

UNIVERSITÉ SAAD DAHLEB DE BLIDA1

Faculté des sciences

Département d'informatique



MEMOIRE DE MASTER

En Informatique

Option : Ingénierie Des Logiciels

THÈME :

**Prédiction du diagnostic médical
guidée par les ontologies
probabilistes**

Réalisé par
GUEMMOUZ MARWA
KHERBACHE HAFSA

Encadré par
Dr. M. FAREH
Dr. I. RIALI

soutenu le :
Septembre 2021

Mme Zahra. Présidente
Mme Lahiani. Examineur

Remerciements

Ce travail est le fruit de la combinaison d'efforts , de patience et du courage durant toutes les années de notre parcours.

Nous remercions tout d'abord ALLAH le tout puissant qui, par sa grâce nous a permis d'arriver au bout de nos efforts en nous donnant la santé, la force, la volonté et en nous faisant entourer des merveilleuses personnes dont nous tenons à remercier.

Nous tenons à remercier toutes les personnes qui ont contribué a la réalisation de notre travail et ont aidé lors de la rédaction de ce mémoire.

Au premier temps nous remercions notre promoteur de mémoire **Mme M. Fareh**, pour sa patience, sa disponibilité et surtout ses judicieux conseils.Nous tenons à remercier aussi **Mr. I. Riali** pour l'orientation et l'aide qui ont constitué un apport considérable dans ce travail.

Nous avons eu la chance d'intégrer votre équipe et d'apprécier vos qualités, vos valeurs, votre sérieux et votre compétence.

Nous tenons à remercier également **Dr. Ousalem** chirurgien résident au département des maladies cardio-vasculaires service COVID-19, Hopital Frantz Fanon Blida.

Nous voudrions exprimer notre reconnaissance envers nos parents, nos sœurs, nos frères et tous nos proches et amis, qui nous ont apporté leur soutien moral et intellectuel tout au long de notre travail.

GUEMMOUZ MARWA

KHERBACHE HAFSA

Résumé

Dans le domaine médical, les médecins sont régulièrement amenés à prendre plusieurs décisions, certaines de ces décisions sont prises facilement, lorsque le diagnostic est facile, et le traitement choisi est efficace, alors que dans d'autres cas, la bonne décision à prendre n'est pas évidente, et les informations nécessaires pour son diagnostic sont remplies d'incertitude. Les ontologies probabilistes sont utilisées pour représenter les connaissances sur un domaine et l'incertitude associée à ces connaissances d'une manière structurée et raisonnée.

L'objectif de notre projet est de prédire le diagnostic médical, en utilisant les ontologies probabilistes basées sur la théorie de probabilité, et les moyens associés à cette théorie pour la modélisation et la représentation de la connaissance incertaine de ce domaine. Nous utilisons la méthodologie UMP-ST basée sur un réseau bayésien multi entités, la construction de ce dernier se réalise par la définition de l'ensemble de ses composants MFragments et ses différents types de noeuds, ainsi que les distributions de probabilité liées aux noeuds. Une fois l'ontologie probabiliste est construite, nous l'utilisons pour la prédiction du diagnostic.

Nous appliquons la méthodologie UMP-ST sur l'étude de cas du diagnostic du COVID-19, afin de réaliser un système qui va aider le médecin à prendre la bonne décision.

Mots clés :

Connaissance incertaine, réseau Bayésien, MEBN, ontologies Probabiliste, inférence.

Abstract

In the medical field, doctors are regularly called upon to make several decisions, some of these decisions are taken easily, when the diagnosis is easy, and the treatment chosen is effective, while in other cases, the right decision to be taken is not obvious, and the information needed for its diagnosis is fraught with uncertainty. Probabilistic ontologies are used to represent knowledge about a domain and the uncertainty associated with that knowledge in a structured and reasoned way.

The objective of our project is to predict medical diagnosis, using probabilistic ontologies based on probability theory, and the means associated with this theory for the modeling and representation of uncertain knowledge in this field. We use the UMP-ST methodology based on a multi-entity Bayesian network, the construction of the latter is carried out by defining all of its MFragments components and its different types of nodes, as well than the probability distributions linked to the nodes. Once the probabilistic ontology is constructed, we use it for diagnostic prediction.

We apply the UMP-ST methodology on the case study of the diagnosis of COVID 19, in order to realize a system that will help the doctors to make the right decision.

Key words :

Uncertain knowledge, Bayesian network, MEBN, Probabilistic ontologies, inference.

ملخص:

في المجال الطبي ، يتم استدعاء الأطباء بانتظام لاتخاذ عدة قرارات ، يتم اتخاذ بعض هذه القرارات بسهولة ، عندما يكون التشخيص سهلاً ، ويكون العلاج المختار فعالاً ، بينما في حالات أخرى ، يكون القرار الصحيح الذي يجب اتخاذه ليس واضحاً ، والمعلومات اللازمة لتشخيصه مشحونة بعدم اليقين. تُستخدم الأنطولوجيا الاحتمالية لتمثيل المعرفة حول المجال وعدم اليقين المرتبط بهذه المعرفة بطريقة منظمة ومنطقية.

الهدف من مشروعنا هو التنبؤ بالتشخيص الطبي ، باستخدام علم الوجود الاحتمالي على أساس نظرية الاحتمالات ، والوسائل المرتبطة بهذه النظرية لنمذجة وتمثيل المعرفة غير المؤكدة في هذا المجال. نحن نستخدم منهجية UMP-ST على أساس شبكة Bayesian متعددة الكيانات ، ويتم إنشاء هذه الأخيرة من خلال تحديد جميع مكونات MFragments وأنواعها المختلفة من العقد ، وكذلك من التوزيعات الاحتمالية المرتبطة بالعقد. بمجرد بناء علم الوجود الاحتمالي ، نستخدمه للتنبؤ التشخيصي.

نطبق منهجية UMP-ST على دراسة الحالة لتشخيص COVID 19 ، من أجل تحقيق نظام يساعد الطبيب على اتخاذ القرار الصحيح.

الكلمات المفتاحية:

معرفة غير مؤكدة ، شبكة Bayesian ، MEBN ، الأنطولوجيا الاحتمالية ، الاستدلال.

Table des matières

Table des figures	8
Liste des tableaux	10
Introduction générale	11
1 ONTOLOGIES CLASSIQUES ET ONTOLOGIES PROBABILISTES	15
1.1 Introduction	15
1.2 Les ontologies en ingénierie de connaissance	15
1.3 Définition des ontologies	16
1.4 Les composants d'une ontologie	16
1.5 Rôle des ontologies	16
1.6 Langages des ontologies	17
1.6.1 Resource Description Framework (RDF) et (RDFS)	17
1.6.2 OWL	17
1.7 Connaissance incertaine	18
1.7.1 Classes d'incertitude d'une ontologie	18
1.7.2 Approches de l'incertitude recommandées par le W3C	19
1.8 Réseau bayésien (RB)	20
1.8.1 Apprentissage des paramètres	20
1.8.2 Apprentissage de la structure	21
1.9 Ontologies probabilistes	21
1.9.1 Définition	21
1.9.2 Langage des ontologies probabilistes	22
1.9.2.1 MEBN	22
1.9.2.2 PR-OWL	23
1.10 Méthodologie UMP-ST	24
1.11 Entre ontologie classique et probabiliste	25
1.12 Conclusion	25

2	ÉTUDE DES TRAVAUX CONNEXES EXISTANTS	27
2.1	Introduction	27
2.2	Diagnostic médical	27
2.2.1	Démarche diagnostique	28
2.2.2	Démarche décisionnelle	28
2.2.3	Incertitude médicale	28
2.2.4	Signe et Symptôme	29
2.2.4.1	Signe	29
2.2.4.2	Symptôme	30
2.3	Travaux existants	30
2.3.1	Travaux basés sur une ontologie médicale	31
2.3.1.1	La maladie d'autisme	31
2.3.1.2	Les maladies infectieuses	32
2.3.1.3	Les maladies infectieuses bactériennes	33
2.3.1.4	La maladie cancer du foie	34
2.3.2	Travaux basés sur les réseaux bayésiens	34
2.3.2.1	La maladie d'Alzheimer	34
2.3.2.2	La maladie de Cancer du sein	35
2.3.2.3	Les maladies cardiovasculaires	36
2.3.2.4	Les maladies des poumons	36
2.4	Comparaison des travaux	37
2.5	Analyse des travaux	38
2.6	Conclusion	38
3	CONCEPTION DU SYSTÈME	39
3.1	Introduction	39
3.2	Motivation de l'étude de cas	39
3.3	COVID-19	40
3.4	Mode de transmission	40
3.5	Symptômes	41
3.6	Facteurs de risque	43
3.7	Data set	43
3.8	Schéma global	45
3.9	Création de l'ontologie probabiliste	46
3.9.1	Les Exigences (requirement)	46
3.9.1.1	Les besoins(goals)	46
3.9.2	Analyse et conception	47
3.9.2.1	Les entités	47
3.9.2.2	Les règles	49

3.9.2.3	Les groupes	52
3.9.3	Implémentation	53
3.9.3.1	Mapping	53
3.9.3.2	Probabilités	58
3.9.4	Test	60
3.9.4.1	Évaluation	60
3.10	Résultat final de l'ontologie probabiliste	63
3.11	Conclusion	67
4	IMPLEMENTATION ET TEST DU SYSTÈME	69
4.1	Introduction	69
4.2	Outils de développement de notre système	69
4.3	La mise en oeuvre du système	71
4.4	Les interfaces	71
4.4.1	Interface d'accueil	71
4.4.2	Add patient	73
4.4.3	Delete patient	74
4.4.4	Consultation	75
4.4.5	Liste des patients	76
4.4.6	Prédiction	77
4.4.7	Détail de la prédiction	78
4.5	La validation du système	78
4.6	Conclusion	81
	Bibliographie	85

Table des figures

1.1	Représentation des classes d'incertitude d'une ontologie (37)	18
1.2	PR-OWL MTheory Concepts [8]	23
1.3	Flux POMC [1]	24
2.1	Travaux existants	31
3.1	Interprétation du résultats IgM/IgG (35)	42
3.2	Schéma global	45
3.3	La structure du réseau bayésien	51
3.4	Réseau bayésien	51
3.5	Diagramme de Classe	52
3.6	MFrag de notre domaine	53
3.7	Variable ordinaire P	54
3.8	Noeud résident "blood sugar"	54
3.9	Noeud de contexte "IsA()"	54
3.10	Noeud d'entrée "HGB"	55
3.11	Le MFrag Patient	55
3.12	Le MFrag Risk-factors	56
3.13	Le MFrag Symptom	56
3.14	Le MFrag Cardio	57
3.15	Le MFrag FNS	57
3.16	Le MFrag des serology.	58
3.17	Le MFrag Diagnostic	58
3.18	Le LPD pour le noeud ISTOUCHEDBY(patient)	59
3.19	Le SSBN pour le noeud IsTouchedBy (patient)	60
3.20	SSBN du test 1	62
3.21	SSBN du test 2	63
3.22	La hiérarchie des classes de l'ontologie	64
3.23	La hiérarchie des propriétés d'objet de l'ontologie	65

3.24	La hiérarchie des propriétés d'objet de l'ontologie	66
3.25	La hiérarchie des propriétés des données de l'ontologie classique	66
3.26	La hiérarchie des propriétés des données de l'ontologie probabiliste	67
4.1	Interface d'accueil	72
4.2	Interface d'ajout	73
4.3	Interface de suppression	74
4.4	Interface de consultation	75
4.5	Interface liste des patients	76
4.6	Interface de prédiction	77
4.7	interface Détail de prédiction	78
4.8	La matrice de confusion	79
4.9	Histogramme : precision, recall, F-mesure, accuracy	81

Liste des tableaux

1.1	Comparaison entre les deux langages [18]	25
2.1	Tableau comparatif	38
3.1	Entités et leurs relations	48
3.2	Les attributs de l'entité Risk factors	48
3.3	Les attributs de l'entité Symptom	48
3.4	Les attributs de l'entité Serology	49
3.5	Les attributs de l'entité Cardio	49
3.6	Les attributs de l'entité FNS	49
3.7	Les relations entre entités	49
3.8	Catégorisation des attributs.	50
4.1	Matrice de confusion	80

Introduction générale

Contexte

Le milieu professionnel de la pratique médicale se caractérise par une grande complexité, de nombreuses zones d'ombre, et l'incertitude comme composante incontournable de tous les niveaux de la prise en charge d'un malade.

Les données sont souvent imprécises, ambiguës et incomplètes et les signes sont similaires d'une maladie à une autre, et différents d'un malade à l'autre, les résultats des examens ont une fiabilité imparfaite et leur utilisation est incertaine, ce qui emmène à des traitements pas totalement surs.

Toutes ces conditions mettent le médecin en complexe incertitude au moment de la prise de décision, donc le diagnostic est particulièrement douteux, c'est pour cela que le domaine médical donne une grande importance au facteur de l'incertitude, car les connaissances sont généralement incomplètes ou d'une nature incertaine, alors difficiles à l'exploration, cette imprécision emmène à des erreurs cliniques liées aux variations de nombreuses maladies.

Les réseaux bayésiens sont proposés pour l'aide à la prise de décision, et ils permettent un raisonnement probabiliste. Ils sont utilisés dans le diagnostic médical pour aider les médecins afin de faire un diagnostic avec minimum d'erreurs.

Ainsi que les ontologies probabilistes basées sur les réseaux bayésiens multi entités qui permettent de décrire les connaissances incertaines de manière raisonnée et structurée et de modéliser l'incertitude par des facteurs de probabilité et des liens de causalité. Ces méthodes ont été employé dans le domaine médical à travers plusieurs langages comme PR-OWL, donnant ainsi place à la modélisation des connaissances probabilistes à travers les ontologies.

Problématique

Dans le domaine médical les données sont complexes et incertaines, ce qui met le médecin dans une situation délicate face a la prise de décision et provoque du stress chez les patients.

Les ontologies classiques permettent la représentation des connaissances. Ce sont des représentations qui regroupent un ensemble de concepts et de relations décrivant un domaine particulier. Mais leur défaut principal est l'incapacité de représenter et de raisonner sur l'incertitude.

Cette incertitude peut se manifester par exemple, lors de la prédiction qu'un patient est atteint par une maladie, dans notre cas d'étude la maladie est la COVID-19, cela conduit à poser les questions suivantes :

1. Comment modéliser ces connaissances incertaines dans le domaine médical ?
2. Comment construire une ontologie probabiliste pour traiter ces connaissances incertaines qui servent à inférer le diagnostic médical ?
3. La mise en œuvre du système de prédiction commence par l'identification d'un ensemble des signes pertinents pour produire un diagnostic, quelles sont les étapes à suivre et les outils nécessaires pour établir ce système d'aide au diagnostic médical basé sur les ontologies probabilistes ?

Objectif

L'existence des connaissances incertaines empêche la prise des décisions adéquates dans le domaine médical.

Notre objectif est de concevoir et de réaliser un système d'aide à la prédiction du diagnostic médical, dans notre étude de cas nous avons traité la maladie COVID-19 (Corona virus disease 19), en se basant sur une ontologie probabiliste pour la modélisation des connaissances incertaines. Cette ontologie est construite à l'aide de la méthodologie UMP-ST (Uncertainty Modeling Process for Semantic Technology).

Nous allons utiliser les réseaux bayésiens pour notre étude de cas, et plus particulièrement le réseau bayésien multi entités, il permet de représenter la connaissance incertaine pour générer une ontologie probabiliste PR-OWL (PProbabilistic Ontology Web Language) à l'aide d'un standard de représentation d'ontologie OWL (Web Ontology Language).

Organisation du mémoire

Afin d'atteindre le but de notre travail, l'organisation de notre mémoire sera comme suit :

La première partie :

Contient les chapitres 1 et 2, qui présentent le contexte de notre travail, en définissant les ontologies classiques, les ontologies probabilistes, les réseaux bayésiens ainsi qu'une étude des travaux connexes.

- **Chapitre 1 : Ontologies classiques et ontologies probabilistes :** Dans ce chapitre nous allons définir les ontologies classiques et probabilistes, leurs composants, leurs rôles ainsi que leurs langages de représentation , nous présentons aussi les réseaux bayésiens et la méthodologie suivit dans notre processus UMP-ST comme approche pour modéliser les ontologies probabilistes.

- **Chapitre 2 : Étude des travaux connexes existants :** Dans ce chapitre nous allons introduire le diagnostic médical ainsi que l’incertitude médicale . Nous parlerons aussi de quelques travaux existants en utilisant les ontologies classiques et les réseaux bayésiens, en comparant entre ces travaux .

La deuxième partie :

Après la présentation du travail, cette partie concerne la conception et le développement du système.

- **Chapitre 3 : Conception du système :** Ce chapitre présente la motivation de notre cas d’étude COVID-19, notre data set, ainsi que la création de notre ontologie probabiliste tout en suivant les différentes étapes de la méthodologie UMP-ST.

- **Chapitre 4 : Implémentation et test du système :** Ce chapitre est le résultat de la conception pour la mise en œuvre de notre ontologie probabiliste PR-OWL, ainsi que la démonstration de notre interface, pour finir avec le test et la validation de notre système.

Ce mémoire sera finalisé par une conclusion générale reprenant les principaux points abordés, et présentant les perspectives envisagées pour l’amélioration de notre travail.

Chapitre 1

ONTOLOGIES CLASSIQUES ET ONTOLOGIES PROBABILISTES

1.1 Introduction

En informatique et en science de l'information, une ontologie est l'ensemble structuré des termes et concepts représentant le sens d'un champs d'informations. Elle est employée pour raisonner à propos des objets du domaine concerné.

Dans ce chapitre nous allons traiter les ontologies classiques et probabilistes, nous allons aborder leurs définitions, leurs différentes caractéristiques, et leur utilisation, en mentionnant l'incertitude, les réseaux bayésiens et leurs influence sur les ontologies.

A la fin de chapitre, nous allons présenter les langages de représentation de ces deux types d'ontologies.

1.2 Les ontologies en ingénierie de connaissance

Au début des années 1990, le terme 'ontologie' s'est propagé dans la communauté ingénierie des connaissances, et depuis il a un très grand succès.

Les ontologies ont donné une solution au problème de la normalisation de la communication et d'acquisition des connaissances, l'idée est d'utiliser un système expert qui sépare une base de connaissances et un moteur d'inférence. Les systèmes de base de connaissance (SBC) proposaient alors de spécifier, d'un côté, des connaissances du domaine modélisé et, de l'autre, des connaissances de raisonnement décrivant les règles heuristiques d'utilisation de ces connaissances du domaine. [6]

1.3 Définition des ontologies

Depuis son adaptation dans le domaine de la science de l'information, le terme **ontologie** a reçu plusieurs définitions : Dans le domaine de l'intelligence artificielle **Neeches et al** [26], ont présenté leur vision : « Une ontologie définit les termes et relations de base pour définir les extensions du vocabulaire ».

En 1993, **Gruber** a proposé la définition la plus utilisée en IA jusqu'à présent : « An ontology is an explicit specification of a conceptualization ». [15]

Swartout et al [32] la définissent comme suit : « une ontologie est un ensemble de termes structurés de façon hiérarchique, conçue afin de décrire un domaine et qui peut servir de charpente à une base de connaissance ».

Borst [2] a modifié légèrement la définition : « Une ontologie est définie comme étant une spécification formelle d'une conceptualisation partagée ».

Puis la même notion a été développée par **Gomez** : « une ontologie fournit les moyens de décrire de façon explicite la conceptualisation des connaissances représentées dans une base de connaissances » [14].

1.4 Les composants d'une ontologie

Gomez a dit que les connaissances traduites par une ontologie sont à piloter à l'aide des éléments suivants : Concepts, Relations, Fonctions, Axiomes, Instances. [13].

- **Les concepts (termes/classe)** : représentent un groupe de différentes instances qui ont les mêmes caractéristiques.
- **Les relations** : c'est les associations pertinentes existant entre les concepts, qui permettent de construire la structure entre les concepts.
- **Les fonctions** : des cas particuliers de relations, dans lesquelles un élément de la relation, est défini en fonction des N-1 éléments précédents.
- **Les axiomes** : toutes les règles utilisées afin de définir la description des concepts et des relations permettant de représenter leur sémantique.
- **Les instances (individu)** : l'unité de base que l'ontologie décrit ou pourrait potentiellement les décrire. Les individus peuvent modéliser des objets concrets (Ex : personne) ou des objets plus abstraits (ex : le travail d'une personne). [22]

1.5 Rôle des ontologies

Gandon [12] dit que : « Une ontologie informatique offre un cadre unificateur et fournit des primitives améliorant la communication entre les personnes, entre les personnes et les systèmes, et entre les systèmes », En s'intéressant aux rôles des ontologies au niveau de l'ingénierie de connaissance. D'où les principaux sont :

- La communication :

Entre humain une ontologie informelle mais cohérente (définie sans ambiguïté) sera suffisante pour le partage de la connaissance. Mais si la communication est entre humain système on aura besoin d'une ontologie cohérente, semi-formelle pour le partage de cette même connaissance. [33]; [16]; [25]; [24].

- L'interopérabilité (entre systèmes) :

L'ontologie joue le rôle d'intermédiaire pour la traduction entre les objets des différents systèmes dans ce contexte Valéry Psyché et ses collègues [29] ont dit : « l'ontologie, assurerait la traduction entre les différentes méthodes de modélisation, paradigmes, langages et systèmes. Le degré de formalité requis serait rigoureusement formel ».

- La réutilisation de la connaissance :

Toute représentation formelle d'une ontologie pourra être réutilisée par d'autres systèmes, cela grâce à la mise en correspondance des contenus sémantique et/ou partage-able qui offre cette dernière.

1.6 Langages des ontologies

1.6.1 Resource Description Framework (RDF) et (RDFS)

Proposé par le W3C (World Wide Web Consortium) le RDF est implémenté en XML (Extensible Markup Language), Avec une entité de base triple (sujet, prédicat, objet). D'où :

- Le sujet représente la ressource à décrire.
- Le prédicat représente un type de propriété applicable à cette ressource.
- L'objet représente une donnée ou autre ressource. [35]

C'est un langage de représentation d'information web de manière flexible, significative pour que les données soient échangées sans perte de sémantique.

RDF Schéma (RDFS) a étendu RDF avec des moyens de spécification du vocabulaire du domaine et les structures d'objets. Ces techniques permettent d'enrichir le Web avec une sémantique traitable par machine, donnant ainsi naissance à ce qui a été surnommé le Web sémantique.

1.6.2 OWL

Le langage d'ontologie Web OWL est conçu pour être utilisé par les applications qui doivent traiter le contenu des informations au lieu de simplement présenter des informations à des humains. OWL facilite l'interopérabilité machine du contenu Web par rapport à celle prise en charge par XML, RDF et RDF Schéma (RDF-S) en fournissant un vocabulaire supplémentaire

ainsi qu'une sémantique formelle. OWL propose trois sous-langues de plus en plus expressives : **OWL Lite, OWL DL et OWL Full.**

1.7 Connaissance incertaine

Le principal défaut des ontologies classiques est leurs incapacités de raisonner sur l'incertitude.

La définition du terme incertitude se diffère d'un domaine à un autre.

Le W3C a développé une ontologie simple pour démontrer certaines fonctionnalités de base d'échange d'informations incertaines, et obtenir une couverture relativement complète des fonctionnalités liées au raisonnement d'incertitude sur les informations disponibles sur le World Wide Web. (Si dessous est présenté le niveau le plus élevé).

1.7.1 Classes d'incertitude d'une ontologie

La figure ci dessus présente les classes d'incertitude d'une ontologie.

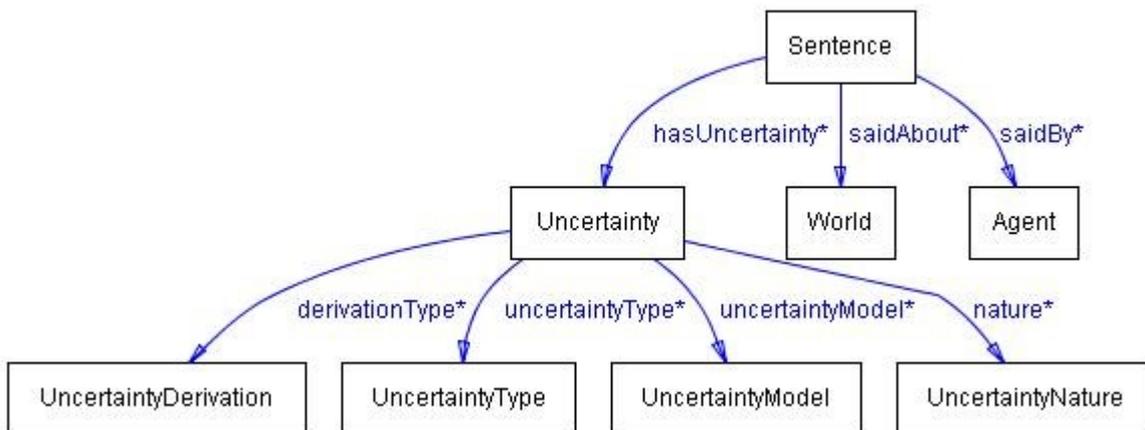


FIGURE 1.1 – Représentation des classes d'incertitude d'une ontologie (37)

- **Sentence** : Une expression dans un langage logique qui évalue une valeur de vérité. On suppose alors que les informations seront présentées sous forme de phrases. L'incertitude sera donc associée aux phrases.

- **World** : Cela représente le monde dont la phrase est dite.

- **Agent** : représentant celui qui fait la déclaration. Il peut s'agir d'un humain ou d'un agent informatique (machine).

- **Uncertainty** : c'est une déclaration sur l'incertitude associée à la phrase.

- **Uncertainty nature** : Cela saisit 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.

• **Uncertainty Derivation** : elle est soit objective (dérivée de manière formelle), soit subjective (peut être une supposition).

• **Uncertainty type** :

- Ambiguïté : les références du terme dans une phrase sur le monde ne sont pas clairement spécifiées et il n'est donc pas possible de déterminer si la phrase est satisfaite.

- Aléatoire, on parle de l'aléatoire lorsqu'on ne connaît pas une loi générale qui explique la réalisation des phénomènes.

- Imprécision : il n'y a pas de correspondance précise entre les termes de la phrase et les référents dans le monde.

- Incohérence : il y'a aucun monde qui satisfait cette affirmation.

- Incomplétude : les informations sur le monde sont incomplètes.

• **Uncertainty Model** : Cette classe contient des informations sur les théories mathématiques pour les types d'incertitude. Les types spécifiques de théories. Ex (probabilité, ensemble flou ..) (37) .

1.7.2 Approches de l'incertitude recommandées par le W3C

D'après le sous groupe du W3C les approches les plus utilisés sont :

- (a) **Théorie des probabilités** : il fournit un langage de représentation mathématiquement solide et un calcul formel pour les degrés rationnels de croyance, ce qui donne à différents agents la liberté d'avoir des croyances différentes sur une hypothèse donnée, d'où la représentation des connaissances incertaines et imparfaites qui peuvent provenir de divers agents. Parmi eux :
 - Réseaux bayésiens (BN) : un langage graphique puissant pour représenter les relations probabilistes entre un grand nombre d'hypothèses incertaines. Ils ont été appliqués à une grande variété de problèmes, y compris le diagnostic médical, les systèmes de classification.
- (b) **Logique floue** : contrairement au formalisme probabiliste qui traite l'incertitude, les formalismes flous permettent de représenter et de traiter des degrés d'appartenance sur des informations vagues (ou imprécises).
- (c) **Fonctions de croyance** : Les fonctions de croyance sont étroitement liées aux probabilités. Les croyances en une hypothèse sont calculées comme la somme des masses de tous les ensembles qu'elle englobe. Une fonction de croyance diffère d'un modèle de probabilité bayésien en ce que l'on ne conditionne pas les parties de la preuve pour lesquelles aucune probabilité n'est spécifiée. Cette capacité à modéliser explicitement le degré d'ignorance rend la théorie très attrayante. (37)

1.8 Réseau bayésien (RB)

La représentation des connaissances et le raisonnement à partir de ces représentations ont donné naissance aux modèles graphiques probabilistes, et plus précisément les réseaux bayésiens, initiés par **Judea Pearl** [20] dans les années **1980**, ce sont révélés des outils très pratiques pour la représentation de connaissances incertaines, et le raisonnement à partir d'informations incomplètes.

Un réseau bayésien $\mathbf{B} = (\mathbf{G}, \mathbf{O})$ est défini par :

- $\mathbf{G} = (\mathbf{X}, \mathbf{E})$, graphe dirigé sans circuit dont les sommets sont associés à un ensemble de variables aléatoires $\mathbf{X} = X_1, \dots, X_n$.
- $\mathbf{O} = P(X_i | Pa(X_i))$, ensemble des probabilités de chaque noeud X_i conditionnellement à l'état de ses parents $Pa(X_i)$ dans \mathbf{G} .

Ainsi, la partie graphique du réseau bayésien indique les dépendances (ou indépendances) entre les variables et donne un outil visuel de représentation des connaissances. De plus, l'utilisation de probabilités permet de prendre en compte l'incertain, en quantifiant les dépendances entre les variables.

D'après **Leray**, Les réseaux bayésiens sont des outils de simulation qui permettent à l'expert d'observer le comportement de son système dans des contextes qu'il n'est pas forcément capable de tester lui-même.[20]

L'apprentissage d'un réseau bayésien à partir d'un jeu de données contenant un ensemble d'événements se fait en deux étapes :

- Apprentissage des paramètres.
- Apprentissage de la structure.

1.8.1 Apprentissage des paramètres

soit à partir des données complètes ou incomplètes.

(a) A partir des données complètes

dans ce cas on cherche à estimer la distribution des probabilités à partir des données disponibles en utilisant :

- **Apprentissage statique** : si 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 événement par la fréquence d'apparition de l'événement dans la base de données.

- **Apprentissage bayésien** : L'estimation bayésienne suit un principe un 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 des données incomplètes

Dans les applications pratiques, les bases de données sont très souvent incomplètes. Cer-

taines variables ne sont observées que partiellement ou même jamais, que ce soit en raison de faute humain ou machine.

Parmi les méthodes d'estimation des paramètres des données incomplètes il existe l'algorithme itératif EM (Expectation Maximisation).

1.8.2 Apprentissage de la structure

L'apprentissage de la structure d'un réseau bayésien à partir des données revient à trouver un graphe qui soit un modèle d'indépendance associé à une distribution de probabilité dont nous possédons un échantillon. Il faut définir les noeuds du réseau, et les liens (généralement de causalité) entre ces noeuds.

Une première approche, proposée initialement par **Spirites** d'un côté, et **Pearl et Verma**[34] de l'autre, consiste à rechercher les différentes indépendances conditionnelles qui existent entre les variables. Les autres approches tentent de quantifier l'adéquation d'un réseau bayésien au problème à résoudre, c'est-à-dire d'associer un score à chaque réseau bayésien. Puis elles recherchent la structure qui donnera le meilleur score dans l'espace B des graphes a cycliques dirigés.

1.9 Ontologies probabilistes

1.9.1 Définition

L'ontologie probabiliste permet d'intégrer l'incertitude a des connaissances de domaine, afin de les décrire de manière raisonnée, structurée et partageable, idéalement dans un format qui peut être lu et traité par un ordinateur. [17].

Et d'après Costa et al.[7], une ontologie probabiliste est « an explicit, formal knowledge representation that expresses knowledge about a domain of application », ceci comprend :

- Tous types d'entités qui existent dans un domaine avec leur propriétés et relations.
- Processus et événements qui se produisent avec ces entités.
- Régularités statistiques qui caractérisent le domaine.
- Connaissances non concluantes, ambiguës, incomplètes, non fiables.
- Incertitude concernant toutes les formes de connaissance ci-dessus.

Parmi les motivations d'utilisation d'ontologies probabilistes, nous citons :

- Décrire en détail la connaissance et l'incertitude associée dans un environnement informatique.
- Plus les données sont grandes plus l'incertitude règne entre les liens.
- Cette incertitude n'est pas prise en compte par la logique déterministe.
- Faire des prédictions en se basant sur des observations historiques.

- Décrire l'incertitude d'une manière structurée et traitée par ordinateur.

1.9.2 Langage des ontologies probabilistes

La communauté du Web sémantique reconnaît de plus en plus la nécessité de représenter et raisonner avec incertitude. Plusieurs représentations candidates pour le raisonnement d'incertitude dans le Web sémantique et probabiliste ont été proposées, parmi eux on a : PR-OWL qui utilise les réseaux bayésiens MEBN.

1.9.2.1 MEBN

Les réseaux bayésiens multi-entités (MEBN) étendent les réseaux bayésiens (BN) pour atteindre les commandes puissantes et expressives.

MEBN représente la connaissance comme une collection de fragments MEBN (MFrag), qui sont organisés en théories MEBN (MTheories). Un MFrag contient des variables aléatoires (RV) et un graphe fragmenté représentant les dépendances parmi ces VR. Il représente un modèle de connaissances reproductible qui peut être instancié autant de fois que nécessaire, pour former une situation spécifique, et ainsi peut être considéré comme un modèle pour la construction et la combinaison de fragments d'un réseau bayésien.

Il est instancié en liant ses arguments aux identifiants d'entités de domaine pour créer des instances de ses VR.

Il existe trois types de nœuds : **contexte, résident et entrée**. Nœuds de contexte représentent les conditions qui doivent être satisfaites pour les distributions représentées dans le MFrag postulé. Les nœuds d'entrée peuvent être influencé par les distributions des autres nœuds dans un MFrag, mais leurs distributions est défini dans leurs MFrag d'origine.

Les distributions pour les nœuds résidents sont définies au sein du MFrag en spécifiant des distributions locales conditionnées aux valeurs des instances de leurs parents dans le graphe fragmenté. Un ensemble de MFrag représente une distribution conjointe sur des instances de ses variables aléatoires.

MEBN fournit un moyen compact de représentation des structures répétées dans un BN. Un important avantage de MEBN est qu'il n'y a pas de limite fixe sur le nombre d'instances RV, et les instances de variables aléatoires sont instanciées dynamiquement selon les besoins.

An MTheory est un ensemble de MFrag qui remplissent collectivement des conditions de cohérence garantissant l'existence d'une distribution de probabilité conjointe unique sur ses instances de variables aléatoires. Pour appliquer une théorie M à raisonner sur des scénarios

particuliers, il faut un système avec des informations spécifiques sur les instances d'entités individuelles impliquées dans le scénario.

Dès la réception de ces informations, l'inférence bayésienne peut être utilisée à la fois pour répondre aux questions d'intérêt. L'inférence bayésienne est utilisée pour effectuer à la fois une inférence spécifique au problème et pour l'apprentissage à partir de données d'une manière saine et cohérente. [4]

1.9.2.2 PR-OWL

PR-OWL est une extension d'OWL, ayant la capacité de faire des raisonnements probabilistes basés sur les réseaux bayésiens avec entités multiples (MEBN : Multi Entity Bayesian Network). PR-OWL a un grand niveau de flexibilité qui ne peut être atteint qu'en utilisant la sémantique sous-jacente de la logique bayésienne du premier ordre, qui ne fait pas partie du standard de OWL par conséquent PR-OWL ne peut être réalisé qu'en étendant la sémantique et la syntaxe abstraite de OWL.

La première étape pour la construction d'une ontologie probabiliste est d'importer dans n'importe quel éditeur OWL un fichier OWL contenant le PR-OWL classes, sous-classes et propriétés. Il n'est pas nécessaire que tous les concepts d'une ontologie probabiliste soient probabilistes mais d'avoir « une partie probabiliste » qui fait référence aux concepts rédigés à l'aide de définition PR-OWL et qui forment collectivement un MTheory partielle.

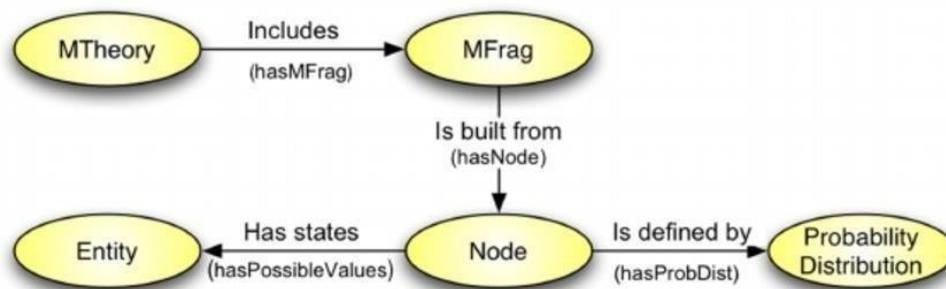


FIGURE 1.2 – PR-OWL MTheory Concepts [8]

La figure 1.2 représente un aperçu des concepts généraux impliqués dans la définition d'une MTheory en PR-OWL. Dans ce diagramme **COSTA et al** [8] ont représenté les classes générales par les ovales et les principales relations entre classes par des flèches. Chaque MTheory est un regroupement des MFrag qui contient des individus.

Les individus de la classe MFrag sont constitués de nœuds, qui peuvent être résidents, d'entrée ou nœuds de contexte (non représentés sur la figure). Chaque individu de la classe Node est une variable aléatoire qui a un ensemble mutuellement exclusif et collectivement exhaustif de

possibles États. Dans PR-OWL, la propriété d'objet `hasPossibleValues` relie chaque nœud à son état possible, qui est des individus de la classe Entité.

1.10 Méthodologie UMP-ST

Le processus de modélisation d'incertitude pour les technologies sémantiques (UMP-ST) a été développé par Carvalho, comme la seule et unique méthodologie pour la modélisation d'ontologies probabilistes. Il couvre la première étape de l'UMP-ST, permettant de modéliser le domaine.

L'outil résout trois problèmes principaux : la complexité de la création d'ontologie probabiliste ; la difficulté de maintenir et faire évoluer les ontologies probabilistes existantes ; et l'absence des outils des ontologies probabilistes.

UMP-ST se compose de quatre disciplines : **Exigences, Analyse et Conception, Implémentation et Test.**

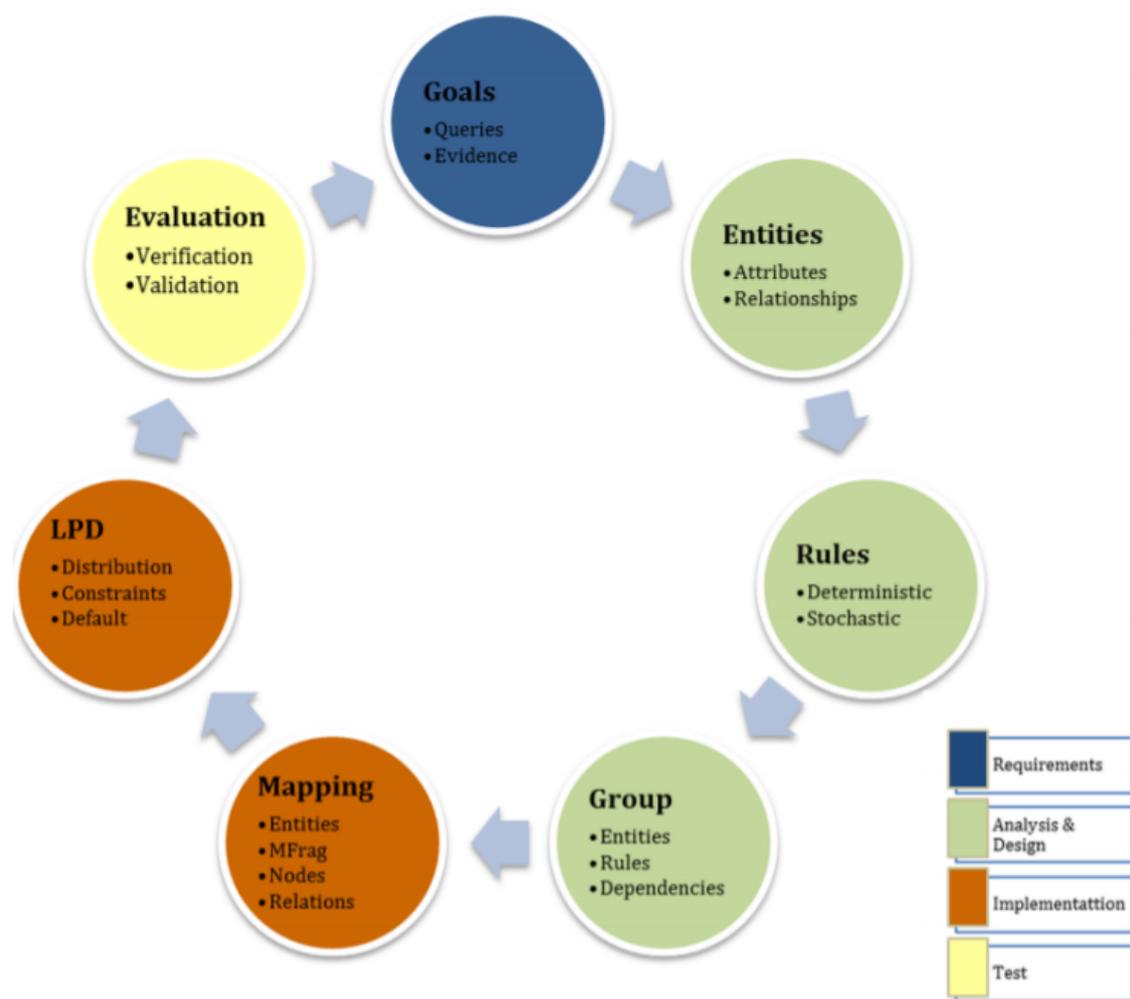


FIGURE 1.3 – Flux POMC [1]

La figure 1.3 présente le cycle de modélisation d'ontologie probabiliste POMC (probabilistic cycle of ontological modeling), qui décrit les principaux résultats de chaque discipline de l'UMP-ST et l'ordre naturel dans lequel chacun est produit. Le produit final d'un cycle est une version de l'ontologie qui répond à un ensemble d'exigences. Chaque étape est décrite comme suit :

1. **La discipline Exigences (The Requirements discipline) :** Définit les objectifs qui doivent être atteints par raisonnement dans PO.
2. **La discipline Analyse et Conception (Analysis and Design discipline) :** Définit les classes d'entités, leurs attributs, leurs relations et leurs règles.
3. **La discipline de mise en œuvre (the Implementation discipline) :** Les entités, les relations, les attributs, les règles et les groupes seront mappés à leurs concepts correspondants dans un formalisme spécifique qui permet une représentation de l'incertitude.
4. **La discipline Test (Test discipline) :** l'utilisateur validera et vérifiera le modèle dans lequel il est observé si l'ontologie résultante a un comportement attendu et si toutes les exigences ont été mises en œuvre.

1.11 Entre ontologie classique et probabiliste

L'ontologie probabiliste peut raisonner sur l'incertitude et permet de fournir un outil de modélisation contrairement à l'ontologie classique qui diminue la performance chaque fois que les données sont grandes « Plus les données sont grandes plus l'incertitude règne entre les liens ». Sans oublier que les ontologies probabilistes ont la capacité de décrire l'incertitude d'une manière structurée et traitée par ordinateur.

Nous présentons dans le tableau une comparaison des deux exemples de langage (OWL, PR-OWL) qui appartient aux deux différentes ontologies (classique, probabiliste) consécutivement :

OWL	PR-OWL
Standardisé	Non standardisé
Langage déterministe	Langage probabiliste
Ne tient pas compte de l'incertitude	Tient compte de l'incertitude
Base logique	Réseau bayésien
Compatible :XML	Compatible :XML
Formalisme :Sujet, Prédicat, Objet	Formalisme :MTheory, MFrag

TABLE 1.1 – Comparaison entre les deux langages [18]

1.12 Conclusion

Dans ce chapitre nous avons traité les ontologies classiques et les ontologies probabilistes, leurs composants ainsi que leurs langages en introduisant la notion d'incertitude et des réseaux

bayésiens.

Ce chapitre nous donne une vue générale sur les ontologies probabilistes et leurs capacités de traiter l'incertitude ce qui nous emmène à présenter les travaux existants dans le chapitre suivant.

ÉTUDE DES TRAVAUX CONNEXES EXISTANTS

2.1 Introduction

Le diagnostic médical est le processus qui constitue l'identification des maladies des patients en fonction des signes, symptômes et antécédents médicaux . Le médecin procède à une première évaluation de la situation clinique du patient en considérant les pathologies compatibles avec symptômes et signes, puis il peut procéder avec des diagnostics différents.

Malheureusement, la quantité d'informations qui doit être prise en compte dans le diagnostic médical est énorme et donc difficile à gérer par les esprits humains [3] plusieurs ontologies déjà existantes permettent la communication et la gestion des données avec minimum d'ambiguïté pour faire des diagnostics fiables pour plusieurs types de maladies.

Une maladie est une Altération de l'état de santé se manifestant par un ensemble de signes et de symptômes perceptibles directement ou non, correspondant à des troubles généraux ou localisés, fonctionnels ou lésionnels, dus à des causes internes ou externes et comportant une évolution. Savoir faire le diagnostic des maladies ainsi que leurs traitements le plutôt possible est un facteur important pour la guérison du malade.

Ce chapitre présente le diagnostic médical , ainsi que quelques travaux connexes au diagnostic médical basé sur les ontologies existantes, afin de faire une analyse de ces derniers.

2.2 Diagnostic médical

En médecine, le diagnostic est le processus permettant d'établir la présence d'une maladie ou d'un dysfonctionnement chez un patient. Cela permet de décider d'une conduite à tenir (surveillance simple, médicaments, chirurgie...)

Le diagnostic est la conclusion d'un cheminement intellectuel pouvant se reposer sur trois étapes de la consultation médicale :

- **L'anamnèse** : l'anamnèse correspond à l'interrogatoire médical. C'est le moment où le patient va décrire ses symptômes.
- **L'examen physique** : a comme but de rechercher des signes physiques des maladies.
- **Les examens complémentaires.**

La réalisation d'un diagnostic passe par les démarches suivantes :

2.2.1 Démarche diagnostique

Le médecin généraliste est confronté dans son exercice quotidien à plusieurs contraintes :

- Il est face à des troubles de santé au stade précoce de leur évolution, qui correspondent rarement à un tableau complet d'une maladie.
- Il a des moyens diagnostiques limités, sans plateau technique.
- Il est amené à prendre des décisions dans un temps court (18 minutes en moyenne).

Afin de prendre des décisions adaptées, le médecin doit nommer précisément chaque situation clinique qu'il prend en charge et évaluer le risque inhérent à chacune des situations cliniques.

2.2.2 Démarche décisionnelle

Penser que les décisions ne se prennent qu'en fonction du résultat de consultation et de ses risques graves est une illusion. En effet, l'efficacité des décisions médicales dépend aussi d'autres éléments liés au patient, au médecin et à leur environnement.

Une fois dénommée la situation clinique qu'il prend en charge, le médecin doit alors tenir compte des autres éléments utiles à une décision adaptée : les facteurs de risque, les recommandations médicales, le contexte environnemental, les déterminants du patient et ceux du médecin. Le médecin peut alors expliquer au patient ce qu'il faudrait faire.

2.2.3 Incertitude médicale

Mal gérée ou mal tolérée, l'incertitude en médecine a des conséquences néfastes pour le patient et/ou le médecin. Dans la littérature, il n'existe pas de modèle explicatif global des différentes dimensions de l'incertitude confrontée dans le soin.

L'incertitude médicale peut toucher deux facteurs, premièrement en fonction de la personne malade, deuxièmement en fonction de l'étape de soin reçu. L'incertitude porte sur l'occurrence de la maladie, entre autres le patient pourra modifier le risque par son propre comportement.

L'incertitude est transmise du patient au médecin qui a les moyens nécessaires pour enlever ou diminuer le doute de la maladie par un jugement suivi d'un diagnostic médical.

On trouve deux critères de l'incertitude selon Schmidt, l'incomplétude des états de la nature de la maladie et du diagnostic, et les facteurs de risque ainsi que leur environnement.

2.2.4 Signe et Symptôme

Les signes sont les caractéristiques cliniques que l'on peut voir ou mesurer.

Un symptôme est un signe, un trouble observable par un patient, qui est la manifestation d'une maladie et de son évolution. Les signes et les symptômes dépendront du type et de la gravité d'une maladie, du niveau d'activité et de la quantité de traitement donnée jusque là.

2.2.4.1 Signe

En principe, le signe est une manifestation objective, distincte du symptôme. Ou alors, le signe est appelé symptôme lorsqu'il est mis en rapport avec un état pathologique. Par son interprétation, le médecin déterminera si tel ou tel signe peut avoir valeur de symptôme. La description subjective faite par le patient est appelée « plainte du patient » ou « motif de consultation ».

- Types des signes

- **Signe positif ou négatif** : un signe est dit positif si c'est sa présence qui a une signification. Un signe est dit négatif si c'est son absence qui a une signification.
- **Signe général** : un signe qui concerne l'organisme dans sa globalité (fièvre, amaigrissement, sueurs...).
- **Signe physique** : signe constaté objectivement (comme l'état de la peau) ou provoqué par une manœuvre d'examen (comme une palpation douloureuse).
- **Signe fonctionnel** : signe ressenti subjectivement par le malade (fatigue, démangeaisons...).
- **Signe constant ou inconstant** : un signe clinique est constant lorsqu'il est toujours retrouvé au cours d'une pathologie donnée. Son absence infirme donc le diagnostic de cette pathologie. Un signe clinique est dit inconstant lorsqu'il peut être ou ne pas être retrouvé au cours d'une pathologie donnée. Son absence n'infirme donc pas le diagnostic de cette pathologie.
- **Signe clinique spécifique ou non spécifique** : un signe clinique est dit spécifique lorsqu'il permettra d'orienter le diagnostic vers un groupe de pathologies dont certaines expressions

cliniques présentent des points communs ou difficilement différenciables. Un signe clinique est dit non spécifique lorsqu'il peut être l'expression clinique de plusieurs maladies différentes.

- **Signe pathognomonique** : la présence d'un tel signe permet à coup sûr, et à lui seul, de poser le diagnostic clinique d'une pathologie. Le signe pathognomonique est le signe le plus spécifique qui soit.

2.2.4.2 Symptôme

Le symptôme est subjectif, et se distingue d'un signe clinique objectif, détecté par le médecin ou lors d'un examen médical (analyses biologiques, imagerie médicale...).

On distingue deux types :

Symptômes communs aux maladies et maladies asymptomatiques : Une maladie se caractérise généralement par différents symptômes et ces symptômes peuvent être communs à plusieurs maladies. Certaines maladies sont dites asymptomatiques, ce qui signifie qu'elles peuvent passer inaperçues aux yeux du patient.

Symptômes au diagnostic de la maladie : Lors d'une consultation médicale, le médecin interroge le patient sur son ressenti afin de connaître ses symptômes. La confrontation des symptômes perçus et des résultats des examens cliniques permet de poser un diagnostic médical.

2.3 Travaux existants

Après l'étude de plusieurs travaux connexes qui traitent l'incertitude sur le diagnostic médical, nous avons pu les classifiés en deux catégories : les travaux qui se basent sur une ontologie médicale, et ceux qui se basent sur les réseaux bayésiens, d'où nous avons présenté pour chaque catégorie les travaux suivants :

1- Travaux basés sur une ontologie médicale

- Une ontologie pour les troubles du spectre autistique (TSA) pour déduire les phénotypes des TSA à partir des données révisées des entretiens de diagnostic d'autisme.[9]
- An ontology-driven clinical decision support system (IDDAP) for infectious disease diagnosis and antibiotic prescription.[31]
- L'ontologie de la brucellose (IDOBUR) en tant d'extension de l'ontologie des maladies infectieuses[21]
- Ontology-Based Approach for Liver Cancer Diagnosis and Treatment [23]

2- Travaux basés sur les réseaux bayésiens

- A Bayesian network decision model for supporting the diagnosis of dementia, Alzheimer's disease and mild cognitive impairment.[30]

- Evaluation of the Diagnostic Power of Thermography in Breast Cancer Using Bayesian Network Classifiers [27]
- An Ontology Driven and Bayesian Network Based Cardiovascular Decision Support Framework [11]
- Modèle probabiliste pour le diagnostic et le phénotypage des maladies pulmonaires obstructives chroniques à l'aide du réseau bayésien [28]

Voici un schéma qui représente les travaux que nous allons traiter

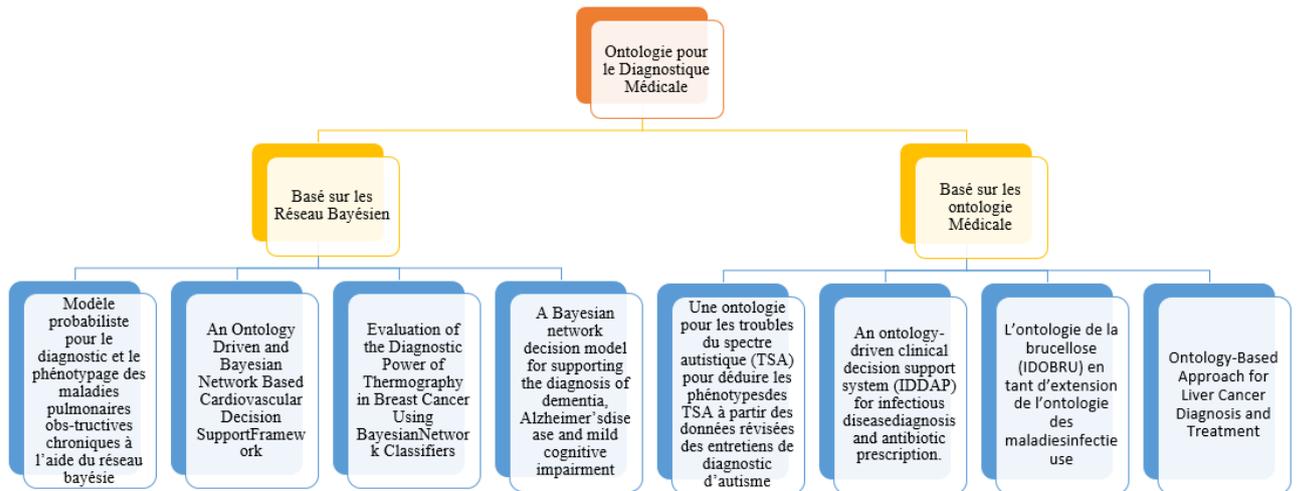


FIGURE 2.1 – Travaux existants

2.3.1 Travaux basés sur une ontologie médicale

2.3.1.1 La maladie d'autisme

- **Titre** : Une ontologie pour les troubles du spectre autistique (TSA) pour déduire les phénotypes des TSA à partir des données révisées des entretiens de diagnostic d'autisme [9].

- **Maladie** : L'autisme, dont les manifestations sont décrites sous l'intitulé de troubles du spectre de l'autisme (TSA), est un trouble du neuro-développement humain caractérisé par des difficultés dans les interactions sociales et la communication, et des comportements et intérêts à caractère restreint, répétitif et stéréotypé.

- **Objectif** : Créer une ontologie qui permettra l'intégration des données et le raisonnement avec les données pour classer les sujets, et sur la base de cette classification déduire de nouvelles connaissances sur les troubles du spectre autistique (TSA) et les troubles neuro-développementaux associés (NDD). En étendant une ontologie de l'autisme existante pour permettre l'inférence automatique des phénotypes de TSA et des critères du Manuel diagnostique et statistique des troubles mentaux (DSM) basés sur les données d'évaluation de l'autisme Diagnostic

Interview-Revised (ADI-R) des sujets.

- **Méthodes** : Des connaissances sur les instruments de diagnostic, les phénotypes de TSA et les facteurs de risque ont été ajoutées pour augmenter une ontologie de l'autisme existante via les définitions de classe du langage Web Ontology et les règles du Web sémantique. Un plugin Protégé personnalisé pour énumérer les axiomes combinatoires OWL afin de prendre en charge les relations plusieurs-à-plusieurs des éléments ADI-R avec les catégories de diagnostic dans le DSM. Un raisonnement pour déduire si 2642 sujets, dont les données ont été obtenues de la Simons Foundation Autism Research Initiative, répondent aux critères de diagnostic DSM-IV-TR (DSM-IV) et DSM-5 sur la base de leurs données ADI-R.

- **Résultats** : Une ontologie étendue en ajoutant 443 classes et 632 règles qui représentent des phénotypes, ainsi que leurs synonymes, facteurs de risque environnementaux et fréquence des comorbidités. L'application des règles sur l'ensemble de données a montré que la méthode produisait des résultats précis.

2.3.1.2 Les maladies infectieuses

- **Titre** : An ontology-driven clinical decision support system (IDDAP) for infectious disease diagnosis and antibiotic prescription.[31]

- **Maladie** : Les maladies infectieuses en termes épidémiologies, il s'agit des maladies qui apparaissent ou dont l'incidence augmente en un lieu donné.[10]

- **Objectif** : Parce que les maladies infectieuses sont courantes, Bien que l'utilisation d'antibiotiques doit en principe être soumise à l'avis d'un médecin, de nombreux patients les prennent sans autorisation médicale et certaines personnes ne peuvent pas facilement ou rapidement consulter un médecin. L'objectif de ce système est le diagnostic des maladies infectieuses et le soutien à la prescription d'antibiotiques.

- **Méthodes** : Cette étude décrit la construction et l'optimisation de la sensibilité et de la spécificité d'un système d'aide à la décision nommé IDDAP, basé sur des ontologies pour le diagnostic des maladies infectieuses et l'antibiothérapie.

L'ontologie de ce système a été construite en rassemblant les ontologies existantes associées aux maladies infectieuses, aux syndromes, aux bactéries et aux médicaments dans le schéma conceptuel hiérarchique de l'ontologie. Premièrement, l'IDDAP identifie une maladie infectieuse potentielle sur la base de l'état pathologique auto-décrit du patient. Ensuite, le système recherche et propose une antibiothérapie appropriée spécifiquement adaptée au patient en fonction de facteurs.

- **Résultats** : L'analyse des performances de l'IDDAP en termes de classificateurs sur la base

des résultats de la courbe des caractéristiques de fonctionnement du récepteur (ROC) 89,91% a révélé les avantages de l'IDDAP lorsqu'il est combiné avec cette ontologie.

2.3.1.3 Les maladies infectieuses bactériennes

- **Titre** : L'ontologie de la brucellose (IDOBUR) en tant d'extension de l'ontologie des maladies infectieuses[21]

- **Maladie** : Les maladies infectieuses regroupent toutes les pathologies provoquées par une bactérie, un virus, un parasite ou un champignon. Une large famille avec des combats plus ou moins difficiles à mener pour l'organisme.

- **Objectif** : Causée par la bactérie Gram-négative intracellulaire *Brucella*. La brucellose est la zoonose bactérienne la plus courante. Des études approfondies sur la brucellose ont produit un grand nombre de publications et de données couvrant divers sujets allant de l'étude génétique de base de *Brucella* aux essais cliniques de vaccins. Pour soutenir l'interopérabilité des données et le raisonnement, une ontologie biomédicale spécifique à la brucellose basée sur la communauté est nécessaire.

- **Méthodes** : L'ontologie de la brucellose, une ontologie biomédicale dans le domaine de la brucellose, est une ontologie d'extension de l'ontologie de base des maladies infectieuses (IDO-core). IDOBUR est importé de 10 ontologies existantes.

IDOBUR a été utilisé pour modéliser différents aspects de la brucellose, y compris l'infection de l'hôte, la transmission de la maladie zoonotique, les symptômes, les facteurs de virulence et la pathogénèse, le diagnostic, la libération intentionnelle, la prévention vaccinale et le traitement. Les études de cas sont généralement utilisées dans notre modélisation IDOBUR. Par exemple, la variation de température diurne chez les patients atteints de *Brucella*, une méthode PCR spécifique à *Brucella* et un traitement de la brucellose recommandé par l'OMS ont été sélectionnés comme cas d'utilisation pour modéliser respectivement les symptômes, le diagnostic et le traitement de la brucellose.

Développé avec OWL, IDOBUR prend en charge le raisonnement ontologique basé sur OWL. Par exemple, en effectuant une requête Description Logic (DL) dans l'éditeur OWL Protégé 4 ou une requête SPARQL dans un serveur IDOBUR SPARQL, un contrôle des facteurs de virulence de *Brucella* a montré que huit d'entre eux sont des antigènes protecteurs connus sur la base des connaissances biologiques capturées dans l'ontologie.

- **Résultats** : IDOBUR est la première ontologie de maladies infectieuses bactériennes signalé développée pour représenter différents aspects de la maladie dans un format logique formel. Il sert de base de connaissances sur la brucellose et prend en charge l'intégration des données sur la brucellose et le raisonnement automatisé.

2.3.1.4 La maladie cancer du foie

-**Titre** :Ontology-Based Approach for Liver Cancer Diagnosis and Treatment [23]

- **Maladie** : Le cancer du foie est une tumeur maligne, ou cancéreuse, qui prend naissance dans les cellules du foie. Le mot « maligne » signifie que la tumeur peut se propager (métastases) à d'autres parties du corps. Quand le cancer apparaît d'abord dans les cellules du foie, il est appelé cancer primitif du foie.

Le cancer du foie est le troisième cancer le plus mortel au monde. Elle caractérise une tumeur maligne qui se développe à travers les cellules hépatiques. Le carcinome hépatocellulaire (CHC) fait partie de ces tumeurs. Le cancer primitif hépatique est la principale cause de décès par cancer.

- **Objectif** : Le but est de traiter le processus de diagnostic des cancers du foie. Pour analyser une grande masse de données médicales, les ontologies sont efficaces ; ils sont efficaces pour améliorer l'analyse des images médicales utilisées pour détecter différentes tumeurs et autres lésions hépatiques. Précisément le HCC. Par conséquent, l'objectif principal de ce travail est de proposer une nouvelle approche basée sur l'ontologie modélisant les tumeurs du CHC en se concentrant sur deux aspects majeurs : le premier se concentre sur la détection des tumeurs en imagerie médicale, et le second se concentre sur sa mise en scène en appliquant différents systèmes de classification.

- **Méthodes** : L'implémentation a été faite en Java en utilisant l'API Jena. Aussi, en développant un prototype OntHCC en utilisant des aspects sémantiques et des règles de raisonnement pour valider le travail. Pour montrer l'efficacité de travail, le teste de l'approche proposée est réalisé sur des jeux de données réels.

- **Résultats** : Les résultats obtenus ont montré un système fiable avec des précisions élevées de rappel 76 %, de précision 85 % et de mesure F 80 %.

2.3.2 Travaux basés sur les réseaux bayésiens

2.3.2.1 La maladie d'Alzheimer

-**Titre** :A Bayesian network decision model for supporting the diagnosis of dementia, Alzheimers disease and mild cognitive impairment.[30]

- **Maladie** : La démence est un état clinique caractérisé par une perte de fonction dans plusieurs domaines cognitifs. Il existe différents types spécifiques de démence, présentant souvent des symptômes légèrement différents. La plus fréquente est la maladie d'Alzheimer (MA), qui représente entre 60% et 80% des cas de démence . La MA est une maladie dégénérative causant

des lésions cérébrales. Les premiers symptômes cliniques de la MA sont souvent liés à une difficulté à se souvenir de nouvelles informations, et les symptômes ultérieurs comprennent une altération du jugement, une désorientation, une confusion, des changements de comportement et des difficultés à parler et à marcher.

- **Objectif** : Cet article propose un modèle de décision en réseau bayésien pour soutenir le diagnostic de la démence, de la MA et des troubles cognitifs légers (MCI). Les réseaux bayésiens sont bien adaptés pour représenter l'incertitude et la causalité, qui sont toutes deux présentes dans les domaines cliniques.

- **Méthodes** : Le réseau bayésien proposé a été modélisé en utilisant une combinaison de connaissances d'experts et de modélisation orientée données. La structure du réseau a été construite sur la base des critères diagnostiques actuels et des contributions de médecins experts dans ce domaine. Les paramètres du réseau ont été estimés à l'aide d'un algorithme d'apprentissage supervisé à partir d'un ensemble de données de cas cliniques réels. L'ensemble de données contient des données provenant de patients et de témoins normaux du Duke University Medical Center (Washington, États-Unis) et du Center for Alzheimer's Disease and Related Disorders (à l'Institut de psychiatrie de l'Université fédérale de Rio de Janeiro, Brésil). Les attributs de l'ensemble de données comprennent des facteurs de prédisposition, des résultats de tests neuropsychologiques, des données démographiques sur les patients, des symptômes et des signes. Le modèle de décision a été évalué à l'aide de méthodes quantitatives et d'une analyse de sensibilité.

- **Résultats** : En conclusion, le réseau bayésien proposé a montré de meilleurs résultats pour le diagnostic de la démence, de la MA et du MCI par rapport à la plupart des autres classificateurs bien connus. De plus, il fournit des informations supplémentaires utiles aux médecins, comme la contribution de certains facteurs au diagnostic.

2.3.2.2 La maladie de Cancer du sein

- **Titre** : Evaluation of the Diagnostic Power of Thermography in Breast Cancer Using Bayesian Network Classifiers [27]

- **Maladie** : Le cancer du sein est l'une des principales causes de décès chez les femmes dans le monde. De nos jours, il existe différentes techniques pour réaliser le diagnostic : mammographie, échographie, IRM, biopsies et, plus récemment, thermographie.

- **Objectif** : Dans cet article, ils évaluèrent le pouvoir diagnostique de la thermographie dans le cancer du sein en utilisant des classificateurs de réseau bayésien. Nous montrerons comment les informations fournies par l'image thermique peuvent être utilisées pour caractériser les patients suspectés d'avoir un cancer.

- **Méthodes** : Pour leur expériences, ils ont utilisé une base de données du monde réel , composée de 98 cas : 77 cas sont des patientes atteintes d'un cancer du sein 78,57% et 21 cas sont des patientes en bonne santé 21,43%. Tous les résultats (malades ou sains) ont été confirmés par une biopsie ouverte.puis on utilise un classificateur de réseau bayésien.

- **Résultats** : Nous avons constaté que le cadre des réseaux bayésiens fournit un bon modèle pour analyser ce type de données : il peut montrer visuellement les interactions entre les attributs et le résultat ainsi que les interactions entre les attributs et mesurer numériquement l'impact de chaque attribut sur la classe.

2.3.2.3 Les maladies cardiovasculaires

- **Titre** :An Ontology Driven and Bayesian Network Based Cardiovascular Decision Support Framework [11]

- **Maladie** : Les maladies cardiovasculaires sont l'ensemble de troubles affectant le cœur et les vaisseaux sanguins, qui comprend.

- **Objectif** : cet article présente un cadre d'aide à la décision cardiovasculaire basé sur des principes clés d'ingénierie d'ontologie et un réseau bayésien.

- **Méthodes** : ils ont séparer les connaissances du domaine (connaissances de l'expert clinique et directives de pratique clinique) des informations probabilistes. Puis, ils ont développés les composants clés du cadre d'aide à la décision cardiovasculaire proposé en utilisant une approche axée sur l'ontologie. ils ont également utilisé une approche de réseau bayésien (BN) pour modéliser l'incertitude clinique dans les dossiers de santé électroniques (DSE).

- **Résultats** : Les résultats ont démontré des probabilités élevées pour la cible activités (facteurs de risque), qui peuvent aider les infirmières du RACPC dans la prise de décision clinique sur la base des preuves observées présentées par le réseau bayésien.

2.3.2.4 Les maladies des poumons

- **Titre** : Modèle probabiliste pour le diagnostic et le phénotypage des maladies pulmonaires obstructives chroniques à l'aide du réseau bayésien [28]

- **Maladie** : ce sont les maladies affectant les poumons touchant le tissu alvéolaire, les petites bronches terminales, d'origine infectieuse (bactérienne, virale, parasitaire) ou non.

- **Objectif** : L'organisation mondiale de la santé (OMS) rapporte que la maladie pulmonaire obstructive chronique (MPOC) est l'une des principales causes de décès dans le monde, avec 90 des décès dus à la MPOC dans les pays à revenu moyen et faible. La prévalence de la MPOC devrait augmenter de plus de 30% au cours de la prochaine décennie, à moins que des mesures urgentes ne soient prises pour réduire les facteurs de risque associés. Un spiromètre est le prin-

principal outil de diagnostic utilisé pour diagnostiquer et séparer la MPOC et l'asthme. Cependant, l'utilisation de la spirométrie comporte un certain nombre de défis qui les rendent moins utiles. Ce contraste déclenche le désir de connaître le niveau de prévalence de la BPCO dans les pays en développement en plus de la conception d'un dispositif de diagnostic qui pourrait être utilisé dans les sociétés à ressources limitées. Par conséquent, cette recherche vise à fournir un modèle pour le diagnostic de la BPCO et la classification des cas en phéno-types à l'aide d'un réseau bayésien

- **Méthodes** : La construction du modèle a été réalisée en développant la structure du réseau bayésien et en instanciant les paramètres pour chaque variable. Les performances du modèle sont validées à l'aide d'une application de réseau neuronal basée sur l'algorithme LevenbergMarquardt.

- **Résultats** : Les résultats montrent qu'un réseau bayésien a réussi à différencier la MPOC de l'asthme et à classer les cas de MPOC en phénotypes. En outre, cette étude vise à déterminer les exigences logicielles pour la conception et le développement d'un dispositif médical pour le diagnostic de la BPCO. Un tel outil peut être utilisé pour la recherche de cas ciblée dans les communautés aux ressources limitées pour l'identification et le traitement précoces de la MPOC.

2.4 Comparaison des travaux

Par la suite nous présentons une comparaison entre les méthodes étudiées, selon les critères suivants :

- **Nature des données** : Dans cette colonne nous donnons la nature des données utilisées au niveau de chaque travail .
- **Formalismes utilisés** : Dans cette colonne, nous présentons les formalismes utilisés, et les langages d'ontologie utilisées.
- **Incertitude** : cette colonne présente le traitement de l'incertitude .
- ne traite pas l'incertitude.
- + traite l'incertitude.

Travaux	Nature des données	Formalisme utilisé	Incertitude
[Delaby et al.2014] [9]	Complètes	OWL	-
[Shen et al.2018][31]	Complètes	Combinaison des ontologie	-
[Lin et al.2011][21]	Complètes	OWL	-
[Seixas et al.2014][30]	Incorporation des connaissance	RB	+
[Farooq et al.2012] [11]	Incomplètes	RB	+
[Nicandro et al.2013] [27]	Incomplètes	Classifieur RB	+
[Olwendo et al. 2016] [28]	Incomplètes	RB	+
[Messaoudi et al.2019] [23]	Complètes	OWL	-

TABLE 2.1 – Tableau comparatif

2.5 Analyse des travaux

Après l'étude des travaux existants, nous constatons que les travaux utilisant les ontologies classiques ne peuvent pas traiter l'incertitude et elles ont que des concepts et des relations précises.

Les réseaux bayésiens sont une technique bien établie pour gérer l'incertitude, ils exploitent le raisonnement probabiliste pour fournir des informations sur les relations de causalités pour un ensemble de variables modélisant un domaine donné. Ils sont crédibles pour l'aide à la décision.

D'après notre étude sur les ontologies traitant le diagnostic médical, on a constaté qu'il n'existe pas d'ontologies probabilistes proposées pour le diagnostic médial et qui traite l'incertitude, cela nous a motivé à construire une ontologie probabiliste, pour la représentation de la base de connaissance et le traitement de l'incertitude en utilisant les réseaux bayésiens multi-entités. Ce qu'il représente le but du troisième chapitre.

2.6 Conclusion

Dans ce chapitre nous avons présenté le diagnostic médical, ainsi que les travaux connexes au diagnostic médical basé sur les ontologies et les réseaux bayésiens, ensuite, nous avons fait une comparaison de ces derniers afin de déduire leurs limites et l'utilité des ontologies probabilistes.

Ce chapitre nous a permet de réaliser l'importance du diagnostic médical en utilisant les ontologies. Dans le chapitre suivant nous allons présenter la conception de notre système ainsi que la motivation de notre choix d'étude de cas.

Chapitre 3

CONCEPTION DU SYSTÈME

3.1 Introduction

L'objectif principal des ontologies probabilistes est d'aider à encadrer l'incertitude. C'est pour ça elles peuvent régler le problème des concepts probabilistes dans plusieurs domaines, parmi-eux le diagnostic médical. En effet, les ontologies probabilistes constituent maintenant un axe de recherche très prometteur, sur lequel nous avons situé notre travail.

Le but de ce chapitre est d'exprimer notre motivation pour le choix de la maladie à traiter, et de présenter les différentes phases du processus de développement de l'ontologie probabiliste.

3.2 Motivation de l'étude de cas

Le 31 décembre 2019 est le jour de l'apparition du COVID-19 pour la 1ère fois dans la région de Wuhan. Le monde ne réalise pas l'importance de ce jour et son impact sur la vie future, jusqu'au 11 mars 2020 l'épidémie de COVID-19 est déclarée pandémie par l'OMS (L'organisation mondiale de la santé), et là, le monde s'arrête.

Le COVID-19 est la crise globale de notre époque et le plus grand défi auquel tout le monde est confronté depuis la deuxième guerre mondiale. Car elle n'est pas qu'une crise sanitaire, mais elle a des impacts sociaux, économiques et politiques dévastateurs qui mettent sous pression les pays qu'elle touche.

Cette pandémie a entraîné des pertes dynamiques de la vie humaine dans le monde entier. Avec ses 2 millions de morts du COVID-19 dans le monde, 2020 restera gravé comme un moment sombre dans l'histoire humaine. Les perturbations économiques et sociales causées par la pandémie sont dévastatrices : des dizaines de millions de personnes risquent de sombrer dans l'extrême pauvreté, tandis que le nombre de personnes sous-alimentées, actuellement estimé à près de 690 millions, pourrait augmenter jusqu'à 132 millions d'ici la fin de l'année (36). Parmi les impacts de COVID-19 sur la vie sociale on trouve :

50% des étudiants qui sont affectés par la fermeture des établissements scolaires.

Plus de 135 millions de personnes seront touchées par la pauvreté d'ici 2030.

Des millions d'entreprises sont confrontées à une menace existentielle. Près de la moitié des 3,3 milliards de travailleurs dans le monde risquent de perdre leurs moyens de subsistance. Les travailleurs de l'économie informelle sont particulièrement vulnérables car la majorité n'a pas de protection sociale et d'accès à des soins de santé de qualité et a perdu l'accès aux actifs productifs. Sans les moyens de gagner un revenu pendant les périodes de confinement, beaucoup sont incapables de se nourrir et de nourrir leur famille. Pour la plupart, pas de revenu signifie pas de nourriture.

Notre application a pour but d'aider les médecins à faire le diagnostic médical à temps, afin de sauver le maximum de vie.

3.3 COVID-19

Les Coronavirus sont des virus issus de la famille des coronaviridae, dont la particularité est de posséder un génome à ARN très long (l'acide ribonucléique permet le transport du message génétique et la synthèse des protéines). Ces germes sont entourés d'une capsule de protéines en forme de couronne, d'où leur nom. Les coronavirus (surnommés CoV) sont une famille de virus plus ou moins sévères : selon le site de l'organisation mondiale de la santé, ils peuvent provoquer de simples rhumes ou des pathologies plus lourdes telles que le syndrome respiratoire du Moyen-Orient (MERS) et le syndrome respiratoire aigu sévère (SRAS).

Aujourd'hui, on en connaît six espèces. Le nouveau coronavirus découvert en 2019 a été baptisé SARS-CoV-2 et la maladie qu'il entraîne, COVID-19. Les coronavirus peuvent être responsables de plusieurs maladies respiratoires et digestives chez plusieurs mammifères. Chez l'humain, ces infections peuvent être :

- Asymptomatiques
- Responsables de pathologies bénignes comme un rhume ou des syndromes grippaux.
- Responsables de complications respiratoires telles que la pneumonie chez les patients immuno-déprimés ou les nourrissons.
- Responsables de syndrome respiratoire grave, entraînant des épidémies.

3.4 Mode de transmission

Ces virus se transmettent par :

- Le mode de transmission principal du coronavirus est celui **des postillons, des gouttelettes de salive** expulsées par une personne positive au virus lorsqu'elle tousse ou elle éternue, chante ou parle. L'OMS explique que "l'on considère qu'un contact étroit avec une personne malade est suffisant pour transmettre la maladie : même lieu de vie, contact direct à moins d'un mètre lors d'une toux, d'un éternuement ou une discussion en l'absence de mesures de protection".

• **Voie aérienne** : D'après l'OMS, il n'y a à ce jour aucune preuve concrète permettant d'affirmer que le coronavirus se transmet par l'air. Mais selon Julian Tang, "l'absence de preuve n'est pas une preuve d'absence". Ce virologue et plus de 230 autres scientifiques internationaux ont appelé "la communauté médicale et les organismes nationaux et internationaux compétents à reconnaître le potentiel de transmission aérienne du COVID-19" Le 21 juillet, des chercheurs américains ont apporté de nouvelles preuves à ce sujet. De l'air a été prélevé des chambres de cinq patients hospitalisés pour une infection à SARS-CoV-2. Des microgouttelettes de moins de cinq microns, voire d'un micron de diamètre contenant du virus ont pu être récoltées. Le virus a pu être répliqué, c'est-à-dire que de nouvelles unités du coronavirus ont pu être produites. "Il est par conséquent infectieux", a affirmé Joshua Santarpia, co-auteur de l'étude.

• **Au contact de sécrétion**

• **Au contact d'objets contaminés**

La période d'incubation, soit l'apparition des premiers symptômes, est de 3 à 6 jours. Et en général, le traitement se concentre sur les symptômes. (35)

3.5 Symptômes

Symptômes cliniques :

• Les Symptômes les plus fréquents :

- **Fever** : La fièvre est l'un des symptômes majeurs de l'existence d'une infection, lorsque la température est supérieure à 38°C le matin et 38.3°C le soir.

- **Asthenia** : La faiblesse générale se caractérise par une diminution du pouvoir de fonctionnement de l'organisme. Le coronavirus peut entraîner une fatigue intense chez les personnes infectées qui peuvent même la ressentir plusieurs semaines après la disparition de l'infection. De plus, le coronavirus entraînerait également un risque de syndrome de fatigue chronique chez les anciens malades.

- **Diarrhea** : La diarrhée aiguë se manifeste par des selles liquides plus de 3 fois par jour. Elle est fréquemment due à une infection intestinale par des virus.

- **Headache** : Les maux de tête peuvent prendre plusieurs formes : céphalée par tension nerveuse, migraines, céphalée de la sinusite et algie vasculaire de la face.

- **Dry cough** : Une toux sèche est une toux sans expectoration (sans émission de sécrétions). Elle est non productive.

Selon l'OMS, les symptômes des formes graves de la COVID-19 sont :

- **Shortness of breath** : L'essoufflement, ou « dyspnée » est une sensation d'inconfort ou de difficulté respiratoire qui survient pour une activité n'entraînant normalement aucune gêne (effort plus ou moins important, repos, position allongée, etc.).

Symptômes biologiques :

- **Blood urea** : L'urée est un produit de la dégradation des protéines qui est éliminé par les urines. Les taux d'urée dans le sang sont donc un reflet de la fonction rénale, et, dans certaines conditions,

de l'apport alimentaire en protéines ainsi que du fonctionnement du foie. Avec une norme de 0.10-0.50 g/l

-Creatinine : La créatine, protéine indispensable pour la production d'énergie par les muscles. En temps normal, la créatinine est transportée par le sang puis éliminée par les reins, dans les urines. L'augmentation de ce taux dans le sang signifie que la fonction rénale est altérée. Avec une norme de $F < 11H < 13$

-Blood Sugar : La glycémie est la concentration de glucose dans le sang, ou plus exactement dans le plasma sanguin. Avec une norme de 0.7 - 1.10.

-D-Dimers : Sont des fragments de fibrine, la principale composante des caillots sanguins. Les D-dimères apparaissent dans le sang, entre autres, quand des caillots se dissolvent. Inférieur 500 microgrammes par litre

-WBC : Sur la numération formule sanguine (NFS), le taux normal de lymphocytes soit être compris entre 1500 et 4000/ mm³ soit entre 20 et 40 du nombre total de globules blancs.

-HGB : L'hémoglobine (Hb)sanguine correspond à la quantité d'hémoglobine contenue dans 100 ml de sang. Elle varie en fonction du sexe et les valeurs normales sont : chez l'homme : 13 à 18 g/dl, chez la femme : 12 à 16 g/dl.

- SpO₂ : La saturation du sang s'exprime en pourcentage et va permettre d'avoir une estimation de l'état d'un patient. La valeur normale est située entre 90% et 100%.

-IgM/IgG : Sont des immunoglobulines produites par le système immunitaire pour assurer une protection contre le SRAS-CoV-2. Les IgM et IgG anti-SARS-CoV-2 peuvent donc être détectées dans les échantillons des patients affectés. Le niveau d'anticorps IgM commence à augmenter environ une semaine après l'infection initiale, tandis que les IgG apparaissent plus tard que les IgM (généralement dans les 14 jours suivant l'infection), ce qui signifie que les IgG servent d'indicateur d'une infection antérieure.

Resultats	Interprétation
IgM+ / IgG+	Infection récente au SARS-CoV-2
IgM+ / IgG-	Infection récente au SARS-CoV-2
IgM- / IgG+	Infection antérieure au SARS-CoV-2
IgM- / IgG-	Pas d'infection ou pas d'anticorps détectables pendant le début de l'infection

FIGURE 3.1 – Interprétation du résultats IgM/IgG (35)

- **CRP :** C'est le dosage de la protéine C. Les valeurs normaux du taux de CRP se situent sous la barre de 6 mg / l de sang.

- **TP :** La mesure du taux de prothrombine est un examen biologique qui permet d'évaluer l'efficacité de la coagulation du sang au sein de l'organisme. Exprimé en pourcentage, il est obtenu en mesurant le temps de Quick. Cette mesure évalue la vitesse de coagulation de sang,

avec une norme de $>70\%$

- **INR** : (International Normalized Ratio) est un indicateur de la coagulation sanguine. Il se mesure en comparant le temps de coagulation du patient à celui d'un patient témoin "international". Avec une norme entre 0,8 et 1,2.

Symptômes radiologiques :

La TDM thoracique sans injection en coupes fines est l'examen de référence : images en verre dépoli, bilatérales, prédominant en périphérie dans les lobes inférieurs avec possible pleurésie et lymphadénopathies ; syndrome interstitiel.

3.6 Facteurs de risque

Le Haut Comité de Santé Publique a identifié les personnes à risque de développer une forme grave d'infection à SARS-CoV-2 :

- Les personnes âgées de 70 ans et plus (même si les patients entre 50 ans et 70 ans doivent être surveillés de façon plus rapprochée).
- Les patients présentant une insuffisance rénale chronique dialysée.
- Les patients aux antécédents (ATCD) cardiovasculaires : hypertension artérielle compliquée, ATCD d'accident vasculaire cérébral ou de coronaropathie, chirurgie cardiaque, insuffisance cardiaque stade NYHA III ou IV.
- Les diabétiques insulino-dépendants non équilibrés ou présentant des complications secondaires à leur pathologie.
- HTA : L'hypertension artérielle (HTA) persistante est une pathologie cardiovasculaire définie par une pression artérielle trop élevée.
- Les personnes présentant une obésité morbide (indice de masse corporelle $> 40\text{kg/m}^2$).
- Maladie chronique des poumons.

3.7 Data set

Le diagnostic de la pandémie du COVID-19 est un effort multidisciplinaire qui nécessite des données importantes. Notre dataset est composée de 300 cas réels provenant des malades de l'hôpital Franz-fanoun (unité COVID-19). Les attributs des données se composent des facteurs de risque, ensemble de résultat de tests radiologiques, test sanguin, des symptômes et des signes cliniques des malades. L'inexistence d'un dataset de ce genre nous a poussé à le construire nous-même.

La création de notre dataset a commencé après avoir eu l'autorisation du directeur des activités médicaux et paramédicaux de l'hôpital pour accéder au service COVID-19 en plein pandémie, le processus de collecte des données nécessite l'aide d'un expert du domaine puis le traitement des dossiers des malades.

Cela nous a pris énormément de temps et de déplacement à l'hôpital pour pouvoir trier les dossiers dans l'archive selon notre besoin. Après avoir discuté avec l'expert du domaine, nous avons préparé une fiche à remplir pour chaque dossier, qui contient les informations nécessaires à fournir pour notre dataset.

Premièrement les symptômes cliniques du patient, les résultats biologiques et radiologiques ainsi que les facteurs de risque.

Le traitement des archives était délicat, beaucoup de dossiers manquaient d'informations, le médecin nous a guidé vers les données les plus indispensables pour pouvoir créer un dataset fiable avec les cas positifs et négatifs de COVID-19.

Après la collecte, nous sommes passés à la création et le remplissage du dataset avec les informations des patients en précisant pour chaque patient les données et le résultat de son cas.

Informations sur les attributs de notre dataset :

- ▷ Age : in years.
- ▷ Sex : f/h.
- ▷ Fever(yes/no).
- ▷ Asthenia(yes/no).
- ▷ Diarrhea(yes/no).
- ▷ Dry cough(yes/no).
- ▷ Headache(yes/no).
- ▷ Shortness of breath(yes/no).
- ▷ Blood urea : in g/l.
- ▷ Creatinine .
- ▷ Blood sugar in g/l.
- ▷ D-dimers in microgramme.
- ▷ WBC in mm³.
- ▷ HGB in g/dl.
- ▷ Spo2 in %.
- ▷ IgM .
- ▷ IgG .
- ▷ CRP in mg/l .
- ▷ TP in %.
- ▷ INR .
- ▷ TDM .
- ▷ Has Cardiovascular history(yes/no).
- ▷ Diabetes(yes/no).
- ▷ Chronic lung disease(yes/no).
- ▷ Renal failure(yes/no).
- ▷ Morbid obesity(yes/no).
- ▷ High blood pressure(yes/no).

3.8 Schéma global

Notre travail sert à réaliser un système de prédiction sur la maladie COVID-19 en utilisant une ontologie probabiliste pour l'aide au diagnostic médical. Le processus de construction du système passe par deux phases, une phase de construction de l'ontologie probabiliste et une phase d'utilisation.

Nous présentons notre schéma dans la figure suivante.

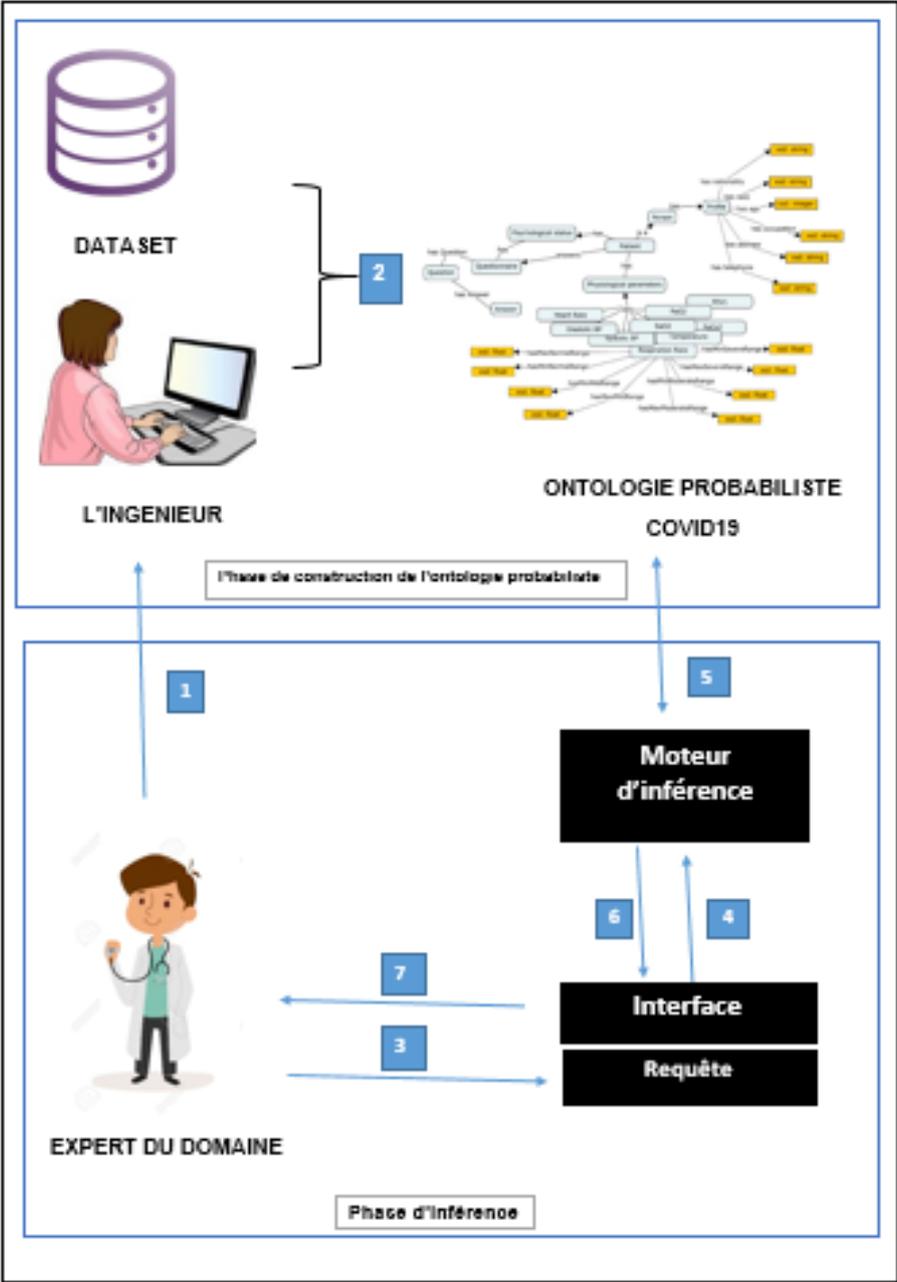


FIGURE 3.2 – Schéma global

Description du schéma :

La réalisation de notre système passe par :

1. Le médecin fournit les informations nécessaires à l'ingénieur concernant la maladie.
2. L'ingénieur utilise le data-set et les informations fournies par le médecin pour la création d'une ontologie probabiliste COVID19.
3. Le médecin remplit la fiche avec les informations du malade et lance la requête.
4. En utilisant les données de la fiche remplie par le médecin l'inférence est lancée.
5. Pour réaliser l'inférence on utilise l'ontologie COVID19.
6. Le résultat de la requête faite par le médecin est reçu sous forme de probabilité.
7. Le médecin reçoit le résultat du diagnostic.

3.9 Création de l'ontologie probabiliste

Bien que les Ontologies Probabilistes (OP) soient l'une des meilleures approches pour modéliser l'incertitude, le processus s'est avéré extrêmement difficile, ce qui a motivé la création du processus de modélisation d'incertitude pour les technologies sémantiques (UMP-ST), un processus qui guide les utilisateurs dans la modélisation des OP.

Pour la création de notre ontologie probabiliste, nous utilisons la méthodologie UMP-ST afin de pouvoir la construire. Cette méthode divise la conception d'une ontologie probabiliste en quatre phases :

3.9.1 Les Exigences (requirement)

Dans cette phase UMP-ST doit définir les objectifs à attendre dans la conception, et cela en posant des questions que le modèle doit savoir répondre. Pour chaque objectif (goals), il est nécessaire de définir des requêtes qui sauront répondre à l'aide des preuves.[1]

3.9.1.1 Les besoins(goals)

Dans notre cas l'ensemble des objectifs, requêtes et preuves qui doivent être traités concerne notre domaine afin de diagnostiquer un malade par la COVID-19 sont :

- **Objectif** : identifier si le patient est atteint par la COVID-19 ?
- **Requête 1** : Le patient a-t-il les symptômes du COVID-19 ?
- **Preuve** :
 - Le patient a une fièvre.
 - Le patient a une Diarrhée.

- Le patient a des maux de tête.
- Le patient a une toux sèche.
- Le patient a une l'asthénie.
- Le patient a un essoufflement.
- Le patient a des troubles dans le taux de la glycémie.
- Le patient a des troubles dans le taux de l' urée.
- Le patient a des troubles dans le taux du créatininémie.
- Le patient a des troubles dans le D-Demères.
- Le patient a des troubles dans le taux de WBC.
- Le patient a des troubles dans le taux de HGB.
- Le patient a une dégradation dans le taux de SpO2.
- Le patient a des troubles dans le taux de I.N.R /TP.
- Le patient a des troubles dans le taux de la CRP.
- Le patient a des troubles dans le taux de IgM/IgG.
- Le patient a des troubles dans le taux de la TDM.

- **Requête 2** : Le patient a-t-il les facteurs de risque du COVID-19 ?

- **Preuve** :

- Le patient est âgé.
- Le patient est diabétique.
- Le patient est cardiovasculaire.
- Le patient présente une obésité morbide.
- Le patient a fait une insuffisance rénale chronique dialysée.
- Le patient présente une maladie HTA.
- Le patient a une chronique des poumons.

3.9.2 Analyse et conception

Cette phase décrit les classes d'entités, leurs attributs, leurs relations et les règles qui leurs sont applicables dans notre domaine. Cette définition est indépendante du langage utilisé pour implémenter le modèle. [4]

3.9.2.1 Les entités

ce sont les classes de l'ontologie

- Patient.
- Risk factors.
- Symptom.

- Cardio.
- Serology.
- F.N.S.

- Propriétés des entités :

Dans cette partie nous allons mentionner les différents attributs et les relations entre leurs entités.

Attribut	Domaine	Range
Age	Patient	Int
Sex	Patient	String

TABLE 3.1 – Entités et leurs relations

Attribut	Domaine	Range
Diabetes	Risk factors	Boolean
Cardiovascular history	Risk factors	Boolean
Chronic lung disease	Risk factors	Boolean
Renal failure	Risk factors	Boolean
Morbid obesity	Risk factors	Boolean
High blood pressure	Risk factors	Boolean

TABLE 3.2 – Les attributs de l’entité Risk factors

Attribut	Domaine	Range
Fever	symptom	Boolean
Dry cough	symptom	Boolean
Headache	symptom	Boolean
Asthenia	symptom	Boolean
Diarrhea	symptom	Boolean
Shortness of breath	symptom	Boolean
Blood urea	symptom	Float
Blood sugar	symptom	Float
Creatinine	symptom	Float
D-dimers	symptom	Float
Crp	symptom	Float
SpO2	symptom	Float
TDM	symptom	Float

TABLE 3.3 – Les attributs de l’entité Symptom

Attribut	Domaine	Range
IgG	Serology	Float
IgM	Serology	Float

TABLE 3.4 – Les attributs de l’entité Serology

Attribut	Domaine	Range
TP	Cardio	Float
INR	Cardio	Float

TABLE 3.5 – Les attributs de l’entité Cardio

Attribut	Domaine	Range
WBC	FNS	Float
HGB	FNS	Float

TABLE 3.6 – Les attributs de l’entité FNS

Attribut	Domaine	Range
Has risk factors	Patient	Risk factors
Has symptom	Patient	Symptom
Has symptomS	Patient	Serology
Has symptomf	Patient	FNS
Has symptomc	Patient	Cardio

TABLE 3.7 – Les relations entre entités

3.9.2.2 Les règles

Afin d’utiliser les réseaux bayésiens multi-entité (MEBN) qui sont la base du langage PR-OWL nous devons avoir des données de type nominal.

I. Catégorisation des attributs

Le tableau présente les intervalles pour chaque attribut.

Il est nécessaire de déterminer l’ensemble des intervalles pour la création du MEBN.

Attribut	Intervalle	Valeurs
Age	$[0 - 50]$	Young
	$50 <$	Old
Blood urea	< 0.1	Low
	$[0.1 - 0.5]$	Normal
	$0.5 <$	High
Blood Sugar	< 0.7	Low
	$[0.7 - 1.1]$	Normal
	$1.1 <$	High
Creatinine	$F < 11$	normalF
	$H < 13$	NormalH
	$F > 11$	HighF
	$H > 13$	HighH
TP	> 70	Normal
	< 70	Low
D-dimers	< 500	Normal
	> 500	High
SpO2	< 90	Low
	$[90 - 100]$	Normal
	$100 <$	High
IgG	< 1	Normal
	> 1	High
IgM	< 1	Normal
	> 1	High
INR	< 0.8	Low
	$[0.8 - 1.2]$	Normal
	$1.2 <$	High
CRP	< 6	Normal
	> 6	High
WBC	< 1500	Low
	$[1500 - 14000]$	Normal
	$14000 <$	High
HGB	$[13 - 18]$	Normalh
	$[12 - 13]$	Normalf
	< 13	Lowh
	< 12	Lowf
	> 18	Highh
	> 16	Highf

TABLE 3.8 – Catégorisation des attributs.

II. Construction de la structure de réseau bayésien

La figure montre la structure de notre réseau bayésien construit après la discussion avec l'expert du domaine.

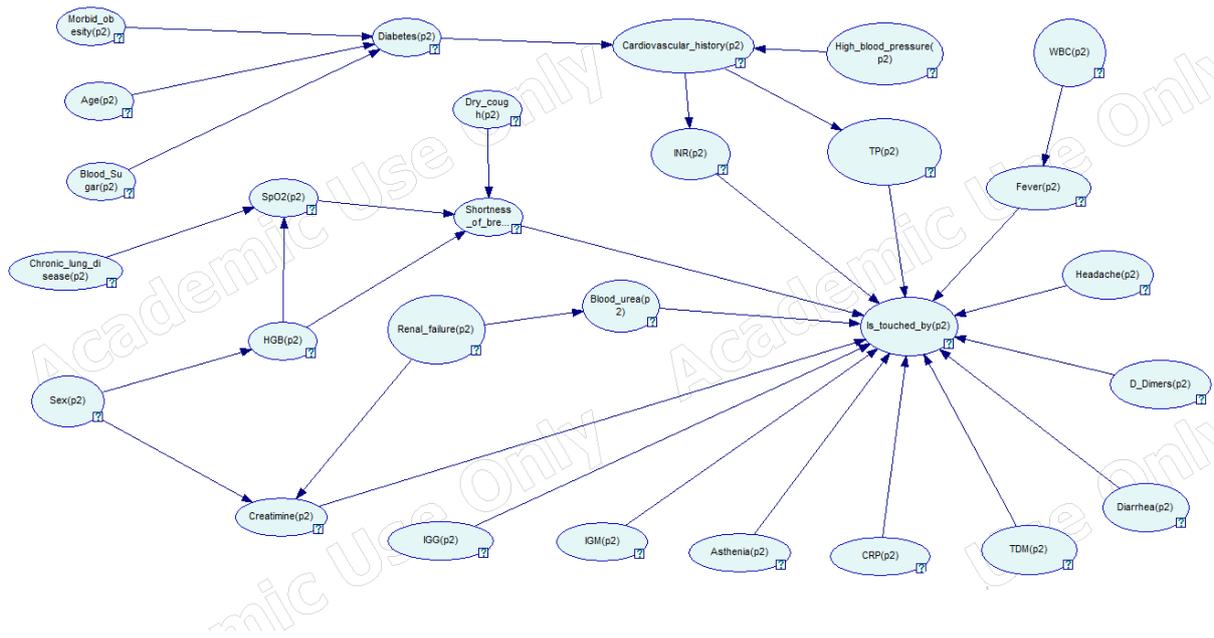


FIGURE 3.3 – La structure du réseau bayésien

III. Application de l’algorithme EM et le calcul de probabilité

L’application de l’algorithme EM sur notre dataset, nous a donné une estimation maximale sur les valeurs manquantes du dataset , pour pouvoir calculer la probabilité sur chaque noeud du RB.

La figure montre les résultats de probabilité de chaque noeud.

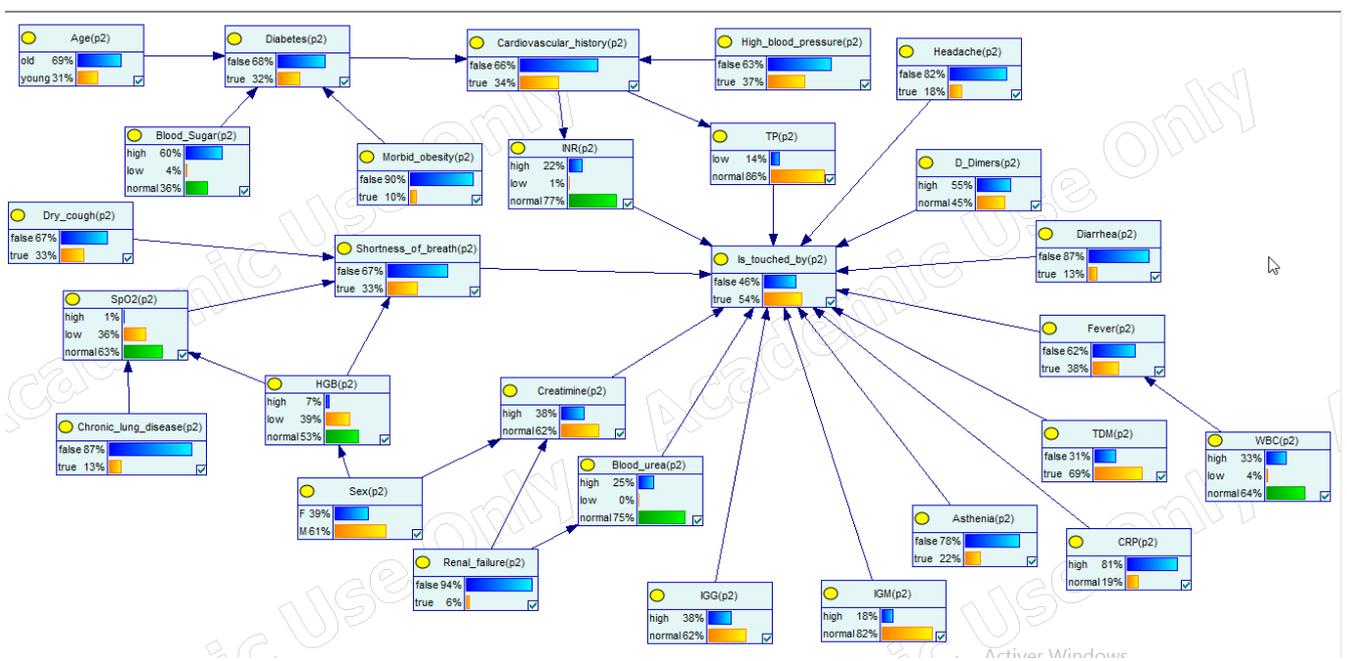


FIGURE 3.4 – Réseau bayésien

3.9.2.3 Les groupes

Cette partie sert à identifier les entités principales, leurs attributs et les relations de notre ontologie probabiliste COVID-19.

Diagramme de classe

Le diagramme de classes sert à représenter les besoins de notre système afin de pouvoir faire la construction de notre ontologie probabiliste.

La figure montre le diagramme de classe de notre système.

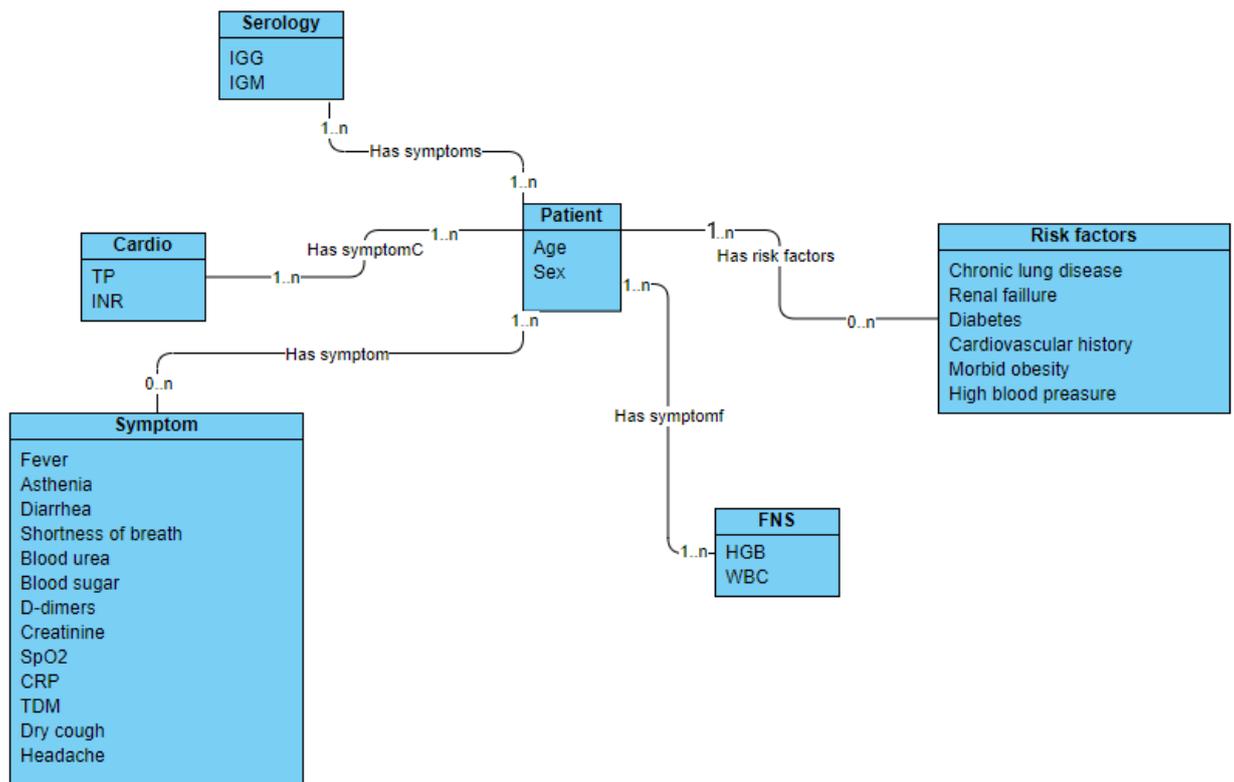


FIGURE 3.5 – Diagramme de Classe

3.9.3 Implémentation

Après avoir terminé la partie analyse et conception, nous allons commencer l'implémentation de notre modèle par un langage spécifique qui permet de raisonner sur les connaissances d'incertitude. Cette partie décrit comment modéliser l'ontologie probabiliste COVID-19 dans PR-OWL.

3.9.3.1 Mapping

Pour générer automatiquement un modèle dans PR-OWL qui nous permet de faire le raisonnement probabiliste sur des connaissances incertaines ; nous avons commencé par mapper les entités, leurs attributs et leurs relations ,un modèle de probabilité est défini comme **Mfrags** , un ensemble de Mfrags constitue la théorie de MEBN **Mtheory** .

Mtheory :

Elle est composée d'un ensemble de Mfrag satisfait des contraintes de cohérence qui garantissent l'existence d'une distribution conjointe unique sur ses variables aléatoires.

La création du Mtheory est la première étape pour construire notre ontologie COVID-19, qui va contenir les Mfrags que notre ontologie aura besoin.

Mfrags :

Les Mfrag représentent des informations de probabilité sur un groupe de variables aléatoires. Après la construction des Mtheory , nous avons commencé à créer les MFrag qui se composent de nœuds résidents, de nœuds d'entrées et de nœuds de contexte.

La figure ci-dessous montre les MFrag de notre domaine.

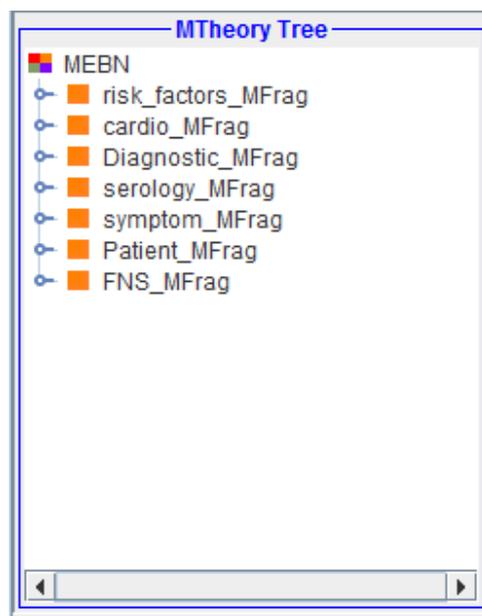


FIGURE 3.6 – MFrag de notre domaine

Chaque MFrag se compose de :

- **Variables ordinaires :**

Une variable ordinaire est déclarée avec la relation 'est-un' (IsA), c'est une instantiation d'une entité (classe OWL). Par exemple : IsA (p, Patient) tel que P est notre variable ordinaire et Patient est l'entité (la classe OWL).

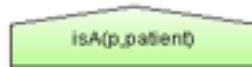


FIGURE 3.7 – Variable ordinaire P

- **ResidentNode (noeuds résidents) :**

Ce sont les variables aléatoires réelles qui forment le coeur du sujet d'un MFrag. La logique MEBN exige que la distribution probabiliste locale de chaque noeud résident doit être unique et explicitement défini dans son MFrag original. Les valeurs possibles d'un noeud résident peuvent être une entité existante. Voici un exemple de noeud résident dans le MFrag "Symptom Mfrag".[19]



FIGURE 3.8 – Noeud résident "blood sugar"

- **ContextNode (noeud de contexte) :**

Ce sont des variables aléatoires booléennes représentant des conditions qui doivent être satisfaites pour faire une distribution dans un MFrag valide. Les noeuds de contexte peuvent représenter plusieurs types de modèles d'incertitude sophistiqués, tel que l'incertitude sur les relations entre les entités. .[19]



FIGURE 3.9 – Noeud de contexte "IsA()"

- **InputNode (noeud d'entrée) :**

Ces noeuds sont essentiellement des « pointeurs » faisant référence aux noeuds résidents d'un autre MFrag. Les noeuds d'entrée fournissent également un mécanisme permettant la réutilisation des noeuds résidents entre les MFrag. Les noeuds d'entrée influencent la distribution de

probabilité des noeuds résidents qui sont leurs enfants dans un MFrage donné, mais leurs propres distributions sont définies ailleurs (c'est à dire dans leurs propres MFrags d'origine). Par exemple le noeud "HGB" dans le Mfrage "diagnostic", fait référence au noeud résident du Mfrage "FNS". [19].



FIGURE 3.10 – Noeud d'entrée "HGB"

Les relations existantes dans un réseau bayésien :

Nous distinguant deux types de noeuds dans une relation de causalité qui vont être inclus dans la structure du réseau :

- Relation de causalité entre les noeuds d'entrées avec des noeuds résidents.
- Relation de causalité entre les noeuds résidents avec d'autres noeuds résidents.[19]

Après la définition des différents composants du Mfrage, nous allons traiter chaque Mfrage de notre domaine en détail

- Mfrage de Patient-Mfrage contient les différents attributs de patient selon les besoins dans notre domaine :

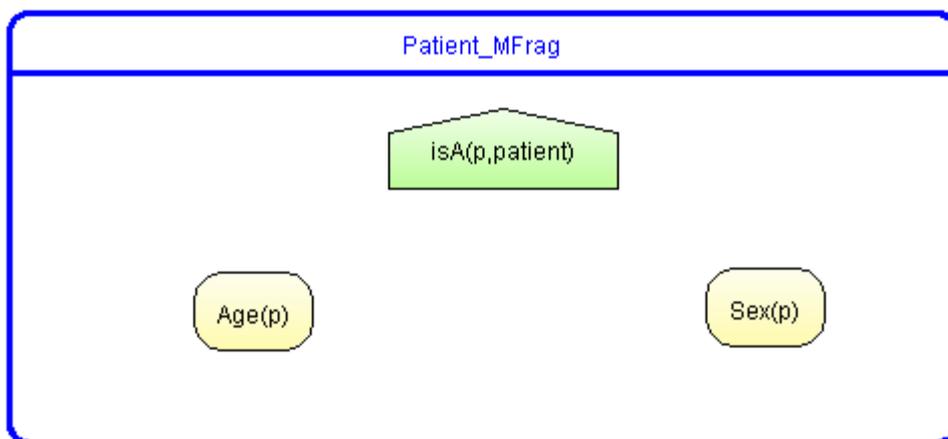


FIGURE 3.11 – Le Mfrage Patient

- Le Mfrage risk-factors contient les différents facteurs de risque de la maladie COVID-19 :

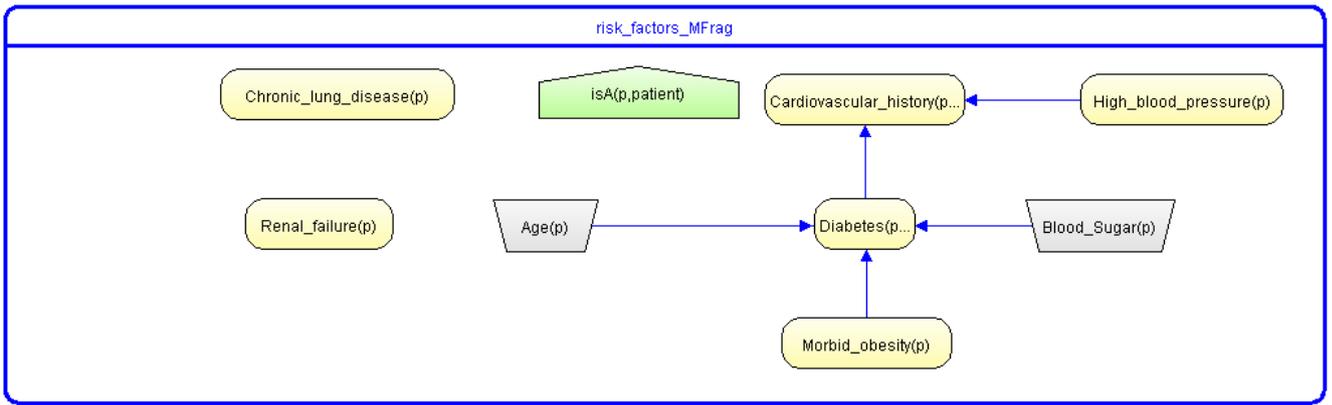


FIGURE 3.12 – Le MFrag Risk-factors

- Le Mfrags symptom contient les différents symptômes cliniques de notre domaine :

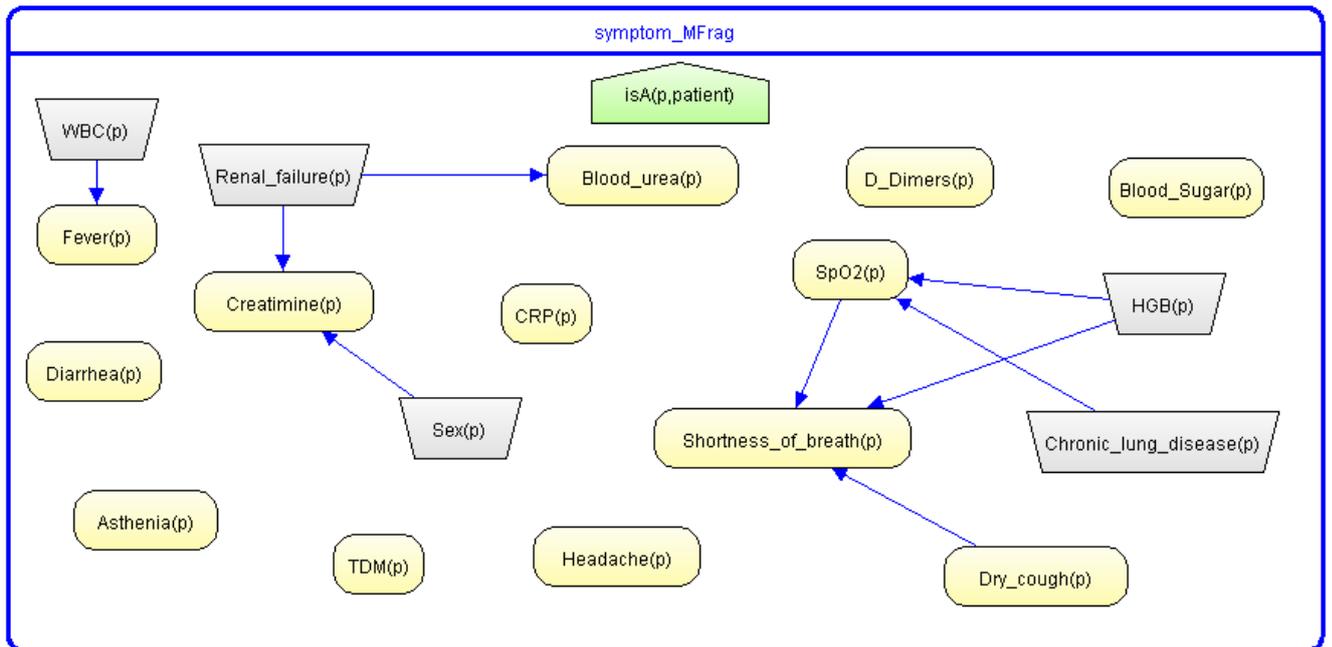


FIGURE 3.13 – Le MFrag Symptom

- Le MFrag Cardio est le MFrag qui représente les attributs des symptômes cardiaques :

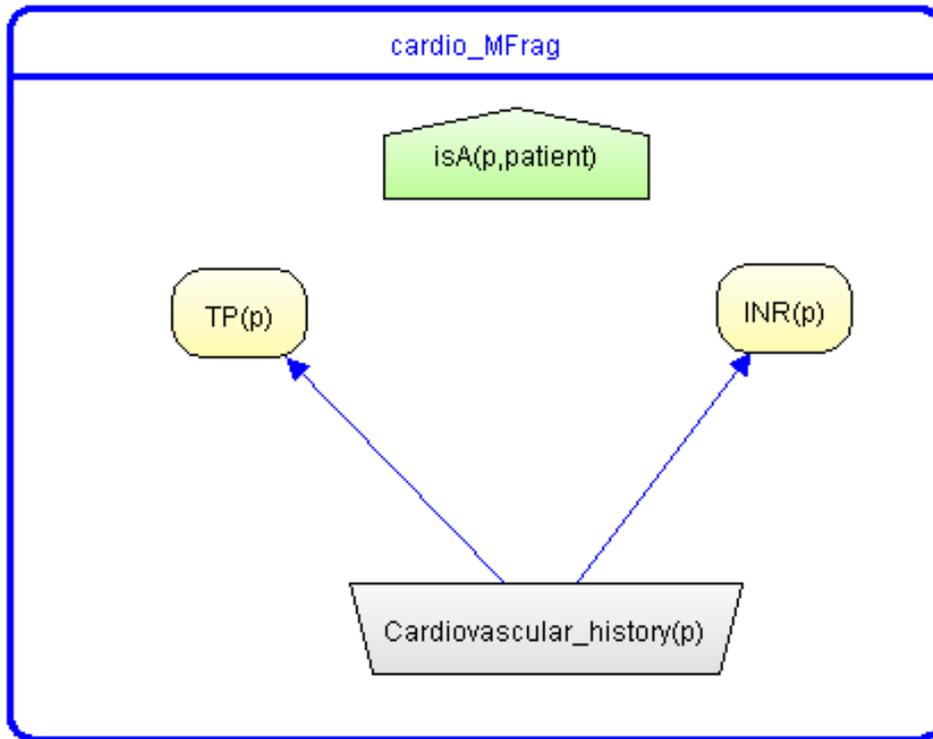


FIGURE 3.14 – Le MFrag Cardio

- Le MFrag FNS contient les symptômes du test FNS :

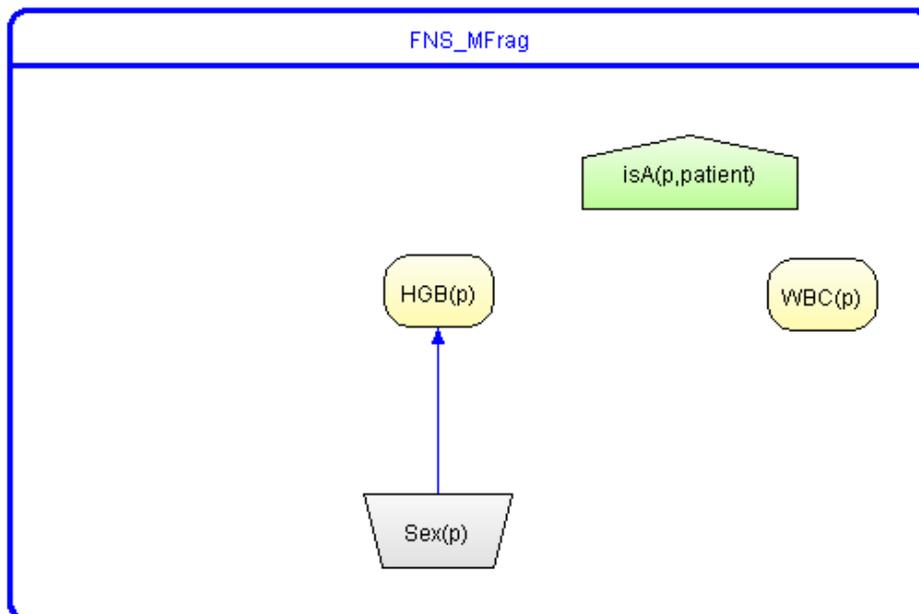


FIGURE 3.15 – Le MFrag FNS

- Le MFrag Serology contient les symptômes du test sérologie :

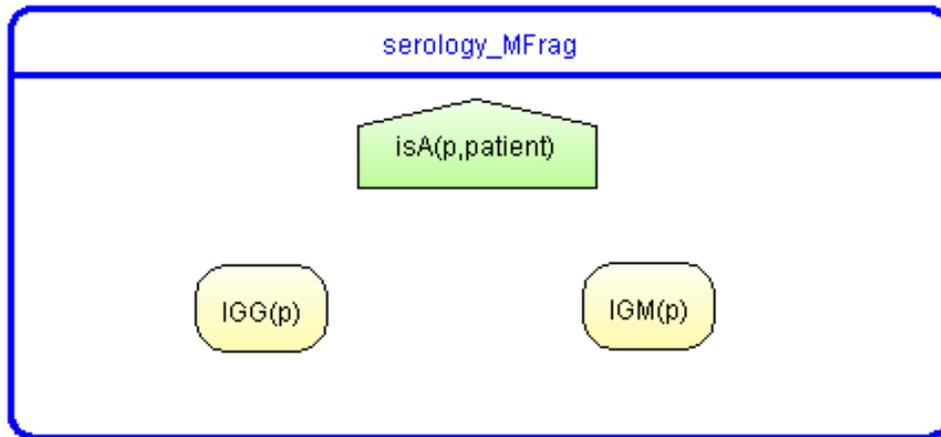


FIGURE 3.16 – Le MFrag des serology.

- MFrag de Diagnostic-MFrag contient tous les symptômes et les facteurs de risque concernant le COVID-19 :

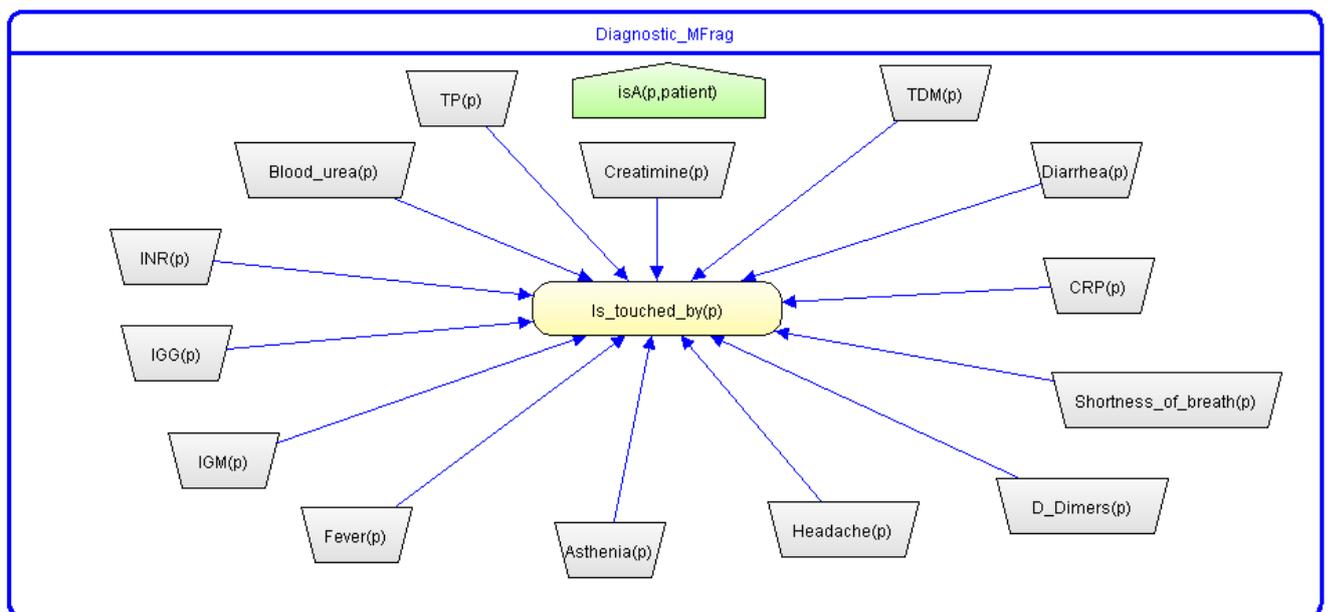


FIGURE 3.17 – Le MFrag Diagnostic

3.9.3.2 Probabilités

La dernière phase de la construction de l'ontologie probabiliste dans UNbbyaes consiste à déterminer la distribution de probabilité locale (LPD) de toutes les variables aléatoires résidentes (VR) à partir des règles définies au préalable.

Chaque MFrag définit les LPD pour ses nœuds peuplés.

Ces distributions sont régies par la satisfaction de contraintes par la VR contextuelle.

Les distributions locales peuvent dépendre des valeurs de VR dont la distribution est définie dans

les MFrag dans lesquels elles sont résident.

La figure suivante montre le LPD de nœud résident IsTouchedBy. Cette distribution suit la grammaire UnBBayes-MEBN pour définir les LPD.

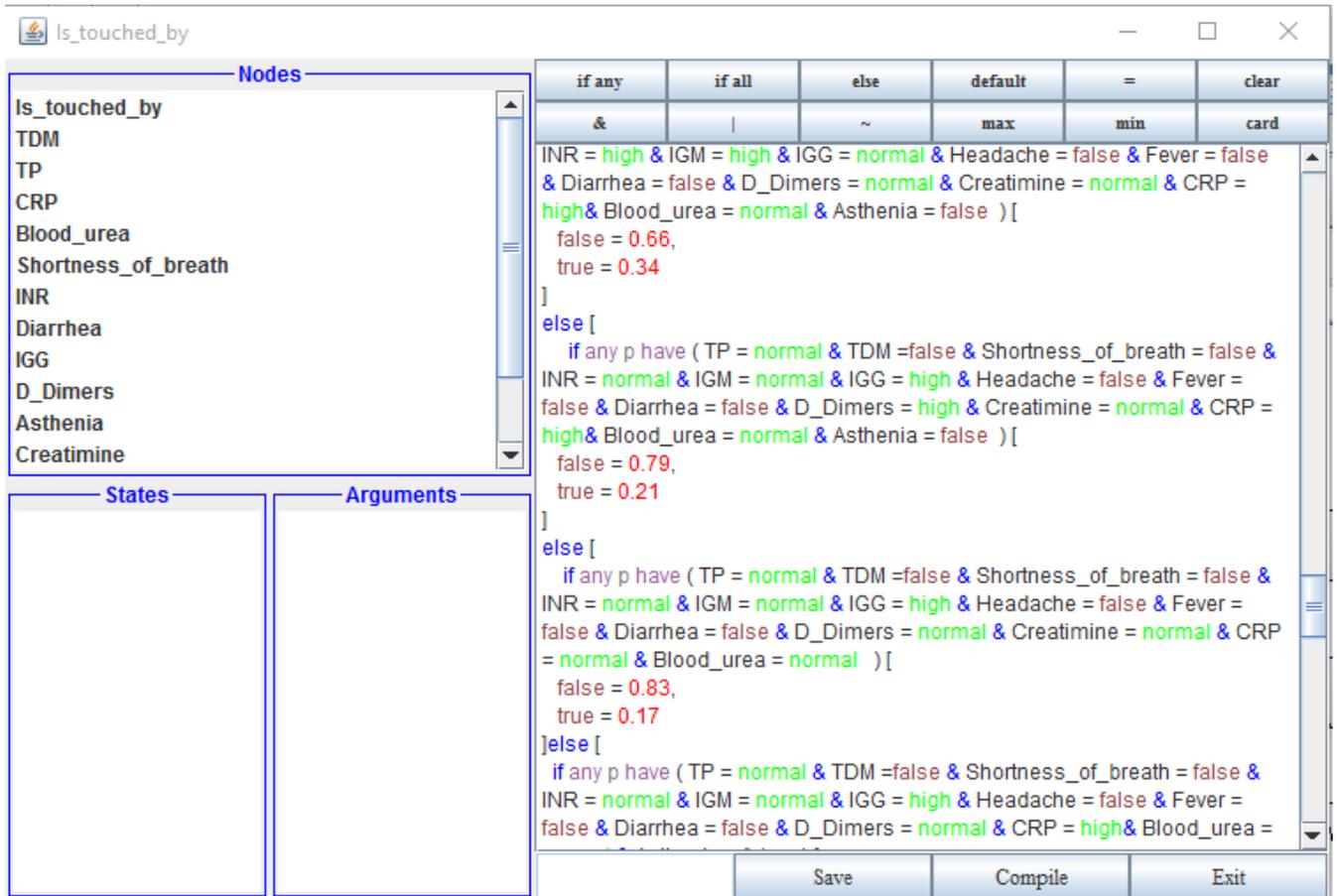


FIGURE 3.18 – Le LPD pour le nœud ISTOUCHEDBY(patient)

Le résultat de ce processus est un réseau bayésien spécifique à la situation (SSBN), qui est un réseau bayésien minimal suffisant pour obtenir la distribution postérieure pour un ensemble d'instances de variables aléatoires cibles, compte tenu d'un ensemble d'occurrences de variables aléatoires.

La Figure suivante présente les résultats d'une seule combinaison d'états pour les VR patient sans préciser ses évidences.

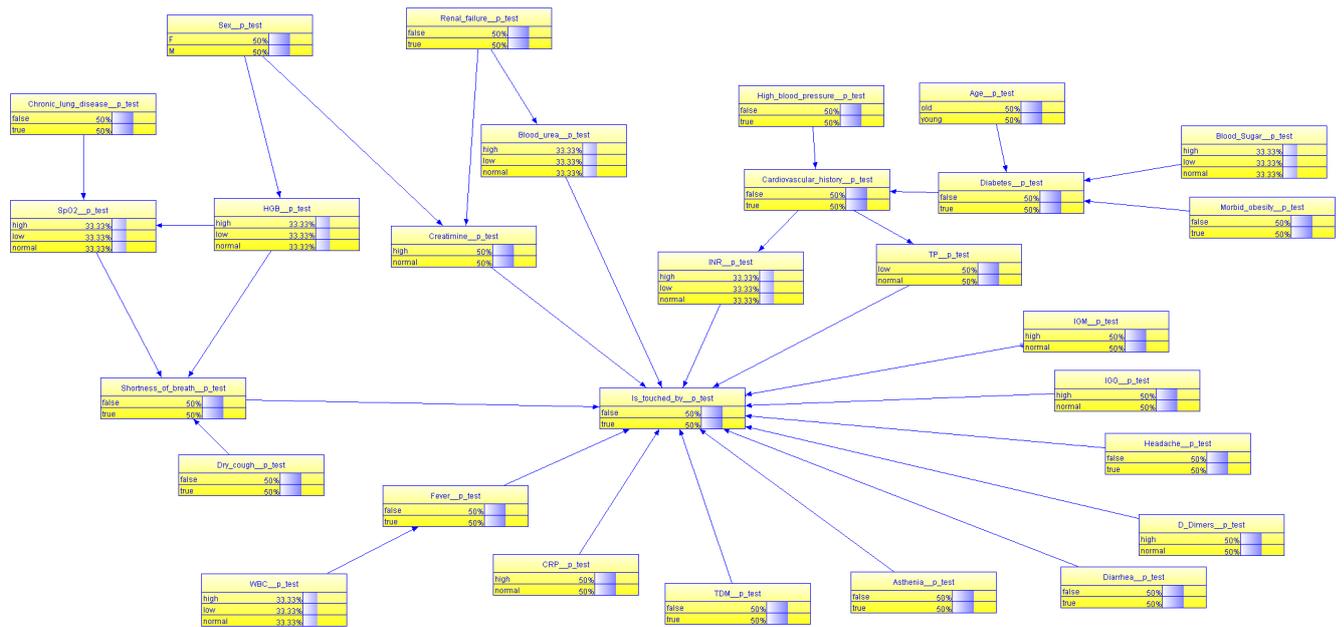


FIGURE 3.19 – Le SSBN pour le nœud IsTouchedBy (patient)

3.9.4 Test

Le but essentiel de la partie de test est d'estimer le bon fonctionnement de notre système et de trouver les défauts afin de pouvoir les améliorer.

Selon [5], il existe deux types d'évaluation à effectuer **la vérification et la validation**.

D'une part, la vérification concerne si notre système a été construit correctement?, elle concerne la fourniture de toutes les fonctionnalités promises au client, cela implique généralement l'examen des exigences, on va vérifier si l'ontologie développée permet de répondre à nos besoins, à travers un scénario réel.

D'autre part, la validation concerne le bon comportement du système. La validation est le test réel du système et il est effectué après la vérification [5].

3.9.4.1 Évaluation

Nous nous concentrerons sur l'évaluation basée sur des cas, en identifiant différents scénarios pour tester notre modèle.

Nous voulons vérifier comment la variable centrée se comporte avec un ensemble de données distinct. Dans le cas de PR-OWL, nous pouvons analyser le comportement des variables aléatoires d'intérêt fournies par MFRag, cela est nécessaire pour saisir la cohérence locale du modèle.

• Vérification

La vérification concerne l'établissement de «le système a été construit correctement», c'est-à-dire que les éléments du système sont conformes à leurs spécifications de performance définies.[5]

• Validation

La validation vise à établir que le «bon système a été construit», c'est-à-dire qu'il réalise son utilisation prévue dans son environnement opérationnel.[5]

Lors de la définition d'un scénario, il faut définir l'hypothèse à tester et le résultat attendu et les preuves a utilisées.

Pour commencer nous avons crée les instances de la classe Patient.

Scénario 1

1. Hypothèse à tester

IstouchedBy(patient)

2. Résultat attendu

Le patient n'est pas touché par la COVID-19.

3. Évidence

SYMPTÔMES :

- Le patient n'a pas une Diarrhée.
- Le patient n'a pas des maux de tête.
- Le patient n'a pas une l'asthénie.
- Le patient a un taux de D-Deméres normal.
- Le patient a un taux de CRP normal.
- Le patient a un taux de WBC normal.
- Le patient a un taux de IgM/IgG normal.
- Le patient a une TDM normal.

FACTEURS DE RISQUE :

- Le patient n'est pas âgé.

Après la création des instances, nous allons charger les données du patient décrites au premier test dans la base de connaissances, afin de générer le SSBN .

La Figure suivante montre le résultat obtenu :

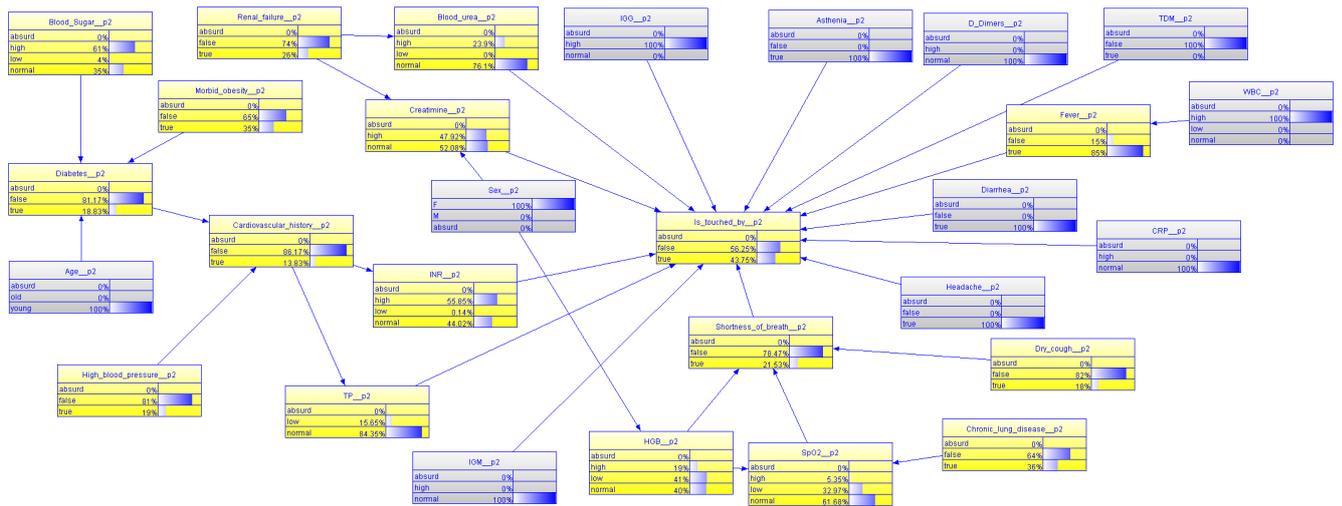


FIGURE 3.20 – SSBN du test 1

Le résultat du test est comme suit :

Le patient n'a pas le risque d'avoir la maladie d'après les pourcentages **FALSE = 56,25%**, **TRUE = 43,75%**

Scénario 2

1. Hypothèse à tester

IstouchedBy(patient)

2. Résultat attendu

Le patient est touché par la COVID-19.

3. Évidence

SYMPTÔMES :

- Le patient a des troubles dans le D-Demères.
- Le patient a des troubles dans le taux de WBC.
- Le patient a des troubles dans le taux de CRP.
- Le patient a des troubles de TDM .
- Le patient a des troubles dans le taux de IgG/IgM .

FACTEURS DE RISQUE :

- Le patient a fait une insuffisance rénale chronique dialysée.
- Le patient a une maladie des poumons.

Après la création des instances, nous allons charger les données du patient décrites au deuxième test dans la base de connaissances, afin de générer le SSBN .

La Figure suivante montre le résultat obtenu :

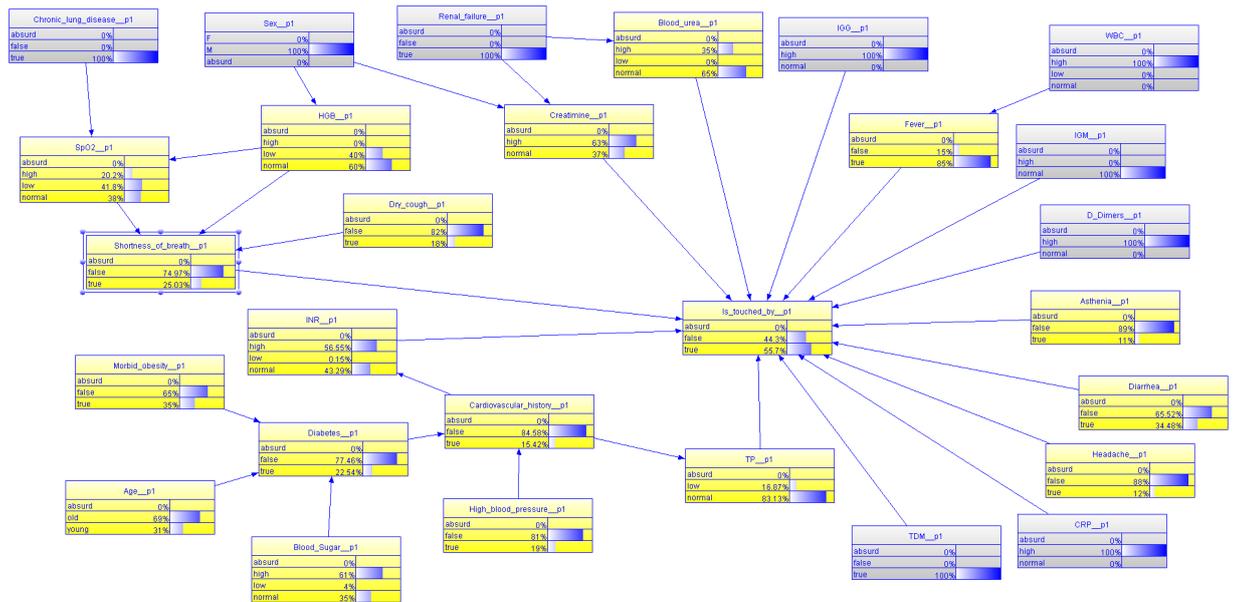


FIGURE 3.21 – SSBN du test 2

Le résultat du test est comme suit :

Le patient a le risque d’avoir la maladie d’après les pourcentages **FALSE = 44,3%**, **TRUE = 55,7%**

3.10 Résultat final de l’ontologie probabiliste

Nous avons obtenu une ontologie probabiliste qui porte le MTheory avec ses différents MFragS en suivant les étapes de la méthodologie UMP-ST.

Ci dessus la hiérarchie des classes de l’ontologie classique (OWL) et de l’ontologie probabiliste (PR-OWL).

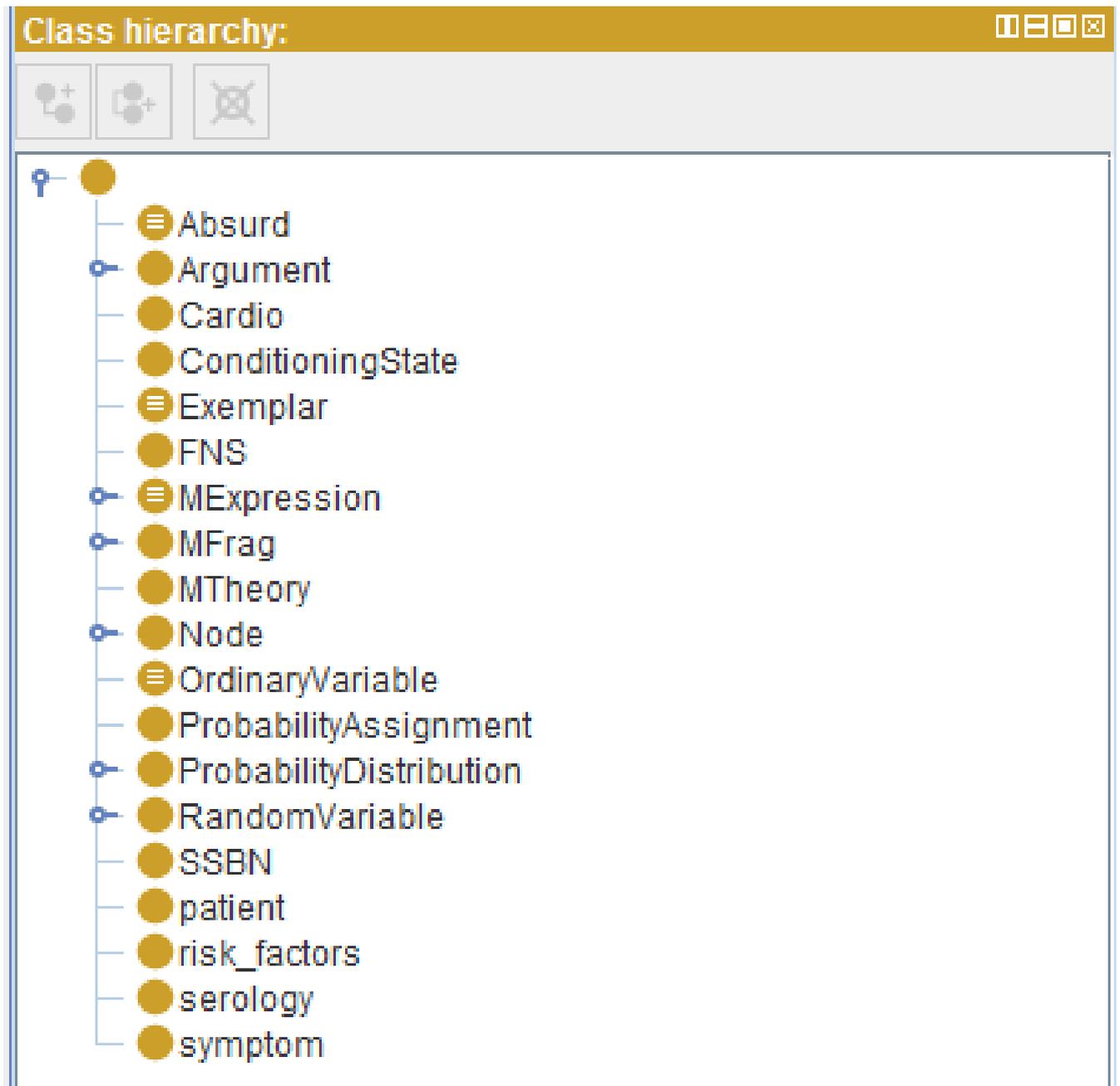


FIGURE 3.22 – La hiérarchie des classes de l'ontologie

La hiérarchie des propriétés de classes :

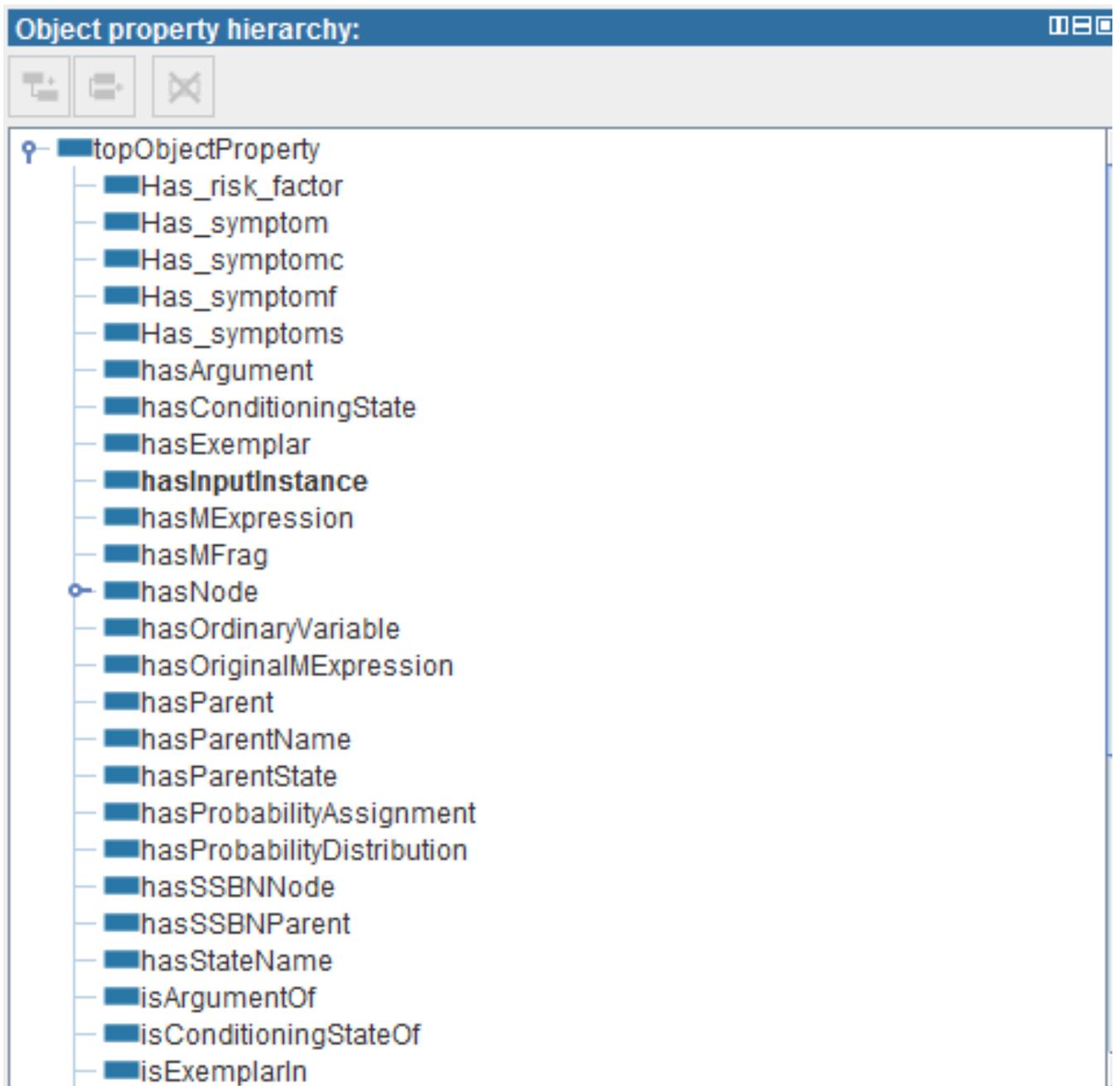


FIGURE 3.23 – La hiérarchie des propriétés d’objet de l’ontologie

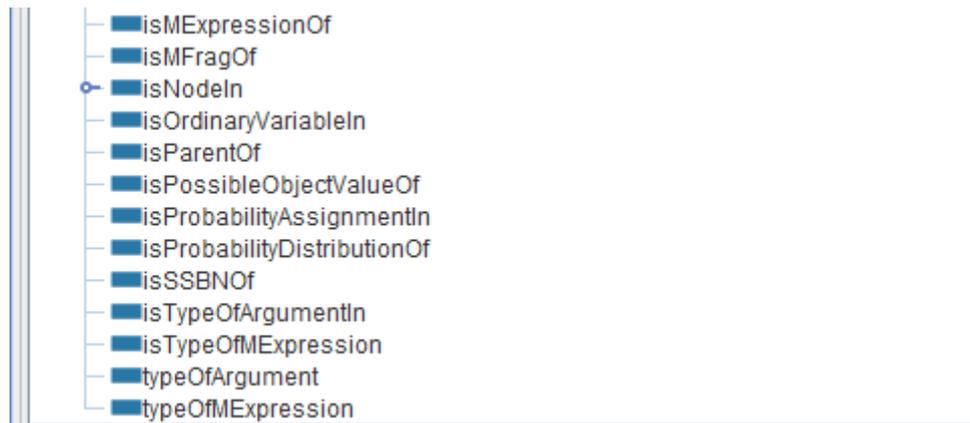


FIGURE 3.24 – La hiérarchie des propriétés d’objet de l’ontologie

La hiérarchie des propriétés des données.

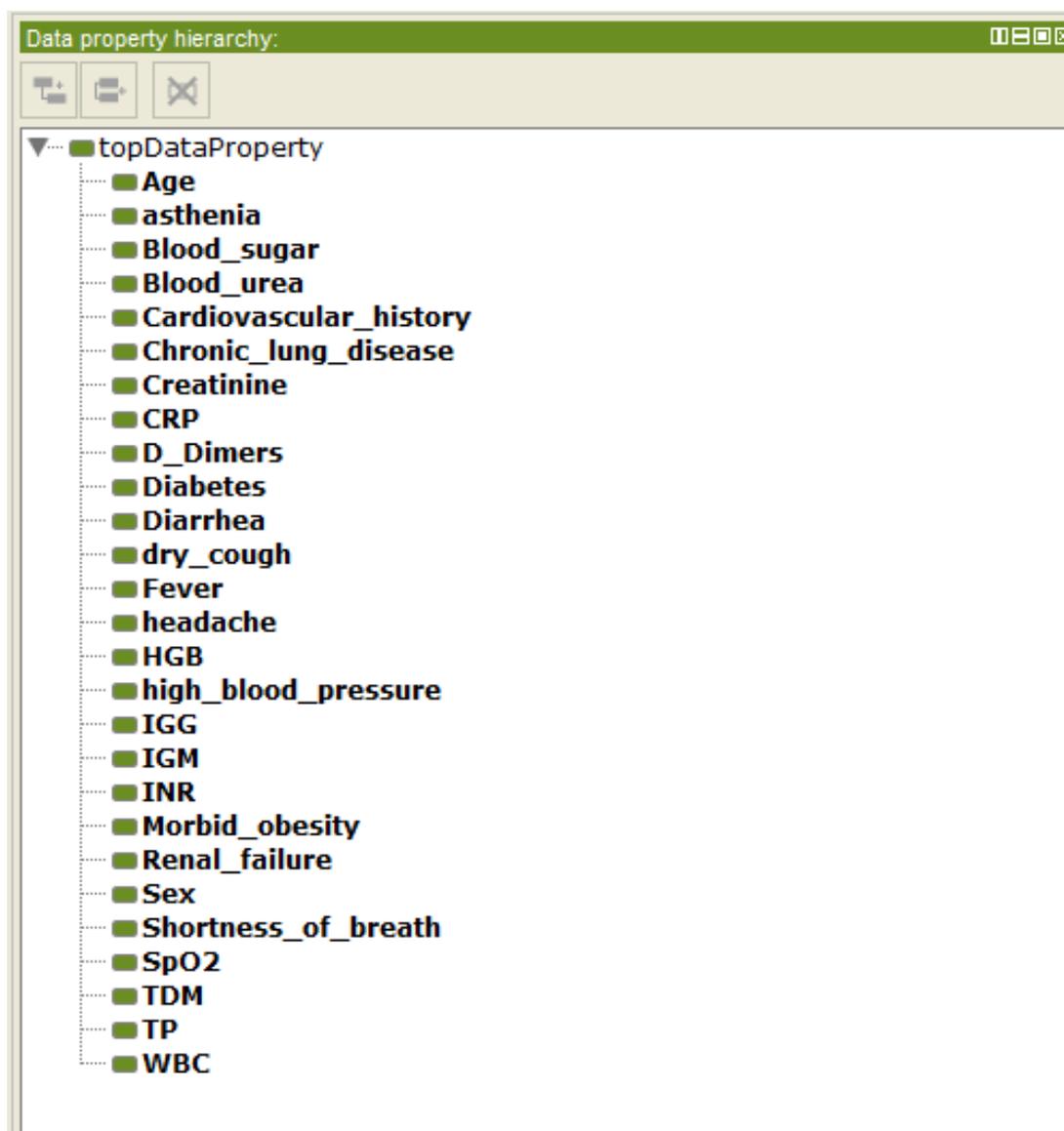


FIGURE 3.25 – La hiérarchie des propriétés des données de l’ontologie classique

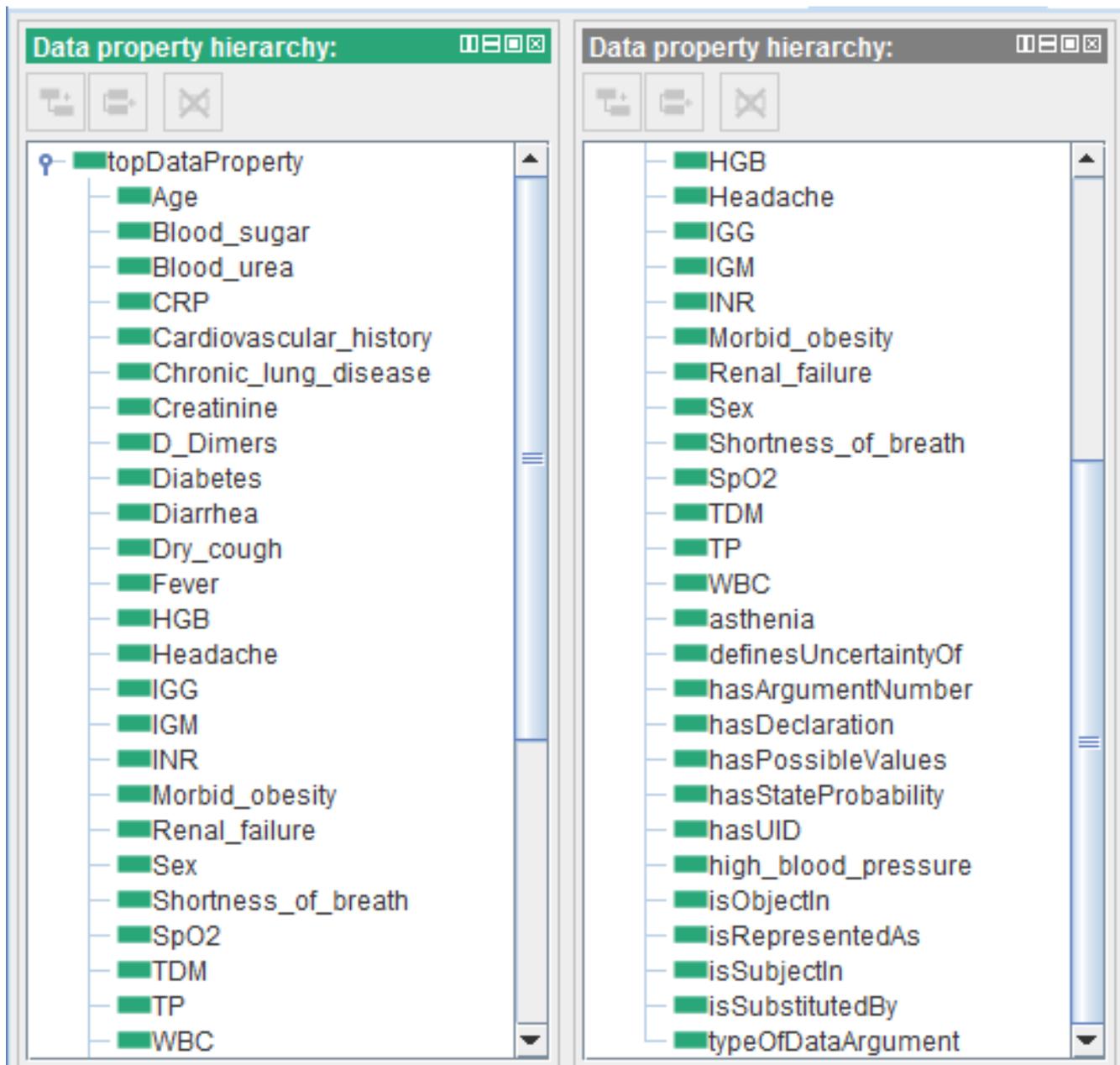


FIGURE 3.26 – La hiérarchie des propriétés des données de l'ontologie probabiliste

3.11 Conclusion

Dans ce chapitre nous avons introduit notre motivation pour la maladie COVID-19 ainsi que le processus de construction de notre dataset et de l'ontologie probabiliste en suivant les étapes de la méthodologie UMP-ST.

Dans le chapitre suivant nous passerons à la réalisation et l'implémentation de notre système de prédiction pour aider les medecins à diagnostiquer la maladie COVID-19.

Chapitre 4

IMPLEMENTATION ET TEST DU SYSTÈME

4.1 Introduction

Après avoir passer par toutes les étapes de la méthodologie UMP-ST, nous allons commencer l'implémentation de notre système qui permet de faire l'inférence, en utilisant UnBBayes. La démonstration visuelle de l'implémentation de notre système va être présenter à travers des captures d'écran des interfaces du système.

4.2 Outils de développement de notre système

Voici la liste des outils de développement que nous avons utilisé pour notre système.

1. Protégé

C'est un Éditeur et framework d'ontologies gratuit et open source pour la création des ontologies.

Il a été créé à l'université Stanford et est très populaire dans le domaine du Web sémantique et au niveau de la recherche en informatique.

Protégé peut lire et sauvegarder des ontologies dans la plupart des formats d'ontologies : RDF, RDFS, OWL.

Protégé est basé sur Java, est extensible et fournit un environnement plug-and-play qui en fait une base flexible pour le prototypage rapide et le développement d'applications.

Permet de créer des ontologies dans les frameworks Frames et Web Ontology Language (OWL) et d'importer, modifier et enregistrer des ontologies existantes écrites en OWL ou RDF (Resource Definition Framework).

Il permet aussi de créer de nouvelles ontologies. Enregistrer les ontologies dans plusieurs formats, y compris les expressions XML de RDF et OWL.

Visualiser les ontologies sous forme graphique, montrant les relations fonctionnelles entre les classes.

Remplir les ontologies avec des instances concrètes de classes. Exécuter des raisonneurs qui peuvent effectuer des inférences sur une ontologie.

L'outil Protégé possède une interface utilisateur graphique (GUI) lui permettant de manipuler aisément tout les éléments d'une ontologie : classe, propriété, instance, . . . etc. Protégé peut être utilisé dans n'importe quel domaine où les concepts peuvent être modélisés en une hiérarchie des classes.

2. **Protege-OWL API**

Est une bibliothèque Java open source pour le langage d'ontologie Web (OWL) et RDF(S). L'API fournit des classes et des méthodes pour charger et enregistrer des fichiers OWL, pour interroger et manipuler des modèles de données OWL et pour effectuer un raisonnement basé sur les moteurs Description Logic. De plus, l'API est optimisée pour la mise en œuvre d'interfaces utilisateur graphiques.

L'API est conçue pour être utilisée dans deux contextes :

- Pour le développement de composants qui sont exécutés dans l'interface utilisateur de l'éditeur Protege-OWL.
- Pour le développement d'applications autonomes (par exemple, des applications Swing, des servlets ou des plug-ins Eclipse)

3. **OWL API**

L'API OWL est une API Java et une implémentation de référence pour créer, manipuler et sérialiser des ontologies OWL. La dernière version de l'API est orientée vers OWL 2.

OWL-API s'agit d'un ensemble d'interfaces riches en fonctionnalité, permettant une manipulation flexible d'ontologies. De nombreux projets utilisent OWL-API comme outil de développement.(38)

4. **Eclipse et Java**

Eclipse est un projet, décliné et organisé en un ensemble de sous-projets de développements logiciels, de la fondation Eclipse visant à développer un environnement de production de logiciels libre qui soit extensible, universel et polyvalent, en s'appuyant principalement sur Java.

Il offre de nombreuses fonctionnalités pratiques. Ainsi qu'un environnement «Java Development Kit » JDK.

5. **Unbbayes**

UnBBayes est un Framework de réseau probabiliste écrit en Java. Il a à la fois une interface

graphique et une API avec inférence, échantillonnage, apprentissage et évaluation. Il supporte BN, ID, MSBN, OOBN, HBN, MEBN / PR-OWL, structure, paramètre et apprentissage progressif.(39)

6. GeNIE smile

GeNIE est un environnement de développement polyvalent et convivial pour les modèles graphiques décisionnels.

Le progiciel GeNIE (Graphical Network Interface) permet de créer intuitivement des modèles de « théorie de la décision » à l'aide de l'interface graphique click-and-drop.

GeNIE est l'interface graphique de SMILE, un « moteur d'inférence bayésien » entièrement portable développé par le Decision Systems Laboratory et testé de manière approfondie sur le terrain depuis 1998.

GeNIE 2.0 est la dernière version de GeNIE.

SMILE (Structural Modeling, Inference, and Learning Engine) est une bibliothèque portable entièrement indépendante de la plate-forme de classes C++ mettant en œuvre des méthodes graphiques de théorie de la décision, telles que les « réseaux bayésiens », les diagrammes d'influence et les modèles d'équations structurelles, directement susceptibles d'être inclus dans des systèmes intelligents.

4.3 La mise en oeuvre du système

Après avoir effectué les étapes de la création de l'ontologie probabiliste en suivant la méthodologie UMP-ST, nous allons utiliser un fichier avec une extension UBF qui porte la MTheory avec ses différents MFrag du domaine ainsi que les différents Nœuds résidents de chaque MFrag , qui va générer automatiquement un fichier PLM , qui représente notre base de connaissance et qui contient le nœud « Avoir risque », avec une nouvelle prédiction d'un nouveau patient, qui va être stocker dans la base de connaissance et sert à prédire la décision.

4.4 Les interfaces

Nous allons présenter les interfaces de notre application.

4.4.1 Interface d'accueil

La première interface à afficher se présente dans la figure 4.1, elle contient les actions principales de notre application.

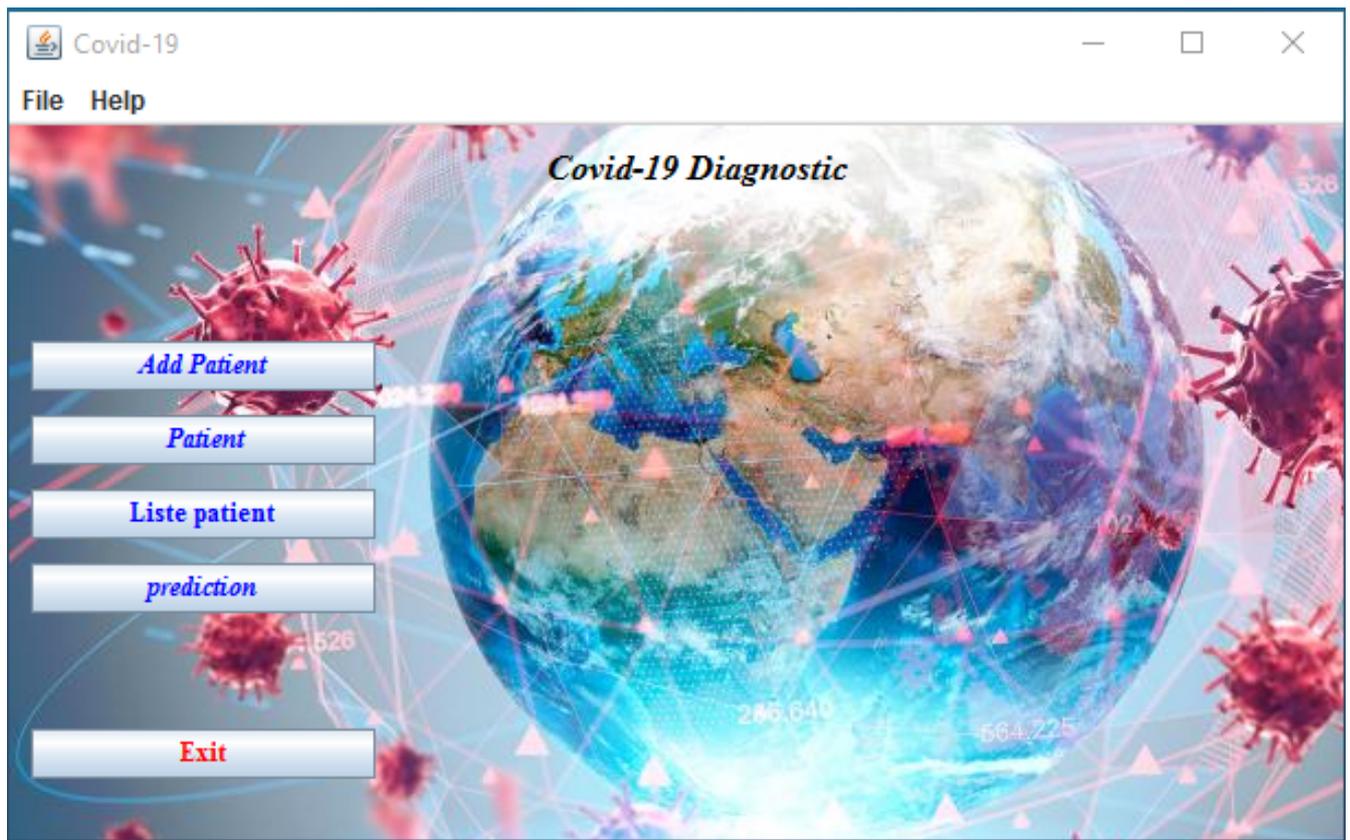


FIGURE 4.1 – Interface d'accueil

Cette interface permet d'effectuer plusieurs fonctionnalités :

- *Ajouter un nouveau patient en introduisant ses informations.
- *Consulter la liste des patients.
- *Prédire le diagnostic d'un patient.

4.4.2 Add patient

Permet d'ajouter un nouveau patient en remplissant ses informations.

The screenshot displays a software window titled "Add Patient". The form contains the following fields and options:

- Name:** Input field containing "bouroubi".
- Age:** Input field containing "52".
- Sex:** Dropdown menu set to "Man".
- Symptom:** A list of symptoms with checkboxes:
 - Asthenia
 - Diarrhea
 - Dry_Cough
 - TDM
 - Headache
 - Fever
 - Shortness_of_breath
- Risk_factors:** A list of risk factors with checkboxes:
 - Chronic_lung_disease
 - High_blood_pressure
 - Renal_failure
 - Morbid_obesity
 - Diabetes
 - Cardiovascular_history
- Other fields:** Blood_Sugar (1.5), D_Dimers (700), CRP, IGG (1.5), IGM (0.6), HGB, TP, INR, Creatimine (12.2), and Blood_urea.

At the bottom right, there are two yellow buttons: "Add" and "Back". A "Message" dialog box is overlaid in the center, displaying an information icon and the text "The patient bouroubi is registered", with an "OK" button below it.

FIGURE 4.2 – Interface d'ajout

4.4.3 Delete patient

Permet de supprimer un patient de notre base de connaissance.

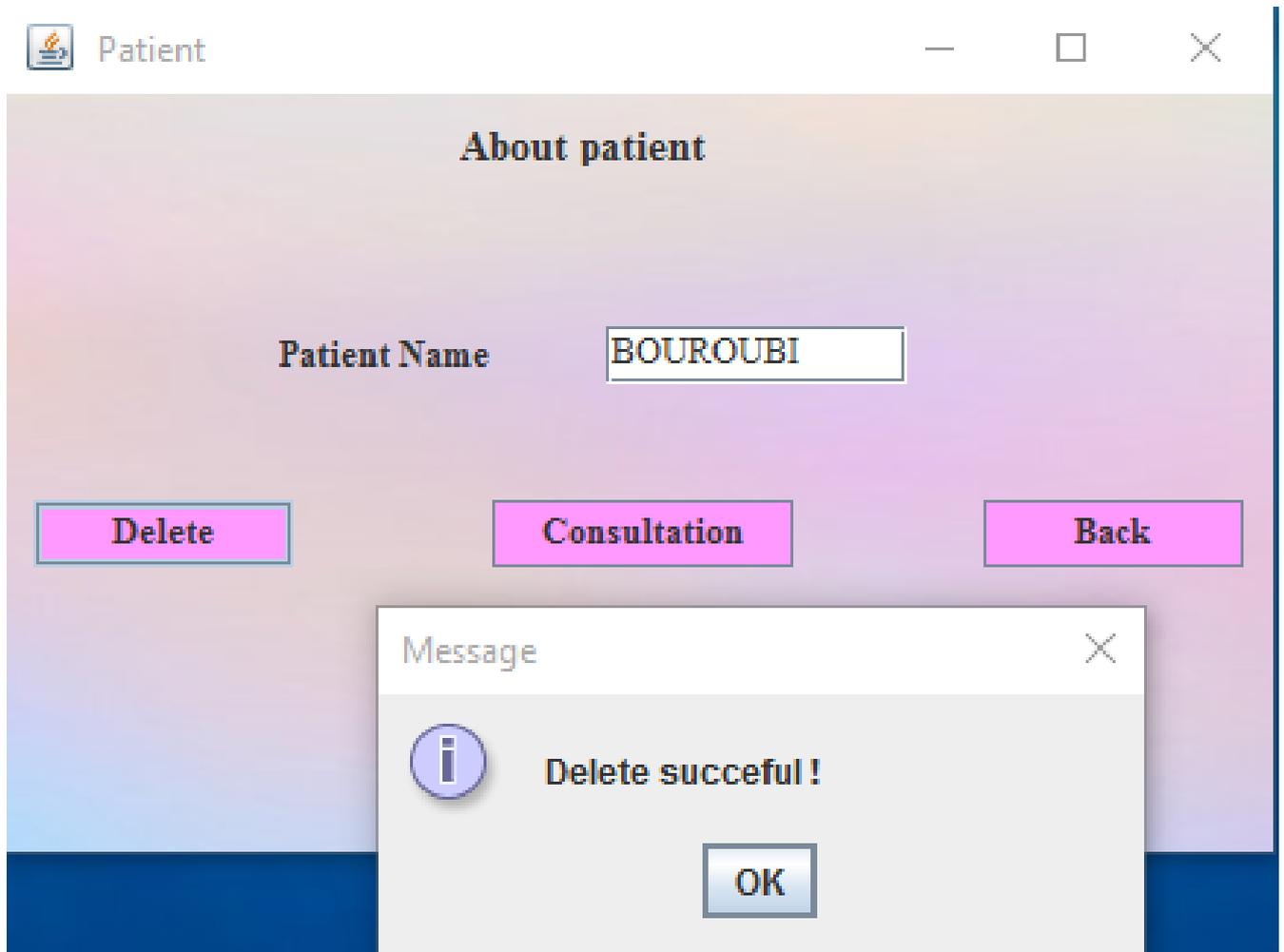


FIGURE 4.3 – Interface de suppression

4.4.4 Consultation

Permet de consulter les informations concernant un patient.

Consult BOUROUBI Information			
Symptom		Risk factors	
Blood_Sugar	High	TDM	True
D_Dimers	High	Creatimine	Normal
CRP		Spo2	Low
IGG	High	Blood_urea	
IGM	Normal	diarrhea	False
Headache	False	Dry_Cough	False
TP		Asthenia	True
INR		Fever	True
HGB		Shortness_of_breath	False
WBC	Normal		
Age	old	Sex	Homme

Back

FIGURE 4.4 – Interface de consultation

4.4.5 Liste des patients

Permet d'afficher tous les patients (par leur nom) existants dans la base de connaissances.

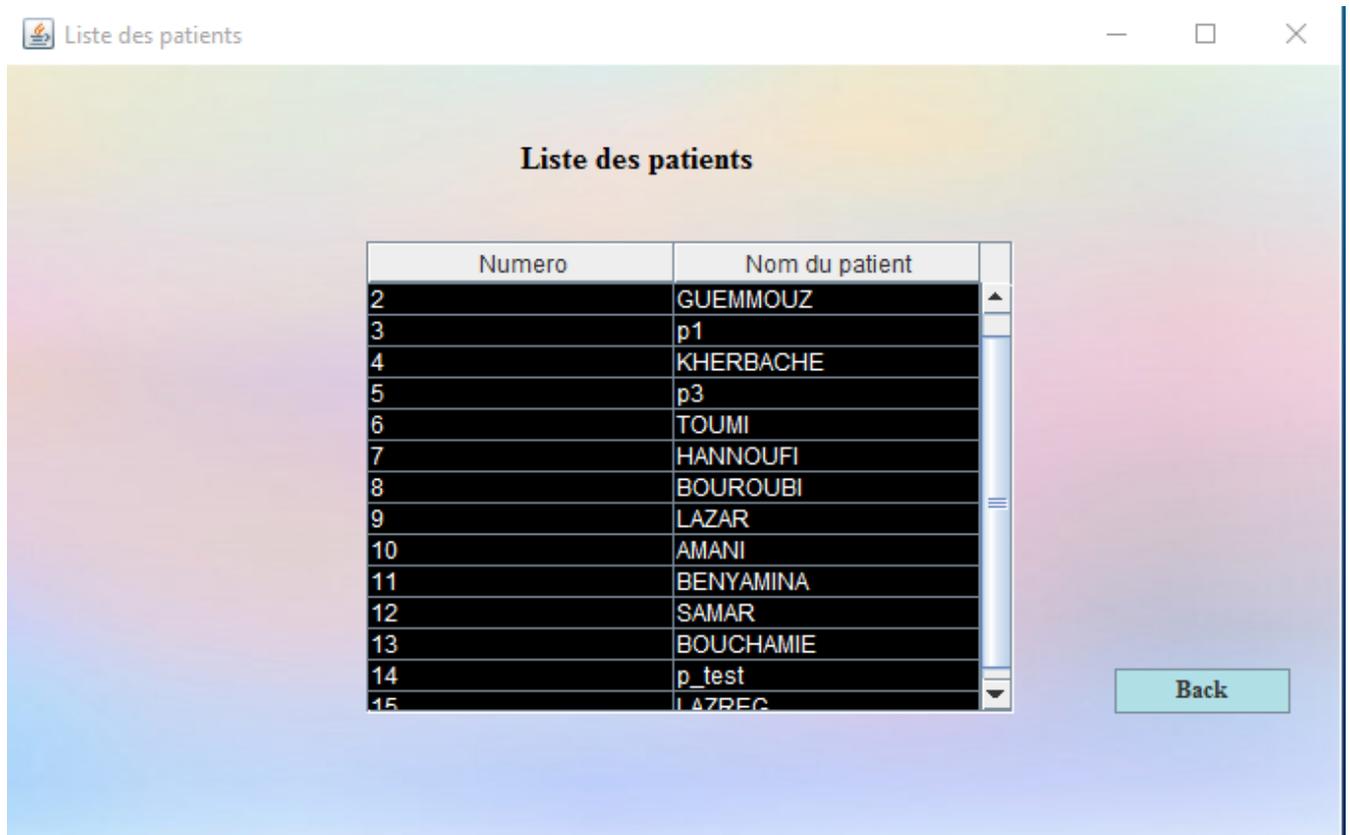


FIGURE 4.5 – Interface liste des patients

4.4.6 Prédiction

Permet de prédire le diagnostic d'un patient.

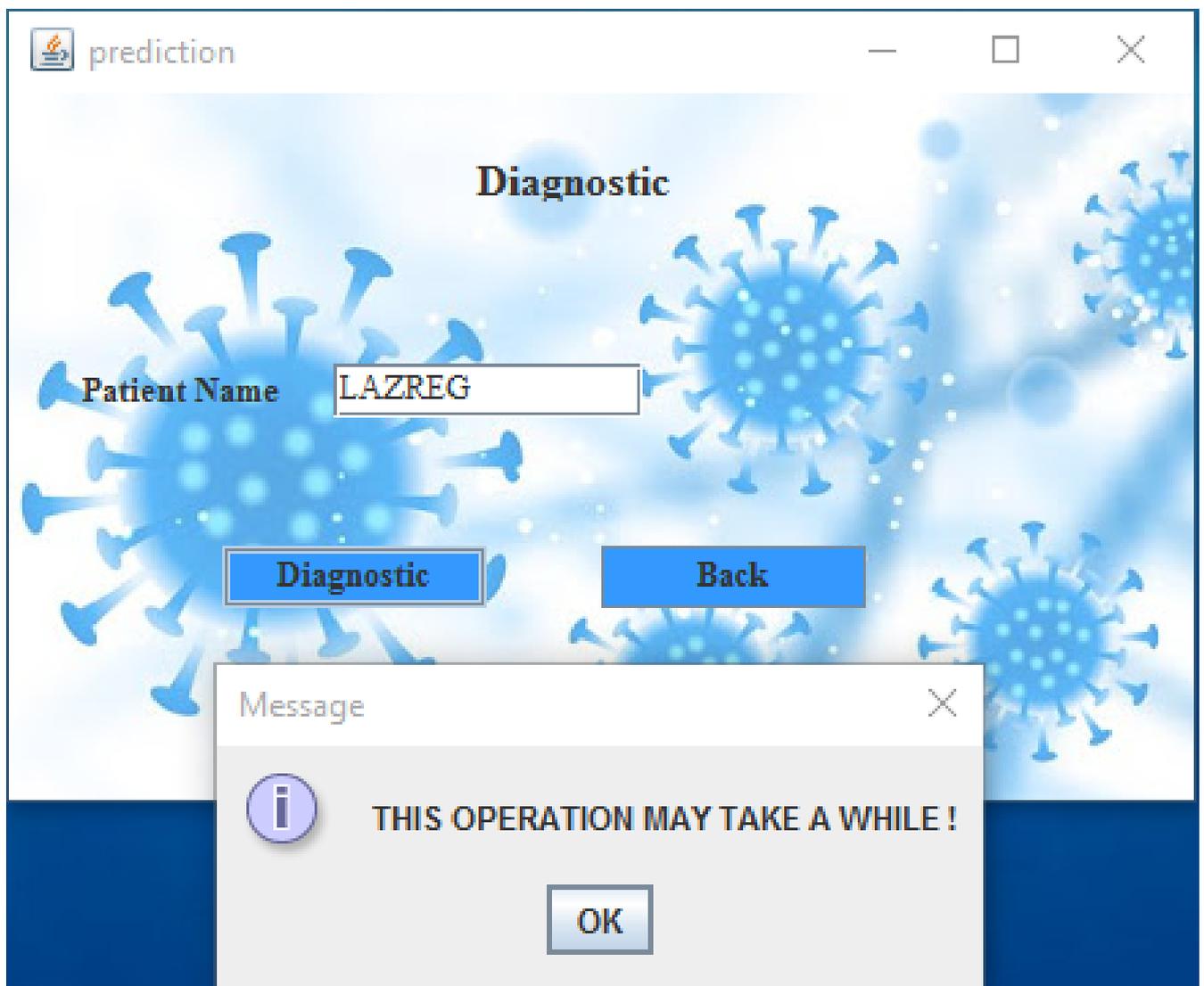


FIGURE 4.6 – Interface de prédiction

4.4.7 Détail de la prédiction

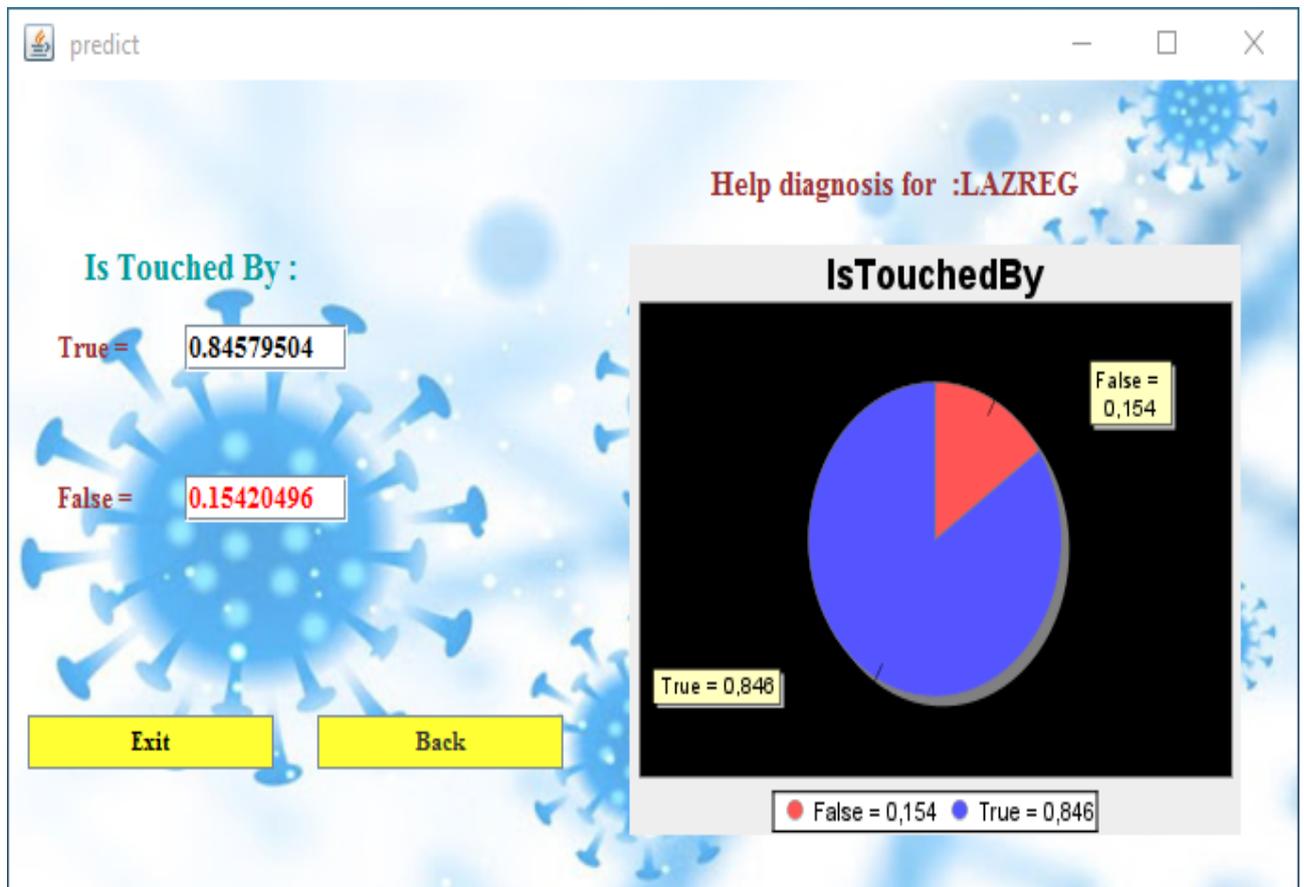


FIGURE 4.7 – interface Détail de prédiction

Cette interface permet d’afficher le détail de la prédiction du patient, et représenter le résultat par un camembert avec la probabilité si le patient est malade ou pas.

4.5 La validation du système

La validation du système consiste à prédire l’efficacité et la fiabilité d’un modèle. Nous allons utiliser **la matrice de confusion** pour valider notre système.

La matrice de confusion est un résumé des résultats de prédictions sur un problème. Les prédictions correctes et incorrectes sont mises en lumière et réparties par classe. Les résultats sont ainsi comparés avec les valeurs réelles.

Cette matrice permet de comprendre de quelle façon le modèle est confus lorsqu’il effectue des prédictions.

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

FIGURE 4.8 – La matrice de confusion

TP (True Positives) : les cas où la prédiction est positive, et où la valeur réelle est effectivement positive.

TN (True Negatives) : les cas où la prédiction est négative, et où la valeur réelle est effectivement négative.

FP (False Positive) : les cas où la prédiction est positive, mais où la valeur réelle est négative.

FN (False Negative) : les cas où la prédiction est négative, mais où la valeur réelle est positive.

Après nous allons calculer les mesures suivantes :

Précision : Proportion de solutions trouvées qui sont pertinentes. Mesure la capacité du système à refuser les solutions non-pertinentes.

Rappel : Proportion des solutions pertinentes qui sont trouvées. Mesure la capacité du système à donner toutes les solutions pertinentes.

F-mesure : Moyenne harmonique de la précision et du rappel. Mesure la capacité du système à donner toutes les solutions pertinentes et à refuser les autres en donnant la performance du système.

Accuracy : Elle est calculée comme la somme des éléments diagonaux de la matrice de confusion,

divisée par le nombre d'individus afin d'avoir une valeur entre 0 et 1.

$$\begin{aligned}
 \textit{precision} &= \frac{TP}{TP + FP} \\
 \textit{recall} &= \frac{TP}{TP + FN} \\
 F1 &= \frac{2 \times \textit{precision} \times \textit{recall}}{\textit{precision} + \textit{recall}} \\
 \textit{accuracy} &= \frac{TP + TN}{TP + FN + TN + FP}
 \end{aligned}$$

Dans notre cas on a obtenu les résultats suivants :

	True	False
True	47	2
False	0	32

TABLE 4.1 – Matrice de confusion

Calcul des mesures :

Précision = 0.95

Recall = 1

F1 Score = 0.97

Accuracy = 0.97

- Le cas (True Positive) : sur les 47 patients qui peuvent avoir un risque de COVID-19, 47 cas de résultats obtenus ont été détectés similaires au data-set.
- Le cas (False Positive) : sur les 34 patients qui n'ont pas la COVID-19, 2 cas de résultats obtenus ont été détectés non similaires au data-set.
- Le cas (False Negative) : sur les 47 patients qui ont la COVID-19, 0 cas de résultats obtenus a été détecté n'est pas similaires au data-set.
- Le cas (True Negative) : sur les 34 patients qui n'ont pas la COVID-19, 32 cas de résultats obtenus ont été détectés similaires au data-set.

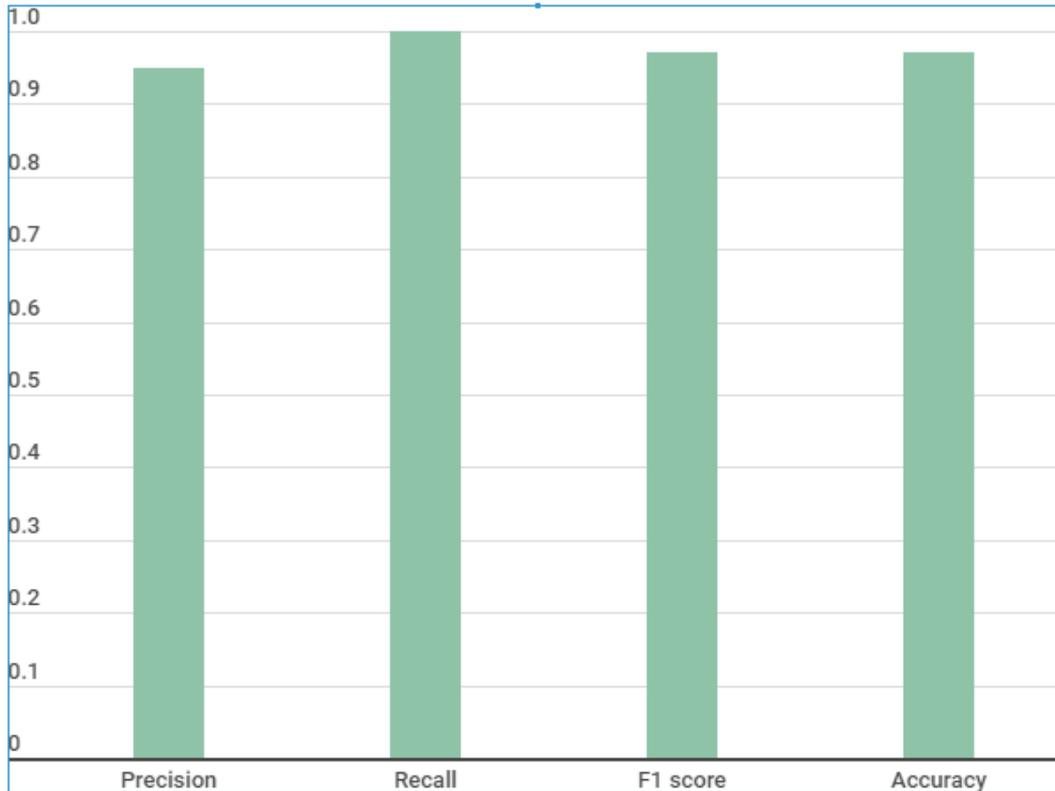


FIGURE 4.9 – Histogramme : precision, recall, F-mesure, accuracy

On remarque que notre système donne de très bonnes valeurs de performance sur les instances du dataset utilisées pour le test.

- Accuracy : Le taux d'accuracy du système est élevé avec un taux d'erreurs minimum, ce que signifie de bonnes valeurs de précision et de recall.
- Precision : Le système s'est entraîné sur une partie du data-set, ce qui a donné une bonne précision.
- Recall : Le taux de rappel est élevé à cause du taux d'erreurs minime .
- F1 score : Le taux de la F-mesure est élevé à cause du taux élevé de la précision.

4.6 Conclusion

Nous avons introduit notre système de prédiction à travers ce chapitre, en important notre ontologie probabiliste et appliquer l'inférence pour la prédiction, en illustrant par des captures de notre application.

Pour tester notre système nous avons présenté les différentes mesures de validation afin d'évaluer la pertinence du système.

Conclusion et perspectives

Dans ce mémoire nous avons présenté un système d'aide au diagnostic médical en se basant sur les ontologies probabilistes pour la modélisation et le traitement de l'incertitude. Le défaut principal des ontologies classiques est leur incapacité de traiter les connaissances incertaines, Cependant, les ontologies probabilistes peuvent présenter un moyen fiable pour la prédiction et le diagnostic médical, en offrant de bons résultats et en assurant un traitement d'incertitude.

Premièrement, nous avons présenté les ontologies classiques et probabilistes ainsi que les connaissances incertaines. Ensuite, nous avons suivi les étapes de la méthodologie UMP-ST pour créer notre ontologie probabiliste.

L'ontologie probabiliste PR-OWL est compatible avec OWL, cette ontologie basée sur les réseaux bayésiens multi entités est capable de prédire tout en tenant compte des incertitudes autour des données.

La structure du MEBN a été réalisée à l'aide de l'expert du domaine qui nous a fourni les majeurs symptômes de la maladie ainsi que les facteurs de risque, et l'apprentissage des paramètres à partir de notre dataset avec des données manquantes, la raison pour laquelle on a utilisé l'algorithme EM.

Afin d'atteindre le but de notre travail et de prédire si un malade est touché par la COVID-19, nous avons représenté les symptômes et les facteurs de risque sous forme de SSBN.

Pour la réalisation de notre conception, nous avons implémenté notre système et présenté ses différentes interfaces. Puis nous avons validé notre système en appliquant plusieurs tests avec des données réelles montrant l'efficacité des ontologies probabilistes dans la modélisation et le raisonnement sous incertitude, ce qui nous a donné de bons résultats.

Nos perspectives est de tester notre système sur un grand nombre de cas, et de rendre le système capable de suivre le développement des différents variants du COVID-19, afin de traiter les différents symptômes et facteurs de risques liés à ces variants. Et de minimiser le temps d'exécution de notre système.

Bibliographie

- [1] D. M. Azevedo, M. Ladeira, L. L. Santos, and R. N. Carvalho. Automatic generation of probabilistic ontologies from ump-st model. In *URSW@ ISWC*, pages 24–35, 2016.
- [2] W. Borst. Construction of engineering ontologies for knowledge sharing and reuse, centre for telematica and information technology, university of twente. Technical report, Enschede, technical report, 1997.
- [3] G. Bucci, V. Sandrucci, and E. Vicario. Ontologies and bayesian networks in medical diagnosis. In *2011 44th Hawaii International Conference on System Sciences*, pages 1–8. IEEE, 2011.
- [4] R. N. Carvalho. *Probabilistic ontology : representation and modeling methodology*. George Mason University, 2011.
- [5] R. N. Carvalho, L. L. dos Santos, M. Ladeira, H. A. Da Rocha, and G. L. Mendes. Ump-st plug-in : documenting, maintaining and evolving probabilistic ontologies using unbbayes framework. In *Uncertainty reasoning for the semantic web III*, pages 1–20. Springer, 2012.
- [6] J. Charlet, B. Bachimont, and R. Troncy. Ontologies pour le web sémantique. *Revue I3, numéro Hors Série «Web sémantique*, pages 43–63, 2004.
- [7] P. C. G. Costa, C. d’Amato, N. Fanizzi, K. B. Laskey, K. J. Laskey, M. Nickles, and M. Pool. *Uncertainty Reasoning for the Semantic Web I : ISWC International Workshop, URSW 2005-2007, Revised Selected and Invited Papers*, volume 5327. Springer, 2008.
- [8] P. C. G. Da Costa, K. B. Laskey, and K. J. Laskey. Pr-owl : A bayesian ontology language for the semantic web. In *Uncertainty Reasoning for the Semantic Web I*, pages 88–107. Springer, 2006.
- [9] E. Delaby. *Dissection de l’architecture génétique de l’autisme par analyse des variations du nombre de copies de gènes*. PhD thesis, Université Pierre et Marie Curie-Paris VI, 2014.

- [10] J.-C. Desenclos and H. De Valk. Les maladies infectieuses émergentes : importance en santé publique, aspects épidémiologiques, déterminants et prévention. *Médecine et maladies infectieuses*, 35(2) :49–61, 2005.
- [11] K. Farooq, A. Hussain, S. Leslie, C. Eckl, C. MacRae, and W. Slack. An ontology driven and bayesian network based cardiovascular decision support framework. In *International Conference on Brain Inspired Cognitive Systems*, pages 31–41. Springer, 2012.
- [12] F. Gandon. Ontologies informatiques. *Interstices*, 2006.
- [13] A. Gómez-Pérez. Ontological engineering : A state of the art. 1999.
- [14] A. Gómez-Pérez. Ontology evaluation. In *Handbook on ontologies*, pages 251–273. Springer, 2004.
- [15] T. R. Gruber. A translation approach to portable ontology specifications. *Knowledge acquisition*, 5(2) :199–220, 1993.
- [16] M. Gruninger. Ontology : applications and design. *Commun. ACM*, 45(2), 2002.
- [17] E. Hlel, S. Jamoussi, and A. B. Hamadou. Intégration d’un réseau bayésien dans une ontologie. In *IC-25èmes Journées francophones d’Ingénierie des Connaissances*, pages 295–297, 2014.
- [18] O. Jules. Modèle de confiance et ontologie probabiliste pilotés par réseaux bayésiens pour la gestion des accords de services dans l’environnement de services infonuagiques. 2015.
- [19] K. B. Laskey, P. C. Costa, and T. Janssen. Probabilistic ontologies for multi-int fusion. *Ontologies and Semantic Technologies for Intelligence*, 213(147) :2, 2010.
- [20] P. Leray. Réseaux bayésiens : apprentissage et modélisation de systèmes complexes. *habilitation à diriger les recherches, Université de Rouen*, 2006.
- [21] Y. Lin, Z. Xiang, and Y. He. Brucellosis ontology (idobru) as an extension of the infectious disease ontology. *Journal of biomedical semantics*, 2(1) :1–18, 2011.
- [22] P. Lord. Components of an ontology. *Ontogenesis*, 2010.
- [23] R. Messaoudi, F. Jaziri, A. Mtibaa, M. Grand-Brochier, H. M. Ali, A. Amouri, H. Fourati, P. Chabrot, F. Gargouri, and A. Vacavant. Ontology-based approach for liver cancer diagnosis and treatment. *Journal of digital imaging*, 32(1) :116–130, 2019.
- [24] R. Mizoguchi and J. Bourdeau. Using ontological engineering to overcome common ai-ed problems. *Journal of Artificial Intelligence and Education*, 11 :107–121, 2000.

- [25] R. Mizoguchi and J. Bourdeau. Theory-aware authoring environment-ontological engineering approach. In *International Conference on Computers in Education, 2002. Proceedings.*, pages 1553–1554. IEEE, 2002.
- [26] R. Neches, R. E. Fikes, T. Finin, T. Gruber, R. Patil, T. Senator, and W. R. Swartout. Enabling technology for knowledge sharing. *AI magazine*, 12(3) :36–36, 1991.
- [27] C.-R. Nicandro, M.-M. Efrén, A.-A. Maria Yaneli, M.-D.-C.-M. Enrique, A.-M. Hector Gabriel, P.-C. Nancy, G.-H. Alejandro, H.-R. Guillermo de Jesus, and B.-M. Rocio Erandi. Evaluation of the diagnostic power of thermography in breast cancer using bayesian network classifiers. *Computational and mathematical methods in medicine*, 2013, 2013.
- [28] S.-L. A. K. Olwendo, Amos. Probabilistic model for chronic obstructive pulmonary disease diagnosis and phenotyping using bayesian network. In *Environment and Water Resource Management*, 2016.
- [29] V. Psyché, O. Mendes, and J. Bourdeau. *Apport de l'ingénierie ontologique aux environnements de formation à distance*. PhD thesis, 2003.
- [30] F. L. Seixas, B. Zadrozny, J. Laks, A. Conci, and D. C. M. Saade. A bayesian network decision model for supporting the diagnosis of dementia, alzheimer s disease and mild cognitive impairment. *Computers in biology and medicine*, 51 :140–158, 2014.
- [31] Y. Shen, K. Yuan, D. Chen, J. Colloc, M. Yang, Y. Li, and K. Lei. An ontology-driven clinical decision support system (iddap) for infectious disease diagnosis and antibiotic prescription. *Artificial intelligence in medicine*, 86 :20–32, 2018.
- [32] B. Swartout, R. Patil, K. Knight, and T. 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.
- [33] M. Uschold and M. Gruninger. Ontologies : Principles, methods and applications. *The knowledge engineering review*, 11(2) :93–136, 1996.
- [34] T. Verma and J. Pearl. Equivalence and synthesis of causal models in proceedings of the sixth annual conference on uncertainty in artificial intelligence (p. bonissone, m. henrion, l. kanal and j. lemmer, eds.) 220–227, 1990.
- [35] G. Wu, J. Li, L. Feng, and K. Wang. Identifying potentially important concepts and relations in an ontology. In *International Semantic Web Conference*, pages 33–49. Springer, 2008.