher un message de confirmation si l'utilisateur veut supprimer le patient, après la confirmation de la suppression, nous allons supprimer le patient et ses informations de l'ontologie floue.

**Bouton "Inference"** : permet d'ouvrir une interface d'inférence pour permettre à l'utilisateur de prédire et afficher les probabilités qu'un patient appartient.

### 4.3.2 Interface : Ajouter patient

Cette interface s'affiche après avoir cliquer sur le bouton ajouter de l'interface d'accueil (voir la figure 4.6). L'interface d'ajouter patient permet à l'utilisateur d'ajouter un patient avec son nom et les données des résultats du laboratoire de manière très simple et facile. Afin de permettre au système d'utiliser ces valeurs pour la prédiction, l'utilisateur confirme l'insertion des données

dans l'ontologie floue en cliquant sur le bouton "Ajouter", au final cliquer sur "Accueil" pour revenir à l'accueil (voir la figure 4.7).

**Centre Medical LA SAUVAGÈRE**

**Ajouter Patient**

Nom :  ALT :  CREA :

Age :  AST :  GGT :

Sexe :  BIL :  PROT :

ALB :  CHE :

ALB :  CHOL :

FIGURE 4.7 – Ajouter Patient.

### 4.3.3 Interface : Modifier patient

Cette interface s'affiche après avoir cliquer sur le bouton Modifier de l'interface d'accueil (voir la figure 4.6, L'interface de modification des informations sur un patient permet à l'utilisateur de mettre a jour les information sur un patient de manière très simple et facile. L'utilisateur confirme la modification des données dans l'ontologie floue en cliquant sur le bouton "Modifier", au final cliquer sur "Accueil" pour revenir à l'accueil (voir la figure 4.8).

**Centre Medical LA SAUVAGÈRE**

**Modifier Information**

Nom :  ALT :  CREA :

Age :  AST :  GGT :

Sexe :  BIL :  PROT :

ALB :  CHE :

ALB :  CHOL :

FIGURE 4.8 – Modifier Patient.

### 4.3.4 Interface Inférence patient

- Interface inférence patient : Après avoir ajouté le patient, l'utilisateur se redirige vers l'accueil pour lancer la prédiction, il choisit le patient voulu à partir de la liste et clique sur le bouton "inférence" (voir la figure 4.6. L'interface de prédiction s'affichera, l'utilisateur peut Vérifier les champs et valider la prédiction en cliquant sur le bouton "Prédire" (voir la figure 4.9). Cette instruction permet d'afficher une interface des résultats de la prédiction (voir la figure 4.10).

**Centre Medical LA SAUVAGÈRE**

**Diagnostiquer Patient**

Nom :

Age :

Sexe :

ALT :

AST :

BIL :

CREA :

GGT :

PROT :

ALB :

CHE :

ALB :

CHOL :

FIGURE 4.9 – Diagnostiquer Patient.

**Centre Medical LA SAUVAGÈRE**

**Résultat**

vous avez Blood\_Donor avec : 52.98255173433891  
vous avez Hepatits avec : 10.619844769270351  
vous avez Fibrosis avec : 10.321827128001019  
vous avez Cirrhosis avec : 15.753949240388698  
vous avez syspect\_blood\_donor avec : 10.3218271280  
Vous avez Blood Donor

FIGURE 4.10 – Résultat de la prédiction .

- Interface Résultat : Cette interface permet d'afficher le résultat de la prédiction sous forme de texte qui s'affiche à gauche, la catégorie de classification plus le pourcentage d'appartenance plus le résultat final, pour l'autre type d'affichage de données l'utilisateur doit cliquer sur le bouton "Vision graphique" pour avoir le résultat sous format graphique (voir la figure 4.10).
- Représentation graphique de résultat : Cette interface affiche le résultat de la prédiction sous forme graphique (classe plus pourcentage d'appartenance) (voir la figure 4.11).

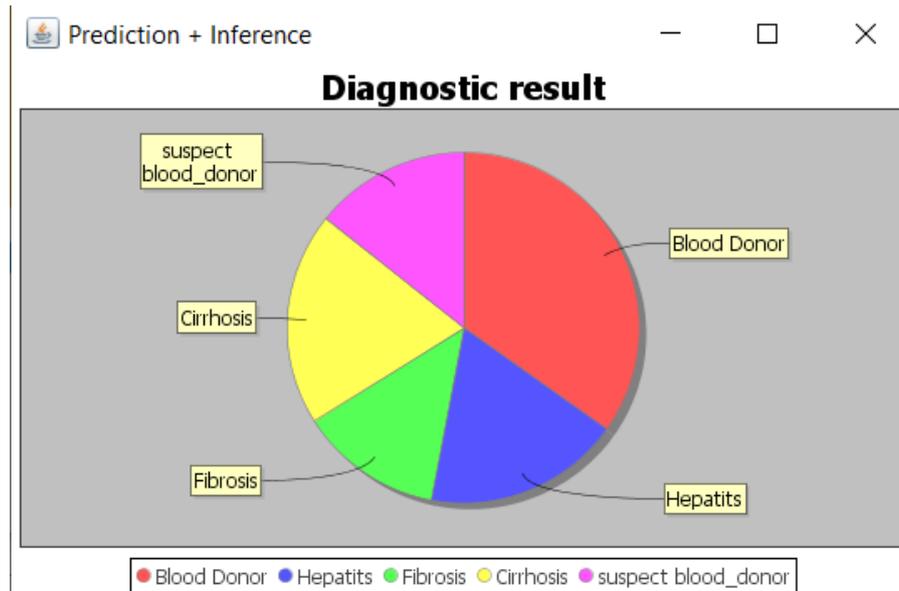


FIGURE 4.11 – Résultat graphique de la prédiction .

## 4.4 Test et validation

Afin de tester les performances de notre système, une étude quantitative a été menée. Nous présentons dans le reste de cette section les résultats obtenus.

### 4.4.1 Mesures d'évaluation

Une matrice de confusion ou tableau de contingence sert à évaluer la qualité d'une classification. Elle est obtenue en comparant les données classées avec des données de référence qui doivent être différentes de celles ayant servi à réaliser la classification.

On classe les résultats en 4 catégories :

- **True Positive (TP)** : la prédiction et la valeur réelle sont positives.
- **True Negative (TN)** : la prédiction et la valeur réelle sont négatives.
- **False Positive (FP)** : la prédiction est positive alors que la valeur réelle est négative.
- **False Negative (FN)** : la prédiction est négative alors que la valeur réelle est positive.

		Valeur prédite				
Valeur réelle		Blood Donor	Hepatitis	Fibrosis	Cirrhosis	suspect blood donor
	Blood Donor	a	b	c	d	e
	Hepatitis	f	g	h	i	j
	Fibrosis	k	l	m	n	o
	Cirrhosis	p	q	r	s	t
	suspect blood donor	u	v	w	x	y

TABLE 4.1 – Matrice de confusion

	Blood Donor	Hepatitis	Fibrosis	Cirrhosis	suspect blood donor
TP	a	g	m	s	y
FP	b+c+d+e	f+h+i+j	k+l+n+o	p+q+r+t	u+v+w+x
TN	g+h+i+j+ l+m+n+o+ q+r+s+t+ v+w+x+y	a+c+d+e+ k+m+n+o+ p+r+s+t+ u+w+x+y	a+b+d+e+ f+g+h+j+ p+q+s+t+ u+v+x+y	a+b+c+e+ f+g+h+j+ k+l+m+o+ u+v+w+y	a+b+c+d+ f+g+h+i+ k+l+m+n+ p+q+r+s
FN	f+k+p+u	b+l+q+v	c+h+r+w	d+i+n+x	e+j+o+t

TABLE 4.2 – Illustration de (TP, FP, TN, FN) a pour chaque classe [9]

On peut ainsi calculer toutes les métriques nécessaires à l'analyse de ce tableau.

On retrouve ici les manières les plus communes de tirer des informations intéressantes de ce genre de tableau, on appelle cet indicateur :

- **Accuracy**, un des critères permettant d'évaluer les modèles de classification. De façon non formelle, elle désigne la proportion des prédictions correctes effectuées par le modèle. Formellement, la justesse est définie ainsi :

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN}$$

#### 4.4.2 Résultats obtenus

Dans notre expérimentation, nous avons opté pour la technique de validation croisée, qui permet d'utiliser l'ensemble de données complet pour l'entraînement et la validation **simultanément**. En effet, nous divisons le jeu de données en k parties (folds). Pour chaque itération k, nous sélectionnons une partie k et nous l'utilisons comme partie test. Le reste (l'union des autres parties k-1) est utilisé pour l'entraînement.

Nous avons mené des séries de tests, pour chaque fois nous appliquons l'inférence bayésienne floue (l'inférence de notre système) et l'inférence en se basant sur d'autres classificateurs de machine learning (LogisticRegression, SupportVector Classifier, DecisionTree, RandomForest, GaussianNaive Bayes) basées sur les informations de chaque patient enregistrées dans la partie k de l'ensemble de données. Après cela, nous comparons les correspondances obtenues par chaque et celles par une classification manuelle (donnée par les médecins).

Après cela, nous calculons la précision pour évaluer la qualité de la classification résultante de chaque type d'inférence.

K	Notre système	Classifieurs de Machine Learning				
		Logistic Regression	Support Vector Classifier	Decision Tree	Random Forest	Gaussian Naive Bayes
2	95,4%	54,41%	45,91%	65,68%	74,18%	68,30%
3	95,15%	62,62%	37,37%	85,85%	79,79%	67,92%
4	94,93%	59,37%	45,48%	89,31%	81,94%	76,04%
5	94,51%	65,71%	42,85%	77,14%	86,57%	77,14%
6	94,31%	60%	46,11%	85,55%	82,77%	85%
7	94,14%	62,85%	42,85%	88,57%	85,71%	88,57%

TABLE 4.3 – Comparaison des 'accuracy' (%) obtenu par notre système et les classifieurs de Machine Learning

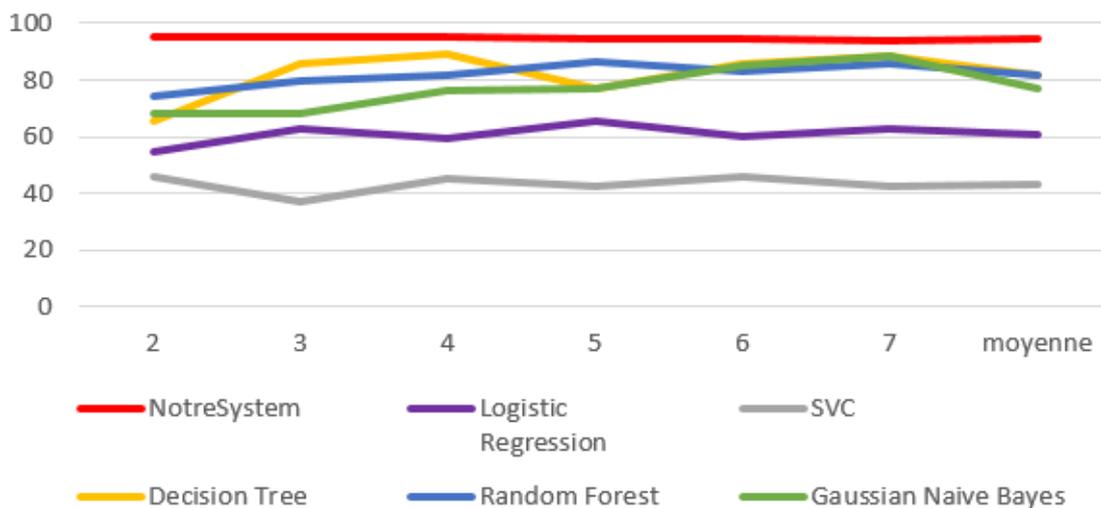


FIGURE 4.12 – Résultats obtenus par chaque classifieur

Nous avons obtenu pour notre système une moyenne d'accuracy qu'égal à 94.74 %. En effet, la moyenne obtenue pour notre système est supérieure à celle des autres algorithmes de classification de machine learning.

Pour conclure, un bon système c'est un système qui a une accuracy plus proche de 100 % est c'est le cas pour notre système( ce qui signifie que la prédiction basée sur notre système est proche de la prédiction du médecin ).

## 4.5 Conclusion

Dans ce chapitre, nous avons présenté en détail les étapes dans lesquelles nous avons créé notre ontologie et expliqué les outils utilisés dans ce travail. Finalement, nous avons confirmé la validité de notre système.



---

# Conclusion générale

---

L'objet de notre travail étant la proposition et l'implémentation d'un système d'aide à la décision médicale à base de connaissances incertaines et de l'utilisation des ontologies floues et les réseaux bayésiens flous.

Dans notre travail, nous avons montré comment peut-on définir et créer un système d'aide au diagnostic médical. Ce système est fondé sur un modèle de réseau bayésien flou et une ontologie floue.

Nous avons proposé une représentation sémantique des connaissances à l'aide de l'ontologie floue. Nous avons suivi un processus de développement d'ontologie complet dans la mesure où partant des données brutes représentées comme un réservoir de connaissances, il permet d'arriver à une ontologie opérationnelle. Pour ce faire, la méthodologie « FODM » a été employée afin d'explicitier et de guider la construction de l'ontologie floue.

À l'aide du l'expert du domaine, nous avons modélisé et construit la structure du RBF en énumérant les facteurs de risques ainsi que les liens de causalité entre eux. L'apprentissage des paramètres a été fait à partir d'une base d'apprentissage (dataset utilisé) qui ne contient plus de données manquantes en utilise l'approche maximum de vraisemblance (MV).

Le système que nous avons mis en place permet la prédiction si le patient est atteint par la maladie de l'hépatite C ou non, tout en sachant ses informations biologiques et ses données pour faire l'inférence.

Pour compléter ce projet et dans le souci d'élargir et d'améliorer notre travail, nous proposons l'amélioration de ce système d'aide au diagnostic médical fondé sur un modèle de réseau bayésien flou et une ontologie floue par :

- Expérimenter ce système dans des situations réelles pour ajuster les paramètres de l'incertitude.
- Expérimentation du système dans le cadre d'un suivi à domicile pour les patients atteints de maladies chroniques.
- Enrichissement de l'ontologie par d'autres maladies en ajoutant les relations les facteurs d'incertitude appropriés.
- Ajouter d'autres fonctionnalités jugées nécessaires dans ce domaine.



---

# Bibliographie

---

- [1] Anne Nicolle, Pierre Beust, and Vincent Perlerin. Un analogue de la mémoire pour un agent logiciel interactif. *Cognito*, 21 :37–66, 2002.
- [2] Audrey Baneyx. *Construire une ontologie de la Pneumologie Aspects théoriques, modèles et expérimentations*. PhD thesis, Université Pierre et Marie Curie-Paris VI, 2007.
- [3] Kenneth J Laskey and Kathryn B Laskey. Uncertainty reasoning for the world wide web : Report on the urw3-xg incubator group. In *URSW*. Citeseer, 2008.
- [4] Mme ZEGHIB Nadia. *Logique floue et SMA : Aide à la décision floue dans les systèmes multi-agents*. PhD thesis, Université Mentouri de Constantine, 2009.
- [5] Lotfi Baghli. *Contribution à la commande de la machine asynchrone, utilisation de la logique floue, des réseaux de neurones et des algorithmes génétiques*. PhD thesis, Université Henri Poincaré-Nancy I, 1999.
- [6] Lotfi A Zadeh. Information and control. *Fuzzy sets*, 8(3) :338–353, 1965.
- [7] Hanène Ghorbel, Afef Bahri, and Rafik Bouaziz. Fuzzy protégé for fuzzy ontology models. *Age*, 12(18) :30, 2009.
- [8] Xin Li, José-Fernán Martínez, and Gregorio Rubio. A new fuzzy ontology development methodology (fodm) proposal. *IEEE Access*, 4 :7111–7124, 2016.
- [9] Everything you should know about confusion matrix for machine learning. <https://www.analyticsvidhya.com/blog/2020/04/confusion-matrix-machine-learning/>. Accessed : 2021-06-25.
- [10] Hépatite c. <https://www.who.int/fr/news-room/fact-sheets/detail/hepatitis-c>. Accessed : 2021-05-11.
- [11] Hépatite c : 320.000 personnes en sont atteintes en algérie. <http://www.santemaghreb.com/actus.asp?id=1186>. Accessed : 2021-05-20.

- [12] ISHAK RIALI. *Modélisation de la connaissance incertaine dans le web sémantique*. PhD thesis, Université Blida1-Saad Dahlab, 2020.
- [13] Lazhari MOUAFFAK. *Développement d'une Ontologie Floue*. PhD thesis, Faculté des Mathématiques et de l'Informatique-UNIVERSITE MOHAMED BOUDIAF DE . . . , 2013.
- [14] Thomas R Gruber. A translation approach to portable ontology specifications. *Knowledge acquisition*, 5(2) :199–220, 1993.
- [15] Nicola Guarino. Understanding, building and using ontologies. *International journal of human-computer studies*, 46(2-3) :293–310, 1997.
- [16] WN Borst. Construction of engineering ontologies, centre for telematica and information technology, 1997.
- [17] Ontologie. <http://websemantique.org/Ontologie>. Accessed : 2021-01-20.
- [18] HACINE GHERBI Ahcine. Construction d'une ontologie pour le web sémantique. *Mémoire Présenté pour l'obtention du Diplôme de Magister*, 2014.
- [19] Rose Dieng, Olivier Corby, Fabien Gandon, Alain Giboin, Joanna Golebiowska, Nada Matta, and Myriam Ribière. *Méthodes et outils pour la gestion des connaissances : une approche pluridisciplinaire du knowledge management*. Dunod, 2001.
- [20] Souheila Khalfi. Construction d'une ontologie pour la prise en charge des patients à domicile. 2009.
- [21] Bill Swartout, Ramesh Patil, Kevin Knight, and Tom Russ. Toward distributed use of large-scale ontologies. In *Proc. of the Tenth Workshop on Knowledge Acquisition for Knowledge-Based Systems*, pages 138–148, 1996.
- [22] Mariano Fernández-López, Asunción Gómez-Pérez, and Natalia Juristo. Methontology : from ontological art towards ontological engineering. 1997.
- [23] J Charlet, B Bachimont, and R Troncy. Le web sémantique, vol. 4 of hors série de la revue information-interaction-intelligence. *Cépaduès*, 4 :7–20, 2005.
- [24] Sami Zghal, Sadok Ben Yahia, Engelbert Mephu Nguifo, and Yahya Slimani. Soda : Une approche structurelle pour l'alignement d'ontologies owl-dl. *ZFO*, 2007.
- [25] Isabelle Bloch and Henri Maitre. Les méthodes de raisonnement dans les images. *Ecole Nationale Supérieure des Télécommunications-CNRS UMR, 5141*, 2004.
- [26] Fernando Bobillo and Umberto Straccia. Fuzzy ontology representation using owl 2. *International journal of approximate reasoning*, 52(7) :1073–1094, 2011.

- [27] Ali Ben Mrad, Mohamed Amine Maalej, Veronique Delcroix, Sylvain Piechowiak, and Mohamed Abid. Intégration de l'evidence floue dans les réseaux bayésiens.
- [28] Judea Pearl. *Probabilistic reasoning in intelligent systems : networks of plausible inference*. Elsevier, 2014.
- [29] Jean-Jacques Boreux, Eric Parent, Jacques Bernier, and Jacques Bernier. *Pratique du calcul bayésien*, volume 118. Springer, 2010.
- [30] Patrick Naïm, Pierre-Henri Wuillemin, Philippe Leray, Olivier Pourret, and Anna Becker. Réseaux bayésiens. *Eyrolles, Paris*, 3 :120, 1999.
- [31] Patrick Naïm, Pierre Henri Wuillemin, and Philippe Leray. Olivier pourret. *Anna becker, Réseaux bayésiens, Eyrolles, 3ème édition, Paris*, 2008.
- [32] Alain Delaplace. *Approche évolutionnaire de l'apprentissage de structure pour les réseaux bayésiens*. PhD thesis, Tours, 2007.
- [33] Olivier François. *De l'identification de structure de réseaux bayésiens à la reconnaissance de formes à partir d'informations complètes ou incomplètes*. PhD thesis, INSA de Rouen, 2006.
- [34] Annie Millet. Méthode de monte carlo. *Université Paris VI et Paris VII, Polycopié*, 2003.
- [35] Ioannis Tsamardinos, Laura E Brown, and Constantin F Aliferis. The max-min hill-climbing bayesian network structure learning algorithm. *Machine learning*, 65(1) :31–78, 2006.
- [36] Wai Lam and Fahiem Bacchus. Learning bayesian belief networks : An approach based on the mdl principle. *Computational intelligence*, 10(3) :269–293, 1994.
- [37] Hao Tang and Shi Liu. Basic theory of fuzzy bayesian networks and its application in machinery fault diagnosis. In *Fourth international conference on fuzzy systems and knowledge discovery (FSKD 2007)*, volume 4, pages 132–137. IEEE, 2007.
- [38] Christopher Fogelberg, Vasile Palade, and Phil Assheton. Belief propagation in fuzzy bayesian networks. In *1st International Workshop on Combinations of Intelligent Methods and Applications (CIMA) at ECAI'08*, pages 19–24, 2008.
- [39] Nicholas Ryhajlo, Liessman Sturlaugson, and John W Sheppard. Diagnostic bayesian networks with fuzzy evidence. In *2013 IEEE AUTOTESTCON*, pages 1–8. IEEE, 2013.
- [40] Vasilios Zarikas, Elpiniki Papageorgiou, Damira Pernebayeva, and Nurislam Tursynbek. Medical decision support tool from a fuzzy-rules driven bayesian network. In *ICAART (2)*, pages 539–549, 2018.

- [41] Firas Zekri, Emna Turki, and Rafik Bouaziz. Alzfuzzyonto : Une ontologie floue pour l'aide à la décision dans le domaine de la maladie d'alzheimer. In *INFORSID*, pages 83–98, 2015.
- [42] Asma Khan, Azeema Sadia, Sohaib Ahmed, Huma Tabassum, and M Shahid Khan. Hepo : The hepatitis ontology for abductive medical diagnostic systems. In *2017 International Conference on Communication, Computing and Digital Systems (C-CODE)*, pages 271–275. IEEE, 2017.
- [43] Daniel-Ioan Curiac, Gabriel Vasile, Ovidiu Baniias, Constantin Volosencu, and Adriana Albu. Bayesian network model for diagnosis of psychiatric diseases. In *Proceedings of the ITI 2009 31st International Conference on Information Technology Interfaces*, pages 61–66. IEEE, 2009.
- [44] Elif Varol Altay and Bilal Alatas. A novel clinical decision support system for liver fibrosis using evolutionary multi-objective method based numerical association analysis. *Medical Hypotheses*, 144 :110028, 2020.
- [45] Aravind B Akella and Sudheer Akella. Applying machine learning to evaluate for fibrosis in chronic hepatitis c. *medRxiv*, 2020.
- [46] World health organization. guidelines on hepatitis b and c testing. <https://www.who.int/hepatitis/publications/guidelineshepatitis-c-b-testing/en/>. Accessed : 2021-05-15.
- [47] Dheeru Dua and Casey Graff. UCI machine learning repository, 2017.
- [48] L'analyse de l'albumine. <https://www.passeportsante.net/fr/Maux/analyses-medicales/Fiche.aspx?doc=albumine-analyse-sang>. Accessed : 2021-06-15.
- [49] Pal,phosphatase alcaline. <https://www.biron.com/fr/glossaire/pal/>. Accessed : 2021-06-15.
- [50] Alt. <https://www.biron.com/fr/glossaire/alt/>. Accessed : 2021-06-15.
- [51] Ast. <https://www.biron.com/fr/glossaire/ast/>. Accessed : 2021-06-15.
- [52] Bilirubine non conjuguée. <https://www.biron.com/fr/glossaire/bilirubine-non-conjuguee/>. Accessed : 2021-06-15.
- [53] cholinesterases. [https://www.chu.ulg.ac.be/jcms/c\\_498493/fr/cholinesterases-sang](https://www.chu.ulg.ac.be/jcms/c_498493/fr/cholinesterases-sang). Accessed : 2021-06-15.

- [54] Le bon et le mauvais cholestérol, et les triglycérides. <https://www.biron.com/fr/centre-du-savoir/petit-guide-biron/cholesterol/>. Accessed : 2021-06-15.
- [55] créatinine. [https://www.chu.ulg.ac.be/jcms/c\\_498839/fr/creatinine-sang](https://www.chu.ulg.ac.be/jcms/c_498839/fr/creatinine-sang). Accessed : 2021-06-15.
- [56] Ggt. <https://www.biron.com/fr/glossaire/ggt/>. Accessed : 2021-06-15.
- [57] Protéines totales. <https://www.biron.com/fr/glossaire/protéines-totales/>. Accessed : 2021-06-15.
- [58] datil. <http://webdiis.unizar.es/~ihvdis/Datil.html>. Accessed : 2021-05-20.
- [59] Weiling Cai, Songcan Chen, and Daoqiang Zhang. Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation. *Pattern recognition*, 40(3) :825–838, 2007.

