

UNIVERSITÉ SAAD DAHLEB DE BLIDA1

Faculté des sciences

Département d'informatique



MEMOIRE DE MASTER

En Informatique

Option : Ingénierie Des Logiciels

THÈME :

**Intégration de la connaissance incertaine
dans une ontologie : Etude sur l'aide au
diagnostic du Covid-19.**

Réalisé par

ABBAS Nour el Houda

BENNILA Mounira

Encadré par

Dr. M. FAREH

Dr. I. RIALI

M. BALA. Président

Mme. BOUDRAA. Examinatrice

Soutenu le : 06 juillet 2022

Remerciements

En tout premier lieu, nous remercions le bon dieu, tout-puissant de nous avoir donné la santé, le courage et la volonté de terminer ce mémoire.

Nous adressons nos remerciements les plus vifs, notre profonde gratitude et notre respect à notre promotrice Mme. FAREH pour sa patience, son temps, ses aides, son partage de connaissances, ses précieux conseils et surtout pour la motivation qu'elle a su nous transmettre. Nous remercions également notre Co-promoteur M. RIALI Ishak pour ses aides, sa disponibilité et ses orientations qui ont tellement contribué à la réalisation de notre projet. Nous avons eu la chance de travailler avec vous et d'apprécier vos qualités, vos valeurs, votre sérieux et votre compétence.

Nous souhaitons adresser tous nos remerciements aux personnes qui ont participé de près ou de loin à la réalisation de ce mémoire, plus particulièrement à : Dr. Belkacemi Walid et son assistant Hassan, Ibnaissa Mohamed Chakib, Abbas Mohamed Redha sans oublier Dr. Oucif Selma, la biochimiste Boubiche khaoula, Hafsa et Redha.

Enfin, nous adressons nos plus sincères remerciements à nos familles : nos parents, nos frères, nos sœurs et tous nos amis, qui nous ont aidé, soutenu et encouragé tout au long de la réalisation de ce mémoire.

Dédicaces

Je dédie ce travail à :

A celle qui m'a arrosé de tendresse et d'affection, ma source d'amour, ma chère maman. Quoi que je fasse ou que je dise je ne saurai point te remercier comme il se doit. Merci pour ta présence, ton amour et tes conseils, merci d'avoir été une amie avant d'être une maman.

A celui qui m'a comblé d'amour et d'affection, qui m'a toujours soutenu et protégé comme une perle rare, mon cher papa. Tout l'encre du monde ne pourrait suffir pour t'exprimer la profondeur des sentiments que j'éprouve pour toi.

A mon seul et unique frère Redha, mon soutien indéfectible. Je ne pourrai jamais imaginer ma vie sans toi, sans nos fous rires, nos coups de gueule et nos aventures. Merci pour tout.

A mon joyeux accompagnant Smokey qui a accompagné mes longues nuits de veille et m'a apporté le soutien dont j'avais besoin.

A l'homme qui aurait été le plus heureux de me voir soutenir, mon cher feu grand-père Lazhar. puisse dieu t'accorder bénédiction et miséricorde.

A ma grand-mère Houria qui n'a cessé de prier pour moi et de me souhaiter le meilleur des avenir. Puisse dieu t'accorder une longue vie pleine de joie et de santé.

A mes meilleures amies Amira, Khaoula et Selma qui m'ont toujours écouté, soutenu et supporté.

A toutes les belles amitiés que j'ai pu construire pendant mon parcours universitaire Ikram, Yasmine, Lylia, Sabrina, Sarah et Ahmed.

A mon binôme Mounira, sans qui le travail n'aurait jamais atteint son objectif. ça a été un plaisir de travailler avec elle.

Enfin, je clôture ma dédicace en l'adressant aux informaticiens de ma famille, Hanane, Hichem, Manel sans oublier ma mère.

Nour el Houda Abbas

Dédicaces

je tiens à dédier ce modeste travail :

À la femme la plus merveilleuse du monde, ma maman chérie, pour ses sacrifices, ses encouragements et son soutien pour que je puisse atteindre mes objectifs.

À mon cher papa, pour son affection, la confiance qu'il m'a accordée et les efforts fournis pour mon éducation et mon bien-être.

À mon frère Abdelhadi et ma sœur Meriem et mon beau-frère Lotfi, pour leurs soutiens moral et leurs précieux conseils tout au long de mes études.

À mes nièces adorées Melina et Nihel et mon petit-neveu adoré Aymen qui me comblent de bonheur et de joie.

À ma meilleure amie Missa que je considère comme une deuxième sœur à qui je peux compter pour me remonter le moral et m'écouter des heures.

À mes chères amies Lyna, Nariman, Ilhem et Siham, pour leurs aides et supports dans les moments difficiles.

À la meilleure déléguée Ikram qui nous a beaucoup aidé.

Enfin, à mon binôme Nour El Houda, avec qui j'ai passé des moments inoubliables qui n'ont pas été toujours faciles, mais grâce à son ambition et son énergie positive, nous avons pu les surmonter.

Mounira Bennila

Résumé

Le domaine médical est caractérisé par son incertitude qui se présente sous plusieurs formes : l'incomplétude, l'aléatoire, l'imprécision etc. Ce qui rend le diagnostic difficile pour les médecins surtout dans le cas où la maladie est récente et les informations nécessaires pour son diagnostic sont remplies d'incertitudes tel que le Covid-19.

Malgré le succès des ontologies qui permettent de représenter les connaissances de manière simple et formelle, elles ne permettent pas la représentation et le raisonnement sous l'incertitude contrairement aux réseaux bayésiens. En vue de tirer partie des capacités des deux systèmes nous nous sommes orientés vers une approche qui consiste à intégrer la connaissance incertaine dans l'ontologie classique via une méta ontologie comportant les composants du réseau bayésien afin de représenter l'incertitude.

L'objectif de notre travail est de concevoir un système d'aide au diagnostic médical, basé sur ce qui a été synthétisé précédemment.

Pour la construction de notre ontologie classique nous avons adopté la méthode Methontology et avons créé notre réseau bayésien hybride en commençant par l'apprentissage de la structure fait avec l'expert du domaine, ensuite l'apprentissage des paramètres qui a nécessité l'utilisation de l'algorithme EM pour les noeuds discrets et le développement de l'algorithme best fit pour les noeuds continus qui consiste à déterminer la meilleure distribution pour chaque noeud continu. Après l'élaboration du réseau bayésien hybride nous avons créé une méta ontologie qui comporte les composants de ce dernier afin de l'intégrer dans l'ontologie classique. Nous avons appliqué ce processus sur l'étude de cas du diagnostic du Covid-19 pour réaliser un système qui aide les médecins à prendre la bonne décision.

Mots clés :

Incertitude, Ontologie, Réseau Bayésien hybride, Inférence Probabiliste, Covid-19

Abstract

The medical field is characterized by its uncertainty, which takes many forms : incompleteness, randomness, imprecision, etc., which makes diagnosis difficult for doctors, especially when the disease is recent and the required information for the diagnosis is uncertain, in which case COVID-19 takes place.

In spite of the success of ontologies that allow us to represent knowledge in a simple and formal way, they do not allow the representation and reasoning under uncertainty, as opposed to Bayesian networks. In order to take advantage of the capacities of both systems, we have oriented ourselves towards an approach that consists of integrating uncertain knowledge into the classical ontology via a meta ontology comprising the components of the Bayesian network in order to represent uncertainty. The objective of our work is to design a medical diagnosis aid system, based on what has been previously synthesized. For the construction of our classical ontology, we adopted the methodology method and created our hybrid Bayesian network based on the phase of structure learning that has been done with the domain expert, followed by a parameters learning phase in which it is necessary to apply the EM algorithm to discrete nodes and the development of the best fit algorithm for the continuous nodes. The latter consists of determining the best distribution for each continuous node. After the development of the hybrid Bayesian network, we created a meta ontology that includes the components of the processed network in order to integrate it into the classical ontology. This process was applied to the case study of the COVID-19 diagnosis to build a system that helps doctors make the right decision.

Key words :

Uncertainty, Ontology, Hybrid bayesian Network, Probabilistic Inference, Covid-19

ملخص :

يتميز المجال الطبي بعدم اليقين الذي يتمثل في عدة أشكال كعدم الاكتمال ، والعشوائية وعدم الدقة ، إلخ. هذا ما يجعل التشخيص صعبًا بالنسبة للأطباء ، خاصة في الحالة التي يكون فيها المرض حديثًا وتكون المعلومات اللازمة لتشخيصه مليئة بالشكوك مثل كوفيد- ١٩ .

على الرغم من نجاح الأنطولوجيا التي تسمح بتمثيل المعرفة بطريقة بسيطة ورسومية ، إلا أنها لا تسمح بالتمثيل والاستدلال في ظل عدم اليقين على عكس شبكات بايز. من أجل الاستفادة من قدرات النظامين ، اخترنا نهجًا يتمثل في دمج المعرفة الغير يقينية في الأنطولوجيا الكلاسيكية من خلال ميطا أنطولوجيا تحمل مكونات شبكة بايز من أجل تمثيل عدم اليقين .

الهدف من عملنا هو تصميم نظام مساعدة تشخيصية طبية، بناء على ما تم ذكره مسبقًا. لبناء الأنطولوجيا الخاصة بنا، اعتمدنا منهجية الميتونتولوجي وأنشأنا شبكة بايز المختلطة بدءًا من تشكيل الهيكل الذي تم تمثيله مع خبير المجال ثم بعد ذلك تعليم المعلمات التي تطلبت استخدام خوارزمية خاصة للعقد المنفصلة وتطوير خوارزمية تدعى الأنسب للعقد المستمرة والتي تعمل على تحديد أفضل توزيع لكل عقدة مستمرة. بعد تطوير شبكة بايز الهجينة ، أنشأنا ميطا أنطولوجيا تتضمن مكونات هذا الأخير من أجل دمجها في الأنطولوجيا الكلاسيكية. طبقنا هذه العملية على دراسة حالة تشخيص كوفيد-١٩ لإنشاء نظام يساعد الأطباء .

الكلمات المفتاحية

عدم اليقين ، أنطولوجيا، شبكة بايز الهجينة ، الاستدلال الاحتمالي ، كوفيد -19

Table des matières

Table des figures	11
Liste des tableaux	13
Introduction générale	15
1 ONTOLOGIES ET RÉSEAUX BAYÉSIENS	19
1.1 Introduction	19
1.2 Ontologie	19
1.2.1 Définition de l'ontologie	19
1.2.2 Composants de l'ontologie	20
1.2.3 Rôle de l'ontologie	21
1.2.4 Langages de l'ontologie	22
1.2.5 Principes d'ingénierie ontologique	22
1.2.6 Cycle de vie de l'ontologie	23
1.2.7 Limites de l'ontologie classique	24
1.3 L'incertitude	25
1.3.1 Définition d'incertitude	25
1.3.2 Types d'incertitude dans le contexte du web sémantique	25
1.4 Réseau bayésien	26
1.4.1 Définition	26
1.4.2 Apprentissage dans le réseau bayésien	27
1.4.3 Inférence dans le réseau bayésien standard	29
1.4.4 Types de réseaux bayésiens	30
1.4.5 Domaines d'application du réseau bayésien	30
1.5 Les travaux connexes qui traitent de l'intégration de l'incertitude au niveau des ontologies	31
1.5.1 Comparaison des travaux	33
1.5.2 Analyse des travaux	35

1.6	Conclusion	37
2	DIAGNOSTIC DU COVID-19	39
2.1	Introduction	39
2.2	Incertitude en médecine	39
2.3	Définition du Covid-19	40
2.4	Impacts du Covid-19	40
2.5	Symptômes	41
2.6	Signes	42
2.6.1	Signe clinique	42
2.6.2	Signe paraclinique	42
2.6.3	Les signes paraclinique du Covid-19	42
2.7	Facteur de risque	43
2.7.1	Facteurs de risque du Covid-19	43
2.8	Tests du Covid-19	43
2.9	Variants du Covid-19	44
2.10	Les travaux connexes au Covid-19	45
2.10.1	Comparaison des travaux	47
2.10.2	Analyse des travaux	51
2.11	Conclusion	53
3	CONCEPTION DU SYSTÈME	55
3.1	Introduction	55
3.2	Motivation de l'étude de cas	55
3.3	Caractéristiques du Système	56
3.4	Schéma global	57
3.5	Description du schéma global	58
3.6	Construction du dataSet	58
3.6.1	Les symptômes et les signes	59
3.6.2	Les facteurs de risque	61
3.7	Création de l'ontologie Classique	62
3.8	Construction du réseau bayésien hybride	69
3.9	Méta ontologie du réseau bayésien	75
3.10	L'intégration de la méta ontologie dans l'ontologie classique	77
3.11	Inférence du réseau bayésien hybride	78
3.12	Conclusion	78
4	IMPLEMENTATION ET TEST DU SYSTÈME	79
4.1	Introduction	79
4.2	Langages de Programmation	79

4.3	Outils de développement	80
4.4	API Utilisés	80
4.5	Résultat de l'ontologie finale	81
4.6	Les interfaces graphiques	82
4.6.1	L'interface d'accueil	82
4.6.2	L'interface patient	83
4.6.3	L'interface ajouter patient	84
4.6.4	L'interface informations patient	85
4.6.5	L'interface modifier informations patient	86
4.6.6	L'interface prédiction	87
4.7	Validation du système	87
4.7.1	Validation du système par l'expert du domaine :	87
4.7.2	Validation du système par les mesures de performances :	91
4.8	Comparaison de notre système avec d'autres modèles probabilistes	93
4.8.1	Analyse des résultats	94
4.9	Conclusion	95
	Conclusion et perspectives	97
	Bibliographie	99

Table des figures

1.1	Le cycle de vie d'une ontologie [13]	23
1.2	Construction d'une ontologie opérationnelle [14]	24
1.3	La modélisation de la table de probabilité [21]	26
1.4	Graphe bayésien	27
3.1	Schéma global	57
3.2	Identification des termes de l'ontologie et leurs définitions	63
3.3	Taxonomie de Risk_factor	65
3.4	Taxonomie de Sign	65
3.5	Taxonomie de Symptom	66
3.6	Diagramme de relations binaires	66
3.7	La structure de l'ontologie classique	67
3.8	Les classes, les object property et les data property de l'ontologie classique	68
3.9	La structure du réseau bayésien	69
3.10	Le best fit du nœud age	71
3.11	Le best fit du nœud INR	71
3.12	Le best fit du nœud CRP	72
3.13	Le best fit du nœud Blood_urea	72
3.14	Le best fit du nœud d-dimers	73
3.15	Le best fit du nœud blood_sugar	73
3.16	Le best fit du nœud Creatinine	74
3.17	Le best fit du nœud SPO2	74
3.18	Le réseau bayésien après l'apprentissage des paramètres	75
3.19	La hiérarchie de la méta ontologie	76
3.20	Les classes, les object property et les data property de l'ontologie finale	77
4.1	Code OWL du noeud Blood sugar	81
4.2	Code OWL du noeud diabète	81
4.3	Interface d'accueil	82

4.4	Interface patient	83
4.5	Interface ajouter patient	84
4.6	Interface Consulter patient	85
4.7	Interface Modifier patient	86
4.8	Interface Prédiction	87
4.9	Résultat du Test 1	89
4.10	Résultat du Test 2	90
4.11	Histogramme 1 : F-score, recall, Precision, Accuracy, Fallout	93
4.12	Histogramme 2 : F-score, recall, Precision, Accuracy	94

Liste des tableaux

1.1	Tableau comparatif des langages, des réseaux bayésiens utilisés et des objectifs	34
1.2	Tableau décrivant les méthodes utilisées et les limites des travaux	35
2.1	Tableau comparatif des objectifs et des méthode utilisées	48
2.2	Tableau comparatif des symptômes et des données utilisées	49
2.3	Les avantages et les limites des travaux	50
3.1	Dictionnaire de concepts	64
3.2	Le tableau des relations binaires	66
4.1	La matrice de confusion	91
4.2	La matrice de confusion de notre système	91

Introduction générale

Contexte

Lorsque le diagnostic est évident, les décisions médicales deviennent faciles à prendre par les praticiens de la santé, à contrario, quand le diagnostic n'est pas évident (diagnostic douteux) et que les symptômes sont communs entre plusieurs maladies, les professionnels de la santé se trouvent alors en difficulté pour se prononcer et établir les prescriptions nécessaires.

Malgré les grandes avancées de la science, le domaine de la santé regorge toujours de connaissances incertaines (manque des informations nécessaires pour un diagnostic fiable), notamment par rapport aux nouvelles maladies, tel que la pandémie du Covid-19.

Souvent, des données imprécises, ambiguës et incomplètes, sont sources de mauvais diagnostics et donc de prescriptions erronées, ce qui implique fatalement des conséquences néfastes pour les malades concernés.

Le développement d'un outil d'aide au diagnostic du Covid-19, sera une réponse concrète au problème précitée. Pour cela, l'utilisation des ontologies a été proposé afin de représenter les connaissances, néanmoins elles ne fournissent pas un support adéquat pour faire face à la connaissance incertaine, ce qui l'empêche de raisonner sous l'incertitude contrairement au réseau bayésien qui permet le raisonnement probabiliste. L'intégration de ce dernier dans une ontologie classique peut être une solution pour la modélisation des connaissances incertaines.

Problématique

Le domaine médical est un domaine rempli d'incertitudes où les connaissances sur une maladie précise tel que le Covid-19 sont incertaines et complexes, ce qui fait que le médecin peut être confronté à des situations délicates au moment de la prise de décision, c'est potentiellement une question de vie ou de mort.

Les ontologies classiques offrent une modélisation formelle des connaissances, cependant, elles ne permettent pas le raisonnement sous l'incertitude qui peut se manifester par exemple, lors d'un diagnostic d'un patient, dans notre cas d'étude la maladie est le Covid-19.

Cela conduit à poser les questions suivantes :

1. Comment représenter et modéliser les connaissances incertaines pour l'aide au diagnostic dans le domaine médical, en utilisant les ontologies et les réseaux bayésiens ?
2. Comment intégrer les connaissances incertaines dans une ontologie et réaliser l'inférence dans le domaine médical ?

Objectif

L'incertitude dans le domaine médical peut constituer pour les praticiens de la santé un handicap pour la bonne prise en charge de leurs patients, d'où l'objectif du présent projet, qui consiste en le développement d'un système d'aide au diagnostic médical de la maladie Covid-19 en se basant sur un réseau bayésien hybride qui servira autant pour réaliser l'inférence que pour l'élaboration d'une méta ontologie, laquelle s'intégrera dans l'ontologie classique afin qu'elle puisse représenter les connaissances incertaines.

Organisation du mémoire

Le projet sera présenté en deux parties :

La première partie : contient les deux premiers chapitres qui présentent le contexte du travail.

- **Chapitre 1 : ONTOLOGIES ET RÉSEAUX BAYÉSIENS :** dans ce chapitre nous allons définir les ontologies, leurs différents composants, rôles, langages de représentation et limites. Nous allons aussi présenter l'incertitude et ses types, ensuite nous allons aborder la notion des réseaux bayésiens qui peuvent présenter l'incertitude de manière claire et lisible, ainsi que leurs types, leurs différentes techniques d'apprentissage de la structure et des paramètres en plus de leurs algorithmes d'inférence. Nous finirons par présenter quelques travaux connexes qui traitent de l'intégration de l'incertitude au niveau des ontologies.

- **Chapitre 2 : DIAGNOSTIC DU COVID-19** : dans ce chapitre, nous allons définir le Covid-19, ses différents variants et citer quelques symptômes, signes, facteurs de risque et tests de laboratoire. Nous finirons par présenter quelques travaux de recherches en rapport avec le Covid-19.

La deuxième partie : Concerne la conception et le développement du système.

- **Chapitre 3 : ANALYSE ET CONCEPTION** : ce chapitre présente la motivation de notre cas d'étude, notre dataset, la construction de notre ontologie classique en suivant les étapes de la méthode Methontology ainsi que la création du réseau bayésien hybride et enfin l'intégration de l'incertitude dans l'ontologie classique.
- **Chapitre 4 : IMPLÉMENTATION ET TEST DU SYSTÈME** : dans ce chapitre nous allons présenter notre système à travers des captures d'écran des interfaces et évaluer ses performances.

Ce mémoire sera clôturer par une conclusion générale qui résume les points principaux abordés et propose certaines perspectives envisagées pour l'amélioration de notre travail.

ONTOLOGIES ET RÉSEAUX BAYÉSIENS

1.1 Introduction

L'ontologie classique est un moyen pour la représentation des connaissances, cela tient à sa capacité à résoudre le problème de représentation et de manipulation des connaissances d'une manière formelle et structurée. Par ailleurs, le réseau bayésien est un système qui représente la connaissance incertaine et permet le calcul des probabilités conditionnelles. Ce dernier propose un formalisme mathématique et des bases théoriques solides pour la modélisation des connaissances incertaines.

Dans ce chapitre nous allons définir la notion d'ontologie, ses différents composants, son rôle, ses langages de représentation et ses limites, ensuite nous allons présenter la notion de l'incertitude et ses types. Nous finirons par aborder la définition du réseau bayésien, ses types, ses différentes techniques d'apprentissage, son inférence en plus de quelques travaux connexes qui traitent l'intégration de l'incertitude au niveau des ontologies.

1.2 Ontologie

Dans cette section, nous présentons l'ontologie, ses rôles ainsi que ses composants, ses langages de représentation et ses limites.

1.2.1 Définition de l'ontologie

Le terme ontologie est un terme grec composé de "ontos" et "logos", qui signifie l'essence de l'être. Ce terme, hérité d'une tradition philosophique intéressée par la science de l'être, est apparu dans le domaine de l'informatique notamment dans le cadre du projet ARPA Knowledge Sharing Effort [1]. Gruber a introduit le concept d'ontologie comme "une spécification explicite" [2], par la suite cette définition a été légèrement modifiée par Borst

[3]. La combinaison des deux définitions peut se résumer en “une spécification explicite et formelle d’une conceptualisation partagée”. Cette définition peut être interprétée ainsi :

Explicite : signifie que le type de concepts et les contraintes de leurs utilisations sont explicitement définis.

Formelle : fait référence au fait que la spécification doit être lisible par une machine.

Conceptualisation : désigne un modèle abstrait d’un certain phénomène du monde reposant sur l’identification des concepts pertinents de ce dernier.

Partagée : capture la connaissance consensuelle, qui n’est pas propre à un individu mais validée par un groupe.

L’ontologie représente donc un modèle conceptuel qui permet le développement des systèmes de connaissances qui peuvent être partagés et réutilisés permettant, par conséquent, l’interopérabilité entre plusieurs sources d’informations et connaissances.

1.2.2 Composants de l’ontologie

Comme nous l’avons abordé, l’ontologie fournit un vocabulaire du domaine et détermine de manière plus ou moins formelle le sens des termes et des relations entre eux, ainsi, les connaissances intégrées dans une ontologie sont formalisées en se basant sur cinq types de composants [2] :

- **Concept** : représente un objet, une idée ou bien une notion abstraite. Selon Uschold [4] un concept se compose de trois parties : un (ou plusieurs) terme(s) qui permettent d’identifier le concept, une notion (ou intention du concept) qui désigne la sémantique du concept défini au travers de ses propriétés et attributs, et enfin un ensemble d’objets (ou extension du concept) qui réfère à toutes les instances du concept.
- **Relation** : les relations dans une ontologie correspondent aux types d’interaction entre les concepts d’un domaine. Par ce fait, elles se définissent à la fois par leur signature mais aussi par le contenu sémantique entre les concepts qu’elles unissent, elles permettent de structurer les connaissances et de définir les interrelations entre les concepts.
- **Fonction** : ce sont un cas particulier de relations dans lesquelles un élément est défini en fonction des éléments précédents.
- **Instances (ou individus)** : elles constituent la définition extensionnelle d’une ontologie.

- **Axiomes** : sont des assertions toujours vraies à propos des abstractions (concepts et relations) du domaine traduit par l'ontologie. Les axiomes sont des expressions en langage logique ayant pour but de déterminer des restrictions sur la valeur des propriétés, définir les arguments d'une relation, etc. Ils sont utilisés aussi pour vérifier la consistance de l'ontologie.

1.2.3 Rôle de l'ontologie

La notion d'ontologie est apparue dans le but de satisfaire plusieurs besoins dans différents domaines d'application. Nous allons citer dans ce qui suit quelques exemples sur les objectifs de l'utilisation de l'ontologie :

- **La connaissance du domaine** : L'ontologie sert à représenter les connaissances du domaine d'un Système à Base de Connaissances (SBC) d'une manière claire et simple. Elle fait en sorte de décrire les objets, leurs propriétés et la manière dont ils peuvent se combiner afin de construire des connaissances complètes du domaine.
- **La communication** : L'ontologie améliore la communication non seulement entre machines mais également entre humains et machines. Elle sert à créer au sein d'un groupe ou d'une organisation un vocabulaire conceptuel commun [5].
- **L'interopérabilité** : L'ontologie assure l'interopérabilité qui se définit par la capacité des systèmes à s'adapter et à collaborer entre eux [5].
- **L'indexation et la recherche d'information** : L'ontologie peut être utilisé pour déterminer les index conceptuels qui décrivent les ressources sur le web [6].

Par ailleurs, l'ontologie a également un rôle très important dans la réutilisation de l'information et l'intégration des ressources hétérogènes. Elle peut être très utiles dans de nombreux domaines d'application tels que la recherche d'information, l'aide à l'analyse... etc.

1.2.4 Langages de l'ontologie

Il existe plusieurs langages qui ont été conçus pour la description de l'ontologie, parmi ces derniers, nous citons :

- **RDFS (Resource Description Framework) :**

RDFS (Resource Description Framework Schema) est un langage de spécification de schémas associé à RDF, il est introduit comme étant une couche au-dessus du RDF. Il définit les classes, les propriétés, les sous classes, les superclasses, les sous-propriétés, et les super-propriétés [7].

- **OWL (Ontology Web Language) :**

C'est un langage qui suit la syntaxe XML, il a un vocabulaire développé qui lui permet d'offrir aux machines la capacité d'interpréter le contenu web en plus de sa capacité de faire le raisonnement qui permet de déduire des informations qui ne sont pas explicitement présentes dans l'ontologie. OWL propose trois sous-langages qui sont : OWL Lite, OWL DL et OWL Full [8].

1.2.5 Principes d'ingénierie ontologique

Il existe plusieurs règles, principes et critères qui assurent le bon développement de l'ontologie :

- **La clarté :** L'ontologie doit fournir des définitions claires et objectives, c'est pour cela qu'il est important d'associer aux entités des définitions en langage naturel [2].
- **La cohérence :** Les axiomes et les concepts doivent être logiquement consistants. Autrement dit, les définitions des classes de l'ontologie ne doivent pas conduire à une contradiction [9].
- **La complétude :** L'ontologie doit couvrir tous les concepts pertinents du domaine d'intérêt [10].
- **L'extensibilité monotone maximale :** La structure utilisée doit prendre en considération les ajouts futurs et surtout assurer que ce nouvel ajout de concepts ne doit pas entraîner la révision des concepts déjà existants [2].

- **Minimiser la distance sémantique** : La distance sémantique entre les concepts qui viennent d'un même parent doit être minimisée le plus possible [11].

1.2.6 Cycle de vie de l'ontologie

Le cycle de vie de l'ontologie permet d'identifier les différentes étapes de sa construction. Dieng et ses collègues [12] ont proposé un cycle de vie inspiré du génie logiciel, il consiste en une étape initiale d'évaluation des besoins, une fois les besoins sont définis l'étape de construction peut démarrer, une étape de diffusion ou il s'agit de tester l'ontologie par rapport au contexte d'usage pour lequel elle a été construite, enfin une étape d'utilisation, cette étape signifie que l'ontologie en question peut être mise à la disposition des utilisateurs. Après chaque utilisation significative, l'ontologie et les besoins sont réévalués et l'ontologie peut être étendue ou reconstruite si nécessaire.

La figure 1.1 représente le cycle de vie d'une ontologie.

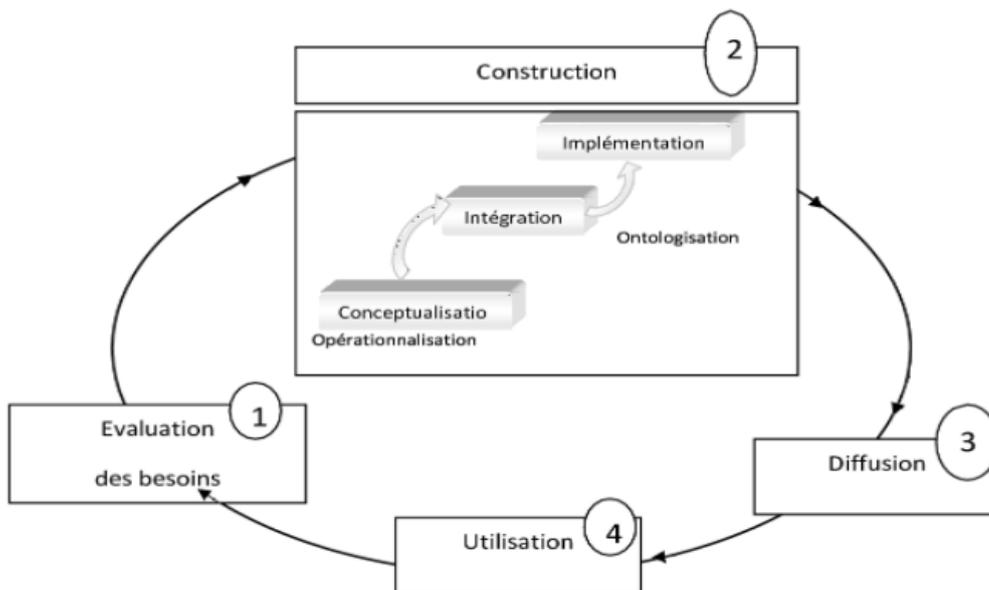


FIGURE 1.1 – Le cycle de vie d'une ontologie [13]

L'étape de construction peut être décomposée en trois principales phases[14] :

1. **La conceptualisation** : permet de faire ressortir, à partir des données brutes, un ensemble de concepts et des relations entre eux décrivant ainsi les connaissances informelles.

2. **L'ontologisation** : Dans cette étape, c'est la formalisation des connaissances, autant que possible, du modèle conceptuel obtenu dans l'étape de conceptualisation.
3. **L'opérationnalisation (ou représentation)** : consiste à donner la main à une machine pour manipuler les connaissances d'une ontologie, c'est-à-dire une transcription de l'ontologie dans un langage formel et opérationnel de représentation de connaissances.

La figure 1.2 représente les trois phases citées ci-dessus :

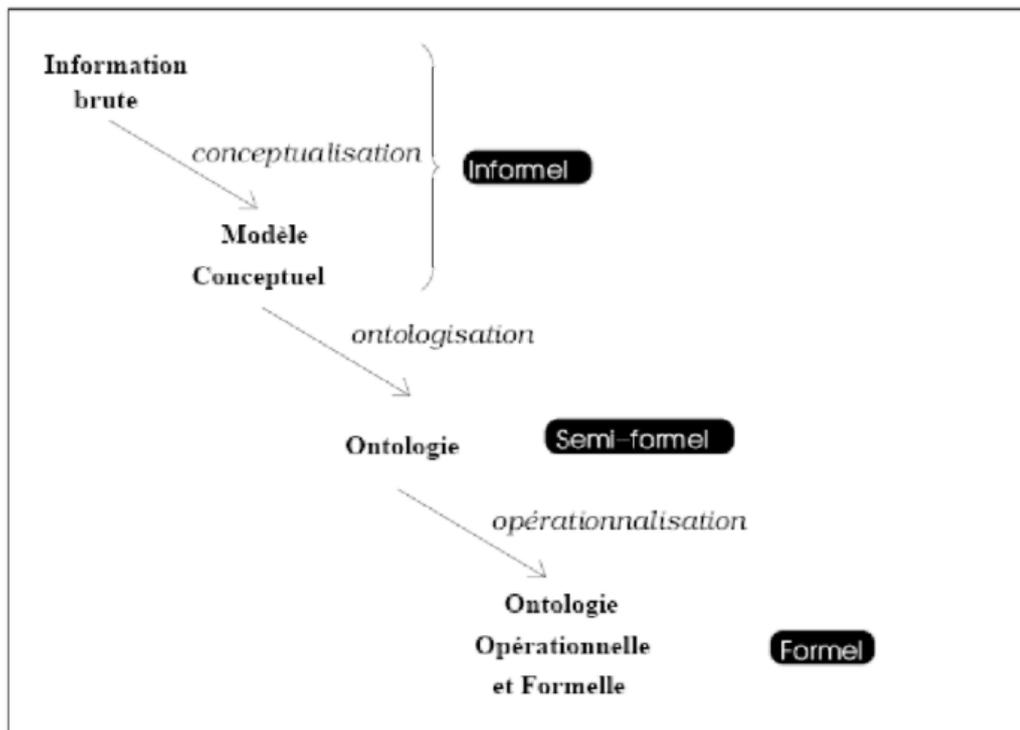


FIGURE 1.2 – Construction d'une ontologie opérationnelle [14]

1.2.7 Limites de l'ontologie classique

Malgré la performance de l'ontologie classique dans la représentation claire et simple de la connaissance, elle se retrouve néanmoins limitée en ce qui concerne la représentation de la connaissance incertaine puisqu'elle ne permet pas la représentation et le raisonnement sous l'incertitude.

1.3 L'incertitude

Avec l'apparition d'internet et du web, l'accès à l'information est de plus en plus facile et moins restreint, mais cela n'empêche pas que la fiabilité de ces informations soit souvent remise en cause. C'est pour cette raison que nous allons nous intéresser par la suite à la notion d'incertitude et ses différents types.

1.3.1 Définition d'incertitude

Il existe plusieurs définitions qui ont été associées à la notion d'incertitude et cela revient à la diversité des contextes dans lesquels elle a été employée.

Selon le dictionnaire Larousse : "Caractère de ce qui est incertain.

ses synonymes sont : imprévisibilité - inconstance - instabilité - précarité - vulnérabilité"

Selon Zadeh [15] " L'incertitude est une propriété inhérente qui caractérise l'information".

Selon Milliken [16] "L'incertitude c'est quand un individu est incapable a prédire quelque chose".

1.3.2 Types d'incertitude dans le contexte du web sémantique

Dans le cadre du web sémantique, l'incertitude est catégorisée en cinq types :

- **L'ambiguïté** : on parle de l'ambiguïté, si l'information n'est pas claire et se rapporte à deux éléments différents pour lesquels une distinction est difficile [17].
- **L'imprécision** : est relative aux informations floues qui ne sont pas exactes [18].
- **L'incomplétude** : signifie l'absence d'une partie de l'information, ou bien d'un jugement incomplet donné par un agent du web sémantique [18].
- **L'inconsistance** : on parle de l'inconsistance ou bien de l'incohérence lorsque les informations sont contradictoires [18].
- **L'aléatoire** : on parle de l'aléatoire lorsqu' il s'agit d'un événement dont la survenance est imprédictible et lorsqu'on a pas une loi générale pour expliquer la réalisation des phénomènes[19].

L'incertitude pourrait être représentée de manière claire et lisible en utilisant un réseau bayésien.

1.4 Réseau bayésien

Dans cette section, nous présentons le réseau bayésien, ses domaines d'utilisation, ses types, ses limites ainsi que l'apprentissage et l'inférence dans le réseau bayésien.

1.4.1 Définition

Le réseau bayésien est un modèle graphique probabiliste qui représente des connaissances incertaines sous forme orienté acyclique.

Il se compose de variables aléatoires représentées par des nœuds, liées entre eux par des arcs orientés. Il est également constitué d'une partie probabiliste qui représente un ensemble de tables de probabilité conditionnelles, où chaque nœud est associé à une table de taille exponentielle au nombre de parents et leur nombre d'états [20].

Exemple :

Caroline est souvent au travail, elle s'inquiète d'un cambriolage durant son absence, ce qui fait qu'elle a installé un système d'alarme chez elle. L'alarme se déclenche soit à cause d'un cambriolage, soit à cause d'un séisme [21].

A : L'alarme de Caroline s'est déclenchée.

C : Un cambriolage s'est passé.

S : Un séisme a eu lieu.

La figure 1.3 représente la table de probabilité conditionnelle.

	Cambriolage,Séisme =			
	O,O	O,N	N,O	N,N
Alarme=O	0.75	0.10	0.99	0.10
Alarme=N	0.25	0.90	0.01	0.90

FIGURE 1.3 – La modélisation de la table de probabilité [21]

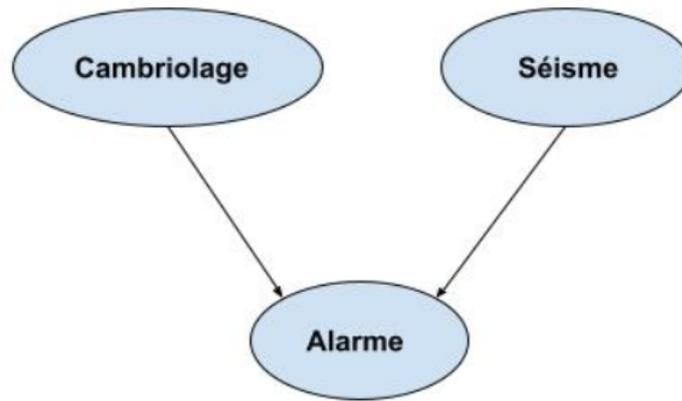


FIGURE 1.4 – Graphe bayésien

L'utilisation du réseau bayésien apporte un plus considérable au niveau de plusieurs tâches, vu qu'elle sert à [22] :

- Représenter graphiquement des connaissances incertaines.
- Faire le raisonnement de données incomplètes.
- Coupler les connaissances des experts et les données disponibles.
- Prévoir le comportement du système.
- Aider au diagnostic.

1.4.2 Apprentissage dans le réseau bayésien

L'apprentissage d'un réseau bayésien à partir d'un ensemble d'observation peut se décomposer en deux étapes : la première est d'apprendre sa structure, se qui consiste à associer un graphe au réseau, une fois sa structure connue, la deuxième se résume à apprendre les bonnes distributions de probabilités conditionnelles des variables.

1. **Apprentissage de la structure** : L'objectif est de trouver la meilleure structure qui représentera le mieux le problème posé. Afin de réaliser l'apprentissage de la structure, il existe deux manières, la première est de trouver la meilleure structure en collaborant avec un expert du domaine, la deuxième est d'apprendre la structure en utilisant des algorithmes précis.

Il existe beaucoup de travaux pour apprendre cette structure à partir des données complètes ou incomplètes. Voici quelques algorithmes basé sur les données complètes et incomplètes :

□ **Algorithmes basés sur un score (Données complètes) :**

Ces algorithmes consistent à parcourir tous les graphes possibles et leur associer

un score puis choisir le graphe ayant le score le plus élevé, cependant, cette méthode est complexe à cause de la taille super exponentielle de l'espace de recherche en fonction du nombre de variables. D'autres méthodes ont été proposées pour résoudre ce problème, telles que : L'arbre de poids maximal, GS (Greedy Search).

□ **Algorithmes de structure (Données incomplètes) :**

Pour réaliser l'apprentissage de la structure à partir des données incomplètes, l'algorithme SEM (Structural EM) a été proposé, cette algorithme combine l'algorithme standard Expectation Maximization (EM) qui permet l'optimisation des paramètres, avec une recherche structurelle pour la sélection du modèle [22].

2. **Apprentissage des paramètres :** L'apprentissage des paramètres se fait à partir des données qui peuvent être complètes ou pas. Il existe différents algorithmes d'apprentissage des paramètres dans les deux cas.

a) **Apprentissage des données complètes :**

On cherche à estimer les distributions de probabilités à partir de données disponibles.

Apprentissage statique : Cette approche, appelée maximum de vraisemblance (MV). C'est la méthode la plus simple et la plus utilisée dans le cas où toutes les variables sont observées, elle consiste à estimer la probabilité d'un événement par la fréquence d'apparition de l'événement dans la base de données [22].

b) **Apprentissage des données incomplètes :**

Dans le cas où les données sont incomplètes et certaines variables ne sont observées que partiellement ou même jamais, que ce soit à cause d'une panne de capteurs, d'une variable mesurable seulement dans un contexte bien précis, d'une personne sondée ayant oublié de répondre à une question, la méthode d'estimation utilisée est l'algorithme itératif EM (Expectation Maximisation) [23].

Algorithme EM (Expectation-Maximisation) : L'algorithme de EM opère deux étapes distinctes à chaque itération :

- i. **Phase Expectation (ou l'étape E) :** elle consiste à estimer des données inconnues, à partir des données observées et de la valeur des paramètres

déterminée à l'itération précédente.

- ii. **Phase Maximisation (ou l'étape M)** : elle consiste à maximiser la vraisemblance, rendue désormais possible en utilisant l'estimation des données inconnues effectuée à l'étape précédente et met à jour la valeur du ou des paramètre(s) pour la prochaine itération.

1.4.3 Inférence dans le réseau bayésien standard

L'inférence est comme une mise à jour de croyance, elle permet de calculer la probabilité a posteriori au niveau de certaines variables en utilisant ce qu'on appelle les observations qui sont des informations qu'on possède déjà sur d'autres variables. Le réseau bayésien met à jour les distributions de probabilité au niveau des variables et cela se fait selon les deux étapes suivantes :

- Associer des valeurs aux variables d'observation.
- Utiliser un algorithme qui peut calculer à posteriori les autres valeurs des variables.

Il existe deux grandes familles d'algorithmes d'inférence qui sont : les algorithmes d'inférence exacte et les algorithmes d'inférence approchée [24].

1. **Inférence exacte** : Cette famille englobe plusieurs algorithmes d'inférence comme l'algorithme d'envoi de message qui a été développé par Judea Pearl connu comme le père de l'inférence dans les réseaux bayésiens. L'idée de cet algorithme consiste pour chaque nœud du réseau, à mettre à jour la distribution de probabilité de proche en proche avec ses voisins pour prendre en compte les observations sur un ou plusieurs nœuds [20].

Sachant que l'inférence exacte est un problème NP complet, d'où la complexité des algorithmes d'inférence dépend de plusieurs facteurs comme le nombre de noeuds du réseau bayésien, sa structure, le nombre de parent d'un noeud, ce qui va amener à utiliser des algorithmes d'inférence approchée.

2. **Inférence approchée** : Du fait des limites de l'inférence exacte qui ne peut pas être utilisée dans des grands réseaux multiplement connectés, il y'a lieu d'opter pour les algorithmes d'inférence approchée qui sont souvent utilisés dans des réseaux bayésiens complexes [22]. Parmi les algorithmes de ce type d'inférence : les algorithmes de Monte-carlo [25] qui donnent des réponses approchées dont l'exactitude dépend du nombre d'échantillons générés.

1.4.4 Types de réseaux bayésiens

Il existe plusieurs extensions de réseaux bayésiens. Parmi les extensions les plus connues :

- **Le réseau bayésien standard** : c'est un réseau bayésien classique, qui est un modèle graphique probabiliste représentant la connaissance et assurant le calcul des probabilités conditionnelles [24]. Il dispose de plusieurs extensions, dont les plus importantes sont :

1. **Le réseau bayésien dynamique** : il permet l'évolution temporelle des variables en fonction d'une séquence discrète. Un réseau bayésien dynamique peut être introduit comme un couple (B1,B2) :

B1 : représente le réseau bayésien standard qui permet la distribution des variables aléatoires qui concerne le temps T0.

B2 : représente le réseau bayésien dynamique qui décrit la transition du pas de temps T-1 au pas de temps T. Ce type de réseau bayésien est inspiré des chaînes Markov cachées qui font que les distributions conditionnelles au temps T dépendent de l'état au temps T-1 [24].

2. **Le réseau bayésien orienté objet** : il permet la combinaison avec la notion d'objet. Il est utilisée lorsque une sous-structure apparaît plusieurs fois [24].
3. **Le réseau bayésien distribué** : il permet la division du réseau en plusieurs sous-réseaux afin de maîtriser le problème de la complexité temporelle et spatiale ainsi que d'assurer un traitement plus facile [24].
4. **Le réseau bayésien flou** : il permet la combinaison avec la théorie des ensembles flous qui est une théorie mathématique du domaine de l'algèbre abstraite. Cette théorie permet de représenter mathématiquement l'imprécision relative à certaines classes d'objets et sert de base à la logique floue [24].
5. **Le réseau bayésien multi-entités (MEBN)** : il permet d'intégrer la logique du premier ordre avec la théorie des probabilités bayésiennes.
Dans ce type de réseau bayésien la connaissance est représentée comme une collection de fragments (MFragments) qui sont organisés en théorie MEBN (MTheories) [26].
6. **Le réseau bayésien hybride** : un réseau bayésien est dit hybride s'il contient un ensemble de variables aléatoires discrètes et continues représentées par les noeuds d'un graphe [27].

1.4.5 Domaines d'application du réseau bayésien

Le réseau bayésien est appliqué dans différents domaines, l'un des principaux domaines d'utilisation est le domaine médical. Il est populaire pour son habilité à transmettre des

prédictions pour des diagnostics de manière probabiliste.

Le réseau bayésien est appliqué aussi dans le domaine de la robotique, il est utilisé pour la création des modèles pour les robots, de manière à ce qu'ils puissent apprendre à se déplacer dans leurs environnements malgré les incertitudes comme les obstacles imprévus. Il y a d'autres domaines d'application pour le réseau bayésien comme le diagnostic informatique tels que la détection et l'analyse des problèmes informatiques, la prédiction de trafic routier, la maîtrise des risques ainsi que la détection des spams et des fraudes...etc [28].

1.5 Les travaux connexes qui traitent de l'intégration de l'incertitude au niveau des ontologies

Dans la littérature plusieurs approches ont été proposées dans le cadre de l'intégration de l'incertitude au niveau des ontologies en se basant sur les réseaux bayésiens. Nous présentons ci-après (par ordre chronologique) quelques approches :

- **(García-Crespo et al., 2010) [29]** : Un système d'aide à la décision dans le domaine de la médecine interne appelé également *Ontology-driven differential diagnosis based on logical inference and probabilistic refinements (ODDIN)* réalisé par García-crespo. Ce système s'appuie sur :
 - Une ontologie (ad hoc) qui contient les critères diagnostiques (signes, symptômes, les tests) et les descriptions des maladies, elle contient aussi des critères diagnostiques faisant référence à d'autres maladies. Cette ontologie est faite en OWL-DL.
 - Un ensemble de règles qui sont stockées dans un fichier différent de celui de l'ontologie et développé avec le langage « JenaRules », ce langage est fondé sur le langage RDF(S) et il utilise une représentation sous forme de triplets RDF.
 - Un fichier externe qui contient un ensemble de probabilités a priori ou prévalences et des probabilités conditionnelles.
 - L'utilisation d'un moteur d'inférence Jena afin de gérer les règles écrites en JenaRules, un moteur d'inférence Pellet pour l'ontologie ainsi qu'une inférence bayésienne pour classer ou ordonner les diagnostics produits.
- **(Laskey et al., 2011) [30]** : Laskey et ses collègues procèdent par extension du langage d'ontologie OWL qui est PROWL (OWL probabiliste). Cette dernière fournit des constructeurs spécifiques afin de modéliser les connaissances à caractères incertains autrement dit de nature probabiliste dans les ontologies et un moteur d'inférence probabiliste. La logique de base de PROWL est celle de la Multi-Entity Bayesian Network (MEBN) qui est un système formel pour la représentation des connaissances intégrant la logique de premier ordre (LPO) avec la théorie bayésienne

des probabilités (BPT). La représentation du MEBN est basée sur les entités, les attributs et les relations de l'ontologie. Les connaissances probabilistes sont exprimées sous forme de collection de MEBN fragments (MFragments). Le Mfragment est une structure représentant des connaissances probabilistes (probabilités conditionnelles) pour un ensemble d'hypothèses et les informations associées par exemple les maladies et les signes. PROWL permet de capturer les connaissances d'un domaine sous forme d'un ensemble de MFragments et utilise les informations probabilistes qui y sont enregistrées afin de produire des réponses aux requêtes.

- **(Fenz, 2012) [31]** : Cette approche proposée par Fenz consiste à construire un réseau bayésien à partir d'une ontologie existante, une table de probabilités conditionnelles (CPT) et le logiciel Norsys Netica Bayesian Network (NNBN). Ce dernier élabore un réseau bayésien à partir des classes, des relations et des instances contenues dans l'ontologie. La construction d'un réseau bayésien s'appuie uniquement sur les classes, les relations et les instances contenues dans l'ontologie et se déroule en quatre étapes :
 1. La sélection des classes, des instances et des relations pertinentes. Cette étape est importante car elle permet à l'expert du domaine de se rassurer qu'il n'y aura pas d'informations redondantes dans le réseau bayésien.
 2. La création de la structure du réseau bayésien est faite automatiquement à partir des classes, des relations et des instances extraites précédemment grâce au logiciel NNBN.
 3. La construction de la table de probabilité conditionnelle (CPT).
 4. L'intégration dans le réseau bayésien des connaissances factuelles existant dans l'ontologie.
- **(Chang et al., 2015) [32]** : Chang et ses collègues ont proposé cette approche, ils ont utilisé le OWL-DL pour construire un système d'aide à la décision dans le domaine médical plus précisément la dépression, il consiste à prédire si le patient a une dépression à partir d'un questionnaire rempli par ce dernier. Afin de construire ce modèle, une ontologie est utilisée pour formaliser un concept spécifique, où chaque nœud représente un individu. Afin de déduire les relations entre les individus, ils ont défini les règles d'inférence qui peuvent être exécutées par le moteur d'inférence SMILE. Cependant, le DOM (depression ontology model) ne peut pas représenter la probabilité d'une instance, il est nécessaire de fusionner le modèle avec le réseau bayésien pour que le DOM accepte les probabilités. Ils ont utilisé OntoBayes pour fusionner le DOM avec le réseau bayésien afin de déduire les probabilités des nœuds. En effet, OntoBayes qui est proposé par Yang et Calmet (Yi Yang Calmet, 2005) sert à intégrer les réseaux bayésiens (RB) dans le langage Web ontologique (OWL) afin de préserver les avantages des deux, ce modèle utilise le langage OWL annoté par les probabilités et les dépendances pour représenter les informations incertaines

dans les structures RB. Ces extensions améliorent la représentation des connaissances dans OWL et permettent aux agents d'agir à la fois dans l'incertitude et dans des environnements ouverts, structurés et complexes.

- **(Hu and Kerschberg, 2018) [33]** : Cette approche proposée par Hu et Kerschberg présente une nouvelle méthodologie permettant d'utiliser les connaissances causales pour étendre et améliorer une ontologie médicale hiérarchique standard. Tout d'abord, ils ont obtenu la structure hiérarchique des variables, des symptômes du patient en se basant sur le dictionnaire médical de la terminologie des activités réglementaires (Medical Dictionary for Regulatory Activities Terminology (MedDRA)) qui est un exemple de Authoritative Medical Ontologies (AMO). Ensuite, ils ont utilisé un réseau bayésien causal (CBN) en utilisant Max-Min Hill-Climbing (MMHC) qui est un algorithme d'apprentissage hybride basé sur les contraintes et les scores et sur l'ensemble de données préexistantes de patients de l'étude Sequenced Treatment Alternatives to Relieve Depression (STAR*D) du National Institutes of Mental Health (NIMH). Enfin, ils ont utilisé les liens de causalité découverts dans le CBN pour faire évoluer l'ontologie et sa hiérarchie.
- **(Li et al., 2019) [34]** : Uncertain Knowledge Reasoning Based on the Fuzzy Multi Entity Bayesian Networks proposé par Li et ses collègues. C'est une nouvelle méthode de raisonnement de la probabilité floue en fonction du Fuzzy MEBN basée sur le langage Fuzzy PR-OWL qui décrit la sémantique floue et les relations incertaines et donne une interprétation sémantique. Dans cette approche, ils ont utilisé un réseau bayésien multi entité flou qui est basé sur la logique floue du premier ordre afin de traiter l'incertitude qui se présente dans leurs contexte sous forme de l'incomplétude et l'imprécision. Li et ses collègues ont proposé un algorithme de propagation des croyances floues (fuzzy belief propagation algorithm) en intégrant l'algorithme de propagation des croyances à la théorie des probabilités floues afin de l'appliquer au cadre de représentation et de raisonnement de l'ontologie PR-OWL floue.

1.5.1 Comparaison des travaux

À travers les tableaux 1.1 et 1.2, nous présentons une comparaison entre les travaux cités selon les critères suivants :

- Le langage utilisé.
- L'objectif.
- Le type de réseau bayésien.
- Le principe de l'approche.

- Les limites.

/ : cela n'est pas spécifié dans l'article de base.

Le tableau 1.1 représente le langage, le réseau bayésien utilisé ainsi que l'objectif de chaque travail.

Travaux	Langage utilisé	Type de RB	Objectif
(García-Crespo et al., 2010) [29]	JenaRules/ OWL-DL	/	Création d'un système d'aide à la décision dans le domaine de la médecine interne
(Laskey et al., 2011) [30]	PR-OWL	MEBN	Création d'une ontologie probabiliste dans le domaine maritime.
(Fenz, 2012) [31]	OWL	Standard	Création d'un modèle qui intègre le raisonnement probabiliste
(Chang et al., 2015) [32]	OWL-DL	Standard	Création d'un système qui permet l'aide au diagnostic de la dépression
(Hu and Kerschberg, 2018) [33]	OWL	RB causal	Évolution des ontologies médicales.
(Li et al., 2019) [34]	Fuzzy PR-OWL	Fuzzy MEBN	Présentation d'un nouveau cadre de raisonnement basé sur la représentation du PR-OWL flou.

TABLE 1.1 – Tableau comparatif des langages, des réseaux bayésiens utilisés et des objectifs

Le tableau 1.2 représente le principe et les limites de chaque travail.

Travaux	Principe de l'approche	Limites
(García-Crespo et al., 2010) [29]	Utilisation d'une ontologie (ad hoc), un ensemble de règles, un moteur d'inférence pour gérer ces règles et un fichier externe qui contient un ensemble de probabilités conditionnelles	Son modèle de représentation ontologique n'intègre pas les signes absents et les probabilités sont exprimées hors du modèle ontologique.
(Laskey et al., 2011) [30]	Utilisation d'une ontologie probabiliste en se basant sur les réseaux bayésiens multi-entités	Le PR-OWL n'est pas encore recommandé par le W3C [31], temps d'exécution est lent et le MEBN ne traite que les données discrètes.
(Fenz, 2012) [31]	Construction d'un réseau bayésien à partir d'une ontologie existante, une CPT et le NNBN	Les réseaux bayésiens standards ne permettent pas une bonne expressivité par rapport aux autres types des réseaux bayésiens et il ne traitent que les données discrètes.
(Chang et al., 2015) [32]	Utilisation de la méthode OntoBayes afin de fusionner le DOM avec le RB	Les réseaux bayésiens standards ne permettent pas une bonne expressivité par rapport aux autres types des réseaux bayésiens et il ne traitent que les données discrètes.
(Hu and Kerschberg, 2018) [33]	Combinaison d'un réseau bayésien causal avec une ontologie (MedDRA) en utilisant l'algorithme MMHC	Le réseau bayésien n'ajoute pas une grande valeur à celle du réseau bayésien classique et il ne traite que les données discrètes.
(Li et al., 2019) [34]	Utilisation d'une ontologie de probabilités floues en se basant sur les réseaux bayésiens multi-entités flous	Le MEBN engendre un temps de réponse très lent.

TABLE 1.2 – Tableau décrivant les méthodes utilisées et les limites des travaux

1.5.2 Analyse des travaux

Après avoir fait une comparaison entre ces travaux, nous constatons que :

- Chaque approche se base sur une méthode différente en utilisant des langages différents avec un objectif précis :
 - García-Crespo et ses collègues [29] ont créé un système d'aide à la décision ODDIN dans l'objectif de l'utiliser dans le domaine de la médecine interne, ce système se base sur l'inférence logique et les raffinements probabilistes, en utilisant les langages JenaRules et OWL-DL.
 - Laskey et ses collègues [30] ont opté pour la création d'une ontologie probabiliste en se basant sur les réseaux bayésiens multi-entités en utilisant PR-OWL, dans le

but de l'utiliser dans le domaine maritime.

- Fenz [31] a construit un réseau bayésien à partir d'une ontologie, une table de probabilités conditionnelles et le logiciel Norsys Netica Bayesian Network (NNBN) en utilisant OWL afin de créer un modèle qui intègre le raisonnement probabiliste.
 - Chang et ses collègues [32] ont combiné les ontologies et les réseaux bayésiens standards en utilisant la méthode OntoBayes tout en utilisant le OWL-DL afin de créer un système qui permet l'aide au diagnostic de la dépression.
 - Hu et Kerschberg [33] ont basé sur un réseau bayésien causal afin d'utiliser les liens découverts par ce dernier dans le but d'évoluer et améliorer une ontologie médicale. Dans cette approche, ils ont utilisé l'algorithme MMHC.
 - Li et ses collègues [34] ont utilisé une ontologie probabiliste floue en se basant sur les réseaux bayésiens multi-entités flous qui adopte la logique floue du premier ordre, ils ont utilisé le langage Fuzzy PR-OWL afin de représenter et raisonner les connaissances incomplètes et imprécises.
- Ces approches ont été utilisées dans des domaines différents, le domaine médical pour la plupart des travaux ou même dans le domaine maritime pour Laskey et ses collègues [30].
 - Les approches qui ont utilisé les réseaux bayésiens multi-entité (MEBN) et PR-OWL comme langage avaient comme inconvénient le temps de réponse qui est très lent. De plus, les réseaux bayésiens multi-entités, causaux et standards ne traitent que les données discrètes.
 - L'utilisation d'un réseau bayésien causal n'ajoute pas une grande valeur à celle du réseau bayésien classique. En effet, tous les arcs d'un réseau bayésien causal représentent des relations de causalité entre les noeuds, de même, certains arcs d'un réseau bayésien classique peuvent représenter des relations de causalité.

À partir de notre étude des travaux mentionnés précédemment, nous concluons que pour traiter l'incertitude qui se présente dans notre contexte sous forme de l'aléatoire et de l'incomplétude, nous pouvons combiner les réseaux bayésiens et les ontologies, ce que nous allons voir dans les prochains chapitres.

1.6 Conclusion

À travers ce chapitre nous avons donné une vision générale de l'ontologie, de ses propriétés et caractéristiques. Par la suite nous avons montré qu'elle présente une limite dû à son incapacité de représenter et de raisonner sur des connaissances incertaines et pour cela nous avons abordé le réseau bayésien qui traite l'incertitude, tout en démontrant ses différentes techniques d'apprentissage de la structure et des paramètres, ainsi que son inférence.

À la fin de ce chapitre nous avons présenté quelques travaux liés à l'utilisation de l'ontologie et du réseau bayésien, ensuite nous avons fait une comparaison de ces approches tout en précisant leurs méthodes, leurs objectifs et leurs limites. Dans le chapitre suivant, nous aborderons la maladie qui nous intéresse dans notre travail qui est le Covid-19 ainsi que quelques travaux connexes au Covid-19.

DIAGNOSTIC DU COVID-19

2.1 Introduction

Depuis bientôt deux ans, l'humanité continue de vivre l'une des plus grandes pandémies qu'a connu l'histoire de l'homme moderne : le Covid-19.

Aujourd'hui, avoir de la fièvre peut ne pas être que le symptôme d'une simple grippe, mais celui du Covid-19. En effet, les médecins sont obligés de prendre plusieurs décisions en fonction des symptômes, signes et facteurs de risque. Ce qui n'est pas évident à cause de l'incertitude qui se présente dans notre contexte sous forme de l'aléatoire et de l'incomplétude.

À travers ce chapitre nous allons aborder la notion de l'incertitude dans le domaine médical en général, ensuite nous allons présenter le Covid-19, ses différents variants et citer quelques symptômes, signes, facteurs de risque, tests de laboratoire et quelques travaux de recherches en rapport avec le Covid-19.

2.2 Incertitude en médecine

En médecine, la prise de décision médicale est considérée comme étant une compétence centrale dans la pratique médicale. Son processus consiste à poser un diagnostic, proposer un traitement ou le différer. Toutefois, la décision médicale est caractérisée par l'incertitude qui se définit comme étant « l'incapacité d'avoir la certitude sur l'étiologie des troubles, et leur évolution potentielle vers une maladie » [35], en effet, les compétences humaines limitées, la complexité et les incertitudes de la maladie contribuent à de grandes variations dans la pratique clinique et à des erreurs cliniques.

Il y a plusieurs sources de l'incertitude médicale, d'une part elle peut être liée à la personne malade qui pourra modifier le risque en expliquant de façon confuse les symptômes ressentis, d'autre part, elle peut être liée à l'étape de soin, il existe trois formes d'incertitude de

l'offre de soins :

- Le diagnostic médical consiste à dévoiler toutes les informations nécessaires par une interrogation ou des analyses cliniques, cependant, l'information dévoilée n'est jamais complète car les investigations du médecin sont sélectives.
- Il y a l'incertitude qui concerne la thérapie, cette dernière peut ne pas être efficace même si le diagnostic est correct.
- Il existe l'incertitude qui concerne les connaissances liées à une maladie ou même à une thérapie.

2.3 Définition du Covid-19

Depuis la grippe espagnole de 1918, l'humanité n'a pas vécu une pandémie aussi grave que celle du Covid-19 [36].

Le Covid-19 fait référence à "Coronavirus Disease 2019", cette maladie a été provoquée par un virus de la famille des Coronaviridae qui est connu chez les poules depuis 1937 et chez les être humains depuis 1960, il a émergé en décembre 2019 en Chine [36].

Selon l'organisation mondiale de la santé (OMS) [37], le Covid-19 est une maladie respiratoire pouvant être mortelle surtout chez les personnes âgées et celles qui ont un problème médical tel qu'un diabète, un cancer, une maladie cardio-vasculaire ou une maladie respiratoire chronique. En outre, c'est une maladie contagieuse qui se transmet par contact rapproché avec des personnes infectées.

2.4 Impacts du Covid-19

Vers la fin de l'an 2019, la ville chinoise Wuhan a connu une maladie respiratoire infectieuse émergente très dangereuse et dévastatrice plus tard reconnue comme le Covid-19. Après peu de temps, cette maladie a touché le monde entier. Ayant été déclarée pandémie par l'OMS le 11 mars 2020, cette épidémie a complètement changé le cours de la vie humaine telle qu'on la connaissait.

D'après l'OMS [38], il a été recensé à l'échelle mondiale à la date du 13 juin 2022 : 532 887 351 cas confirmés de Covid-19, dont 6 307 021 décès.

Au 7 juin 2022, un total de 11 854 673 610 doses de vaccin ont été administrées. L'évolution fulgurante de la pandémie a contraint les états ainsi que les organisations mondiales à adopter en urgence le confinement des populations et la limitation de la circulation des biens et des personnes, s'en est suivi un chamboulement total aussi bien au niveau sociétal, social et économique. Ne disposant ni de vaccins ni de médicaments dédiés à la maladie, le monde a dû s'organiser comme il le pouvait en attendant une percée scientifique qui mettrait à sa disposition les solutions médicales nécessaires et suffisantes pour endiguer à

défaut d'éradiquer le fléau.

Les impacts de cette pandémie sont d'ordre multiple et ont touchés le monde dans sa globalité et dans sa diversité [39] :

- **L'intégrité humaine** : décès, hospitalisation, invalidité, séquelles physiques et mentales post Covid.
- **Le ralentissement de l'économie** : baisse des investissements, de la production, des dépenses des ménages, de l'emploi et du commerce.
- **L'augmentation des charges sociales et des déficits des budgets d'états.**
- **La dégradation de l'enseignement** : la fermeture des écoles et des universités a entraîné la rupture d'apprentissage pour centaines de millions d'élèves et d'étudiants.
- **L'augmentation de la pauvreté** : la crise économique engendrée par le Covid-19 a fait basculer des milliers de personnes dans la pauvreté absolue.

2.5 Symptômes

Un symptôme est une manifestation d'une maladie contribuant au diagnostic, autrement dit un trouble qui peut être ressenti par le patient et parfois même constaté par d'autres personnes. [40].

Selon l'OMS [37] :

- **Les symptômes les plus fréquents du Covid-19**
 - Fièvre.
 - Toux sèches.
 - Courbatures.

- **Les symptômes les moins fréquents du Covid-19**
 - Perte du goût et de l'odorat.
 - Conjonctivite.
 - mal de gorge.
 - Diarrhé.
 - Frissons.
 - Nausées ou vomissements.

- **Les symptômes les plus graves du Covid-19**
 - Essoufflement.
 - Perte d'appétit.
 - État confusionnel.

- Douleurs ou sensation d'oppression persistantes dans la poitrine.
- Dé-saturation.
- Persistance de la fièvre.

2.6 Signes

En médecine humaine, un signe est la manifestation objective d'une pathologie, constatée par le médecin. On distingue deux types de signes : le signe clinique et le signe paraclinique [40].

2.6.1 Signe clinique

Il se présente sous trois formes [41] :

- **Signe général** : il est lié à l'organisme dans sa globalité, exemple : sueurs, amaigrissement, fièvre etc. . .).
- **Signe fonctionnel ou symptôme** : c'est un signe ressenti subjectivement par le patient, par exemple (les démangeaisons, la fatigue etc. . .)
- **Signe physique** : c'est un signe constaté par le médecin lors d'un examen clinique (par l'inspection, la palpation etc. . .).

2.6.2 Signe paraclinique

C'est un signe détecté par des examens non cliniques tels que les radios, les scanners, les analyses, les IRM etc [41].

2.6.3 Les signes paraclinique du Covid-19

Selon l'OMS [37], les signes paracliniques sont les suivants :

- Des opacités en verre dépoli, non systématisées, multifocales, bilatérales et asymétriques, de localisation sous pleurale à prédominance basale et postérieure.
- Une lymphocytopénie est fréquemment constatée.
- Le taux de CRP et de ferritine augmente chez les patients malades.
- Le taux de D-dimers augmente.
- Présence d'ARN Covid-19 TDM thoracique en faisant le PCR.
- Présence d'antigène Covid-19 en faisant le test antigénique.

2.7 Facteur de risque

Un facteur de risque est tout attribut, caractéristique augmentant la possibilité de développer une maladie ou de souffrir d'un traumatisme [42].

2.7.1 Facteurs de risque du Covid-19

Selon le centre national d'information sur la biotechnologie(NCBI) [43] les facteurs de risque sont les suivants :

- L'âge
- Les maladies pulmonaires
- L'immunodépression
- L'obésité
- Le diabète
- L'HTA
- Le cancer
- Les maladies cardiovasculaire
- Les maladies rénales et les maladies hépatiques

2.8 Tests du Covid-19

Les tests du Covid-19 selon l'OMS [37] sont les suivants :

Test PCR :

Le test PCR représente l'examen clé pour le diagnostic de l'infection Covid-19. Il permet de déterminer au moment du prélèvement si la personne est porteuse du virus ou pas. Ce test nécessite un prélèvement nasopharyngé qui consiste à introduire un écouvillon très fin dans une narine, ce qui va prendre environ 15 à 30 secondes, les résultats peuvent être disponibles en 24h.

Test antigénique :

Le test antigénique a comme objectif de déterminer si la personne est infectée par le Covid-19 au moment du test ou pas. Il se fait via un prélèvement nasopharyngé tout comme le PCR. Il se différencie sur les molécules recherchés vu que ce dernier vise les

antigènes, autrement dit la présence de protéines du virus Sars-Cov-2 dans l'échantillon. Ce test prend environ 15 à 30 minutes pour donner les résultats.

Test sérologique :

Le test sérologique Covid-19 consiste à rechercher la présence d'anticorps dirigés contre le virus à partir d'une prise de sang, ces anticorps apparaissent dans les jours ou les semaines qui suivent une infection au Sars-Cov-2. Ce test permet de savoir si la personne concernée a rencontré le virus, sans en préciser la date. Par ailleurs, il ne remplace pas le test PCR, cela veut dire qu'il ne permet pas de savoir si la personne est contagieuse. Il existe deux types de ce test :

- Les tests ELISA automatisables : ils ne peuvent être faits qu'au sein d'un laboratoire de biologie médicale, ses résultats sortent en quelques heures.
- Les tests sérologiques rapides dits "TROD" : ils peuvent être réalisés en pharmacie, le résultat est obtenu en 15 à 30 minutes.

Les autotests :

Un autotest est un test à réaliser soi-même, il permet de rechercher la présence du virus Sars-Cov-2 à l'aide d'un prélèvement nasal. Les autotests ont une sensibilité plus faible, ils ne peuvent pas remplacer les tests antigéniques ou le PCR, ils peuvent trouver leur efficacité dans une utilisation répétée. Les autotest délivrent leur résultat en 15 à 30 minutes.

2.9 Variants du Covid-19

Il existe plusieurs variants du Covid-19, nous citons les variants les plus préoccupants selon l'OMS [37] :

Alpha :

Connu comme B.1.1.7, il a été détecté pour la première fois en septembre 2020 au Royaume-Uni. Il s'est propagé dans plus de 50 pays et selon une étude médicale publiée dans la revue médicale BMJ, il est non seulement plus contagieux mais aussi 64% plus mortel que le coronavirus classique.

Bêta :

Surnommé variant sud-africain(B.1.351), il a été identifié pour la première fois en décembre 2020 en Afrique du Sud. Selon les scientifiques, il présente 13 mutations principales dont 8 sur la protéine S. Il est également connu par sa vitesse de propagation, sachant qu'il était rapporté dans 58 pays.

Gamma :

C'est un variant brésilien qui est désigné par l'appellation : 20J/501Y.V3, il a été renommé "Gamma" en mai 2021. Ce variant a provoqué une flambée épidémique à Manaus, la capitale de l'Etat d'Amazonas. Il a été détecté pour la première fois chez un touriste japonais au Brésil en janvier 2021. En seulement un mois, ce variant est devenu dominant dans la région de Manaus et par la suite il a été signalé dans 68 pays.

Delta :

Appelé par B.1.617, il a été détecté dans l'ouest de l'Inde en mai 2021. Il est qualifié de "double mutant" parce qu'il est notamment porteur de deux mutations préoccupantes au niveau de la protéine de pointe ("spike") du virus Sars-CoV-2. Ce variant est signalé dans 80 pays au total.

Omicron :

Connu par B.1.1.529, il a été détecté pour la première fois en Afrique du Sud . Le 26 novembre 2021 l'OMS a déclaré ce dernier comme un variant préoccupant. Ce variant présente un grand nombre de mutations, dont certaines sont inquiétantes. Il est moins virulent que les variants précédents mais beaucoup plus contagieux.

2.10 Les travaux connexes au Covid-19

Pendant ces dernières années, plusieurs travaux ont vu le jour afin d'alléger la charge du Covid-19 sur les systèmes de santé. Ces travaux ont pour objectif de prédire l'infection par le Covid-19, sa propagation et l'évolution du nombre de personnes contaminées etc. Globalement il existe trois approches, qui se basent respectivement sur des réseaux bayésiens, des ontologies avec d'autres classifieurs et des méthodes d'apprentissage profond. Au vu de notre contexte nous nous intéressons aux travaux qui se basent sur les ontologies et les réseaux bayésiens.

□ Travaux basés sur les réseaux bayésiens :

- **(Fenton et al., 2020) [44] :**

Ce travail se concentre sur le calcul de la probabilité qu'une personne ait le Covid-19 en utilisant un réseau bayésien qui va tenir en compte les symptômes, les facteurs de risque et aussi les interactions sociales récentes et cela se fait à l'aide des smartphones qui vont

s'échanger des clés numériques via bluetooth, dans le cas où les propriétaires de ces smartphones se trouvent dans un rayon de 2 mètres pendant plus de 15 min. Si un utilisateur est infecté par le Covid-19, les autres utilisateurs qui sont près de lui vont être mis au courant de cette proximité dangereuse et ils sauront qu'un isolement de 2 semaines est à prévoir.

- **(Butcher et Fenton, 2020)[45] :**

Ce travail s'appuie sur la solution du réseau bayésien de (Fenton et al., 2020) [44] et sur le travail de (Prodhan et al., 2020) [46]. Il a pour objectif de réaliser un modèle qui permet la prédiction de l'infection par le Covid-19. Ce modèle tient compte des symptômes, des facteurs de risques et même d'autres données qui peuvent contribuer à la prédiction tel que la profession, l'origine ethnique et d'autres données démographiques.

- **(Neil et al., 2020) [47] :**

Dans ce travail ils ont utilisé un réseau bayésien pour estimer le taux de prévalence de l'infection par le Covid-19 ainsi que son taux de létalité dans plusieurs pays. L'utilisation du réseau bayésien leur a permis de combiner plusieurs sources de données dans un seul modèle qui donne des estimations statistiques qui reflète mieux l'incertitude.

- **(de Terwangne et al., 2020) [48] :**

Cette étude a pour objectif d'évaluer la précision prédictive de la classification de gravité du Covid-19 faite par l'OMS. Cela se fait en comparant son pouvoir prédictif avec un modèle de prédiction appelé Covid-19 EPI-SCORE qui se base sur un modèle probabiliste (un réseau bayésien).

- **Travaux basés sur les ontologies :**

- **(Oyelade et Ezugwu, 2020) [49] :**

Cette étude s'appuie sur un modèle de raisonnement à base de cas (CBR) qui permet d'utiliser des connaissances spécifiques à des cas précédemment expérimentés et validés. En utilisant une approche de traitement du langage naturel pour analyser les enregistrements

de cas et une ontologie pour modéliser ces derniers, afin de faire une détection et un diagnostic du Covid-19.

- **(Kouamé et al., 2021) [50] :**

Ce travail propose des approches basées sur des mécanisme d'adaptation pour détecter le plus vite possible la présence des symptômes du Covid-19 chez les personnes souffrantes de la maladie COPD afin de leur fournir le soin adapté, en utilisant un modèle ontologique nommé Suspected COPD covidOlogy qui se constitue de 5 ontologies chacune utilisé pour un objectif précis en se basant sur un lanagage de règles (Semantic Web Rule Language).

□ **Parmi les travaux basés sur les méthodes d'apprentissage profond :**

- **(Abbasi-mehr et Paki, 2021)[51] :**

Dans cette étude, ils ont fait une combinaison de trois méthodes d'apprentissage profond (l'attention multi tête, la mémoire à long terme et le réseau de neurones convolutifs(CNN))avec l'algorithme d'optimisation bayésienne, afin de prévoir les cas confirmés du Covid-19.

2.10.1 Comparaison des travaux

À travers les tableaux 2.1, 2.2 et 2.3, nous présentons une comparaison entre les travaux cités selon les critères suivants :

- L'objectif.
- Le modèle utilisé.
- Les données utilisées.
- Les symptômes.
- Les avantages et les limites.

/ : cela n'est pas spécifié dans l'article de base.

Le tableau 2.1 représente l'objectif et le modèle utilisé pour chaque travail.

Travaux	Objectif	Modèle utilisé
(Fenton et al., 2020) [44]	Le diagnostic prédictif du Covid-19.	Réseau bayésien causal.
(Butcher et Fenton, 2020) [45]	Le diagnostic prédictif du Covid-19.	Réseau bayésien causal.
(Neil et al., 2020) [47]	L'estimation du taux de prévalence de l'infection par le Covid-19 et de sa létalité dans plusieurs pays.	Réseau bayésien classique.
(de Terwangne et al., 2020) [48]	L'évaluation de la précision prédictive de la classification de gravité du Covid-19 faite par l'OMS.	Réseau bayésien classique.
(Oyelade et Ezugwu, 2020) [49]	La détection et l'aide au diagnostic du Covid-19.	Combinaison de la force de traitement du langage naturel avec l'ontologie et le raisonnement à base de cas.
(Kouamé et al., 2021) [50]	Un modèle capable de détecter les malades qui souffrent de la maladie COPD et qui sont susceptibles d'être infectés par le Covid-19 afin de leur offrir un bon soin avant qu'il soit trop tard.	Ontologie + un mécanisme d'adaptation et de raisonnement à base de cas.
(Abbasi-mehr et Paki, 2021) [51]	La prédiction des cas confirmés du Covid-19.	Combinaison de trois méthodes d'apprentissage profond avec un algorithme d'optimisation bayésienne.

TABLE 2.1 – Tableau comparatif des objectifs et des méthode utilisées

Le tableau 2.2 représente les symptômes et les données utilisées pour chaque travail :

Travaux	Données utilisées	Symptômes
(Fenton et al., 2020) [44]	Les symptômes, les facteurs de risque, les interactions sociales récentes.	/
(Butcher et Fenton, 2020) [45]	Les symptômes, les données démographiques, les facteurs de risque	Les nausées ou les vomissements, la toux, l'essoufflement, les douleurs abdominales et les douleurs thoraciques.
(Neil et al., 2020) [47]	Les données disponibles sur les enquêtes sérologiques, le nombre de décès, la précision des tests.	/
(de Terwangne et al., 2020) [48]	La classification de la gravité selon l'OMS, l'insuffisance rénale aiguë, l'âge, les niveaux de lactate déshydrogénase (LDH), les lymphocytes et le temps de prothrombine activé (APTT), les symptômes.	La fièvre, la toux, la dyspnée, la douleur thoracique, Dysgueusie ou anosmie, Myalgie, le maux de tête, les nausées, le vomissement, la douleur abdominale, la diarrhée.
(Oyelade et Ezugwu, 2020) [49]	La base de données des cas confirmées du Covid-19.	Fièvre, Anosmie, Pneumonie, Syndrome de détresse respiratoire aiguë (SDRA), insuffisance organique, Dyspnée, Nausées et vomissements, Maux de tête, Diarrhée, Infections des voies respiratoires, Souffle court, Rhinorrhée.
(Kouamé et al., 2021) [50]	Les signes vitaux, les symptômes du Covid-19, les facteurs physiques et psychologiques du patient, les antécédents médicaux, des données sur la maladie (COPD), age, sexe.	Fatigue extrême, toux, maux de tête, difficultés respiratoires, dyspnée, vomissements, douleur abdominale, perte sensorielle, fièvre, frissons, perte de goût, raideur musculaire, diarrhée, gastro-intestinaux.
(Abbasi-mehr et Paki, 2021) [51]	Les données de série chronologique du Covid-19.	/

TABLE 2.2 – Tableau comparatif des symptômes et des données utilisées

Le tableau 2.3 représente les avantages et les limites de chaque travail :

Travaux	Avantages	Limites
(Fenton et al., 2020) [44]	Le RB assure des prédictions précises même avec un manque de données d'observations.	Il n'existe aucun moyen efficace de vérifier la validité des informations saisies par l'utilisateur sans compromettre sa confidentialité.
(Butcher et Fenton, 2020) [45]	Le RB ne dépend pas de l'observation d'un résultat de test clinique pour une prédiction exacte, il donne de bons résultats même avec un manque de données d'observation.	/
(Neil et al., 2020) [47]	L'utilisation d'un RB permet de donner de bons résultats même avec un manque de données d'observation.	Plusieurs données importantes sont inaccessibles, le manque de la transparence.
(de Terwangne et al., 2020) [48]	L'utilisation d'un modèle bayésien permet d'utiliser le score sans disposer de toutes les informations sur les variables du modèle car le théorème de Bayes permet de mettre à jour la sortie de manière incrémentielle à la réception d'éléments de preuve décrivant le malade.	Le manque de données dans les informations anamnestiques, notamment dans les symptômes secondaires.
(Oyelade et Ezugwu, 2020) [49]	le CBR est efficace pour la résolution des problèmes et l'automatisation de l'apprentissage par un agent, il a la capacité de se coupler avec l'apprentissage, il compile les solutions passées, évite les anciennes erreurs, interprète les règles et soutient l'acquisition de connaissances .	Le choix du meilleur modèle de mesure de distance pour le calcul de la similarité des cas est un problème qui nécessite une solution optimale compte tenu de la sensibilité des cas médicaux.
(Kouamé et al., 2021) [50]	l'ontologie permet la manipulation de plusieurs règles, son utilisation est très utile dans l'extraction des connaissances. Elle fournit une expressivité élevée, permet d'être un support de techniques de raisonnement et de réutiliser les représentations existantes.	/
(Abbasi-mehr et Paki, 2021) [51]	La conception de ce modèle est basé sur la stratégie de prévision à sorties multiples qui permet de prévoir plusieurs jours à venir.	Ce modèle a besoin de beaucoup de données afin de fournir des résultats précis et exacts.

TABLE 2.3 – Les avantages et les limites des travaux

2.10.2 Analyse des travaux

Après l'étude des travaux existants qui se base sur les méthodes d'apprentissage profond, sur les ontologies et d'autres sur les réseaux bayésiens, nous constatons que :

- Chaque travail a utilisé des données différentes selon son besoin et son objectif :
 - Fenton et ses collègues [44] ont utilisé comme données : les symptômes, les facteurs de risque et les interactions sociales récentes, afin de prédire l'infection par le Covid-19.
 - Butcher et Fenton [45] ont utilisé comme données : les symptômes, les données démographiques, les facteurs de risque et la profession, dans le but de prédire l'infection par le Covid-19.
 - Neil et son équipe [47] leur travail a été basé sur : les données disponibles sur les enquêtes sérologiques, le nombre de décès et la la précision des tests, afin de faire une estimation du taux de prévalence de l'infection par le Covid-19 et de sa létalité dans plusieurs pays.
 - de Terwangne et ses collègues [48] : Dans leur travail, ils avaient besoin des données suivantes : la classification de la gravité selon l'OMS [37], quelques facteurs de risque tels que (l'insuffisance rénale aiguë et l'âge), les niveaux de lactate déshydrogénase (LDH), les lymphocytes, le temps de prothrombine activé (APTT) et les symptômes, dans l'objectif d'évaluer la précision prédictive de la classification de gravité du Covid-19 faite par l'OMS. [37].
 - Kouamé et son équipe [50] : Afin de détecter les malades qui souffrent de la maladie COPD et qui sont susceptibles d'être infectés par le Covid-19, ils ont utilisés les données suivantes : les signes vitaux, les symptômes du Covid-19, les facteurs physiques et psychologiques du patient, les antécédents médicaux, les données sur la maladie (COPD), l'âge et le sexe.
 - Oyelade Ezugwu [49] : Dans leur travail, ils ont utilisé la base de données de cas confirmés du Covid-19 pour la prédiction du Covid-19.
 - Abbasimehr Paki[51] : Ils ont utilisé les données de série chronologique du Covid-19 pour La prédiction des cas confirmés du Covid-19.
- Chaque travail se base sur divers symptômes et facteurs de risque qui diffèrent d'un modèle à un autre, il y a des symptômes qui sont communs entre les travaux et d'autres qui ont été pris en considération dans seulement quelques uns et qui peuvent avoir un impact sur le résultat obtenu :
 - Butcher et Fenton [45] : les nausées ou les vomissements, la toux, l'essoufflement, les douleurs abdominales et les douleurs thoraciques.

- de Terwangne et ses collègues [48] : La fièvre, la toux, la dyspnée, la douleur thoracique, la dysgueusie ou l'insomnie, la myalgie, le maux de tête, les nausées, le vomissement, la douleur abdominale et la diarrhée.
- Kouamé et son équipe [50] : La fatigue extrême, la toux, le maux de tête, la difficultés respiratoires, la dyspnée, les vomissements, la douleur abdominale, la perte sensorielle, la fièvre, les frissons, la perte de goût, la raideur musculaire, la diarrhée, les troubles gastro-intestinaux.
- Oyelade et Ezugwu [49] : La fièvre, l'anosmie, la pneumonie, le syndrome de détresse respiratoire aiguë, l'insuffisance organique, la dyspnée, les nausées et les vomissements, le maux de tête, la diarrhée, l'infection des voies respiratoires, le souffle court et la rhinorrhée.

- L'utilisation des méthodes d'apprentissage profond nécessite un dataset volumineux et varié afin d'assurer un bon apprentissage.
- Les ontologies sont un outil très efficace et utile pour la représentation des connaissances d'une manière simple et formelle, par contre elle ne permettent pas le raisonnement sous l'incertitude.
- Les réseaux bayésiens ont la capacité de gérer l'incertitude, de plus ils peuvent maintenir des prédictions précises même avec peu de données d'observation [52].
- En ce qui concerne les travaux basés sur les réseaux bayésiens : les deux travaux de Fenton et ses collègues [44] et celui de Butcher et Fenton [45] ont utilisé un réseau bayésien causal, cependant les deux autres travaux de Neil [47] et de Terwangne [48] ont opté pour l'utilisation d'un réseau bayésien classique.

À travers notre analyse des travaux mentionnés ci-dessus nous avons constaté qu'il n'existe pas un modèle qui intègre le réseau bayésien dans l'ontologie pour la prédiction du Covid-19, cela nous a motivé à utiliser une ontologie pour la représentation des connaissances avec un réseau bayésien pour le traitement de l'incertitude. Cela est l'objectif du troisième chapitre.

2.11 Conclusion

Dans ce chapitre nous avons présenté le Covid-19, ses variants, ses symptômes, ses facteurs de risque ainsi que quelques travaux existants qui concernent la prédiction et la propagation du Covid-19 basés sur les réseaux bayésien, les ontologies et les méthodes d'apprentissage profond afin de faire une analyse sur ces derniers et déduire leurs avantages et limites. À travers cette analyse nous avons confirmé l'importance des réseaux bayésiens dans le traitement de l'incertitude ainsi que l'utilité des ontologies dans la représentation des connaissances. Dans le chapitre suivant nous allons présenter la conception de notre système.

CONCEPTION DU SYSTÈME

3.1 Introduction

Après avoir présenté tout ce qui concerne les éléments fondamentaux de notre système, que sont les ontologies et les réseaux bayésiens, ajoutée à cela, la maladie à laquelle notre étude s'intéresse, le Covid-19, nous sommes arrivées à l'étape de la conception à travers laquelle nous allons définir une solution pour notre problématique et proposer un système permettant l'aide au diagnostic du Covid-19.

Dans ce chapitre nous allons justifier notre motivation pour le choix de cette maladie et les raisons pour lesquelles ce projet a été élaboré. En deuxième lieu, nous allons expliquer les étapes de conception, en commençant par la construction du dataset, qui sera suivi par la présentation du scénario global de notre système ainsi que du processus de création de l'ontologie probabiliste.

3.2 Motivation de l'étude de cas

À cause de la pandémie du Covid-19, l'activité humaine s'est presque complètement figée et des millions de personnes se sont retrouvées au chômage et cela suite à la fermeture d'entreprises, la perturbation des marchés et la suppression d'emplois. Par ailleurs, les conséquences économiques étaient si graves qu'elle comprenaient la fermeture de nombreux sites de production et des ruptures d'approvisionnements dans de nombreux secteurs d'activité ce qui a fait basculer plus de 400 millions de personnes dans la pauvreté absolue. D'autres conséquences sociales de la pandémie sont perçues telles que la fermeture des écoles qui a entraîné l'arrêt de l'apprentissage pour des centaines de millions d'élèves et d'étudiants. Pour y remédier des solutions d'enseignement à distance ont été proposées or celles-ci ne sont pas facilement envisageables dans les pays sous développés ou en voie de

développement.

Le but ultime de notre système est de contribuer à la prévention à temps contre le Covid-19 en collaborant avec le corps médical afin de diminuer la perte humaine et minimiser les différents dégâts causés par cette pandémie.

3.3 Caractéristiques du Système

Après l'analyse des travaux connexes présentés dans les deux premiers chapitres, nous avons pu apprécier les avantages et les limites de chaque méthode et de chaque réseau bayésien utilisé, ce qui nous a aidé dans le choix des critères de notre système. Nous commençons par mentionner quelques conclusions et critiques qui concernent les méthodes utilisées afin de justifier le choix que nous allons prendre par la suite :

- L'utilisation des méthodes d'apprentissage profond nécessite un dataset volumineux afin d'assurer un bon apprentissage et donc réaliser une bonne inférence.

- L'utilisation des ontologies a été efficace en termes de représentation des connaissances d'une façon claire, simple et formelle, par contre il lui manque la capacité de raisonner sous l'incertitude.

- L'utilisation des réseaux bayésiens permet le raisonnement sous l'incertitude qui peut se présenter sous plusieurs formes tels que : l'aléatoire qui consiste en l'inexistence d'une loi générale expliquant la réalisation des phénomènes (avoir une fièvre ou une toux sèche n'implique pas nécessairement l'infection par le Covid-19) et l'incomplétude qui signifie l'absence d'une partie de l'information tel que les signes et les symptômes absents.

- Malgré le manque de données d'observation, l'utilisation des réseaux bayésiens assure des prédictions précises [52].

En ce qui concerne les travaux qui se basent sur les réseaux bayésiens, nous avons les critiques suivantes :

- L'utilisation d'un réseau bayésien MEBN engendre un temps d'exécution très lent.
- Le réseau bayésien standard ne permet pas une bonne expressivité par rapport aux autres extensions des réseaux bayésiens.
- Le réseau bayésien causal est un réseau bayésien standard comportant que des liens de causalité.

- Les réseaux bayésiens MEBN, standards, causaux ne traitent que les données discrètes.

Après avoir étudié les limites de chaque méthode nous avons opté pour l'utilisation d'une ontologie pour la représentation des connaissances et d'un réseau bayésien pour le traitement d'incertitude qui se présente dans notre contexte sous forme de l'aléatoire et de l'incomplétude. Le type du réseau bayésien que nous avons choisi est l'hybride, ce dernier a la capacité d'accueillir simultanément les variables discrètes et continues afin d'avoir un système plus précis.

3.4 Schéma global

Notre travail consiste à concevoir un système d'aide au diagnostic médical, en se basant sur une ontologie classique et un réseau bayésien hybride. La construction de notre système passe par plusieurs étapes :

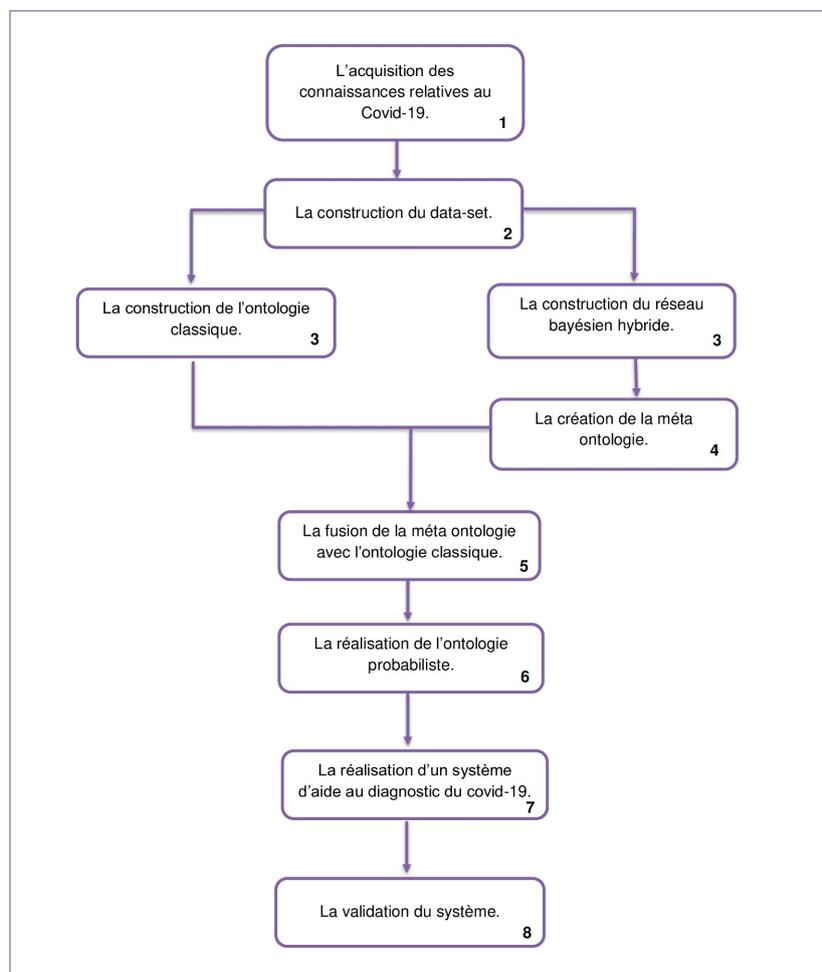


FIGURE 3.1 – Schéma global

3.5 Description du schéma global

1. Afin de récolter les informations nécessaires en lien avec le Covid-19, nous avons contacté un expert du domaine, qui nous a fait bénéficier de son expertise et fourni toutes les connaissances relatives à cette maladie.
2. Après avoir eu toutes les informations nécessaires, nous avons pu construire notre dataset en nous basant sur les cas d'un établissement de santé (clinique).
3. Par la suite, nous avons créé notre ontologie classique en suivant la méthode Méton-tologie, tout en faisant l'apprentissage de la structure de notre réseau bayésien avec l'expert du domaine qui nous a expliqué les différents liens entre les nœuds. Après l'apprentissage de la structure, nous avons fait l'apprentissage des paramètres qui a nécessité l'utilisation de l'algorithme EM pour les nœuds discrets et le développement de l'algorithme best fit pour les nœuds continus.
4. Après avoir terminé avec le réseau bayésien, nous avons créé une méta ontologie qui contient tous les composants d'un réseau bayésien hybride.
5. Afin d'intégrer l'incertitude dans l'ontologie classique nous avons fusionné la méta ontologie avec l'ontologie classique.
6. Après la réalisation de l'étape précédente nous avons obtenu une ontologie probabiliste capable de représenter les connaissances incertaines.
7. Par la suite, nous avons implémenté un système d'aide au diagnostic du Covid-19.
8. A la fin, nous avons validé notre système par l'expert du domaine ainsi que par les mesures de performances.

3.6 Construction du dataSet

L'ensemble de données sur le Covid-19 est indisponible au public sur les sites web, cela nous a poussé à développer un dataset.

Notre dataset est une amélioration de la base de données de [53], après l'avoir enrichi, nous avons atteint neuf-cents cas réels qui proviennent de la clinique Belkacemi Healthcare.

Les composants de notre dataset sont des symptômes, des signes paracliniques et des

facteurs de risque. Après avoir eu l'autorisation des médecins de la clinique, nous avons commencé la construction de notre dataset en collectant un bon nombre de dossiers de patients qui contiennent leurs analyses médicales, radiographies, etc.

Le médecin responsable nous a donné l'accès au logiciel de la clinique afin d'accéder aux symptômes et aux facteurs de risque des patients.

Le processus de collecte des données a nécessité l'aide d'un expert du domaine qui devait nous expliquer comment comprendre les bilans sanguins et faire la différence entre les valeurs normales chez les patients qui n'ont aucune maladie et les valeurs normales chez les patients qui ont différents types de maladies et surtout de faire le lien entre quelques résultats et les maladies chroniques dont le patient peut souffrir.

Cela nous a pris énormément de temps et de va-et-vient à la clinique pour pouvoir trier les dossiers des patients selon notre besoin. La collecte et le remplissage du dataset n'a pas été une mission facile puisque beaucoup de dossiers manquaient d'informations, à cet effet, nous avons essayé de collecter les cas qui contiennent le moins de données manquantes afin de réaliser un dataset fiable avec des cas positifs et négatifs au Covid-19.

3.6.1 Les symptômes et les signes

Les symptômes et les signes jugés pertinents par l'expert du domaine avec qui nous avons travaillé sont définis par l'OMS [37] comme suit :

- **Headache (true/false)** : ou maux de tête, C'est une manifestation douloureuse souvent liées à des troubles mineurs tels que le stress et la fatigue et rarement liées à des troubles plus graves.
- **Fever (true/false)** : la fièvre est un mécanisme de défense de l'organisme contre les infections, autrement dit c'est une réaction du corps lorsque celui-ci se défend, ou la température de ce dernier dépasse trente-huit degrés Celsius (38°C).
- **Dry cough (true/false)** : ou toux sèche, c'est l'un des symptômes les plus fréquents du Covid-19, elle est provoquée par une irritation au niveau des voies respiratoires.
- **Muscle soreness (true/false)** : les courbatures sont des douleurs musculaires qui peuvent être un symptôme initial d'une infection virale telle que la grippe.
- **Diarrhea (true/false)** : selon l'OMS, la diarrhée est définie par une émission de selles trop fréquentes de trois à quatre selles par jour, de consistance anormale (liquides

ou très molles), elle peut être un symptôme d'une infection intestinale par des virus.

- **Shortness of breath (true/false)** : essoufflement ou «dyspnée», signifie des difficultés respiratoires suite à un niveau réduit d'une activité physique ou un repos, il est un symptôme rencontré souvent dans l'insuffisance cardiaque et les maladies cardiaques.
- **Blood urea (g/l)** : l'urée sanguine est un produit qui résulte d'un processus de dégradation des protéines. L'urée est produite par le foie, elle contient de l'azote, ensuite elle est éliminée par le rein. Elle varie en fonction de l'état du rein et permet d'évaluer la présence d'une insuffisance rénale. Sa valeur normale est comprise entre 0.1 et 0.55 g/L.
- **Creatinine (mg/l)** : est un produit de la dégradation de la créatine stockée dans les muscles, elle est transportée par le sang ensuite éliminée par les reins, son dosage sanguin permet d'estimer la fonction rénale. Les valeurs considérées normales sont comprises entre sept et dix mg/l chez les femmes, et chez les hommes entre neuf et treize mg/l.
- **Blood Sugar (g/l)** : ou glycémie, c'est le taux de glucose (sucre) qui est présent dans la circulation sanguine. Sa valeur normale est entre 0.7 et 1.1 g/L.
- **D-dimères (microgram)** : sont des produits de la dégradation de la fibrine. Dans le cas normal, la concentration de D-dimères dans le sang est inférieure à cinq-cents µg/l (microgrammes par litre) cependant, la présence d'une augmentation importante permet de détecter la présence d'un caillot de sang.
- **SpO2(%)** : La saturation en oxygène s'exprime en pourcentage, elle correspond au taux d'oxygène contenu dans les globules rouges après leur passage dans les poumons. Sa valeur normale est entre quatre-vingt-quinze pour cent et cent pour cent.
- **CRP (C reactive protein) (mg/l)** : est une protéine produite par le foie à la suite d'une inflammation dans l'organisme, son taux normal doit être inférieur à six mg/L.
- **L'INR (International Normalized Ratio)** : est un marqueur qui permet de surveiller certains facteurs sanguins impliqués dans la coagulation. Sa valeur normale est entre

0.8 et 1.2 chez les gens qui ne consomment pas de traitement anticoagulant et entre deux et quatre chez ceux qui reçoivent des AVK.

3.6.2 Les facteurs de risque

Les facteurs pouvant aggraver la situation du malade de Covid-19 pris en compte par l'expert du domaine sont définis par l'OMS [37] comme suit :

- **Age** : La possibilité d'être infecté, hospitalisé ou bien de mourir à cause du Covid-19 augmente de façon exponentielle avec l'âge.
- **Cardiovascular history (true/false)** : Les antécédents (ATCD) cardiovasculaires tels que les maladies cardiovasculaires qui regroupent les pathologies, ces derniers touchent le cœur et l'ensemble des vaisseaux sanguins, comme l'athérosclérose, les troubles du rythme cardiaque, l'insuffisance cardiaque etc. . . , et d'autres antécédents tels que les chirurgies cardiaques.
- **Diabetes (true/false)** : est une maladie chronique caractérisée par une hyperglycémie. Le diabète apparaît lorsque le pancréas ne produit pas suffisamment d'insuline ou que l'organisme n'utilise pas correctement l'insuline qu'il produit.
- **Renal failure (true/false)** : ou L'insuffisance rénale, c'est quand les reins cessent progressivement de fonctionner et les déchets du métabolisme s'accumulent dans le corps, cette maladie est souvent liée au diabète et à l'HTA.
- **Morbid obesity (true/false)** : L'obésité se définit comme une accumulation anormale ou excessive de graisse corporelle qui peut nuire à la santé.
- **HTA (true/false)** : L'hypertension artérielle (HTA) est une maladie caractérisée par une pression artérielle trop élevée.
- **Chronic lung disease (true/false)** : ou les maladies pulmonaires, ce sont des maladies telles que les bronchites chroniques, dilatation des bronches, l'asthme etc. . .

- **Immune system disease (true/false)** : ou les maladies du système immunitaire, ce sont toutes les maladies qui affectent le système immunitaire. On distingue deux groupes : les maladies auto-immunes et les déficits immunitaires.

3.7 Création de l'ontologie Classique

Pour la création de notre ontologie classique nous avons suivi une méthode bien structurée qui couvre le cycle de vie de l'ontologie d'une manière bien détaillée ce qui caractérise particulièrement cette dernière. Cette méthode est appelée Methontology [54], elle est constituée de trois processus qui sont la spécification, l'acquisition des connaissances et la conceptualisation.

Dans ce qui suit nous allons présenter tous les processus mentionnés ci-dessus.

1. Spécification

Le processus de spécification consiste à définir quelques points relatifs à l'ontologie : le domaine de connaissance qu'elle couvre, l'objectif de sa création, ses utilisateurs, ses sources d'informations et son degré de formalisation.

- **Domaine de connaissance :**

Le domaine que l'ontologie va couvrir est le domaine médical, plus précisément la connaissance relative au diagnostic du Covid-19.

- **Objectif :**

L'objectif est d'utiliser l'ontologie comme une base de connaissance afin de réaliser un système d'aide au diagnostic du Covid-19.

- **Utilisateurs :**

Dans notre cas, les utilisateurs de l'ontologie sont les médecins.

- **Les sources d'informations :**

Les sources d'informations sur lesquelles nous nous sommes basés pour arriver à la construction de l'ontologie sont les bilans sanguins et les fiches des patients se trouvant sur le logiciel du médecin.

- **Degré de formalisation :**

Le degré de formalisation de notre ontologie est formel.

2. Acquisition des connaissances :

Afin d'acquérir les connaissances nécessaires pour la construction de notre ontologie, nous avons collaboré avec un expert du domaine.

- **Identification des termes de l'ontologie et leurs définitions :**

La figure 3.2 présente les termes utilisés dans notre ontologie ainsi que leurs descriptions.

Nom de Terme	Description
Patient	Concept d'un malade
Risk_factor	Concept d'un facteur de risque
Cardiovascular_history	Concept des antécédents cardiovasculaire
Chronic_lung_disease	Concept des maladies pulmonaires chronique
Diabetes	Concept du diabète
HTA	Concept de L'hypertension artérielle
Immune_system_disease	Concept des maladies du système immunitaire
Morbid_obesity	Concept de L'obésité
Renal_failure	Concept des maladies rénale
Sign	Concept d'un signe
Blood_sugar	Concept du taux de glucose
Blood_urea	Concept de l'urée sanguine
Creatinine	Concept de la créatinine
CRP	Concept de la protéine Créactive
D_dimers	Concept du d_dimers
INR	Concept du temps de prothrombine
SPO2	Concept de la saturation en oxygène
Symptom	Concept d'un symptômes
Diarrhea	Concept de la diarrhée
Dry_cough	Concept de la toux sèche
Fever	Concept de la fièvre
Headache	Concept des maux de tête
Muscle_soreness	Concept des courbatures
Shortness_of_breath	Concept de l'essoufflement

FIGURE 3.2 – Identification des termes de l'ontologie et leurs d éfinitions

3. Conceptualisation

Le processus de conceptualisation se constitue de plusieurs tâches qui consistent à représenter le dictionnaire des concepts, les taxonomies des concepts, le diagramme et le tableau des relations binaires.

- **Dictionnaire des concepts :**

Le tableau 3.1 présente les concepts de notre ontologie, leurs attributs et leurs relations.

Concept	Attribut	Relation
Patient	<ul style="list-style-type: none"> • Has_age • Has_sex 	<ul style="list-style-type: none"> • Has_syptom • Has_risk_factor • Has_sign
Risk_factor		Has_risk_factor
Cardiovascular_history	Cardiovascular_state	subclass
Chronic_lung_disease	Chronic_state	subclass
Diabetes	Diabetes_state	subclass
HTA	HTA_state	subclass
Immune_system_disease	Immune_state	subclass
Morbid_obesity	Morbid_state	subclass
Renal_failure	Renal_state	subclass
Sign		Has_sign
Blood_sugar	Sugar_value	subclass
Blood_urea	Urea_value	subclass
Creatinine	Creatinine_value	subclass
CRP	CRP_value	subclass
D_dimers	Ddimers_value	subclass
INR	INR_value	subclass
SPO2	SPO2_value	subclass
Symptom		Has_syptom
Diarrhea	Diarrhea_state	subclass
Dry_cough	Dry_state	subclass
Fever	Fever_state	subclass
Headache	Headache_state	subclass
Muscle_soreness	Muscle_state	subclass
Shortness_of_breath	Shortness_state	subclass

TABLE 3.1 – Dictionnaire de concepts

- **Taxonomies des concepts :**

Les taxonomies permettent de représenter les relations hiérarchique entre les classes. Dans ce qui suit nous allons présenter la taxonomie de trois concepts qui sont Risk_factor, Sign et Symptom.

1. **Taxonomie de Risk_factor**

La figure 3.3 représente la taxonomie de la classe Risk_factor.

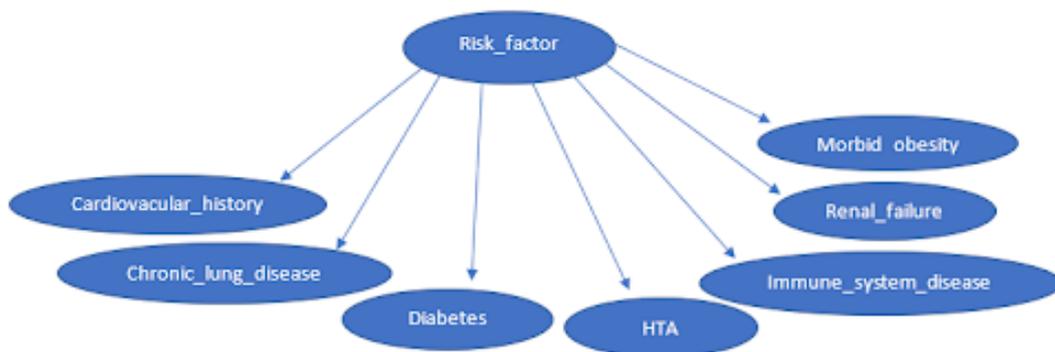


FIGURE 3.3 – Taxonomie de Risk_factor

2. **Taxonomie de Sign**

La figure 3.4 représente la taxonomie de la classe Sign.

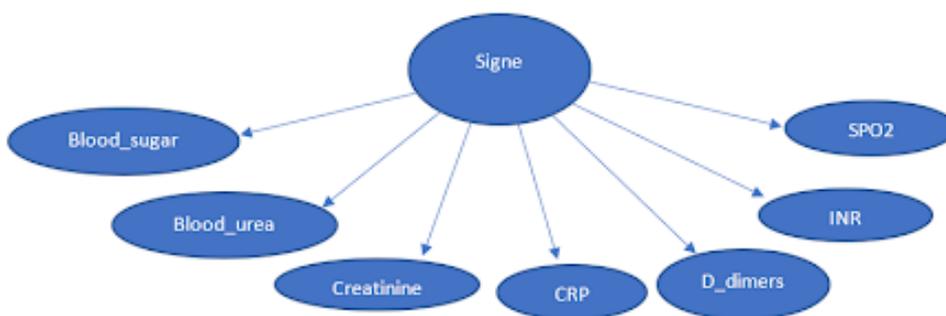


FIGURE 3.4 – Taxonomie de Sign

3. Taxonomie de Symptom

La figure 3.5 représente la taxonomie de la classe Symptom.

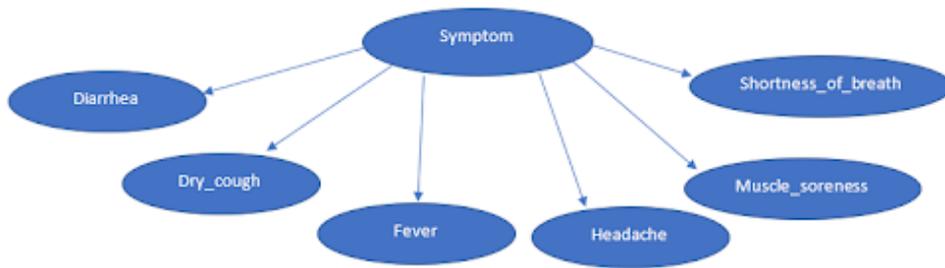


FIGURE 3.5 – Taxonomie de Symptom

- **Diagramme des relations binaires**

À travers le diagramme des relations binaires nous présentons les différentes relations entre les concepts de notre ontologie 3.6.

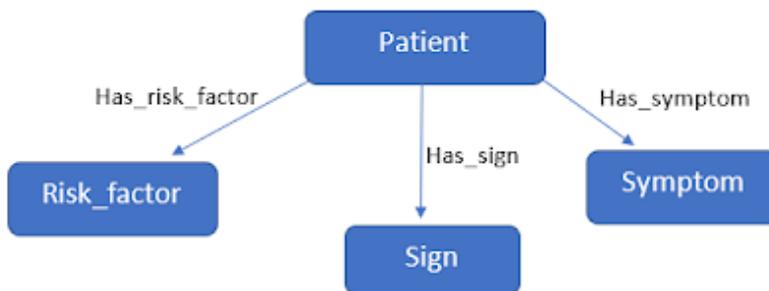


FIGURE 3.6 – Diagramme de relations binaires

- **Le tableau des relations binaires**

Le tableau 3.2 représente les relations entre les concepts sources et cibles.

Nom de la relation	Concept source	Concept cible
Has_symptom	Patient	Symptom
Has_risk_factor	Patient	Risk_factor
Has_sign	Patient	Sign

TABLE 3.2 – Le tableau des relations binaires

En suivant les étapes de la méthode Methontology, nous avons obtenu une ontologie classique, la figure 3.7 représente la structure globale de cette dernière :

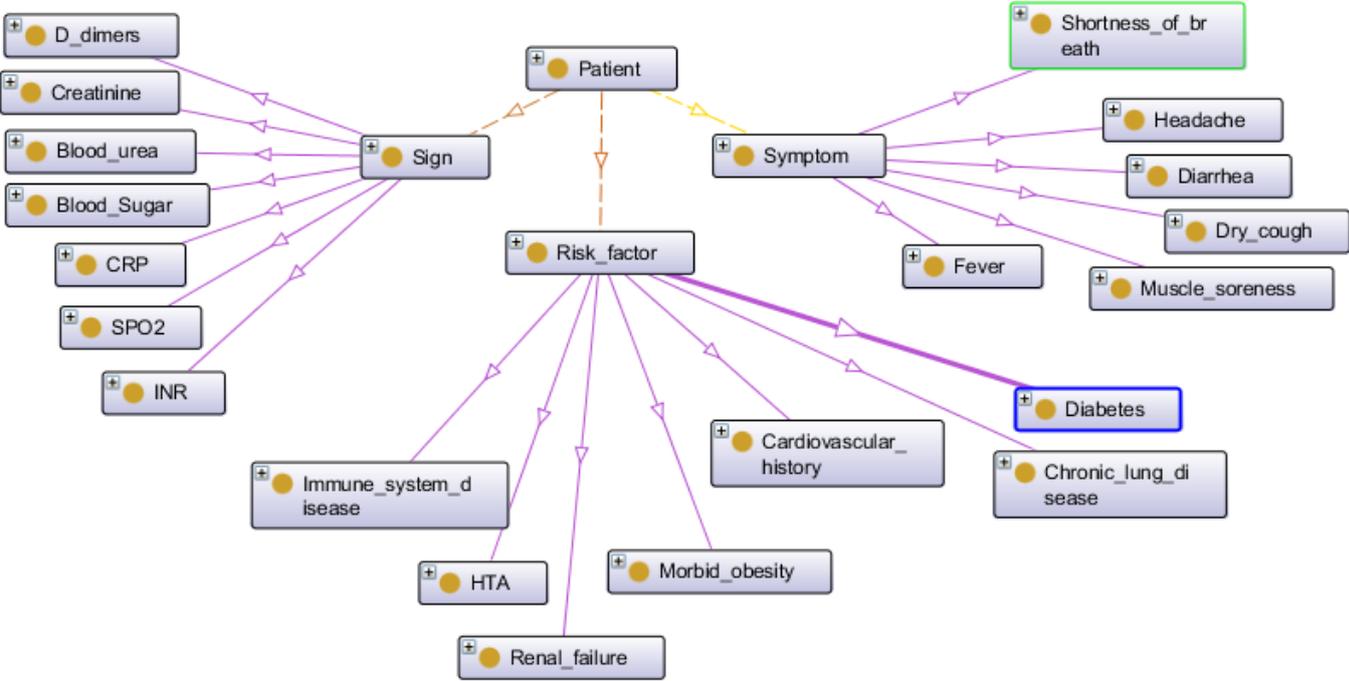


FIGURE 3.7 – La structure de l’ontologie classique

La figure 3.8 montre la hiérarchie des classes, les object property et les data property de l'ontologie classique.

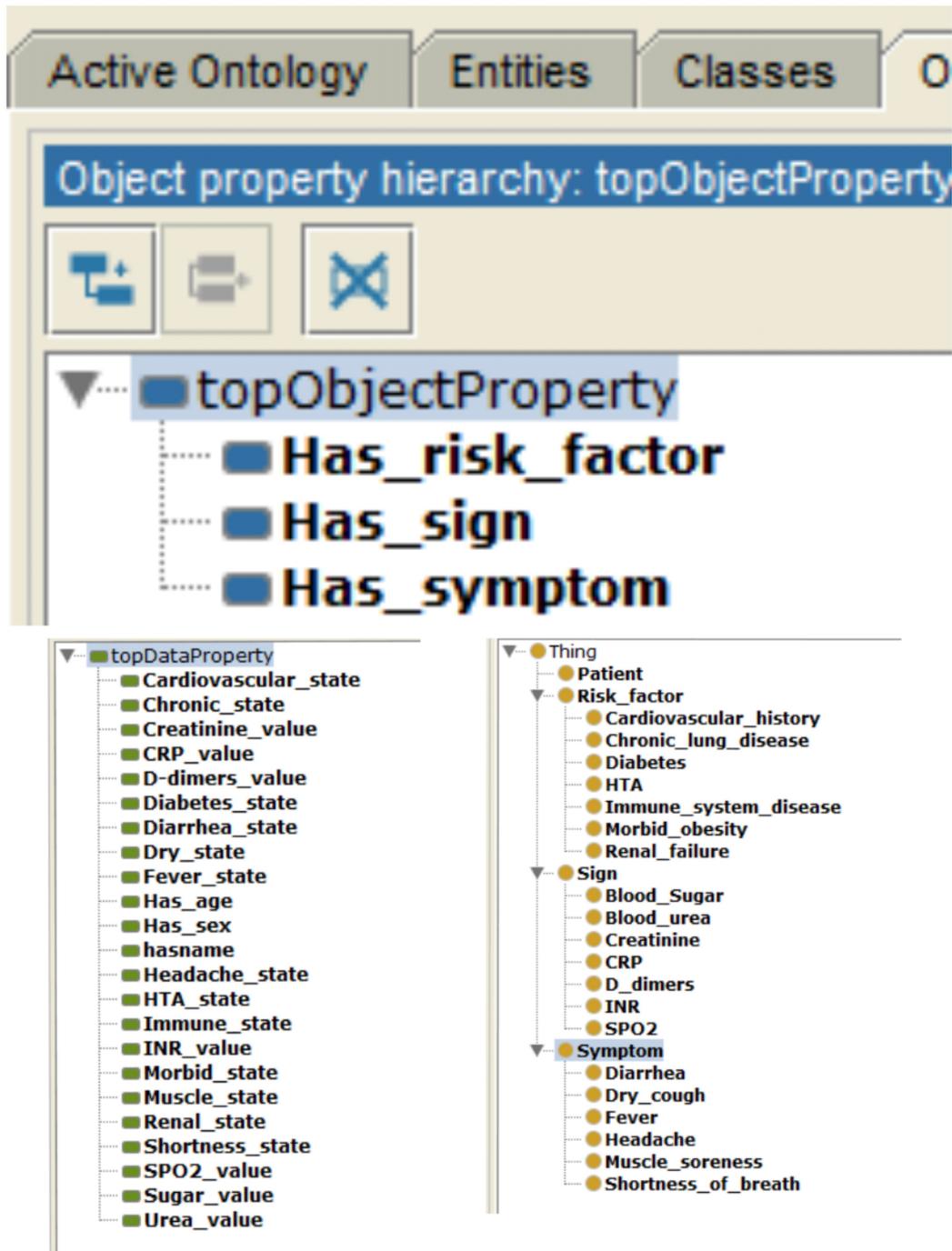


FIGURE 3.8 – Les classes, les object property et les data property de l'ontologie classique

3.8 Construction du réseau bayésien hybride

La construction du réseau bayésien a été réalisé en deux étapes qui ont consisté : Pour la première, en l'apprentissage de la structure, à travers plusieurs discussions avec l'expert du domaine afin de définir les noeuds de notre réseau bayésien et les liens entres ces derniers.

Pour la deuxième, en l'apprentissage des paramètres, qui dépend de la nature des noeuds ; Cette étape consiste à appliquer l'algorithme EM pour les nœuds discrets et à développer l'algorithme best fit (meilleur ajustement de courbe) qui sert à déterminer la distribution la mieux adaptée qui correspond à notre dataset, afin d'associer à chaque noeud continu la distribution adéquate.

1. Apprentissage de la structure :

La figure 3.9 montre la structure de notre réseau bayésien, faite avec l'expert du domaine.

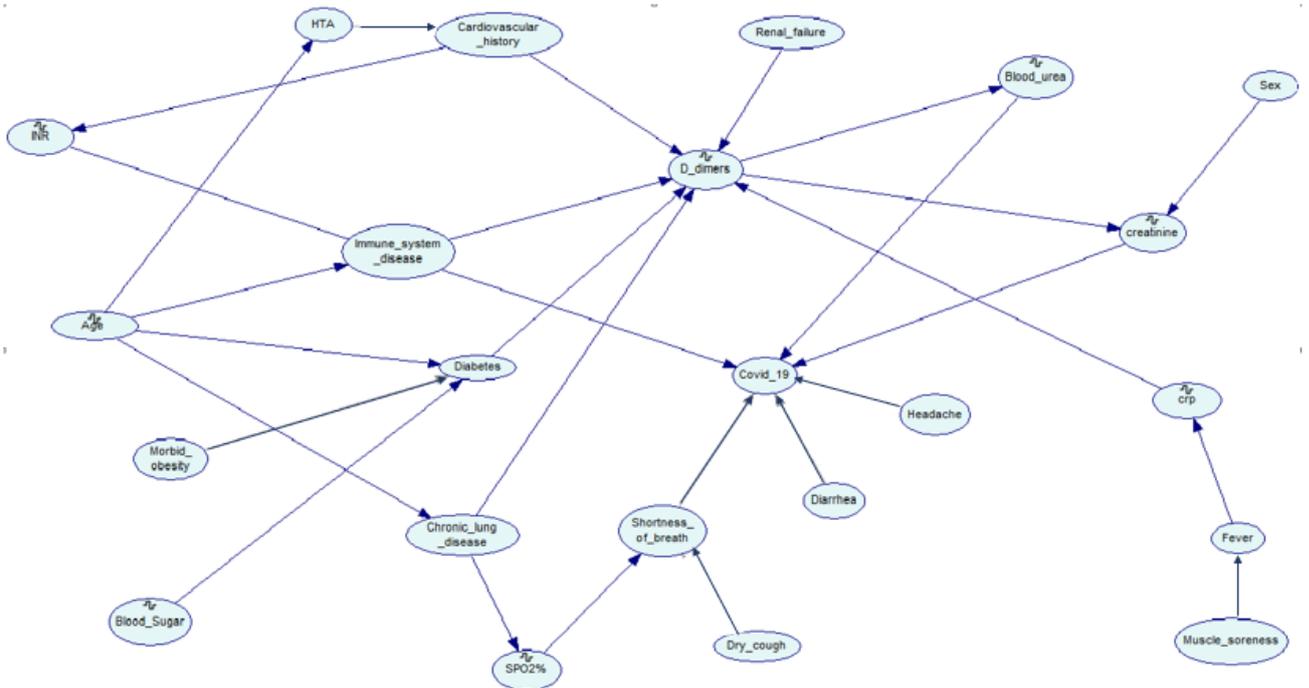


FIGURE 3.9 – La structure du réseau bayésien

2. Apprentissage des paramètres :

Du fait que le réseau bayésien hybride contient des nœuds continus et discrets, ça n'a pas été évident de trouver un moyen pour l'apprentissage des paramètres de ce dernier. C'est pour cela que nous avons traité chaque type de nœuds différemment. Pour les nœuds discrets nous avons appliqué l'algorithme EM et pour les nœuds continus nous avons choisi pour chaque nœud l'équation de la distribution qu'il lui convient. Le choix de l'équation a été fait selon les données des nœuds continus. Pour cela nous avons développé un algorithme qui indique le meilleur fit pour chaque nœud.

Nous allons présenter par la suite l'algorithme du choix de la distributions et le résultat qu'il a donné pour chaque nœud.

• L'algorithme

- a) Lire le dataset.
- b) Tracer l'histogramme de la colonne choisie (par exemple Age).
- c) Préparer les données et les mettre dans un tableau.
- d) Ajuster les différentes distributions et trouver la distribution la mieux adaptée aux données en suivant les étapes :
 - Créer une instance Fitter qui fournit une classe simple pour identifier les distributions tels que normal, gamma...etc.
 - Fournir les données (colonne) et la liste des distributions.
 - Appliquer la méthode fit qui consiste à créer une courbe à partir des fonctions mathématiques et à ajuster les paramètres de ces fonctions afin de se rapprocher de la courbe mesurée.
 - Générer le résumé de la distribution ajustée.
- e) Identifier la meilleure distribution et ses paramètres.

- Après avoir appliqué l'algorithme mentionné ci dessus pour le meilleur fit pour le nœud âge nous avons obtenu la distribution Beta comme meilleur fit avec les paramètres suivants : (3.16861, 2.46467)

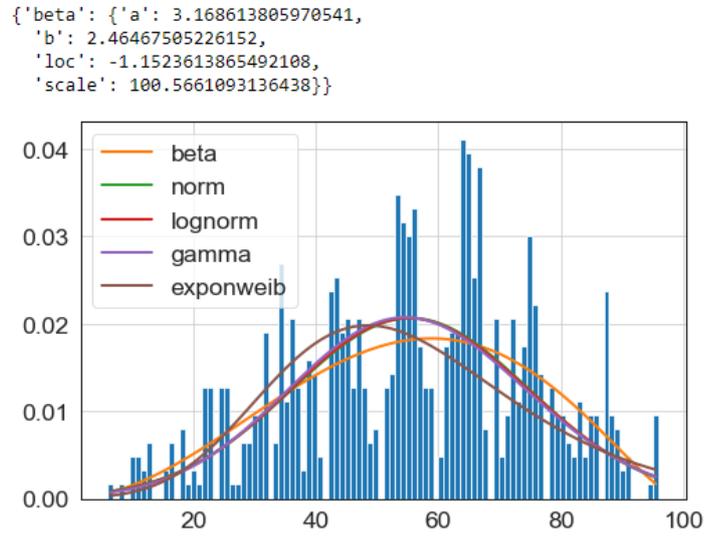


FIGURE 3.10 – Le best fit du nœud age

- Pour le nœud INR le meilleur fit que nous avons obtenu est la distribution Weibull et ses deux paramètres sont : (390.17882, 0.48695)

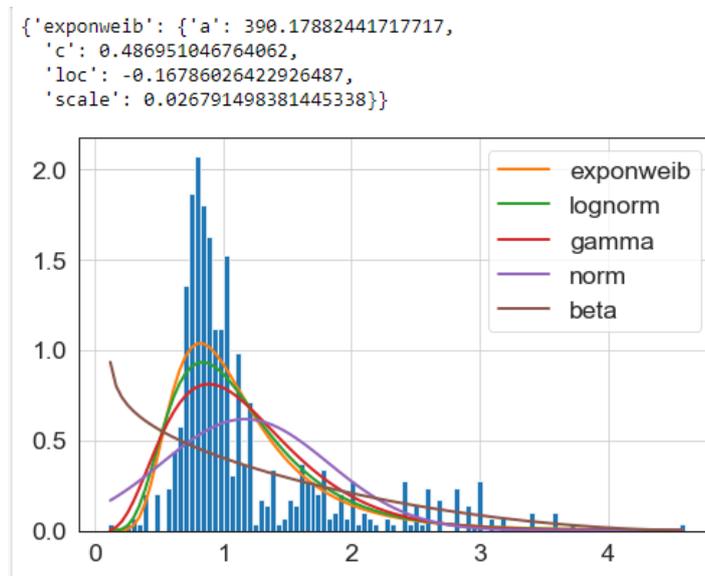


FIGURE 3.11 – Le best fit du nœud INR

- Le meilleur fit obtenu pour le nœud CRP est la distribution Lognormal, ses paramètres sont : (3.68544, 1.31991)

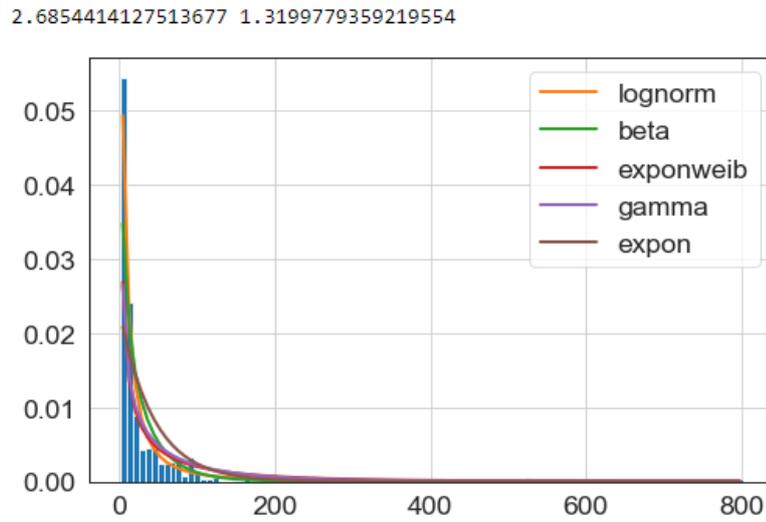


FIGURE 3.12 – Le best fit du nœud CRP

- La distribution Weibull est le meilleur fit pour le nœud Blood_Urea, ses paramètres sont (0.51881, 2.25951)

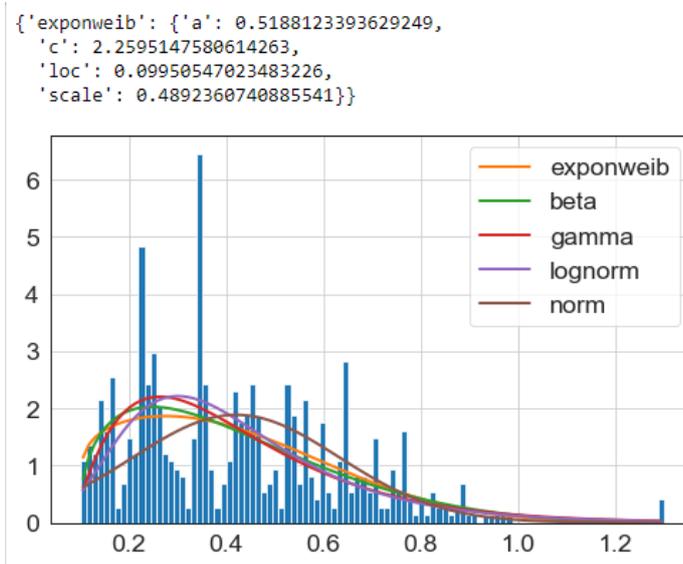


FIGURE 3.13 – Le best fit du nœud Blood_urea

- Nous avons obtenu la distribution Lognormal comme le meilleur fit pour le nœud D_dimers avec $(6.17895, 0.56902)$ comme paramètres.

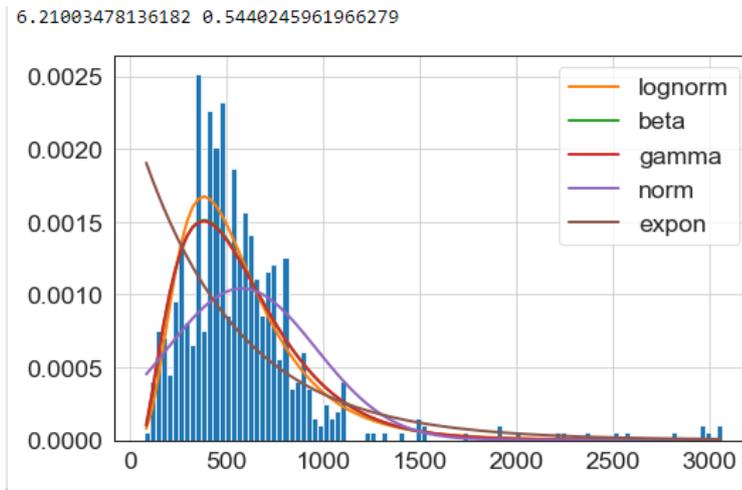


FIGURE 3.14 – Le best fit du nœud d-dimers

- La distribution Weibull avec ses paramètres $(230.13112, 0.37901)$ est le meilleur fit pour le nœud Blood_sugar.

```
{'exponweib': {'a': 230.13112107921015,
               'c': 0.3790169366934928,
               'loc': 0.2706218347025948,
               'scale': 0.007730344160491022}}
```

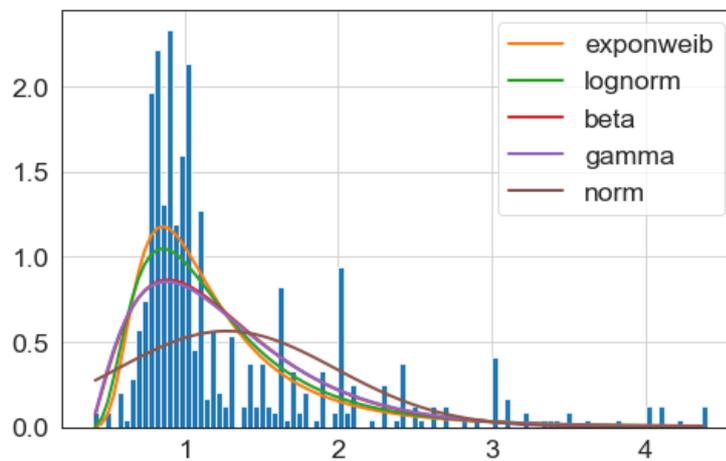


FIGURE 3.15 – Le best fit du nœud blood_sugar

- Pour le nœud créatinine le meilleur fit obtenu est la distribution Lognormal, ses paramètres sont : (2.43007,0.288078)

2.4328300985918 0.295271974119896

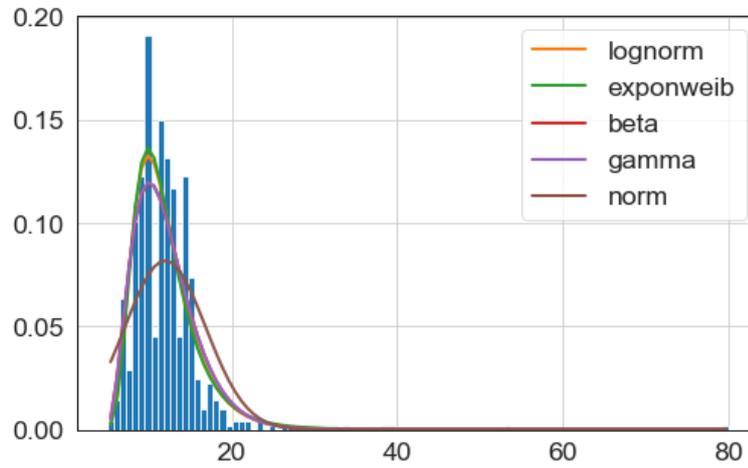


FIGURE 3.16 – Le best fit du nœud Creatinine

- Le meilleur fit pour le nœud SPO2 est la distribution Normal avec les paramètres suivants : (93.1737, 5.68578).

{'norm': {'loc': 94.17370892018779, 'scale': 5.685789098094582}}

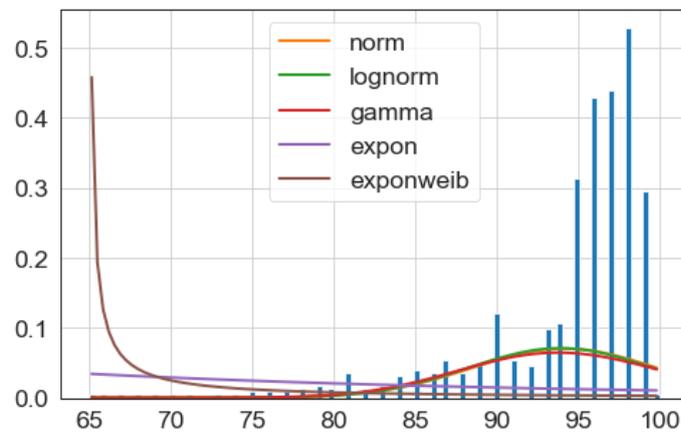


FIGURE 3.17 – Le best fit du nœud SPO2

Le réseau bayésien après l'apprentissage des paramètres :

La figure 3.18 représente notre réseau bayésien hybride après l'apprentissage des paramètres.

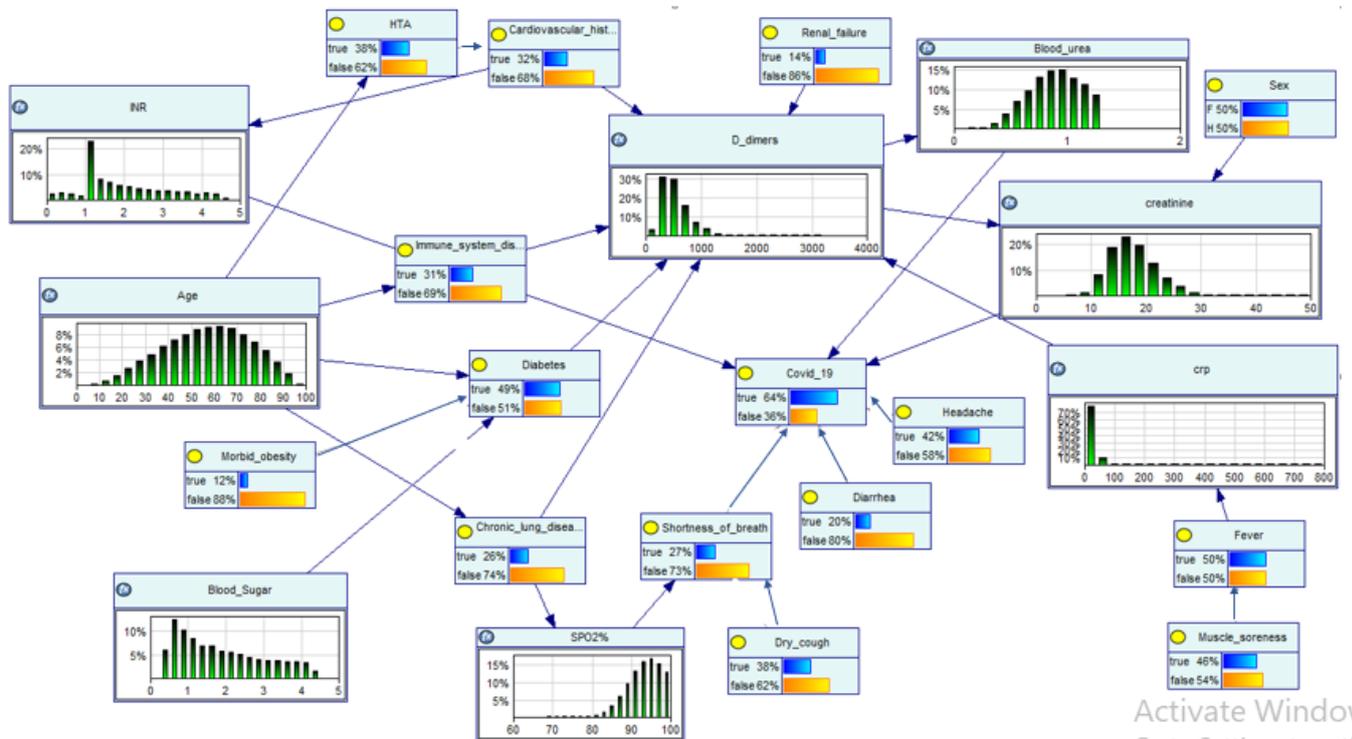


FIGURE 3.18 – Le réseau bayésien après l'apprentissage des paramètres

3.9 Méta ontologie du réseau bayésien

Afin d'intégrer l'incertitude dans notre ontologie classique nous avons créé une méta ontologie comportant tous les composants d'un réseau bayésien hybride, qui peut être réutilisée dans un autre travail. Ensuite, nous l'avons instancié pour avoir une méta ontologie correspondante à notre réseau bayésien hybride, dans le but de la fusionner avec notre ontologie classique, tout cela pour arriver à une ontologie probabiliste capable de représenter les connaissances incertaines.

La figure 3.19 montre la structure globale de la méta ontologie :

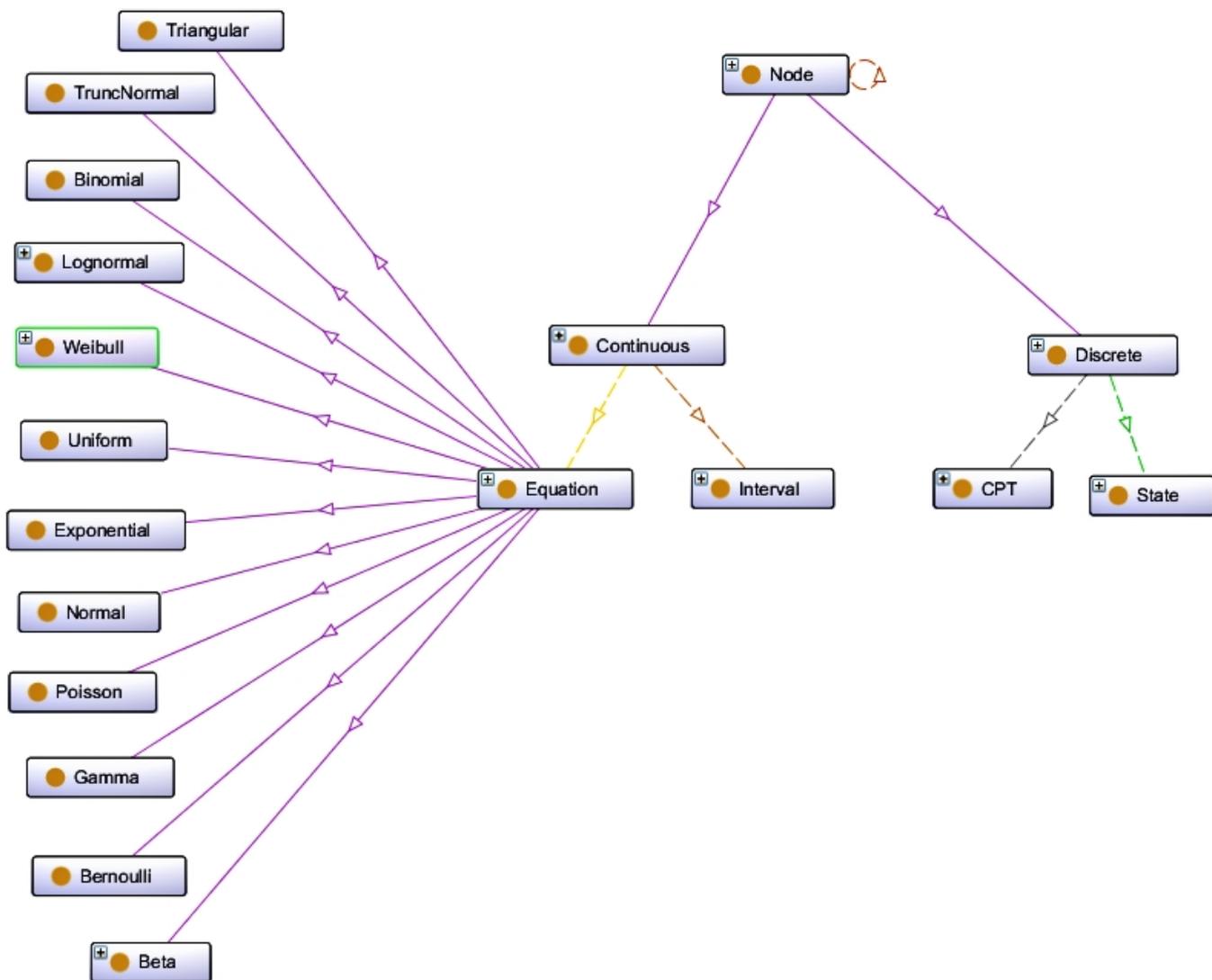


FIGURE 3.19 – La hiérarchie de la méta ontologie

3.10 L'intégration de la méta ontologie dans l'ontologie classique

Après avoir intégré la méta ontologie dans l'ontologie classique, nous avons obtenu les classes, les data property et les object property suivants :

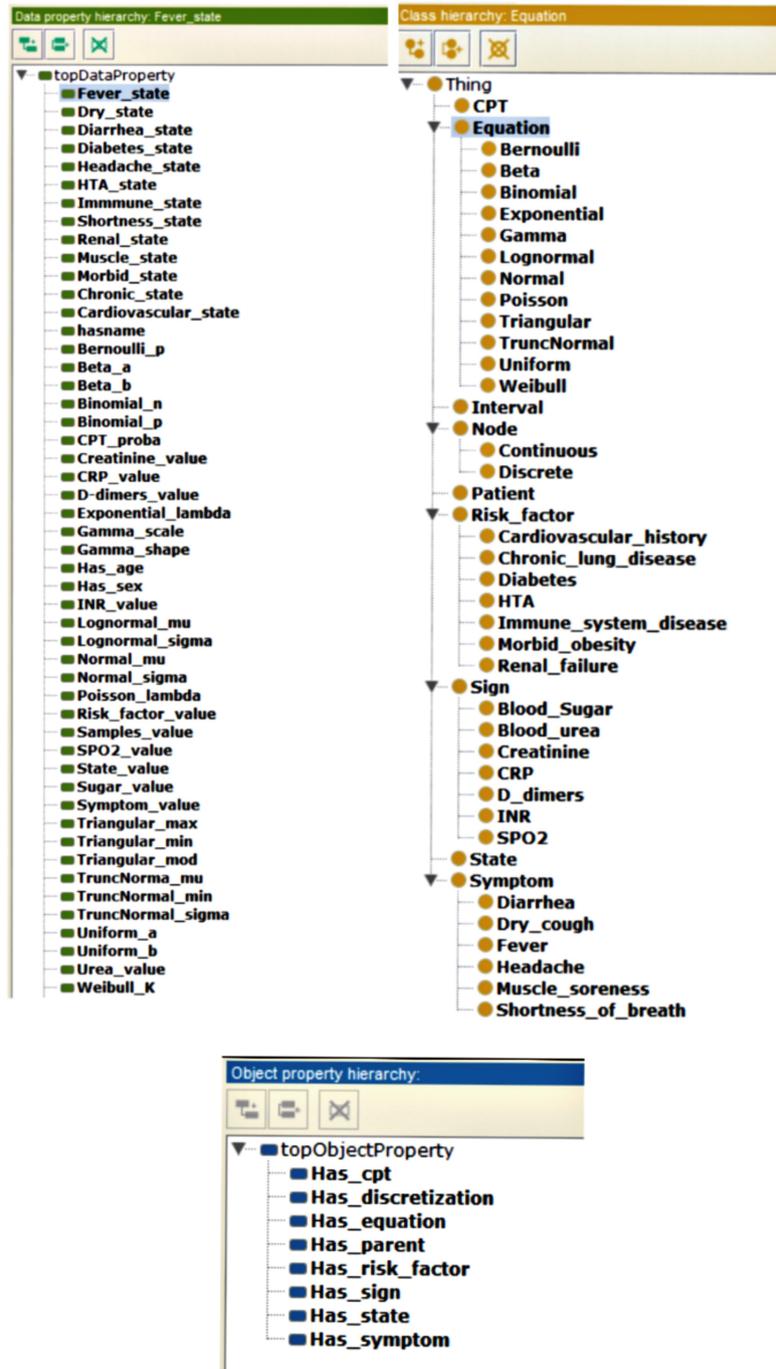


FIGURE 3.20 – Les classes, les object property et les data property de l'ontologie finale

3.11 Inférence du réseau bayésien hybride

L'inférence sur les réseaux bayésiens hybrides passe par 3 étapes présentées par les algorithmes :

1. Hybrid Forward Sampling : Cet algorithme consiste à générer des échantillons de l'ensemble des nœuds continus du réseau bayésien, en se basant sur les équations de distribution introduites pour chaque nœud, afin de trouver les distributions de probabilité pour les nœuds continus. L'algorithme Forward Sampling se base sur la discrétisation du nœud continu et celle de ses parents pour dériver une distribution de probabilité conditionnelle du nœud en question.

2. Autodiscretization :

L'auto-discrétisation est une approche de résolution de réseaux bayésiens hybrides contenant à la fois des variables discrètes et continues. Il offre une étape intermédiaire, dans laquelle il convertit un réseau bayésien hybride en un réseau bayésien discret. Pour toute variable continue, l'utilisateur peut spécifier une discrétisation à suivre dans cette reconversion. Une fois le réseau est converti, il se comporte comme un réseau bayésien discret. Le grand avantage de cet algorithme est qu'il convertit le réseau bayésien hybride d'origine en un réseau bayésien discret uniquement à des fins d'inférence, préservant la liberté de modélisation.

3. Algorithme d'inférence probabiliste exacte :

Après l'application des deux algorithmes d'échantillonnage et d'auto-discrétisation, les algorithmes d'inférence probabiliste exacte peuvent être appliquées sur le réseau bayésien qui résulte des deux algorithmes cités. .

3.12 Conclusion

À travers ce chapitre nous avons abordé notre motivation pour la maladie Covid-19 ainsi que le processus de construction de notre dataset et de l'ontologie probabiliste. Dans le prochain chapitre nous allons procéder à la réalisation et l'implémentation de notre système.

IMPLEMENTATION ET TEST DU SYSTÈME

4.1 Introduction

Dans le chapitre précédent nous avons présenté la construction du réseau bayésien hybride et de l'ontologie classique ainsi que l'intégration du réseau bayésien dans l'ontologie. À travers le présent chapitre nous allons décrire l'implémentation de notre système, pour la description visuelle de l'implémentation de ce dernier, nous allons effectuer des captures d'écran des différentes interfaces, par la suite, nous allons valider notre système.

4.2 Langages de Programmation

1. **Java** : est un langage de programmation orienté objet et aussi une plate-forme informatique qui été créé par Sun Microsystems et présenté officiellement en Mai 1995. Java est exploité dans le monde entier dans le but de développer des applications mobiles, des logiciels, des jeux etc. Il permet d'utiliser efficacement de nombreuses applications et services [55].

Dans notre projet nous l'avons utilisé pour le développement de notre application.

2. **Python** : est un langage de programmation créé par Guido van Rossum en 1989 et publié pour la première fois en 1991. C'est un langage interprété c'est-à-dire il est directement exécuté sans passer par une phase de compilation qui traduit le programme en langage machine, il est considéré comme un langage de haut niveau car il propose des fonctionnalités avancées et automatiques. Par ailleurs, python est simple et très lisible et il est doté d'une bibliothèque de base très fournie [56].

Dans notre projet nous l'avons utilisé pour le développement d'un algorithme pour l'apprentissage des noeuds continus de notre réseau bayésien hybride.

4.3 Outils de développement

Nous allons présenter la liste des outils de développement utilisés pour notre système :

1. **Protégé** : est un éditeur d'ontologie qui permet la création, l'importation, la modification et l'enregistrement des ontologies existantes écrites en OWL ou RDF [57].
2. **Génie smile** : GeNIe Modeler est une interface utilisateur graphique (GUI) pour SMILE Engine créé par le Decision Systems Laboratory et il a été testé de façon minutieuse sur le terrain depuis 1998. GeNIe permet la construction et l'apprentissage de modèles interactifs.
Smile est une bibliothèque de classes C++ qui peut être intégrée dans un logiciel utilisateur existant via son API afin d'améliorer les produits utilisateur avec des capacités de modélisation de décision [58].
3. **Eclipse** : Eclipse IDE est un environnement intégré de développement créé par IBM en 2001, il est écrit en java mais il peut être utilisé pour développer sous n'importe quel langage de programmation. Eclipse est doté de ses nombreuses fonctionnalités pratiques qui facilitent la programmation d'un logiciel [59].

4.4 API Utilisés

1. **API Swing Java** : Swing fait partie de la bibliothèque Java Foundation classes qui permet la construction des composants et des interfaces graphiques, ce dernier a été intégré au jdk depuis sa version 1.2 [60].
2. **OWL API** : est une API Java qui représente un ensemble d'interfaces riches en fonctionnalités, assurant une bonne manipulation d'ontologie [61].
3. **Protégé owl-api** : c'est une bibliothèque Java open source pour le langage d'ontologie Web (OWL) et RDF(S). L'API offre des classes et des méthodes afin de permettre le chargement et l'enregistrement des fichiers OWL, la manipulation des modèles de données OWL et l'exécution d'un raisonnement basé sur des moteurs Description Logic. De plus, l'API est optimisée pour la mise en place d'interfaces utilisateurs graphiques [62].

4.5 Résultat de l'ontologie finale

L'ontologie finale est le résultat de l'intégration de la méta ontologie dans l'ontologie classique. Nous présentons par la suite le code OWL d'un noeud discret et continu.

La figure 4.1 représente le code OWL d'un noeud continu qui montre l'équation de distribution du noeud et ses paramètres.

```
<!-- http://www.semanticweb.org/nira/ontologies/2022/4/untitled-ontology-10#Blood_sugar_equation -->
<owl:NamedIndividual rdf:about="&untitled-ontology-10;Blood_sugar_equation">
  <rdf:type rdf:resource="&untitled-ontology-10;Weibull"/>
  <untitled-ontology-10:Weibull_K rdf:datatype="&xsd;float">0.379017</untitled-ontology-10:Weibull_K>
  <untitled-ontology-10:Weibull_lambda rdf:datatype="&xsd;float">230.131</untitled-ontology-10:Weibull_lambda>
</owl:NamedIndividual>
```

FIGURE 4.1 – Code OWL du noeud Blood sugar

La figure 4.2 représente le code OWL d'un noeud discret qui comporte sa table de probabilité conditionnelle.

```
<owl:NamedIndividual rdf:about="&untitled-ontology-10;CPT_Diabetes">
  <rdf:type rdf:resource="&untitled-ontology-10;CPT"/>
  <untitled-ontology-10:CPT_proba rdf:datatype="&xsd:string">0.4985241803201001, 0.5014758196799,
0.405536780738231, 0.5944632192617691, 0.2350197181024495, 0.7649802818975505, 0.05845972272900969,
0.9415402772709903, 0.4552459956834521, 0.5447540043165479, 0.167494112984828, 0.8325058870151718,
0.4974855466003639, 0.5025144533996362, 0.6992363884590467, 0.3007636115409532, 0.4788036328018127,
0.5211963671981873, 0.2440844661851784, 0.7559155338148216, 0.5679563363806319, 0.432043663619368,
0.3795380992538835, 0.6204619007461165, 0.3537869964460676, 0.6462130035539323, 0.2342069126444142,
0.7657930873555858, 0.4637739580738594, 0.5362260419261407, 0.4606895312423144, 0.5393104687576857,
0.4680667897728565, 0.5319332102271437, 0.1709516717849127, 0.8290483282150873, 0.1332978862564355,
0.8667021137435645, 0.05097779200540555, 0.9490222079945945, 0.4913538346493325, 0.5086461653506675,
0.4788697154249463, 0.5211302845750537, 0.4977621992184016, 0.5022378007815983, 0.3525406027141841,
0.6474593972858158, 0.1576227091873897, 0.8423772908126104, 0.3328666974141243, 0.6671333025858757,
0.3158929252932715, 0.6841070747067285, 0.1766864181939793, 0.8233135818060208, 0.7479070327333591,
0.2520929672666409, 0.4637184547783426, 0.5362815452216574, 0.8157885413080702, 0.1842114586919298,
0.7871154437799489, 0.2128845562200511, 0.4912309556299367, 0.5087690443700632, 0.3329339787468109,
0.667066021253189, 0.4371071536119655, 0.5628928463880346, 0.2179886069058673, 0.7820113930941327,
0.9377700778608978, 0.06222992213910211, 0.5269195372834491, 0.473080462716551, 0.8443084780905867,
0.1556915219094132, 0.740041545309719, 0.2599584546902809</untitled-ontology-10:CPT_proba>
</owl:NamedIndividual>
```

FIGURE 4.2 – Code OWL du noeud diabète

4.6 Les interfaces graphiques

Nous allons présenter les différentes interfaces de notre application :

4.6.1 L'interface d'accueil

La figure 4.3 est l'illustration de la première interface de notre système, elle contient trois boutons, le premier est le bouton 'Patient' qui permet d'accéder à une interface contenant tout ce qui concerne le patient. Le deuxième est le bouton 'Aider' qui permet d'accéder à des interfaces qui expliquent très bien le fonctionnement du système. Le dernier bouton 'Quitter' qui permet de fermer le système.



FIGURE 4.3 – Interface d'accueil

4.6.2 L'interface patient

Sur l'interface patient, nous trouvons cinq boutons qui permettent d'ajouter un nouveau patient, de le modifier, de le supprimer, de consulter ses informations et de restituer le résultat de son diagnostic. À travers cette interface, le médecin peut également consulter la liste de ses patients.

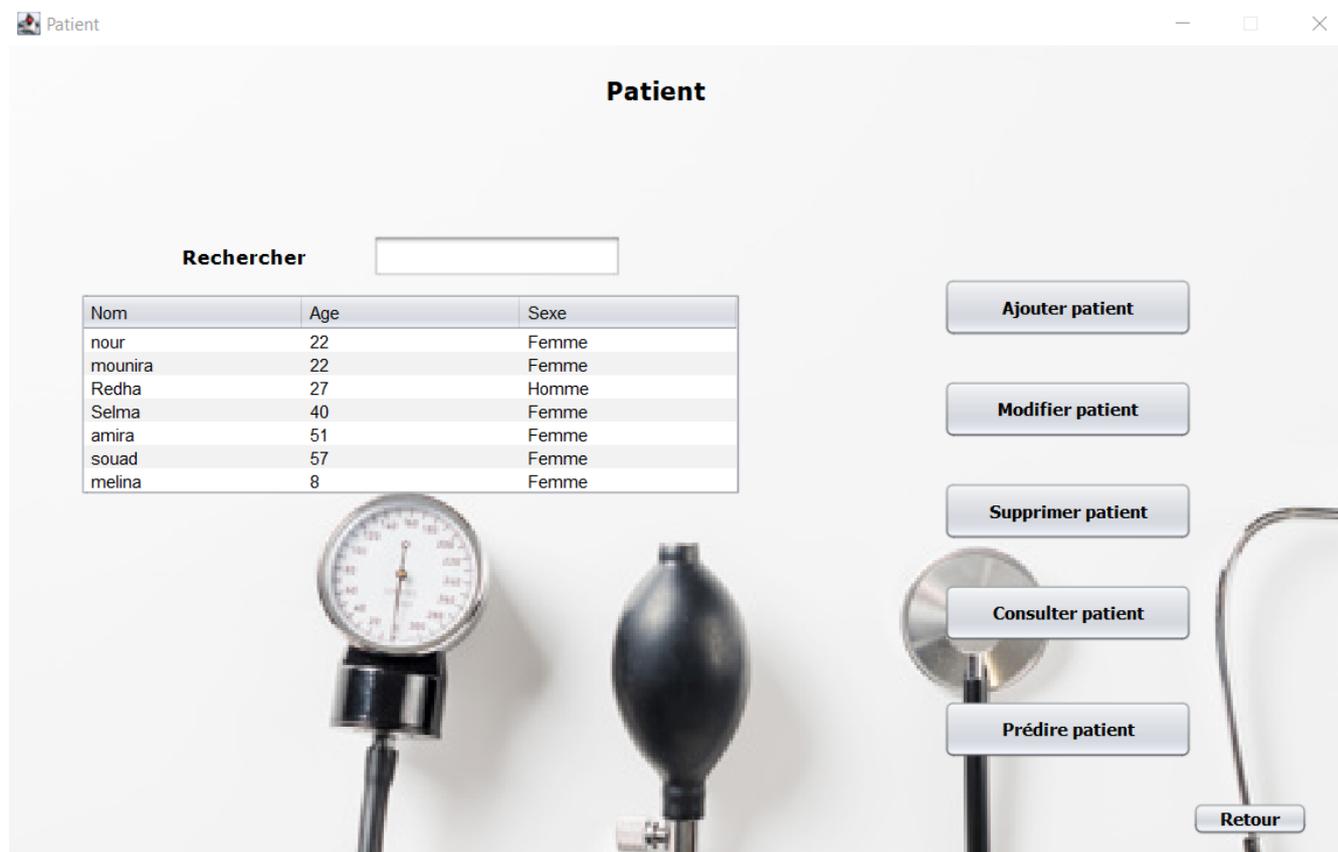


FIGURE 4.4 – Interface patient

4.6.3 L'interface ajouter patient

Cette interface permet d'ajouter un nouveau patient.

Ajouter patient

Nom : Age : Sexe :

Symptômes :

- Fièvre
- Toux sèche
- Courbatures
- Diarrhée
- Essoufflement
- Maux de tête

Facteurs de risque :

- Maladies du système immunitaire
- Antécédents cardiovasculaires
- Maladies pulmonaires
- insuffisance rénale
- Obésité
- Diabète
- HTA

Signes :

Urée sanguine SPO2%

Créatinine CRP

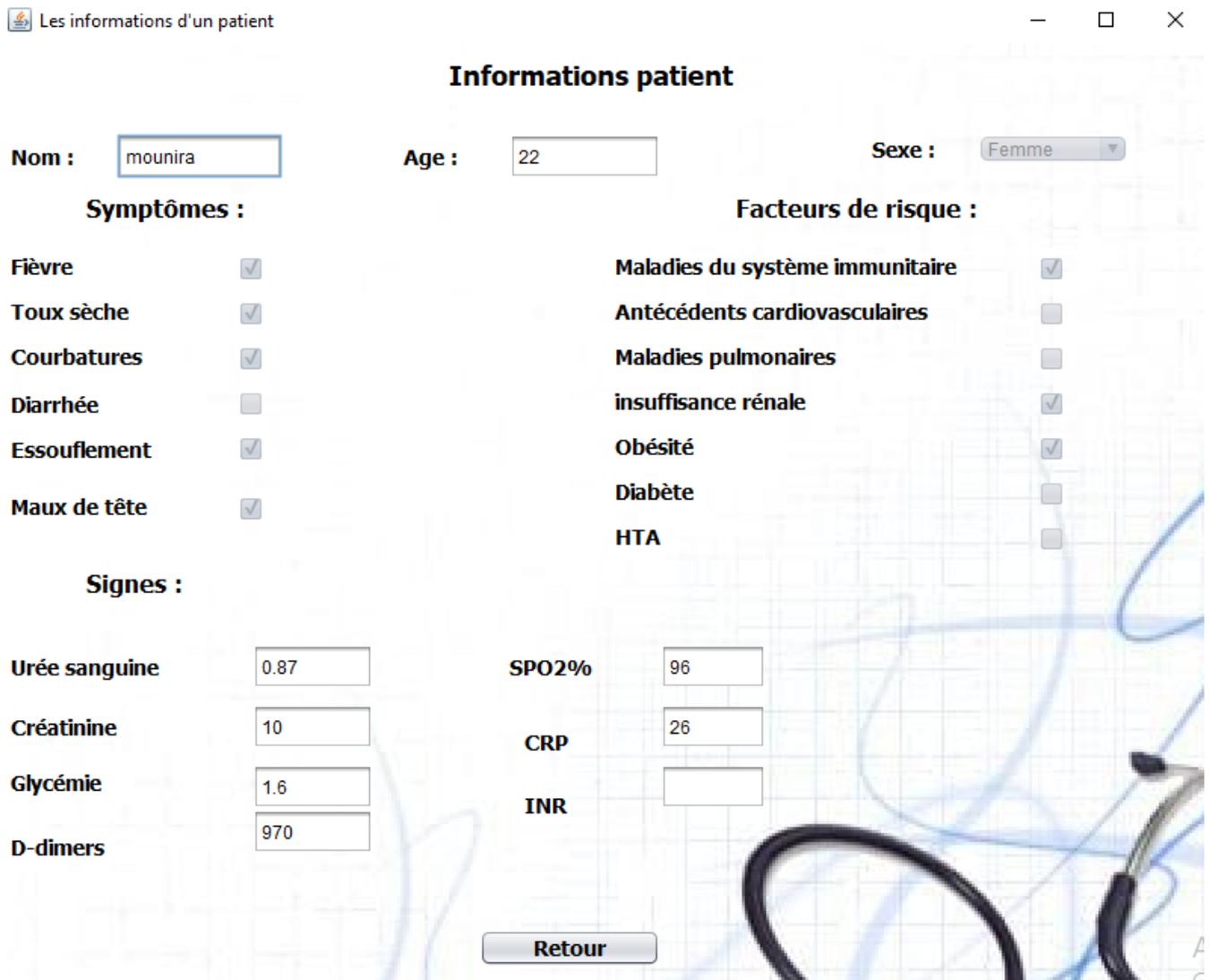
Glycémie INR

D-dimères

FIGURE 4.5 – Interface ajouter patient

4.6.4 L'interface informations patient

Cette interface affiche les informations du patient en question.



The screenshot shows a web application window titled "Les informations d'un patient" with standard window controls. The main content area is titled "Informations patient" and contains the following fields and sections:

- Nom :** Input field containing "mounira".
- Age :** Input field containing "22".
- Sexe :** Dropdown menu set to "Femme".
- Symptômes :** A list of symptoms with checkboxes:
 - Fièvre
 - Toux sèche
 - Courbatures
 - Diarrhée
 - Essoufflement
 - Maux de tête
- Facteurs de risque :** A list of risk factors with checkboxes:
 - Maladies du système immunitaire
 - Antécédents cardiovasculaires
 - Maladies pulmonaires
 - insuffisance rénale
 - Obésité
 - Diabète
 - HTA
- Signes :** A list of lab tests with input fields:
 - Urée sanguine: 0.87
 - Créatinine: 10
 - Glycémie: 1.6
 - D-dimères: 970
 - SPO2%: 96
 - CRP: 26
 - INR: (empty field)
- Retour** button at the bottom center.

FIGURE 4.6 – Interface Consulter patient

4.6.5 L'interface modifier informations patient

Cette interface permet de modifier les informations du patient sélectionné.

Les informations d'un patient

Modifier informations patient

Nom : **Age :** **Sexe :**

Symptômes :

Fièvre

Toux sèche

Courbatures

Diarrhée

Essoufflement

Maux de tête

Signes :

Urée sanguine

Créatinine

Glycémie

D-dimères

Facteurs de risque :

Maladies du système immunitaire

Antécédents cardiovasculaires

Maladies pulmonaires

insuffisance rénale

Obésité

Diabète

HTA

FIGURE 4.7 – Interface Modifier patient

4.6.6 L'interface prédiction

Cette interface permet d'afficher le résultat du diagnostic d'un patient.

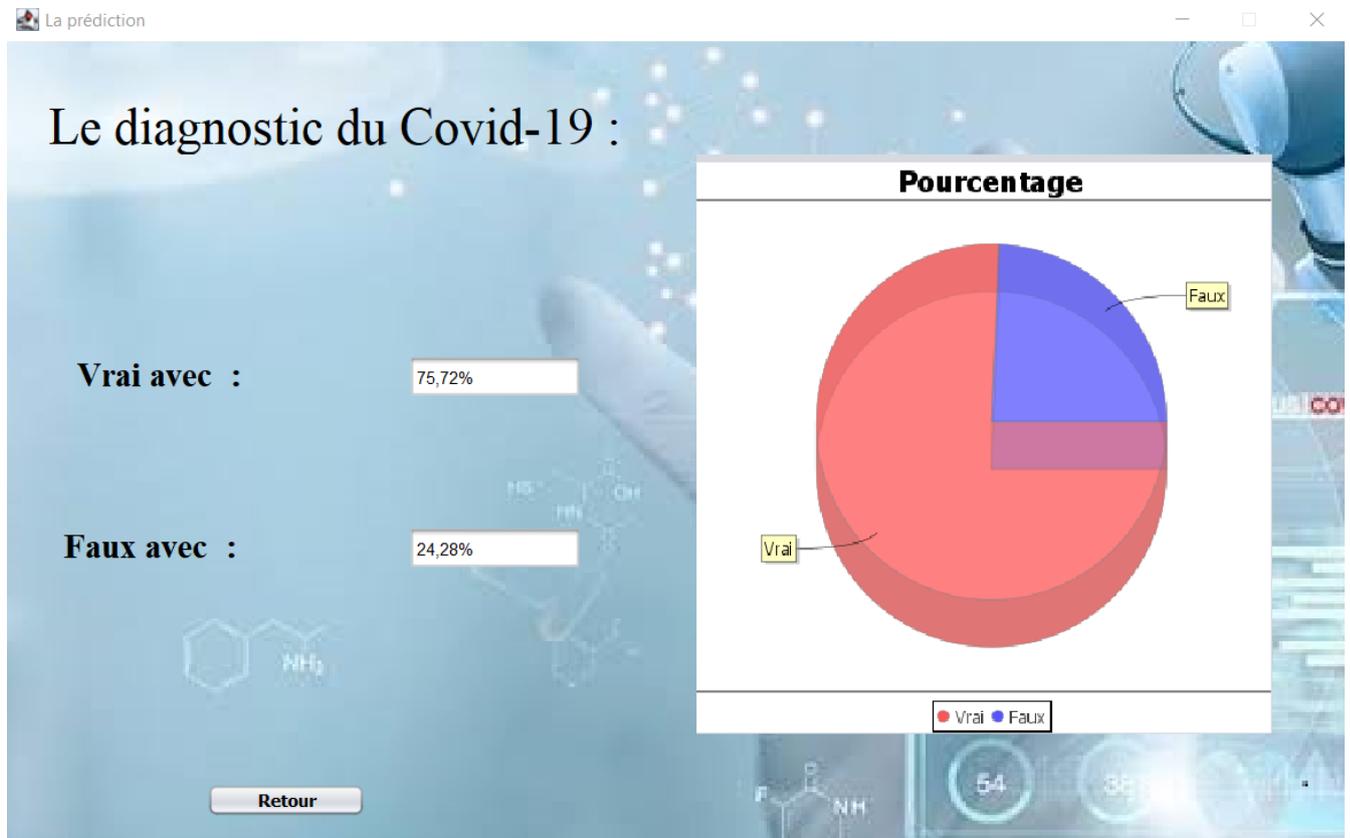


FIGURE 4.8 – Interface Prédiction

4.7 Validation du système

La validation de notre système se fait en deux étapes qui consistent en :

- La validation par notre expert du domaine.
- La validation par les mesures de performance.

4.7.1 Validation du système par l'expert du domaine :

Après l'implémentation de notre système, nous avons présenté ce dernier à notre expert du domaine, afin de le tester par des cas réels et de le valider. Nous présentons par la suite les résultats des tests obtenus.

1. Test 1 :

- **Résultat attendu :** le patient est infecté par le Covid-19.
- **Evidences :**
 - a) Les symptômes :
 - Le patient a des maux de tête.
 - Le patient a des courbatures.
 - Le patient souffre de la fièvre.
 - Le patient souffre de la toux sèche.
 - Le patient n'a pas la diarrhée.
 - Le patient ne souffre pas d'un essoufflement.
 - b) Les signes :
 - La valeur de l'urée sanguine n'est pas déclaré.
 - La valeur de créatinine est de 13.9 mg/l chez le patient.
 - La valeur de glycémie est de 3 g/L chez le patient.
 - La valeur de D_dimers est de 485 µg/l (valeur normale).
 - Le taux de SPO2 est de 85% chez le patient.
 - La valeur de CRP est de 79 mg/l chez le patient.
 - La valeur de L'inr n'est pas déclaré.
 - c) Les facteurs de risque :
 - Le patient est âgé de 53 ans.
 - Le patient est une femme.
 - Le patient a des antécédents cardiovasculaire.
 - Le patient n'a pas le diabète.
 - Le patient ne souffre pas d'insuffisance rénale.
 - Le patient n'a pas le HTA (hypertension).
 - Le patient ne souffre pas d'une maladie pulmonaire.
 - Le patient n'a pas des maladies du système immunitaire.

Le résultat du test présenté dans la figure 4.9 est comme suit :

Le patient est infecté par le Covid-19 d'après les pourcentages **TRUE = 69%** et **FALSE = 31%**.

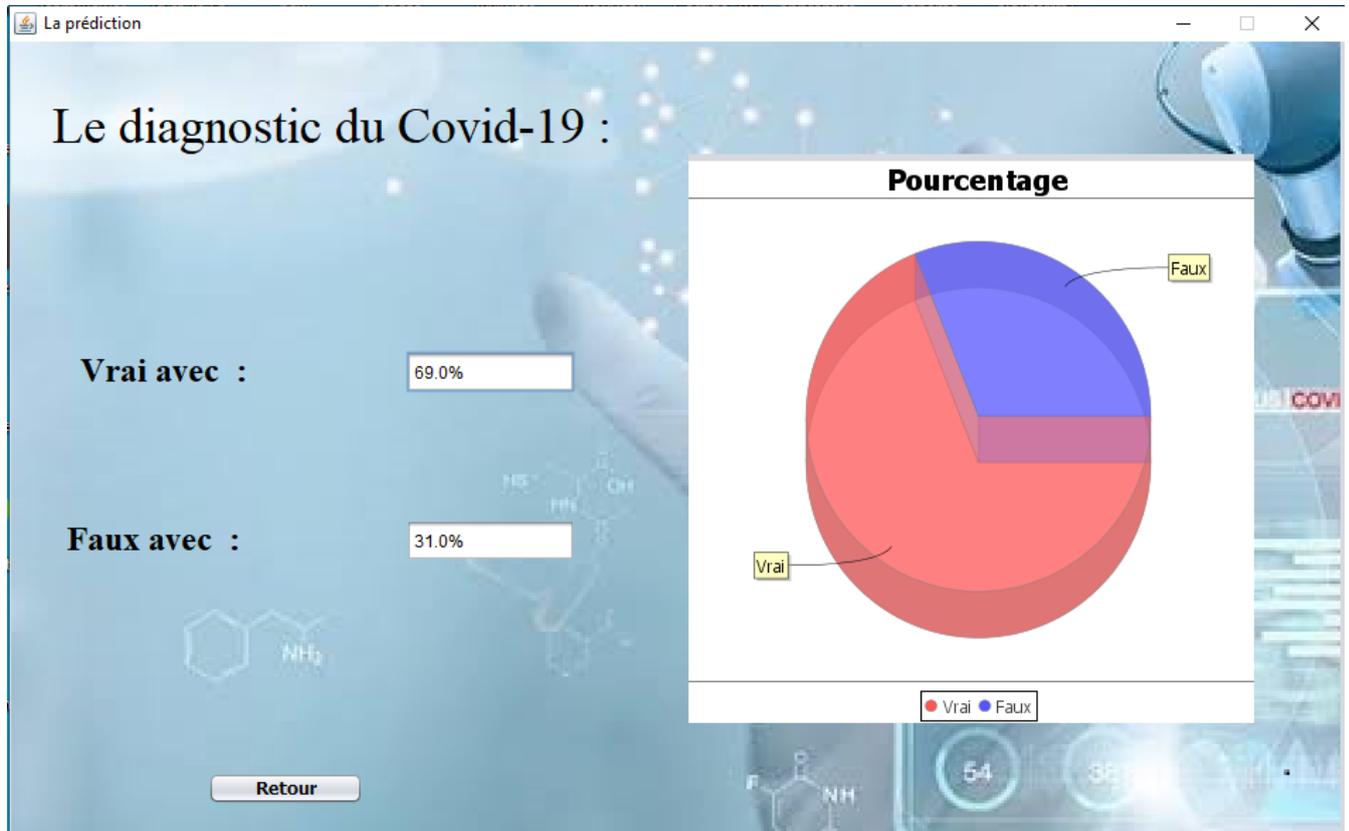


FIGURE 4.9 – Résultat du Test 1

2. Test 2 :

- **Résultat attendu :** le patient n'est pas infecté par le Covid-19

- **Evidences :**

- a) Les symptômes :

- Le patient n'a pas de maux de tête.
 - Le patient souffre de la fièvre.
 - Le patient souffre de la toux sèche.
 - Le patient a des courbatures.
 - Le patient n'a pas une diarrhée.
 - Le patient n'a pas un essoufflement.

- b) Les signes :

- La valeur de l'urée sanguine est de 0.26 g/L chez le patient.

- La valeur de créatinine n'est pas déclaré.
- La valeur de glycémie est de 0.8 g/L chez le patient.
- La valeur de D-dimers n'est pas déclaré.
- Le taux de SPO2 est 98% chez le patient.
- La valeur de CRP est de 7.1 mg/l (valeur anormale).
- La valeur de l'INR n'est pas déclaré.

c) Les facteurs de risque :

- Le patient est âgé de 6 ans.
- Le patient est une fille.
- Le patient ne souffre d'aucune maladie.

Le résultat du test présenté dans la figure 4.10 est comme suit :

Le patient n'est pas infecté par le Covid-19 d'après les pourcentages **TRUE = 35%** et **False = 65%**.

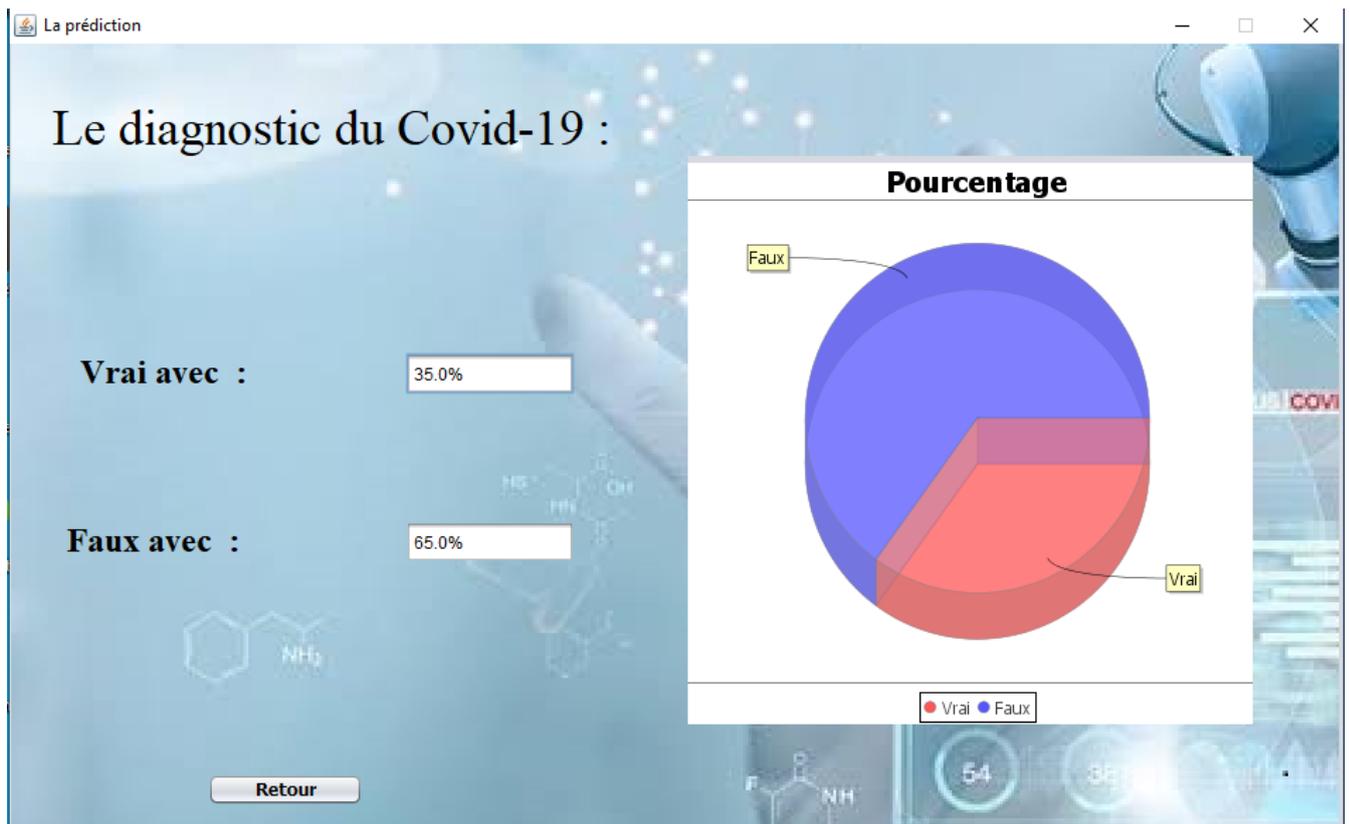


FIGURE 4.10 – Résultat du Test 2

Le médecin a testé dix cas positifs et autant de cas négatifs, tous les résultats y compris les deux que nous avons présenté ont été compatibles avec son diagnostic, ce qui prouve la performance de notre système.

4.7.2 Validation du système par les mesures de performances :

La validation a pour objectif de tester l'efficacité et la fiabilité du système en plus de la précision de ses résultats par rapport aux données précédemment introduites à partir du dataset. Afin de valider notre système nous avons utilisé la matrice de confusion qui résume les prédictions correctes et incorrectes.

	Predicted : True	Predicted : False
Actual : True	TP	FN
Actual : False	FP	TN

TABLE 4.1 – La matrice de confusion

- TP : est le nombre de patients touchés par le Covid-19 qui sont détectés correctement.
- TN : est le nombre de patients qui ne sont pas touchés par le Covid-19 et qui sont détectés correctement.
- FN : est le nombre de patients qui sont touchés par le Covid-19 et qui ne sont pas détectés correctement.
- FP : est le nombre de patients qui ne sont pas touchés par le Covid-19 et qui ne sont pas détectés correctement.

La table de confusion de notre système :

	Predicted : True	Predicted : False
Actual : True	140	18
Actual : False	4	18

TABLE 4.2 – La matrice de confusion de notre système

- **Le cas (True Positive)** : sur les 158 patients confirmés atteint par le Covid-19, 140 cas ont été correctement détectés. Ce qui représente un pourcentage de 89% des cas
- **Le cas (False Positive)** : sur les 22 patients confirmés non atteint par le Covid-19, les résultats restitués n'étaient erronés que pour 4 cas. Ce qui représente un pourcentage de 18% des cas.
- **Le cas (False Negative)** : sur les 158 patients confirmés atteint par le Covid-19, 18 cas ont été incorrectement détectés. Ce qui représente un pourcentage de 11% des cas.
- **Le cas (True Negative)** : sur les 22 patients confirmés non atteint par le Covid-19, 18 cas ont été correctement détecté. Ce qui représente un pourcentage de 81% des cas.

Par la suite nous allons calculer les mesures suivantes :

- **Precision** : appelée aussi la valeur prédictive positive qui permet de savoir quelle proportion d'identifications positives était effectivement correcte.
- **Recall** : permet de savoir la proportion de résultats positifs réels qui ont été identifiés correctement [63].
- **F-score** : est calculé à partir de la précision et du rappel. Elle permet de mesurer la capacité du système à donner toutes les solutions correctes et à refuser ceux qui ne le sont pas [63].
- **Accuracy** : est le niveau d'exactitude d'une mesure par rapport à sa valeur réelle. Elle est définie comme la somme des éléments diagonaux de la matrice de confusion divisée par le nombre d'individus [63].
- **Fallout** : est le pourcentage de l'erreur du système [63].

Ces mesures sont définies comme suit :

$$Precision = TP / (TP + FP)$$

$$Recall = TP / (TP + FN)$$

$$F - score = (2 \times Precision \times Recall) / (Precision + Recall)$$

$$Accuracy = (TP + TN) / (TP + FN + TN + FP)$$

$$Fallout = FP / (FP + TP)$$

Les résultats obtenus : **Precision** = 0.97 , **Recall** = 0.89 , **F1** = 0.93 , **Accuracy** = 0.88 , **Fallout** = 0.03

figure 4.11 représente les résultats des mesures de performance obtenus par notre système.

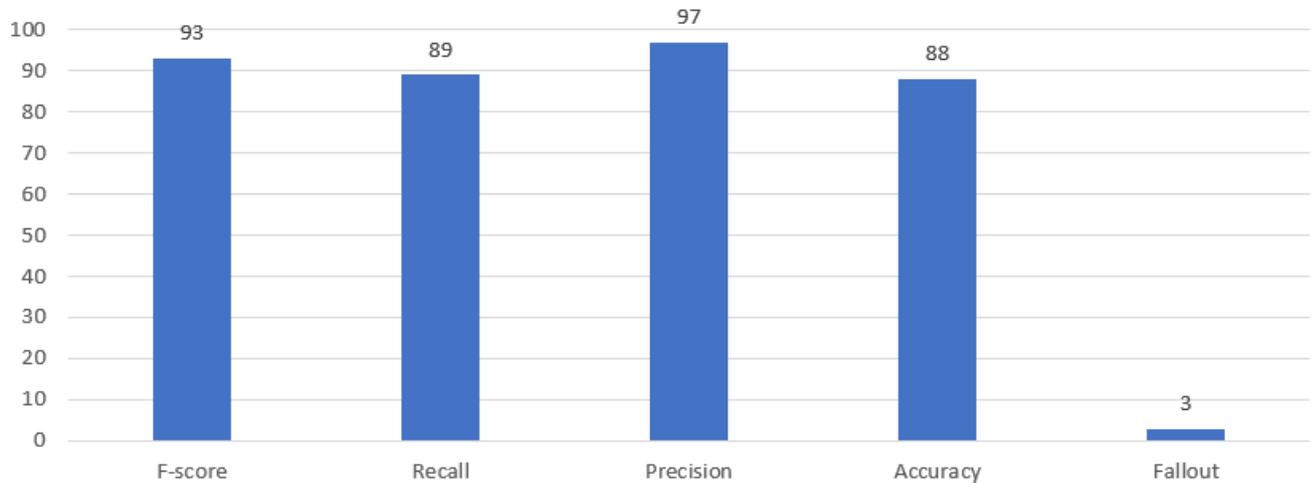


FIGURE 4.11 – Histogramme 1 : F-score, recall, Precision, Accuracy, Fallout

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

- **Recall** : Le taux de Recall est élevé grâce au taux faible de Falout.
- **Precision** : Le taux de précision est élevé, car le système retourne des résultats positifs correctes plus que des résultats positifs incorrectes.
- **Accuracy** : Le taux Accuracy du système est élevé grâce aux valeurs élevées de TP et TN.
- **F-score** : Le taux de F-score est élevé grâce aux taux élevés de Precision et Recall .
- **Fallout** : Le taux de Fallout est faible, cela revient au taux élevé de Precision.

4.8 Comparaison de notre système avec d'autres modèles probabilistes

Afin d'évaluer la performance de notre système nous avons soumis les résultats obtenus par ce dernier à un test comparatif avec les résultats obtenus par deux autres modèles probabilistes à savoir le réseau bayésien classique qui a la même structure que notre réseau bayésien hybride et le classifieur Naïf de Bayes.

La figure 4.12 visualise les résultats des mesures obtenues par notre système et ceux des deux autres classifieurs.

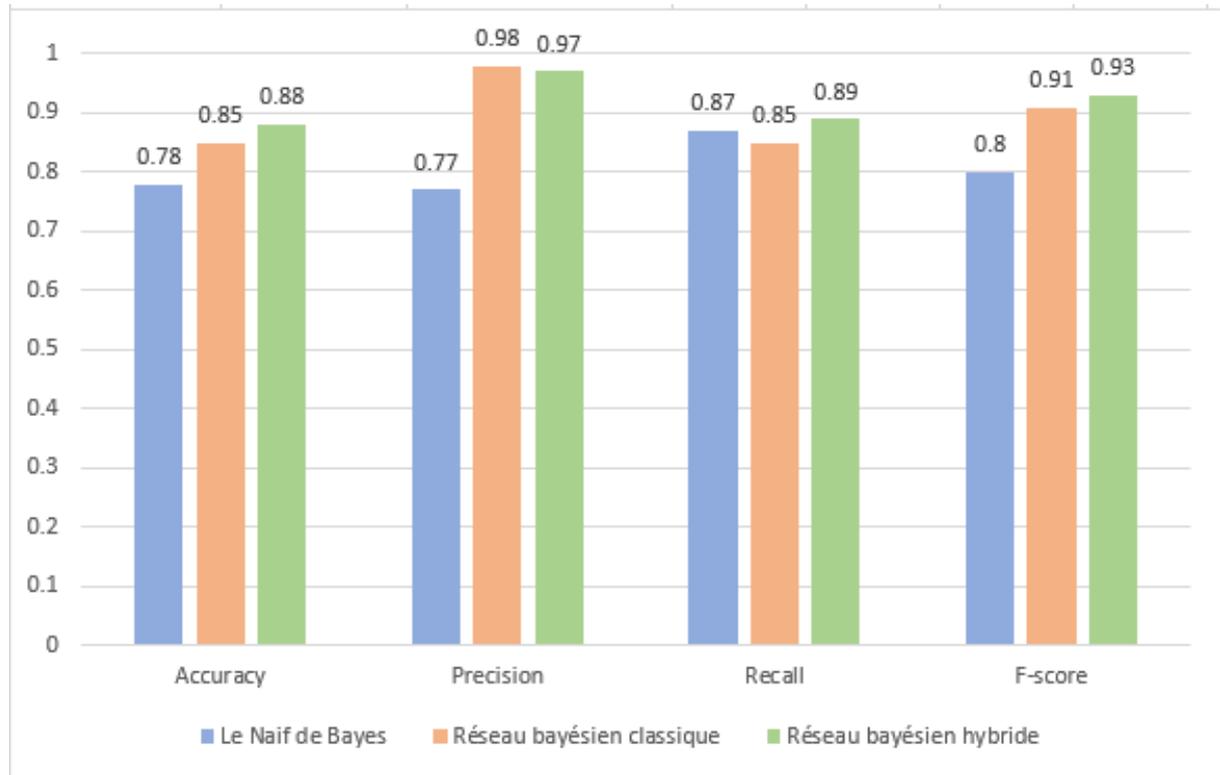


FIGURE 4.12 – Histogramme 2 : F-score, recall, Precision, Accuracy

4.8.1 Analyse des résultats

La comparaison des résultats restitués par les trois systèmes précités, nous permet de faire les évaluations ci-après :

- En terme de Accuracy, notre système restitue un meilleur résultat que les deux autres systèmes, à hauteur de 10% par rapport au réseau bayésien classique et de 3% par rapport au classifieur Naïf de Bayes, cette performance s'explique principalement par les scores élevés de true positive (TP) et true négative (TN) obtenus par notre système.
- Le taux de Recall obtenu par notre système (89%), s'avère lui aussi meilleur que ceux obtenus avec le réseau bayésien classique (85%) et le classifieur Naïf de Bayes (87%), Ceci étant la conséquence de la faible valeur de false negative (FN), ce qui confirme le corollaire avec le faible écart du nombre de cas déclarés avec ceux confirmés positives.
- En terme de Precision, notre système confirme son niveau de performance, en atteignant un résultat supérieur de 20% à celui du classifieur Naïf de Bayes et

pratiquement équivalent à celui du réseau bayésien classique. Les bons pronostics obtenus avec notre système et le réseau bayésien classique se confirment à travers la valeur élevée de true positive (TP) et de la faible valeur de false positive (FP).

- Dans la continuité des performances précitées réalisées par notre système, nous constatons le bons score obtenu par notre système au niveau de F_score, qui devance celui du réseau bayésien classique et du classifieur Naif de Bayes respectivement de 13% et de 2%.

En analysant les mesures obtenues, nous concluons que notre système est plus performant avec un taux de Accuracy, Recall et F_score plus élevé que celui des deux autres classifieurs. Ce qui signifie qu'il a une bonne performance.

4.9 Conclusion

À travers ce dernier chapitre nous avons introduit notre système et présenté les différents outils, langages et API utilisés pour sa construction. Nous avons également présenté les différentes mesures de validation afin d'évaluer sa pertinence.

Conclusion et perspectives

Le domaine de la santé est un domaine rempli de connaissances incertaines et imprécises. Cependant il exige une grande précision vu qu'une seule erreur peut engendrer des pertes humaines ainsi que d'autres dégâts tels que les perturbations économiques et sociales. Afin de diminuer ces dégâts, ce travail s'est focalisé sur la recherche d'une solution informatique d'aide au diagnostic du Covid-19.

Les ontologies permettent de représenter les connaissances d'un domaine précis d'une manière simple et formelle. Cependant, elles ne permettent pas la modélisation et le raisonnement des connaissances incertaines. Afin de traiter ces connaissances nous avons utilisé les réseaux bayésiens qui sont une approche puissante pour traiter l'incertitude, plus précisément les réseaux bayésiens hybrides.

Dans la première partie de ce mémoire nous avons présenté les ontologies qui se sont avérées incapable de représenter les connaissances incertaines ensuite nous avons abordé les réseaux bayésiens.

La deuxième partie a été consacré à l'élaboration de notre dataset en se basant sur des données de cas réels récupérés auprès d'un établissement de santé. Le dataset ainsi élaboré a permis la construction de notre ontologie classique en suivant les étapes de la méthode Méthontology et aussi la construction du réseau bayésien hybride en passant par l'apprentissage de sa structure qui a été réalisé avec l'expert du domaine et ensuite l'apprentissage des paramètres en appliquant l'algorithme EM sur les nœuds discrets et en définissant les équations de distribution pour les noeuds continus à l'aide de l'algorithme best fit.

Pour l'intégration des connaissances incertaines dans l'ontologie classique, nous avons créé une méta ontologie qui contient les composants du réseau bayésien hybride, ensuite nous l'avons intégré dans l'ontologie classique pour arriver à la fin à une ontologie probabiliste.

Afin de réaliser notre conception, nous avons implémenté notre système présenté par

plusieurs interfaces, ensuite nous l'avons testé avec des données réelles afin de le valider et montrer l'efficacité des réseaux bayésiens hybrides.

Nos perspectives consistent à améliorer notre dataset, en ajoutant d'autres cas afin de l'enrichir encore plus. En outre, réaliser une application mobile qui va aider directement les personnes qui doutent d'avoir le Covid-19 et appuyer la validation de notre système à travers la comparaison de ses restitutions avec ceux du réseau bayésien flou.

Bibliographie

1. GRUBER, T. R. *The Role of Common Ontology in Achieving Sharable, Reusable Knowledge Bases in Principles of Knowledge Representation and Reasoning : Proceedings of the Second International Conference* (Morgan Kaufmann, 1991), 601-602.
2. GRUBER, T. R. A translation approach to portable ontology specifications. en. *Knowledge Acquisition* **5**, 199-220. ISSN : 10428143. <https://linkinghub.elsevier.com/retrieve/pii/S1042814383710083> (2022) (juin 1993).
3. BORST, W. N. Construction of Engineering Ontologies for Knowledge Sharing and Reuse. Undefined. <https://research.utwente.nl/en/publications/construction-of-engineering-ontologies-for-knowledge-sharing-and-> (2021) (sept. 1997).
4. USCHOLD, M. & KING, M. *Towards a Methodology for Building Ontologies in In Workshop on Basic Ontological Issues in Knowledge Sharing, held in conjunction with IJCAI-95* (1995).
5. USCHOLD, M. & GRUNINGER, M. Ontologies : Principles, methods and applications. *The knowledge engineering review* **11**. ISBN : 1469-8005 Publisher : Cambridge University Press, 93-136 (1996).
6. OTTENS, K. *Un système multi-agent adaptatif pour la construction d'ontologies à partir de textes* thèse de doct. (Université Paul Sabatier-Toulouse III, 2007).
7. BRICKLEY, D. & GUHA, R. RDF vocabulary description language 1.0 : RDF schema. <http://www.w3.org/TR/rdf-schema/> (jan. 2004).
8. BECHHOFFER, S. *et al.* OWL web ontology language reference. *W3C recommendation* **10**, 1-53 (2004).
9. ABACHA, A. B., DA SILVEIRA, M. & PRUSKI, C. *Une approche pour la validation du contenu d'une ontologie par un système à base de questions/réponses in IC-24èmes Journées francophones d'Ingénierie des Connaissances* (2013).
10. DJEDIDI, R. *Approche d'évolution d'ontologie guidée par des patrons de gestion de changement.* thèse de doct. (Université Paris Sud-Paris XI, 2009), 259.

11. PSYCHE, V., BOURDEAU, J. & MENDES, O. Apport de l'ingénierie ontologique aux environnements de formation à distance. fr. *Sciences et Technologies de l'Information et de la Communication pour l'Éducation et la Formation* **10**, 89-126. ISSN : 1952-8302. https://www.persee.fr/doc/stice_1952-8302_2003_num_10_1_858 (2021) (2003).
12. DIENG-KUNTZ ROSE *et al.* *Méthodes et outils pour la gestion des connaissances : une approche pluridisciplinaire du knowledge management (2ième édition)*, Dunod Edition Informatiques Séries Systèmes d'Information, 2001. 2eme. ISBN : 2 10 006300 6 (2001).
13. FÜRST, F. *Contribution à l'ingénierie des ontologies : une méthode et un outil d'opérationnalisation* thèse de doct. (Nantes, 2004).
14. KOLLI, M. *Intégration d'ontologies dans le cadre du web sémantique : une détection des relations sémantiques basée sur le RÀPC. PhD thesis, Thèse présentée pour obtenir le grade de doctorat en Science en Informatique .Université Mentouri de Constantine.* doctorat (Université Mentouri de Constantine., 2011).
15. ZADEH, L. The concept of a linguistic variable and its application to approximate reasoning—I. en. *Information Sciences* **8**, 199-249. ISSN : 00200255. <https://linkinghub.elsevier.com/retrieve/pii/0020025575900365> (2022) (jan. 1975).
16. MILLIKEN, F. Three Types of Perceived Uncertainty About the Environment : State, Effect, and Response Uncertainty. *Academy of Management review*, 133-143 (jan. 1987).
17. KERBRAT-ORECCHIONI, C. L'ambiguïté : définition, typologie. fre. *MOM Éditions* **33**. Publisher : Persée - Portail des revues scientifiques en SHS, 13-36. https://www.persee.fr/doc/mom_0184-1785_2005_act_33_1_2284 (2021) (2005).
18. DUBOIS, D. & PRADE, H. La problématique scientifique du traitement de l'information. *Information-Interaction-Intelligence* **1**, 79-98 (2001).
19. RIALI, I. *Modélisation de la connaissance incertaine dans le web sémantique* These de doctorat (Saad Dahleb, sept. 2020).
20. JUDEA PEARL. *Probabilistic Reasoning in Intelligent Systems : Networks of Plausible Inference* 1st Edition. Anglais. ISBN : 978-0-08-051489-5 (sept. 1988).
21. ALAIN MILLE. *Cours Cognition et connaissance de l'université Claude Bernard Lyon 1 option : Master 2 Recherche Connaissance et Raisonnement, 2006 – 2007* 2006.
22. NAÏM, P., WUILLEMIN, P.-H., LERAY, P., POURRET, O. & BECKER, A. Réseaux bayésiens. *Eyrolles, Paris* **3**, 120 (1999).

23. NGUYEN, H.-T. *Réseaux bayésiens et apprentissage ensembliste pour l'étude différentielle de réseaux de régulation génétique français*. Thèse de doct. (UNIVERSITÉ DE NANTES, 2012).
24. MRAD, A. B. *Observations probabilistes dans les réseaux bayésiens* thèse de doct. (Université de Valenciennes et du Hainaut-Cambresis ; École nationale d'ingénieurs de Sfax (Tunisie)), 2015).
25. HENRION, M. en. in *Machine Intelligence and Pattern Recognition* 149-163 (Elsevier, 1988). ISBN : 978-0-444-70396-5. <https://linkinghub.elsevier.com/retrieve/pii/B9780444703965500194> (2022).
26. FRANÇOIS, O. *De l'identification de structure de réseaux bayésiens à la reconnaissance de formes à partir d'informations complètes ou incomplètes*. thèse de doct. (INSA de Rouen, 2006).
27. DOAN, T., HADDAWY, P. & NGUYEN, T. *A hybrid Bayesian Network modeling environment in Proceeding of the 1999 National Computer Science and Engineering Conference (NCSEC), Bangkok, Thailand* (Citeseer, 1999).
28. LUCAS, P. J., VAN DER GAAG, L. C. & ABU-HANNA, A. Bayesian networks in biomedicine and health-care. *Artificial intelligence in medicine* **30**. ISBN : 0933-3657 Publisher : Elsevier Science Publishers Ltd. Essex, UK, 201-214 (2004).
29. GARCÍA-CRESPO, Á., RODRÍGUEZ, A., MENCKE, M., GÓMEZ-BERBÍS, J. M. & COLOMO-PALACIOS, R. ODDIN : Ontology-driven differential diagnosis based on logical inference and probabilistic refinements. en. *Expert Systems with Applications* **37**, 2621-2628. ISSN : 09574174. <https://linkinghub.elsevier.com/retrieve/pii/S0957417409007702> (2021) (mars 2010).
30. LASKEY, K. B., HABERLIN, R., CARVALHO, R. N. & da COSTA, P. C. G. *PR-OWL 2 Case Study : A Maritime Domain Probabilistic Ontology*. in *STIDS* (2011), 76-83.
31. FENZ, S. An ontology-based approach for constructing Bayesian networks. en. *Data & Knowledge Engineering* **73**, 73-88. ISSN : 0169023X. <https://linkinghub.elsevier.com/retrieve/pii/S0169023X11001583> (2021) (mars 2012).
32. CHANG, Y.-S., FAN, C.-T., LO, W.-T., HUNG, W.-C. & YUAN, S.-M. Mobile cloud-based depression diagnosis using an ontology and a Bayesian network. en. *Future Generation Computer Systems* **43-44**, 87-98. ISSN : 0167739X. <https://linkinghub.elsevier.com/retrieve/pii/S0167739X1400137X> (2021) (fév. 2015).
33. HU, H. & KERSCHBERG, L. *Evolving Medical Ontologies Based on Causal Inference in 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (IEEE, Barcelona, août 2018), 954-957. ISBN : 978-1-5386-6051-5. <https://ieeexplore.ieee.org/document/8508259/> (2021).

34. LI, D. *et al.* Uncertain Knowledge Reasoning Based on the Fuzzy Multi Entity Bayesian Networks. *Computers, Materials & Continua*. <http://www.techscience.com/cmc/v61n1/23114> (2022) (2019).
35. DRUENNE - REDIS, S. *Connaissances et gestion de l'incertitude diagnostique par les internes en stage ambulatoire. Etude quantitative auprès des internes de médecine générale à Amiens en 2020* thèse de doct. (UNIVERSITE PICARDIE JULES VERNE AMIENS, 2020). http://www.sfm.org/data/generateur/generateur_fiche/1479/fichier_these_s_druenne-redise8d5.pdf.
36. ABDELMALEK, R. Covid-19, chroniques d'une pandémie annoncée Covid-19, chronicles of a forecasted pandemic. *La Tunisie médicale* **98**, 278-281 (2020).
37. *Site officiel de l'Organisation mondiale de la Santé* <https://www.who.int/fr> (2022).
38. *Coronavirus fr.* <https://www.who.int/fr/health-topics/health-systems-governance> (2021).
39. KUDÜS, C. *IMPACTS SOCIO-ÉCONOMIQUES DE LA PANDÉMIE DE COVID-19 DANS LES PAYS MEMBRES DE L'OCI : Perspectives et défis* français. **No : 9**. ISBN : 978-975-6427-99-6. <https://sesricdiag.blob.core.windows.net/sesric-site-blob/files/article/726.pdf?fbclid=IwAR0Wje3WFkaqfotloytksGITRy8Hm9jNewc58pnmY8dUs> (Turquie, 2020).
40. LACOMBE, M., HARLAY, A. & MONCEAUX, P. *Lexique de termes médicaux avec lexique étymologique - Michel Lacombe - 12^e éd.* <https://livre.fnac.com/a192553/Michel-Lacombe-Lexique-de-termes-medicaux> (2021) (déc. 1998).
41. BOTTÉRO, J. *Symptomes, signes, écritures en Mésopotamie ancienne : Extrait de "Divination et rationalité"* French. OCLC : 492948808 (PUF., Paris, 1990).
42. CUCHERAT, M. La notion de facteur de risque. *Le Courrier de l'Arcol* **3**, 146-9 (1999).
43. *National Center for Biotechnology Information* <https://www.ncbi.nlm.nih.gov/> (2022).
44. FENTON, N. E. *et al.* A Bayesian network model for personalised COVID19 risk assessment and contact tracing. *MedRxiv*. Publisher : Cold Spring Harbor Laboratory Press, 2020.07. 15.20154286 (2021).
45. BUTCHER, R. & FENTON, N. *Extending the range of symptoms in a Bayesian Network for the Predictive Diagnosis of COVID-19* en. preprint (Health Informatics, oct. 2020). <http://medrxiv.org/lookup/doi/10.1101/2020.10.22.20217554> (2022).

46. PRODHAN, G. & FENTON, N. *Extending the range of COVID-19 risk factors in a Bayesian network model for personalised risk assessment* en. Pages : 2020.10.20.20215814. Oct. 2020. <https://www.medrxiv.org/content/10.1101/2020.10.20.20215814v1> (2022).
47. NEIL, M., FENTON, N., OSMAN, M. & MCLACHLAN, S. Bayesian network analysis of Covid-19 data reveals higher infection prevalence rates and lower fatality rates than widely reported. en. *Journal of Risk Research* **23**, 866-879. ISSN : 1366-9877, 1466-4461. <https://www.tandfonline.com/doi/full/10.1080/13669877.2020.1778771> (2022) (août 2020).
48. De TERWANGNE, C. *et al.* Predictive Accuracy of COVID-19 World Health Organization (WHO) Severity Classification and Comparison with a Bayesian-Method-Based Severity Score (EPI-SCORE). *Pathogens* **9**, 880. ISSN : 2076-0817. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7692702/> (2022) (oct. 2020).
49. OYELADE, O. N. & EZUGWU, A. E. A case-based reasoning framework for early detection and diagnosis of novel coronavirus. en. *Informatics in Medicine Unlocked* **20**, 100395. ISSN : 23529148. <https://linkinghub.elsevier.com/retrieve/pii/S2352914820303683> (2022) (2020).
50. KOUAMÉ, K.-M. & MCHEICK, H. An Ontological Approach for Early Detection of Suspected COVID-19 among COPD Patients. en. *Applied System Innovation* **4**. Number : 1 Publisher : Multidisciplinary Digital Publishing Institute, 21. ISSN : 2571-5577. <https://www.mdpi.com/2571-5577/4/1/21> (2022) (mars 2021).
51. ABBASIMEHR, H. & PAKI, R. Prediction of COVID-19 confirmed cases combining deep learning methods and Bayesian optimization. en. *Chaos, Solitons & Fractals* **142**, 110511. ISSN : 09600779. <https://linkinghub.elsevier.com/retrieve/pii/S0960077920309036> (2022) (jan. 2021).
52. ZHENG, H.-T., KANG, B.-Y. & KIM, H.-G. en. in *Uncertainty Reasoning for the Semantic Web I* (éd. da COSTA, P. C. G. *et al.*) Series Title : Lecture Notes in Computer Science, 161-173 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2008). ISBN : 978-3-540-89764-4 978-3-540-89765-1. http://link.springer.com/10.1007/978-3-540-89765-1_10 (2022).
53. KHERBACHE, H. & GUEMMOUZ, M. *Prédiction du diagnostic médical guidée par les ontologies probabilistes* français. Thèse de doct. (Saad Dahlab BLIDA, 2021). <https://di.univ-blida.dz/jspui/bitstream/123456789/12434/1/Guemmuoz%5C%20marwa%5C%20et%5C%20Kherbache%5C%20hafsa.pdf>.
54. FERNÁNDEZ-LÓPEZ, M., GÓMEZ-PÉREZ, A. & JURISTO, N. *Methontology : from ontological art towards ontological engineering*. Publisher : American Association for Artificial Intelligence (1997).

55. *Java, 2022* https://www.java.com/fr/download/help/whatis_java.html (2022).
56. *Python 2022* <https://www.python.org/> (2022).
57. *protégé* <https://protege.stanford.edu/> (2022).
58. *SMILE Engine – BayesFusion* <https://www.bayesfusion.com/smile/> (2022).
59. *Eclipse IDE for Java Developers | Eclipse Packages* <https://www.eclipse.org/downloads/packages/release/kepler/sr1/eclipse-ide-java-developers> (2022).
60. *Oracle Java Technologies | Oracle* <https://www.oracle.com/java/technologies/> (2022).
61. *OWL-API : Une API Java pour manipuler des ontologies* <https://www.contentside.com/owl-api-une-api-java-pour-manipuler-des-ontologies/> (2022).
62. *ProtegeOWL API Programmers Guide - Protege Wiki* https://protegewiki.stanford.edu/wiki/ProtegeOWL_API_Programmers_Guide (2022).
63. POWERS, D. & AILAB. Evaluation : From precision, recall and F-measure to ROC, informedness, markedness & correlation. *J. Mach. Learn. Technol* **2**, 2229-3981 (jan. 2011).