

**REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR
ET DE LA RECHERCHE SCIENTIFIQUE**

UNIVERSITE SAAD DAHLAB DE BLIDA
FACULTE DES SCIENCES
DEPARTEMENT D'INFORMATIQUE



MEMOIRE DE FIN D'ETUDES

Pour l'obtention

D'un Diplôme de Master en Informatique

Option : Systèmes Informatiques et Réseaux

THÈME :

**Systeme d'aide au diagnostic
médical en utilisant les réseaux
bayésiens flous**

Réalisé par :

M^{elle}. AMARI HOURIA

Soutenu devant :

Mme:Aroussi

Mme :Cheriguene

Mme. M.Fareh

Mr. Senouci.H

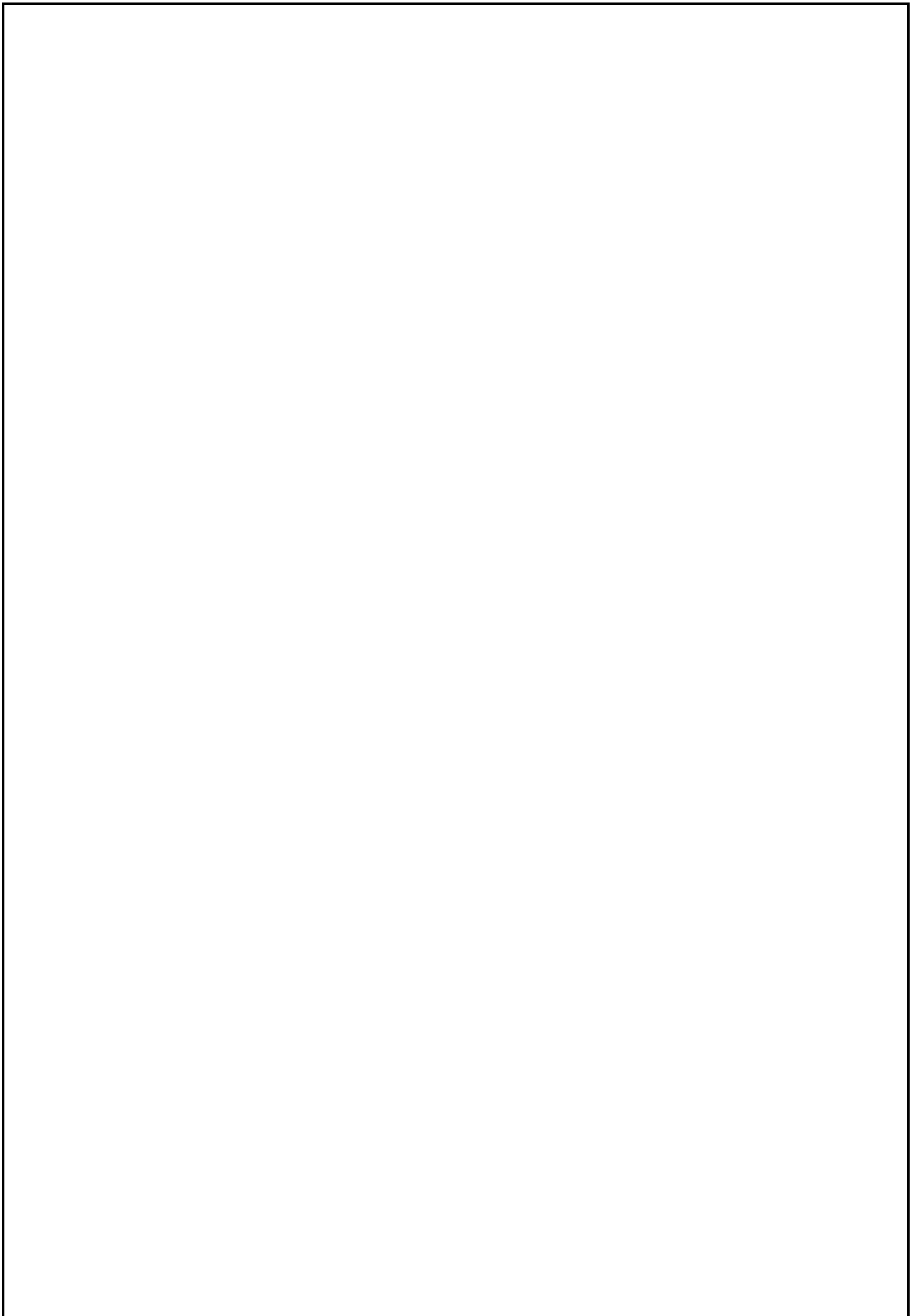
Présidente

Examinateur

Promotrice

Encadreur

2018/2019



Dédicaces

A :

Mes chers parents

Mon mari

Mes frères et sœurs

Ma belle famille

Je dédie ce mémoire

AMARI HOURIA

Remerciement

C'est avec un grand plaisir que je réserve cette page en signe de gratitude et de reconnaissance à tous ceux qui ont assisté ce travail.

Je tiens tout d'abord à remercier Dieu le tout puissant, qui nous a donné la force et la patience d'accomplir ce travail.

Je tiens à remercier, en premier lieu et tout particulièrement, mon promoteur Mme : **M. Fareh**, pour toutes les discussions intéressantes et fructueuses que nous avons eues, mais aussi, pour la patience dont elle a fait part à mon égard et pour son aide.

Je voudrais exprimer mes plus profonds remerciements au docteur Mr **I. RIALI**, ces précieux conseils et suivis durant la réalisation de ce projet.

Je remercie également les membres du jury pour l'intérêt qu'ils ont porté à mon travail.

Et bien sûr, ceux sans qui je ne serais rien : mes parents, mon mari, mes sœurs, ma famille et tous mes amis qui me soutiennent depuis toujours.

ملخص :

يتميز المجال الطبي بعدم اكتمال المعلومات.على الرغم من نجاح شبكات بايز التقليدية في جميع المجالات لنموذج وجعل التفكير مع المعرفة. ومع ذلك، فإنها لا توفر حلولاً في المواقف المعقدة أو المعرفة

غير الدقيقة. شبكات بايز الغامضة هي الأفضل في التكيف مع المشاكل عدم الدقة.

الهدف من هذا العمل هو اجراء نظام للمساعدة في التشخيص الطبي يعتمد على المعرفة غير المؤكدة باستخدام شبكة بايزي الضبابية قمنا ببناء شبكة بايزي لدينا من تعلم الهياكل ومعايير التعلم حتى لاختبار من البيانات المتعلقة بأمراض الكبد.

الكلمات المفتاحية: المعرفة غير المؤكدة ، عدم الدقة ، المنطق الضبابي ، شبكات بايزي أفضابية ، التشخيص الطبي.

Résumé

Le domaine médical est caractérisé par une incertitude médicale de nature radicale et d'incomplétude d'information. Malgré la réussite des réseaux bayésiens classiques utilisés dans tous les domaines pour modéliser et faire des raisonnements avec les connaissances. Cependant, ils ne fournissent pas une solution dans les situations compliquées où la connaissance est imprécise. Les réseaux bayésiens flous sont les mieux adaptés aux problèmes de l'imprécision.

L'objectif de ce travail est de réaliser un système d'aide ou de diagnostic médical basé sur des connaissances incertaines, en utilisant le réseau bayésien flou, nous avons construit notre système à partir de l'apprentissage des structures et l'apprentissage des paramètres jusqu'à la validation à partir d'un ensemble de données concernant la maladie « foie » qui représente notre cas d'étude.

Mots clés : Connaissance incertaine, Imprécisions, Logique floue, Réseaux bayésiens flous, Diagnostic médical.

Abstract

The medical field is characterized by a radical medical uncertainty incompleteness of information. Despite the success of the classic Bayesian networks used in all areas to model and make reason with knowledge. However, they do not provide a solution in complicated situations where the knowledge is imprecise. Bayesian fuzzy networks are suited to problems vagueness.

The objective of this work is to realize a system of help or medical diagnosis based on uncertain knowledge, using the Bayesian fuzzy network; we build our system from learning structures and learning parameters up to testing and validating from liver disease data.

Key words: Uncertain knowledge, Imprecision, Fuzzy logic, Bayesian Fuzzy Networks, Medical diagnosis.

Table de matière

Introduction générale.....	1
Chapitre 1 :Diagnostic médical	3
1. Introduction	4
2. Les incertitudes médicales	4
3. Signe.....	6
4. Symptômes	7
5. La maladie	7
6. Diagnostic médical	8
7. Les réseaux bayésiens et raisonnement abductif.....	11
8. Conclusion.....	11
Chapitre2 :Réseau Bayésien	12
1. Introduction	13
2. Définition de réseau bayésien	13
3. Un exemple de réseau bayésien	14
4. Apprentissage dans les réseaux bayésiens	14
4.1 Apprentissage de structure.....	15
4.1.1. Algorithme de structure (à partir des données complètes).....	15
4.1.2. Algorithme de structure (à partir des données incomplètes).....	16
4.2.Apprentissage des paramètres.....	16
4.2.1. Apprentissage des paramètres (à partir des données complètes).....	16
4.2.2. Apprentissage des paramètres (à partir des données incomplètes).....	17
5. Inférence dans les réseaux bayésiens	17
6. Quelques domaines d'application.....	19
7. La théorie des ensembles floue	19
8. Définition les réseaux bayésiens flous.....	20

9.	Exemple de réseau bayésien flou.....	21
10.	Conclusion	21
Chapitre 3 :La conception		22
1.	Introduction	23
2.	Motivation de l'étude de cas	23
3.	La maladie de foie	Error! Bookmark not defined. 4
4.	Base des données sur la maladie de foie.....	Error! Bookmark not defined. 5
5.	Les informations sur la source	Error! Bookmark not defined. 5
6.	Les attributs	Error! Bookmark not defined. 7
7.	Les informations sur les attributs.....	Error! Bookmark not defined. 7
8.	Création Réseau Bayésien.....	29
	8.1. L'apprentissage de la structure.....	29
	8.1.1. Transformation des attributs réels vers le type nominal....	29
	8.1.2. Les algorithmes pour l'apprentissage de structure.....	31
	8.1.3. Discussion sur algorithmes d'apprentissage de structure...	33
	8.1.4. Structure finale.....	33
	8.2. L'apprentissage des paramètres.....	34
9.	La fuzzification... ..	36
10.	L'Inférence	41
11	Conclusion	42
Chapitre 4 :Implementation et test de système		43
1.	Introduction	44
2.	Outils de développement de notre système.....	44
	2.1.java.....	44
	2.2.Netbeans.....	44
	2.3.Smile.....	45
	2.4.jSmile.....	45
	2.5.GeNie.....	45

3.	Les interfaces.....	47
4.	Test du système.....	50
5.	Validation croisée.....	51
	5.1. Résultat de la validation croisée.....	52
6.	Conclusion.....	52
	Conclusion Générale.....	53
	Références	55

La liste des figures

Figure 1.	Exemple réseau bayésien simple	14
Figure 2.	Exemple de la logique flou.	21
Figure 3.	Structure de réseau bayésien utilisé algorithmme Naïve Bayes.	31
Figure 4.	Structure de réseau bayésien utilisé algorithmme augmented Naive Bayes.....	3
Figure 5.	Structure de réseau bayésien utilisé algorithmme Greedy Thick Thinning.....	33
Figure 6.	Structure après la modification de l'expert.....	34
Figure 7.	Paramètres d'Age après EM.	35
Figure 8.	Paramètres de Total Bilirubin après EM.....	35
Figure 9.	Paramètres de Totale proteins après EM.	35
Figure 10.	Structure après l'apprentissage des paramètres.....	36
Figure11.	L'interface d'accueil.....	47
Figure 12.	Interface de réseau bayésien classique.	48
Figure 13.	Interface inference de réseau bayésien classique.....	48
Figure 14.	Interface de réseau bayésien flou.....	49
Figure 15.	Interface inference de réseau bayésien flou.....	49
Figure67.	Résultats de test.....	50
Figure 18.	Exemple d'un résultat de validation (K=2).....	5

Introduction générale

1. Contexte

Dans le domaine médical, les données sont souvent imprécises, ambiguës et incomplètes: un signe donné peut être présent ou absent dans la même maladie, le même signe peut être présent dans plusieurs maladies différentes. Les résultats des examens supplémentaires ne vous ont pas tué avec une fiabilité imparfaite. Leur interprétation et leur utilisation sont incertaines, l'évolution de la maladie et les résultats du traitement entrepris ne peuvent jamais être totalement sûrs. Tout cela oblige le médecin à prendre ses décisions, malgré l'incertitude qui entoure la prise de décision qui a souvent à voir avec la complexité du diagnostic, compte tenu de la nature incertaine des connaissances médicales, qui contribue à de grandes variations dans la pratique clinique, aux erreurs cliniques et mauvais respect des directives. Plusieurs approches existent dans la littérature médicale pour aider le médecin à prendre de meilleures décisions, parmi ces méthodes existent des réseaux bayésiens flous. Un RBF est une extension d'un réseau bayésien classique, permettant de modéliser simultanément des connaissances imprécises et floues. Il permet également un raisonnement probabiliste basé sur des observations imprécises et / ou vague.

Les travaux présentés sur les maladies hépatiques du réseau bayésien flou concernent des modèles graphiques probabilistes, qui sont d'une part des modèles de représentation de la connaissance et de l'incertitude et d'autre part des modèles de raisonnement.

2. Problématique

L'imprécision dans le diagnostic est fréquemment rencontrée dans la pratique médicale et provoque du stress chez les patients ainsi que les médecins. En effet, les facteurs contribuant à l'imprécision comprennent la variabilité biologique des patients, les préjugés des patients et des médecins, l'erreur dans l'interprétation des tests.

Cette imprécision peut se manifester par exemple, lors de la probabilité qu'un patient est atteint par le foie, cela conduit à poser la question suivante :

- comment construire un réseau bayésien flou pour traiter ces connaissances imprécises qui servent à inférer le diagnostic médical ?

2. L'objectif

Dans ce contexte, l'objectif de ce travail est de mettre en place un système d'aide au diagnostic pour diagnostiquer à partir des observations sur les patients (biologiques, tests... etc.). Nous avons choisi d'utiliser des réseaux bayésiens flous (RBF) qui sont des techniques utiles et efficaces. Ces techniques ont prouvé leur efficacité dans plusieurs domaines, comme le domaine de l'aide à la décision et le domaine de la modélisation de systèmes complexes. L'objectif est de concevoir et mettre en œuvre un système d'aide au diagnostic médical guidé par le FBR à partir d'un ensemble de données concernant la maladie "foie".

4. Organisation du mémoire

L'organisation du mémoire est comme suit :

Chapitre 1 - le diagnostic médical : ce chapitre contient les définitions des incertitudes médicales.

Chapitre 2 - les réseaux bayésiens : ce chapitre a pour objectif de donner une vision sur les réseaux bayésiens standards et flous. Nous parlons sur les différentes méthodes d'apprentissage et d'inférence.

Chapitre 3 - la conception du système : nous présentons les processus de notre système pour construire des réseaux bayésiens flous.

Chapitre 4 – Implémentation et test du système: est dédié à l'implémentation des solutions issues de la conception. Notamment nous présentons dans ce chapitre, les principales interfaces de notre système et test du système.

Enfin, une conclusion générale synthétise notre travail et présente les perspectives envisagées.

Chapitre 1

Le diagnostic médical

1. Introduction

La santé est un domaine où l'incertitude est d'une importance considérable. En effet, un diagnostic médical fiable permettra d'identifier la maladie d'un patient sur la base d'observations de ses signes et symptômes et de faire un bon raisonnement. Ces imperfections gèrent cette connaissance imprécise et incertaine qui doit être présentée dans un graphe de réseau bayésien.

Dans ce chapitre, nous allons représenter les définitions des signes, symptômes, maladies et diagnostic médical.

2. Les incertitudes médicales

La décision thérapeutique est caractérisée par une incertitude médicale de nature radicale d'incomplétude et d'asymétrie d'information. Face à son patient, le médecin est souvent dans l'ignorance totale malgré le poids de son savoir et sa capacité d'expertise. Pourtant, même dans ces situations, le médecin prend une décision. Les modèles de décision basées sur l'espérance d'utilité ne peuvent envisager une telle situation car ils reposent sur des mécanismes d'optimisation nécessitant la connaissance, du moins l'estimation du risque encouru. C'est donc vers les modélisations de la rationalité endogène que l'on se tournera. Parmi celles-ci, la théorie de la décision fondée sur les cas passés apparaît la mieux adaptée au problème posé car elle endogène l'incertain et l'utilise à son profit comme moteur de la nouveauté et de l'apprentissage. La transposition de la théorie à la décision médicale constitue donc, malgré de nombreuses imperfections, une voie de recherche prometteuse.

Dans le domaine de l'incertitude médicale, une double typologie peut être réalisée. D'une part en fonction de la personne à laquelle le risque est associé : le patient (la demande) comme le médecin (l'offre) subissent un aléa. D'autre part en fonction de l'étape du processus de "soins" : étape de prévention de la maladie, étape de dépistage de la maladie, étape de diagnostic et, enfin étape de soins curatifs.

- ✓ Le patient est, bien entendu, confronté à un risque maladie pouvant, dans un premier temps, se résumer à trois états : bonne santé, maladie curable et maladie incurable.

L'incertitude porte sur l'occurrence de la maladie. Or, cette occurrence n'est pas toujours

probabilisable, comme le suggère Fougère (1982) en rappelant, entre autres, que la maladie n'est pas indépendante des facteurs de risque [1].

- ✓ Dans un second temps, afin de lever la simplification précédente, l'incertitude à laquelle est confronté le patient peut-être affinée en distinguant un continuum d'états de la nature, ceci peut soit traduire l'incertitude devant l'occurrence de la maladie, celle-ci pouvant surgir n'importe quand, soit rendre compte du degré de "gravité" de la maladie lorsque celle-ci est déclarée, soit, enfin, faire état de l'infinité des pathologies. Dans les trois cas, c'est l'infinité qui pose problème car, comme le note Ménard (1994), les capacités de jugement individuel ne peuvent l'apprécier. Ceci d'une part parce que "le nombre de situations contingentes à prendre en compte excède les capacités de tout agent" d'autre part parce que le coût de traitement de l'information devient vite démesuré [1].

Ce n'est toutefois pas là la seule incertitude à laquelle le praticien est exposé. On peut relever ainsi trois autres formes d'incertitude de l'offre de soins, en fonction de l'étape du processus de soins :

- ✓ la première est celle de la justesse du diagnostic. Le rôle du diagnostic consiste à dévoiler toute l'information nécessaire à sa l'élaboration au moyen d'une interrogation ou d'une analyse clinique. Toutefois, l'information dévoilée au cours de ce processus n'est jamais complète puisque les investigations du médecin sont sélectives (aucun patient n'est jamais soumis à un examen exhaustif). Or, le patient n'est pas l'expert, il peut donc mettre, volontairement ou non, l'une des caractéristiques de son état ou de son comportement. On est alors dans une situation "classique" de hasard moral ou de sélection adverse.
- ✓ seconde incertitude porte sur la thérapie. Celle-ci n'est pas nécessairement efficace, même si le diagnostic est correct. La prophylaxie est basée sur les résultats d'études épidémiologiques fondées sur des analyses statistiques de nature fréquentâtes. Or la statistique n'a de sens qu'au pluriel. Appliquées à un groupe, fréquences et probabilités fournissent un renseignement théorique précieux, permettant de circonscrire le risque. Mais, appliquées à l'individu, ces notions perdent tout leur sens car elles répondent au risque par une mesure de l'incertitude, là où l'individu espère une certitude. Rien n'assure donc que le traitement curatif que choisira le médecin sera immanquablement adapté à

son patient. Dans le cas le plus fréquent où les soins curatifs sont efficaces, l'incertitude se déplace sur la qualité ces soins. En effet, lorsque, pour une pathologie donnée, les thérapies sont multiples, le patient attend non seulement d'être guéri, mais de l'être dans les meilleures conditions possibles. Or les critères de jugement de qualité sont aussi de nature probabiliste. Ici encore, rien ne prouve que le traitement choisi soit le meilleur pour le patient.

- ✓ La troisième incertitude concerne la connaissance même d'une thérapie ou d'une maladie. En matière médicale, les états de la nature ne sont jamais parfaitement connus. La liste des pathologies, comme celle des thérapies, est éphémère. Le praticien est donc dans l'impossibilité structurelle d'évaluer le risque maladie ou le risque guérison, quand bien même le diagnostic serait exhaustif. Supposer le contraire, comme le font les théories de la décision basées sur l'espérance d'utilité, reviendrait à supposer que le praticien peut juger d'un éventuel risque futur, encore inexistant.

Finalement, deux aspects caractérisent fondamentalement l'incertitude médicale. Celle-ci est fondamentalement radicale en raison d'une part d'une incomplétude des états de la nature (maladie, thérapie, mais aussi diagnostic), d'autre part en raison de son caractère endogène (facteurs de risque). On retrouve bien ici les deux critères de définition de l'incertitude pure. L'environnement dans lequel le médecin prend sa décision est donc à information incomplète et à information imparfaite [1].

3. Signes

Est la manifestation d'une maladie, constatée par un médecin ou tout autre observateur. L'interprétation de cette observation permet au médecin de s'orienter vers un diagnostic.

➤ Signe clinique:

- ✓ Signe général: un signe qui concerne l'organisme dans sa globalité (fièvre etc).
- ✓ Signe fonctionnel : signe ressenti subjectivement par le malade.
- ✓ Signe physique : signe constaté objectivement (comme l'état de la peau) [2].

➤ Signes para cliniques

Sont tous les indices qui permettent de détecter une maladie ou une anomalie par des examens non cliniques (c'est-à-dire tous les examens qu'on ne peut pas réaliser à l'œil nu). Par exemple les radios, les scanner, les IRM, les analyses d'urine ou encore les prises de sang sont des analyses para cliniques. Du coup, les signes para cliniques peuvent être un taux trop élevé de sucre dans le sang, un os déplacé visible sur une radio ou la présence de bactérie dans les urines. [3]

✓ **Signe positif**

Un signe est dit positif si c'est sa présence à une signification. Par exemple une quantité d'immunoglobuline du type A supérieure à 3g/dl est un signe positif pour le myélome multiple.

✓ **Signe négatif**

Un signe est dit négatif si c'est son absence à une signification. Par exemple l'absence de CRAB (C-hypercalcémie, R- insuffisance rénale, A- anémie, B- lésion osseuse).

4. Symptômes

Est un signe, un trouble observable par un patient, qui est la manifestation d'une maladie et de son évolution. Ou subjectif, et se distingue d'un signe clinique, détecté par le médecin ou lors d'un examen médical (analyses biologiques, imagerie médicale...) [4].

5. La maladie

Altération de l'état de santé se manifestant par un ensemble de signes et de symptômes perceptibles directement ou non, correspondant à des troubles généraux ou localisés, fonctionnels ou lésionnels, dus à des causes internes ou externes et comportant une évolution[5].

Toute maladie se définit par une cause, des symptômes, des signes cliniques et para cliniques, une évolution, un pronostic et un traitement.

6. Diagnostic médical

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

Le diagnostic médical n'est pas uniquement réalisé par les médecins. De nombreux autres soignants établissent des diagnostics chaque jour : kinésithérapeutes, dentistes, sages-femmes...

- ✓ Premièrement, le médecin prend connaissance des symptômes se manifestant chez un patient. à partir des symptômes, il formule des hypothèses diagnostiques initiales.
- ✓ Dans un deuxième temps, il procède à un examen initial du patient qui lui permet d'augmenter la part de confiance en certaines hypothèses, et de la diminuer pour d'autres.
- ✓ Le médecin obtient des signes supplémentaires via des examens biologiques ou d'imageries. En même temps, il pose au patient des questions dont les réponses peuvent être utiles à conforter ou rejeter une hypothèse initialement formulée.

Le diagnostic est la conclusion d'un cheminement intellectuel pouvant se reposer sur trois étapes de la consultation médicale : l'anamnèse, l'examen physique et les examens complémentaires.

✓ **L'anamnèse**

Corresponds à l'interrogatoire médical, C'est le moment où le patient va décrire ses symptômes, Mais également le moment où le soignant vont recueillir les informations générales du patient (son traitement, ses allergies, ses antécédents...)

✓ **Examen physique**

L'examen physique a comme but de rechercher des signes physiques des maladies.

Généralement, l'examen est divisé en inspection, palpation, auscultation et percussion.

Là encore, l'examen sera orienté selon le profil du patient, ses plaintes ou ses antécédents.

Une bonne connaissance des maladies (sémiologie) et du patient est primordiale pour savoir quoi chercher et interpréter les signes.

✓ **Examens complémentaires**

Le soignant dans certains cas souhaitera compléter son examen clinique par des examens complémentaires avant de poser son diagnostic final, les examens complémentaires sont l'ensemble des examens prescrits par le médecin et destinés à l'aider aux analyses.

Le processus du diagnostic médical peut relever d'un raisonnement déductif, inductif, abductif, par analogie [6].



➤ **Le raisonnement déductif**

Est un processus qui conduit, sur la base de prémisses fondatrices ou de propositions formulées à partir de données d'observation ou d'expérience, à des conclusions qui en découlent rationnellement par enchaînement et de façon intelligible.

Le **sylogisme aristotélicien** est un exemple de raisonnement déductif qui permet, à partir de deux prémisses, de tirer une conclusion certaine (tableau 1). Ce type de raisonnement ne vaut que par la valeur des prémisses [7].

Exemple	Structure
Tous les médecins sont des professeurs	Prémisse majeure
Or, chahinez est médecin	Prémisse mineure
Donc, chahinez est professeur	Conclusion

Tableau 1. Exemple de syllogisme aristotélicien.

➤ **Le raisonnement inductif**

Est un raisonnement qui permet d'élaborer des lois ou des théories générales à partir des prémisses constituées d'observations particulières répétées.

L'induction parvient à une conclusion d'ordre général qui apporte une connaissance nouvelle. Cette connaissance ne peut prétendre qu'au vraisemblable.

Le raisonnement inductif permet d'étendre nos connaissances avec des méthodes scientifiques rigoureuses. Les connaissances du clinicien ne sont pas toutes du domaine de la science. Il est victime de préjugés, de pseudo-connaissances. Par pseudo-connaissances, on entendra tout fait ne reposant pas sur une vérité scientifique. Au fur et à mesure de son expérience, le clinicien accumule des observations et se fait une représentation de certaines maladies qui repose aussi sur un mode inductif mais moins rigoureux. Ce mode de raisonnement est utile pour façonner les scripts et une meilleure représentation appliquée des prototypes des maladies [7].

➤ **Le raisonnement abductif**

Est une inférence, c'est-à-dire un mouvement de pensée par lequel on passe d'un fait surprenant à une hypothèse susceptible d'en rendre compte. La conclusion d'une abduction n'est qu'une hypothèse qui doit être testée.

Face à un problème mal défini, le clinicien se raccroche à des données connues stockées dans ses connaissances.

Face à un problème mal défini, le chercheur élabore de nouvelles hypothèses qu'il teste en mettant en place un protocole expérimental.

« La théorie générale de l'abduction permet d'unifier la démarche hypothético-déductive bernardienne et le raisonnement clinique diagnostique. L'abduction est une inférence qui a pour point de départ les données descriptives d'un fait ou d'un phénomène et pour point d'arrivée une hypothèse qui s'accorde avec les données.

Ainsi, les 3 modes d'inférence (abduction, déduction, induction) sont utilisés par le clinicien et le chercheur [7].

➤ **Le raisonnement par analogie**

Le raisonnement analogique « qui va du particulier au particulier », qui généralise des cas particuliers allant ainsi du particulier au général il y a un raisonnement analogique qui interprète les cas particuliers à l'aide de cas typiques et généraux.

Le raisonnement analogique qui assimile un cas particulier un autre cas particulier et conclut que si deux choses se ressemblent à certains points du la vue, Elles se ressemblent encore à d'autres points de vue [8].

Les systèmes existants d'aide au diagnostic médical :

Année de publication	Editeur	Titre	Collaborateurs
30 sept. 2008	Groupe d'études de Psychologie de l'Université de Paris.	Bulletin de psychologie	Université de Paris
	Dunod, 1983	Technique Et Science Informatiques: TSI	Association française pour la cybernétique économique et technique
20 nov. 2007	La Documentation française, 1983	La distribution du médicament en France	
16 mai 2011	Vigot Éditions, 1986	Recueil de médecine vétérinaire,	Académie vétérinaire de France, Paris
	Dunod, 2008	Le système de santé en France: Diagnostic et propositions	

Tableau 2 les systèmes existants d' aide au diagnostic médical

7. Conclusion

Le diagnostic médical est un raisonnement scientifique, qui présente de nombreuses difficultés telles que l'incertitude et l'imprécision. Chaque étape du diagnostic médical peut donner lieu à des erreurs qui dans le domaine médical peuvent avoir de très graves conséquences. Dans le chapitre suivant, nous expliquerons les réseaux bayésiens.

Chapitre2

Les réseaux

bayésiens

1. Introduction

La représentation des connaissances et le raisonnement basé sur ces représentations ont donné lieu à de nombreux modèles. Modèles graphiques probabilistes, et plus précisément réseaux bayésiens. L'intérêt particulier des réseaux bayésiens est de prendre en compte simultanément des connaissances a priori expertes. Les différents travaux et recherches scientifiques ont utilisé la théorie des ensembles flous au format des réseaux bayésiens flous.

2. Définition de réseau bayésien

Les réseaux bayésiens (BR) sont un couplage entre la théorie des graphes et la théorie probabiliste [9]. Un modèle graphique probabiliste représentant des variables aléatoires sous la forme d'un graphique orienté acyclique. Il est composé d'un ensemble de variables et d'un ensemble d'arcs entre les variables.

Un réseau bayésien défini par: un graphe orienté acyclique $G = (V, E)$.

Avec V l'ensemble des nœuds du réseau et E l'ensemble des arcs. L'ensemble V est donc un ensemble de variables aléatoires discrètes. Chaque nœud de V est conditionnellement indépendant de ses non-descendants, compte tenu de ses parents immédiats [10].

Les réseaux bayésiens sont principalement utilisés pour le diagnostic (médical et industriel), l'analyse des risques, la détection du spam et le datamining.

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

x :le nœud

$Pa(x)$:le parent de nœud x .

Les réseaux bayésiens sont des modèles graphiques probabilistes représentant des connaissances et permettant de calculer des probabilités conditionnelles à l'aide du théorème de Bayes et d'apporter des solutions à différents types de problèmes.

Ils décrivent la distribution des probabilités associées à un ensemble de variables, dont certaines

sont directement dépendants et d'autres sont conditionnellement indépendants [11].

3. Un exemple de réseau bayésien

Ce matin-là, alors que le temps est clair et sec, M. Holmes sort de sa maison. Il s'aperçoit que la pelouse de son jardin est humide. Il se demande alors s'il a plu pendant la nuit, ou s'il a simplement oublié de débrancher son arroseur automatique. Il jette un coup d'oeil sur la pelouse de son voisin, M. Watson, et s'aperçoit qu'elle est également humide. Il en déduit qu'il a probablement plu et décide de partir au travail sans vérifier son arroseur automatique. La représentation graphique du modèle causal utilisé par M. Holmes est composée de 4 nœuds : Arroseur (A), Pluie (P), Jardin humide (J), Jardin du voisin humide (W). [12].

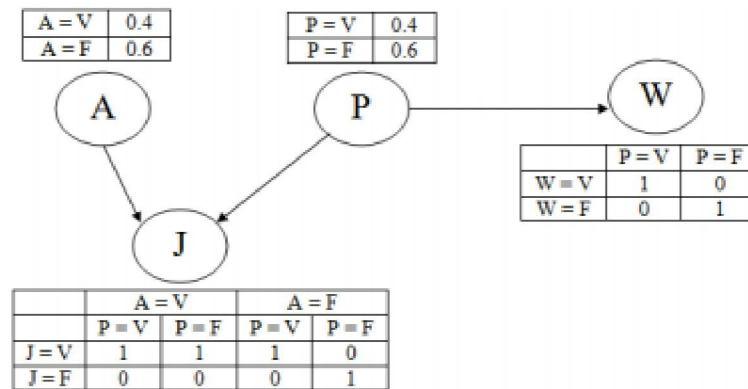


Figure 1. Exemple de réseau bayésien simple .

4. Apprentissage dans les réseaux bayésiens

Nous avons montré précédemment qu'un réseau bayésien est composé du graphe et d'un ensemble de probabilités conditionnelles.

Le problème d'apprentissage est divisé en deux parties: [13]

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

4.1 Apprentissage de structure

Nous avons examiné différentes méthodes d'apprentissage des structures d'un réseau bayésien à partir de données complètes ou incomplètes, ou avec l'aide d'un expert, en supposant que la structure de ce réseau était déjà connue.

Vient maintenant la question d'apprendre cette structure: comment trouver la structure qui représentera le mieux notre problème. Avant d'aborder les deux principales familles d'approches (recherche d'indépendance conditionnelle et méthodes basées sur les scores).

Il existe plusieurs travaux pour apprendre cette structure à partir de données complètes ou incomplètes, nous commençons par quelques algorithmes basés sur des données complètes:

a. Algorithme de structure (à partir des données complètes)

Deux grandes familles d'approches d'indépendances conditionnelles et méthodes basées sur un score.

✓ Algorithme basé sur score

Contrairement à la première famille de méthodes qui tentait de trouver une indépendance conditionnelle entre les variables, les approches suivantes chercheront soit la structure qui maximise un certain score, soit chercher les meilleures structures et combiner leurs résultats. Pour que ces approches basées sur les scores soient pratiques dans la pratique, nous verrons que le score doit être décomposable localement, c'est-à-dire exprimé comme la somme des scores locaux à chaque nœud. Il y a aussi le problème de traverser l'espace des réseaux bayésiens à la recherche de la meilleure structure. Une recherche exhaustive étant impossible, les algorithmes proposés fonctionnent sur un espace réduit (espace arborescent, ordonnancement des nœuds), ou effectuent une recherche gourmande dans cet espace.

✓ Algorithme indépendance conditionnelle

La seconde famille de méthodes Pearl et Verma d'une part avec les algorithmes IC (causalité inductive), Spirtes, Glymour et Scheines d'autre part avec les algorithmes PC (recherche causale). De même, plus récemment, l'algorithme BN-PC de Cheng et al. Ces algorithmes reposent tous

sur un principe identique: construire un graphe non orienté contenant les relations entre les variables, à partir de tests d'indépendance conditionnelle détecter des structures en V (également à l'aide de tests d'indépendance conditionnelle), "propager" les orientations de certains arcs, éventuellement en tenant compte causes artificielles dues à des variables latentes.

b. Algorithme de structure (à partir des données incomplètes)

Le premier problème à résoudre, lorsque les données sont incomplètes, concerne le calcul de la probabilité ou plus généralement du score pour une structure fixe, puis sa maximisation. Concernant la maximisation de cette vraisemblance, le principe de l'algorithme EM pourrait être utilisé pour estimer les paramètres[14].

4.2. Apprentissage des paramètres (à partir des données complètes)

Estimer les distributions de probabilité (ou les paramètres des lois correspondantes) à partir des données disponibles. L'estimation des distributions de probabilité, paramétriques ou non, est un sujet très vaste et complexe, les méthodes les plus utilisées dans le cadre des réseaux bayésiens, selon que les données à notre disposition sont complètes ou non, dans le cas où toutes les variables sont observées, la méthode la plus simple et la plus utilisée est l'estimation statistique qui consiste à estimer la probabilité d'un événement par la fréquence d'occurrence de l'événement dans la base de données. Cette approche, appelée maximum de vraisemblance (MV)[15].

4.2.1. Apprentissage des paramètres (à partir des données incomplètes)

Dans les applications pratiques, les bases de données sont très souvent incomplètes. Certaines variables ne sont que partiellement ou jamais observées, que ce soit à cause d'une défaillance du capteur, une variable qui ne peut être mesurée que dans un contexte spécifique, un répondant qui a oublié de répondre à une question, etc. pour ensuite se concentrer sur l'un des algorithmes les plus utilisés, l'algorithme EM de maximisation [16].

✓ Algorithme EM :

L'algorithme espérance-maximisation (en anglais expectation-maximization algorithm, souvent abrégé EM), proposé par Dempster et al. (1977)[17], est un algorithme itératif qui permet de trouver les paramètres de la probabilité maximale d'un modèle probabiliste lorsque celui-ci dépend de variables latentes non observables. De nombreuses variantes ont ensuite été proposées, formant toute une classe d'algorithmes. L'algorithme d'espérance-maximisation comporte :

- ✓ une étape d'évaluation de l'espérance (E), où l'on calcule l'espérance de la vraisemblance en tenant compte des dernières variables observées.
- ✓ une étape de maximisation (M), où l'on estime le maximum de vraisemblance des paramètres en maximisant la vraisemblance trouvée à l'étape E.

On utilise ensuite les paramètres trouvés en M comme point de départ d'une nouvelle phase d'évaluation de l'espérance, et l'on itère ainsi [18].

5. Inférence dans les réseaux bayésiens

Un réseau bayésien est donc un cadre probabiliste formalisant l'indépendance conditionnelle d'un système. La loi des probabilités conditionnelles permet de calculer des probabilités a priori pour toute variable du réseau. Mais ce qu'on appelle "inférence" dans un réseau bayésien, c'est l'estimation (ou la mise à jour) des distributions de certaines variables lorsque l'on fixe (avec une observation par exemple) des états pour d'autres variables. On peut déduire certaines variables en utilisant la loi des probabilités conditionnelles si la variable instanciée est en amont de la variable à inférer, ou théorème de Bayes dans le cas contraire. Dans le cas où plusieurs variables sont instanciées, les auteurs [19] ont développé des algorithmes dits de «propagation de message» en appliquant les théorèmes de probabilités conditionnelles et de Bayes au voisinage des variables instanciées, et en propageant à leur tour les résultats de leur implémentation à leurs voisins et ainsi de suite.

• Inférence exacte

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

–X: variable de question/requête

- E: l'ensemble des variables d'évidence
- e: un événement particulier
- Y: l'ensemble des variables cachées
- L'ensemble complet des variables est:

$$\mathbf{X}=\{X\} \cup E \cup Y$$

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

✓ **Inférence approximative**

Les méthodes d'inférence exactes qui viennent d'être discutées ne peuvent pas être utilisées pour les grands réseaux. Pour cela, nous considérons des approches approximatives. Nous verrons des algorithmes basés sur un échantillonnage aléatoire (Monte Carlo) dont la précision dépendra d'un certain nombre d'échantillons.

✓ **Méthodes d'échantillonnage directes**

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

6. Quelques domaines d'application

Il existe plusieurs domaines d'application des réseaux bayésiens dont :

Santé : parmi les premières applications de réseau bayésien c'est l'application médicale, il est très utile dans le domaine médical.

Informatique : peut être représentée par certains types d'inférences, ou par un système de planification.

Gestion de connaissances : le domaine de la gestion des connaissances, qui connaît un intérêt croissant, est également un champ d'application potentiel pour les réseaux bayésiens, dans la mesure où ceux-ci offrent un formalisme riche et intuitif de représentation de la connaissance [20].

7. La théorie des ensembles flous

La théorie des sous-ensembles flous est une théorie mathématique du domaine de l'algèbre abstraite. Il a été développé par Lotfi Zadeh en 1965 afin de représenter mathématiquement l'imprécision relative à certaines classes d'objets et sert de base à la logique floue [21]. Des sous-ensembles flous (ou parties floues) ont été introduits afin de modéliser la représentation humaine des connaissances, et ainsi améliorer les performances des systèmes de décision qui utilisent cette modélisation. Au lieu d'appartenir à l'ensemble «vrai» ou «faux» de la logique binaire traditionnelle, la logique floue admet des degrés d'appartenance à un ensemble donné. Le degré d'appartenance à un ensemble flou est matérialisé par un nombre compris entre 0 et 1. Une valeur précise de la fonction d'appartenance liée à une valeur de la variable est notée μ et appelée « facteur d'appartenance ».

Exemple :

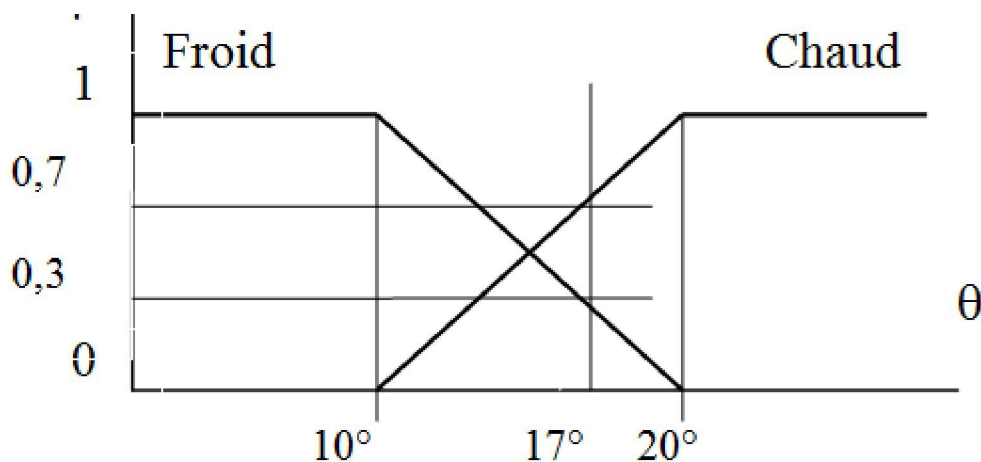


Figure 2. Exemple de la logique floue[21].

Explication de graphe :

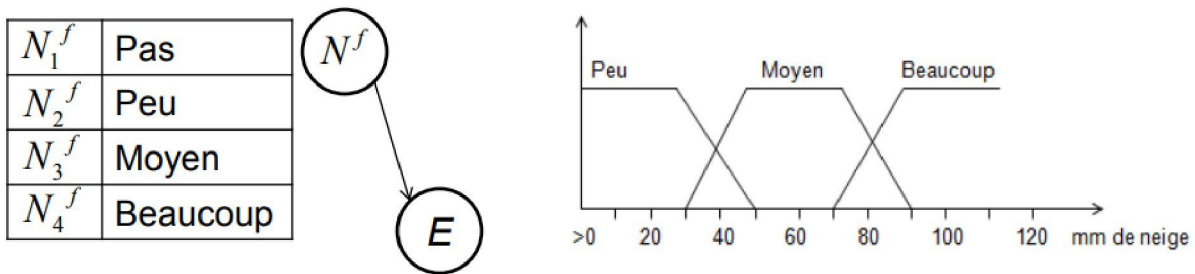
D'après ce graphique, on peut constater que pour une valeur $\theta = 17^\circ$, le facteur d'appartenance à l'ensemble « froid » vaut $\mu_{\text{froid}} = 0,3$ et le facteur d'appartenance à l'ensemble « chaud » vaut $\mu_{\text{chaud}} = 0,7$.

8. Définition les réseaux bayésiens flous

Diverses recherches ont récemment permis d'utiliser la théorie des ensembles flous dans le contexte des réseaux bayésiens. Différentes approches ont été proposées, mais à ce jour aucune n'est parvenue à un consensus. Nous avons cherché à généraliser les réseaux bayésiens aux réseaux bayésiens flous qui remplacent les états discrets par des états flous à travers le réseau. Cela contribue à renforcer la robustesse du système.

Plus précisément, le réseau bayésien flou est une extension du réseau bayésien standard car il se combine avec la théorie des ensembles flous.

9. Exemple de réseau bayésien flou



- Nœud N remplacé par : observation floue.
- fonction floue traduisant les valeurs linguistiques.
- observation de 80mm → degré d'appartenance à N 3 et N 4 [22].

Explication de l'exemple :

D'après ce graphique, on peut constater que pour une valeur 80 mm, le degré d'appartenance à l'ensemble « N3 » vaut Moyen graphique 0,3 et le degré d'appartenance à l'ensemble « N4 » vaut beaucoup graphique 0,7.

9. Conclusion

Le début de ce chapitre présente brièvement les modèles graphiques probabilistes et en particulier les réseaux bayésiens. Très utile pour l'informatisation du domaine médical, dans le chapitre suivant nous expliquerons comment faire un diagnostic médical avec des réseaux bayésiens flous.

Chapitre 3

La conception du système

1. Introduction

De la même manière qu'il vaut mieux concevoir une maison avant de la construire, mieux vaut modéliser un système avant de le réaliser. Un modèle est une simplification de la réalité qui permet de mieux comprendre le système à développer, le diagnostic médical est un processus qui permet aux médecins de poser des diagnostics. Ces prédictions gèrent les connaissances imprécises qui doivent être établies dans un réseau bayésien. La construction du réseau bayésien se fait selon une certaine méthode.

2. Motivation de l'étude de cas

En 2019, une maladie du foie affecte un million de personnes dans le monde, principalement déclenchée par une cirrhose du foie due à l'abus d'alcool et par les virus de l'hépatite B et C. Plus largement, l'alcool est responsable de 9% des décès chez l'homme et de 3% chez la femme.

L'hépatite virale a causé 1,34 million de décès en 2015, un chiffre comparable aux décès dus à la tuberculose et au VIH. Mais à mesure que la mortalité due à la tuberculose et au VIH diminue, la mortalité due à l'hépatite augmente.

Il y a eu environ 1,75 million de nouveaux cas d'infection par le VHC en 2015, ce qui porte le nombre total de personnes vivant avec l'hépatite C dans le monde à 71 millions.

Bien que la mortalité globale due à l'hépatite augmente, le nombre de nouvelles infections par le VHB diminue en raison de l'augmentation de la couverture vaccinale contre le VHB chez les enfants. À l'échelle mondiale, 84% des enfants nés en 2015 ont reçu les 3 doses recommandées de vaccin contre l'hépatite B [23].

3. La maladie de foie

Le foie est le plus grand organe du corps humain. Il appartient au système digestif et assure de nombreuses fonctions vitales pour l'organisme. Il est situé dans la partie supérieure droite de l'abdomen: partiellement protégé par les côtes, il est séparé des poumons et du cœur par le diaphragme et est placé à droite de l'estomac.

Le foie reçoit le sang de deux vaisseaux principaux: l'artère hépatique et la veine porte.

En entrant dans le foie, ces vaisseaux se divisent pour former un réseau très dense de vaisseaux extrêmement minces.

Le sang de la veine transporte des substances de l'estomac et des intestins, de la rate et du pancréas, qui seront transformées lors de leur passage dans le foie. Le sang de l'artère hépatique fournit essentiellement l'oxygène nécessaire à cet organe. Après ses échanges avec les cellules hépatiques, le sang emprunte un réseau de veines de plus en plus important, jusqu'aux trois veines sup hépatiques qui quittent le foie et se jettent dans la veine cave inférieure.

Avec les vaisseaux sanguins, le foie est traversé par un grand nombre de voies biliaires. Ils collectent la bile et la conduisent hors du foie dans le canal hépatique commun, qui sera plus tard appelé canal cholédoque. Ce canal mène au duodénum - la partie supérieure de l'intestin - où la bile est utilisée pour la digestion.

La maladie du foie peut être liée à une inflammation aiguë ou chronique qui est secondaire à une infection virale ou bactérienne et / ou à une attaque chimique et / ou une obstruction vasculaire et / ou une obstruction biliaire et / ou un traumatisme. Les tumeurs hépatiques peuvent être liées à une tumeur qui est devenue foie ou voies biliaire ou liée à une tumeur qui s'est développée à l'extérieur du foie et envahit ensuite le foie [24].

4. Base des données sur la maladie de foie

La base de données concerne les patients qui souffraient probablement d'une maladie du foie,

On représente les informations à propos de notre BD :**Titre: (Indian Liver Patient Dataset)**
ILPD

5. Les informations sur la source

Les informations sur la source se trouvent dans le tableau suivant:

Date	21/05/2012
Source	Bendi Venkata Ramana, <u>ramana.bendi '@' gmail.com</u> , Associate Professor, Department of Information Technology, Aditya Institute of Technology and Management, Tekkali - 532201, Andhra Pradesh, India.
Créateur	Prof. M. Surendra Prasad Babu , <u>drmsprasadbabu '@' yahoo.co.in</u> , Department of Computer Science & Systems Engineering, Andhra University College of Engineering, Visakhapatnam-530 003 Andhra Pradesh, India.
Guide par	.Prof. N. B. Venkateswarlu, <u>venkat_ritch '@' yahoo.com</u> , Department of Computer Science and Engineering, Aditya Institute of Technology and Management, Tekkali - 532201, Andhra Pradesh, India.

+**Tableau 3.** Renseignement sur la source.

6. Les attributs

Nous présentons la liste des attributs qui existe dans la base de données indienne sur les maladies du foie et leurs abréviations dans le tableau suivant :

Attributs	Abréviations	Type
Age du patient	Age	Numérique
Sexe du patient	Gender	Nominale
Bilirubine totale	TB	Numérique
Bilirubine directe	DB	Numérique
Phosphotase alcaline	Alkphos	Numérique
Alamine Aminotransférase	Sgpt	Numérique
Aspartate Aminotransférase	Sgot	Numérique
Protiens totale	TP	Numérique
Albumine	ALB	Numérique
Ratio albumine et globuline	A/G	Numérique

Tableau 4. Attributs et abréviations.

- ✓ **Nombre d'instances :** 583(416 patient hépatique) et (167 patient non hépatique), contient cette base de donnée 441 male patient et 142 female patient.
- ✓ **Nombre d'attributs :** 10(9 numérique, 1 nominal).

7. Les informations sur les attributs

La description de tous les attributs de base donnés qui ont utilisé :

Age (numérique).

Gender of the patient (nominal): Male, Female.

Total bilirubin (numérique) : est un pigment jaune non soluble dans l'eau, résultant de la dégradation de l'hémoglobine. C'est le principal colorant de la bile. Il est produit dans les cellules de la rate et de la moelle osseuse, puis est transporté dans la circulation sanguine par l'albumine

pour atteindre le foie. Pour une analyse de la bilirubine, un test sanguin doit être effectué, qui consiste en un test sanguin veineux. Il est recommandé de ne pas manger ni boire au moins 4 heures avant le test sanguin. Le médecin peut également demander au patient d'arrêter de prendre certains médicaments susceptibles d'affecter les résultats du test de la bilirubine[25].

Direct Bilirubin (numérique) : la bilirubine non conjuguée est un produit de la dégradation de l'hémoglobine dans les globules rouges. Le foie transforme cette bilirubine en bilirubine directe qui peut être éliminée dans l'intestin par la vésicule. Les niveaux de bilirubine totale sont donc à la fois le reflet de la destruction des globules rouges ainsi que du bon fonctionnement du foie, de la vésicule et des voies biliaires.

Alcaline Phosphatase (numérique) : ont des enzymes qui se trouvent dans la plupart des tissus du corps, en particulier les os, le foie, l'intestin, les reins, etc. Elles sont également présentes dans le placenta pendant la grossesse [26].

Alamine Aminotransferase (numérique) : relativement spécifiques pour le foie, les ASAT étaient auparavant désignés par l'acronyme TGO (ou SGOT pour sérum-glutamyl oxaloacétate transférase); ALT sous l'acronyme TGP (ou SGPT pour sérum-glutamyl pyruvate transaminase) [27].

Aspartate Aminotransferase (numérique) : principalement présent dans le foie, les muscles, le cœur, les reins, le cerveau et le pancréas.

Total protiens (numérique):sont en quelque sorte les éléments constitutifs essentiels de nos cellules: ils jouent un rôle dans toutes les réactions du corps.

Albumine (numérique) : est la protéine la plus abondante (60%) dans le sang. Il est produit par les hépatocytes (cellules hépatiques), mais peut également provenir des aliments (on le trouve, par exemple, dans le blanc d'œuf ou dans le lait).

8. Création le Réseau Bayésien

Créer le réseau bayésien de la maladie, à travers la structure d'apprentissage puis les paramètres.

8.1. L'apprentissage de la structure

L'objectif de l'apprentissage de la structure est de trouver une structure du graphique à partir des données disponibles qui représente le mieux un problème. L'apprentissage de la structure d'un réseau bayésien à partir d'un ensemble de données devient de plus en plus difficile à mesure que la taille de l'ensemble de données d'apprentissage diminue.

8.1.1. Transformation des attributs réels vers le type nominal

Cette conversion vous permet de changer les valeurs numériques en valeurs nominales en utilisant ces normes que nous avons reçues de l'expert dans le domaine médical.

Attributs	Intervalles	Nominalization
Age	Min=0 ans max=90ans	Young [0-18]ans Adulte]18-75[ans
Total Bilirubin	Min=0mg/L Max=30mg/L	normal [0-10[mg/L high [10-30[mg/L
Direct Bilirubin	Min=0mg/L Max=10mg/L	normal [0-2[mg/L high]2-10[mg/L
Alkaline Phosphotase	Min E=0 Max E=500	Normal[0-375[High]375-500[
Alkaline Phosphotase	Min a=35 Max a=130	Low[0-35[Normal]35-120[High]120-130[
Alamine Aminotransferase	Min=5UI/L Max=45UI/L	Low]5-10[UI /L Normal[10-35[UI /L High]35-45[UI /L
Aspartate Aminotransferase	Min=10UI/L Max=80UI/L	Low]10-35[UI /L Normal]35-40[UI /L High]40-80[UI /L
Total Protiens	Min=53g/L Max=85g/L	Low]53-62[g/L Normal]62-78[g/L High]78-85[g/L

Albumin	Min=34 g/L Max=50 g/L	Low[34-36]g/L Normal[36-48] g/L High]48-50] g/L
Ratio Albumin and Globulin Ratio	Min=1.2g/L Max=1.8g/L	Low[0-1.2]g/L Normal[1.2-1.8] g/L

Tableau 5. Transformation vers le type nominal.

8.1.2. Les algorithmes pour l'apprentissage de structure

Nous allons montrer les différents algorithmes que nous avons appliqués, afin d'obtenir une structure de réseau bayésienne.

- **Naïve Bayes**

Est un type de classification bayésienne probabiliste simple basée sur le théorème de Bayes avec une forte indépendance (soi-disant naïve) des hypothèses les réseaux bayésiens naïfs sont largement utilisés pour résoudre les problèmes de classification. En effet, la classification est une appartient à chaque objet et les nœuds enfants comme étant des variables observées correspondant aux différents attributs spécifiant cet objet [28].

- **Structure de réseau bayésien après l'application de l'algorithme Naive Bayes**

L'application d'algorithmes Naive Bayes sur la base de données des maladies du foie donne la structure de réseau bayésienne suivante:

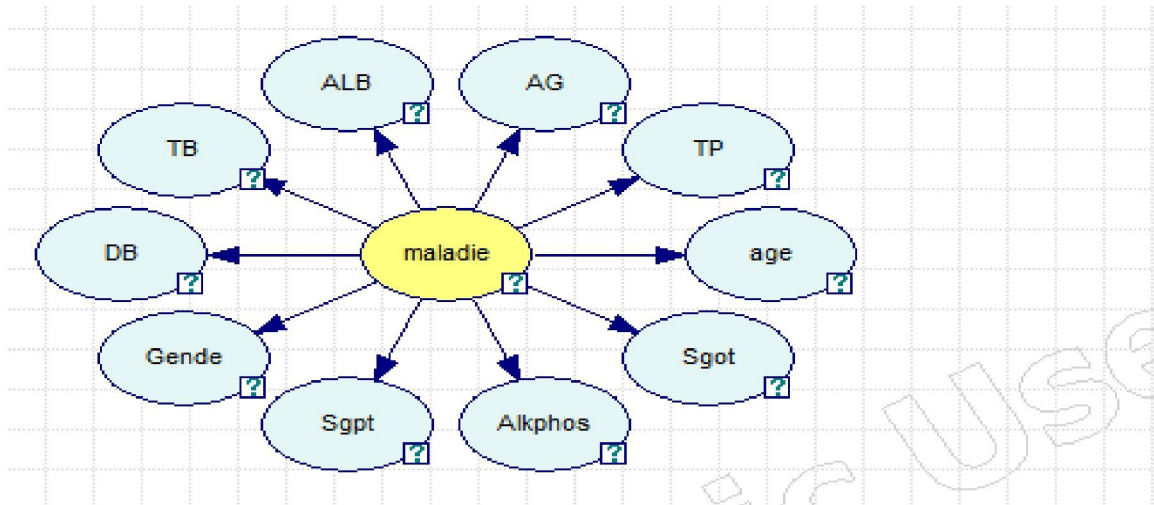


Figure 3. Structure de réseau bayésien, par l'algorithme Naïve Bayes.

- **Augmented Naive Bayes**

L'algorithme d'apprentissage de structure d'augmented Naïve Bayes (ANB) est une méthode d'apprentissage de structure semi-naïve basée sur l'approche de la recherche bayésienne.

- **Structure de réseau bayésien après l'application de l'algorithme d'augmented Naïve Bayes**

L'application d'algorithmes naïfs augmentés de Baye à la base de données sur les maladies du foie donne la structure suivante :

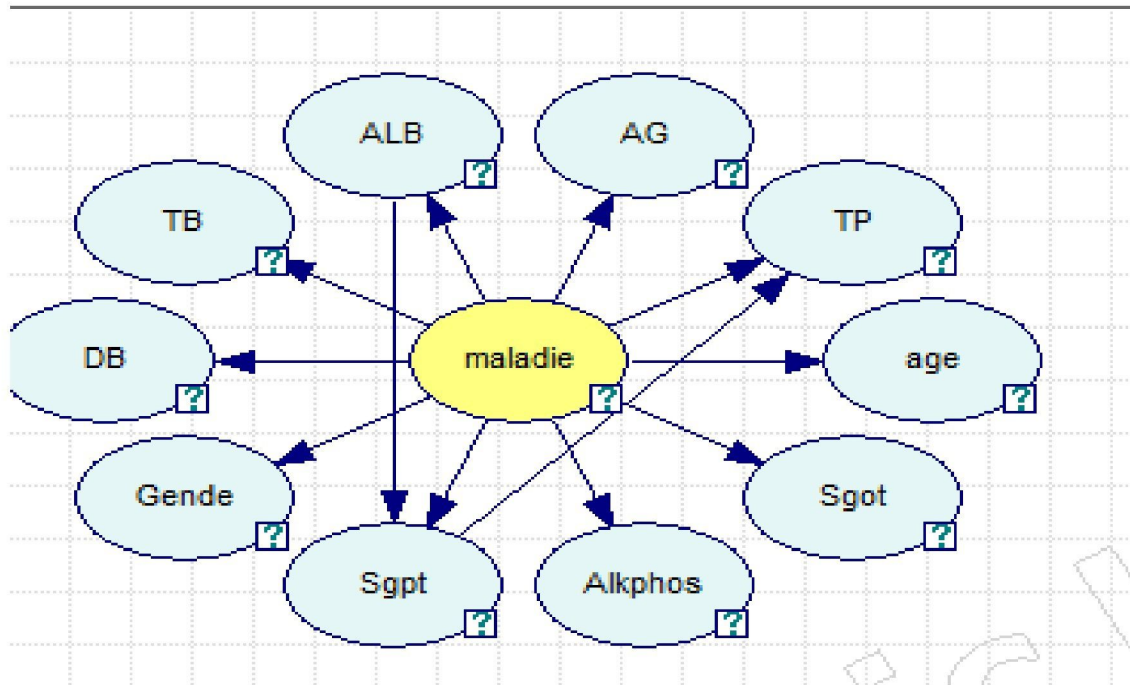


Figure 4. Structure de réseau bayésien par l'algorithme augmented Naïve Bayes.

- **Greedy Thick Thinning**

Cet algorithme d'apprentissage utilise la procédure Greedy Thick Thinning. Il s'agit d'un algorithme d'apprentissage de structure de graphique à usage général, ce qui signifie qu'il essaiera de trouver l'espace graphique complet pour le meilleur graphique.

- **Structure de réseau bayésien après l'application de l'algorithme Greedy Thick Thinning**

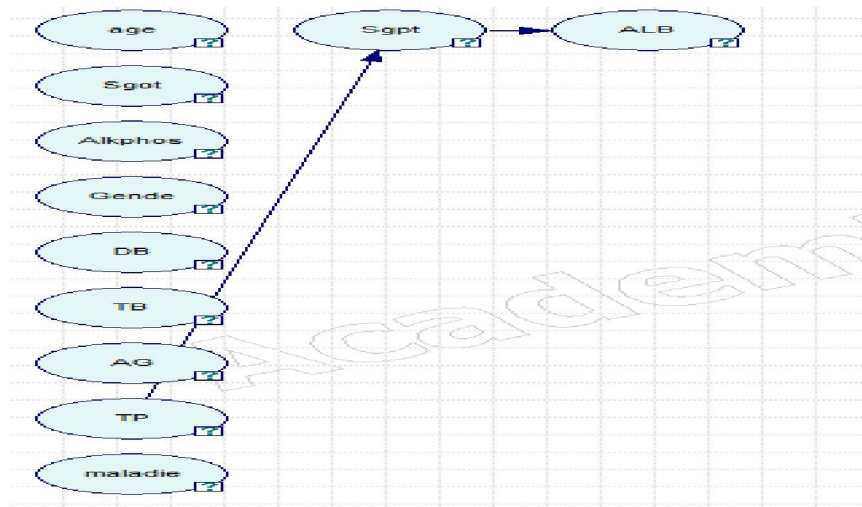


Figure 5. Structure de réseau bayésien par l'algorithme Greedy Thick Thinning.

9.1.3. Discussion sur algorithmes d'apprentissage de structure

Les structures que nous avons créées à l'aide des algorithmes d'apprentissage de structure ne sont pas les mêmes, et les relations entre les attributs ne sont pas nécessairement correctes, c'est-à-dire qu'elles ne sont pas validées par l'expert médical. Nous avons choisi l'algorithme Naïve Bayes, car il donne une structure, après consultation avec l'expert du domaine cette structure est validée, par rapport à la structure du réseau bayésien.

8.1.4. Structure finale

Nous présentons la structure après la modification de l'expert dans la figure suivante :

▶	adulte	0.47859589
	young	0.52140411

Figure 7. Paramètres d'Age après EM.

Dans le cas où les nœuds ont un parent, nous avons le nœud Total Bilirubin son parent est Direct Bilirubin.

	DB	nomal	high
▶	normal	0.48392857	0.57213115
	high	0.51607143	0.42786885

Figure 8. Paramètres de Total Bilirubin après EM.

Dans le cas où les nœuds ont deux parents, nous avons le nœud Total Protéines ses parents sont Ratio Albumin and Globulin Ratio et Albumin.

	ALB	low		normal		high	
	A/G	low	normal	low	normal	low	normal
▶	low	0.33333333	0.33333333	0.47083333	0.53603604	0.539801	0.46436782
	normal	0.33333333	0.33333333	0.0020833...	0.0022522...	0.0024875...	0.0022988...
	high	0.33333333	0.33333333	0.52708333	0.46171171	0.45771144	0.53333333

Figure 9. Paramètres de Totale protéines après EM.

Nous présentons la structure finale après l'apprentissage des paramètres dans la figure suivante :

L'apprentissage des paramètres modifiera le pourcentage d'états de noeud dans notre structure lorsque nous appliquerons l'algorithme EM.

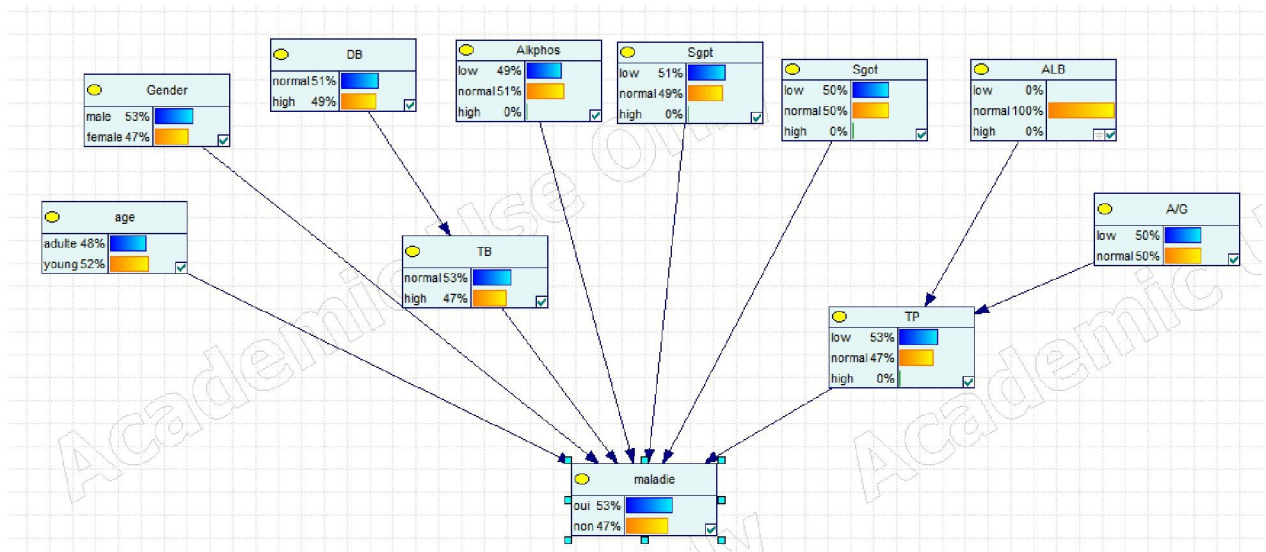


Figure 10. Structure après l'apprentissage des paramètres.

10. La fuzzification

La fuzzification est le processus de transformation d'une valeur scalaire réelle en une valeur floue. Cela se fait avec les différents types de fuzzifiers (fonctions d'appartenance).

Plusieurs variables linguistiques peuvent être utilisées pour caractériser un seul type de données

. Ici, nous choisirons trois variables linguistiques pour qualifier le degré de nœuds flous de la base de données : Low, normal et high. Sauf l'âge, ces variables linguistiques sont Young et adulte. Pour ce faire, vous devez créer une fonction d'appartenance pour chaque variable. Le tableau suivant montre les arguments et les types de chaque état qui appartient aux nœuds flous :

Nœud flou	états	Fonction d'appartenance	
		type	Arguments
Age	adulte	Trapezoid	18, 20, 80,80
		Triangular	18, 20,80
	young	Trapezoid	0, 0, 18,20
		Triangular	0, 10,20
Alkphos	Low	Trapezoid	0, 0, 30,35
		Triangular	0,17.5, 35
	normal	Trapezoid	30, 35, 120,130
		Triangular	30,60,130
	high	Trapezoid	120, 130, 150,150
		Triangular	120, 140,150
Sgpt	Low	Trapezoid	5, 5, 8,20
		Triangular	5, 10,20

	normal	Trapezoid	8, 20, 35,40
		Triangular	8, 25,40
	high	Trapezoid	35, 40, 50,50
		Triangular	35, 45,50
Sgot	Low	Trapezoid	10, 10, 30,35
		Triangular	10,17.5, 35
	normal	Trapezoid	30, 35,40 ,50
		Triangular	30, 45,50
	high	Trapezoid	40, 50, 90,90
		Triangular	40, 60,90
TP	Low	Trapezoid	53, 53, 60,62
		Triangular	53, 58,62

	normal	Trapezoid	60, 62, 70,80	
		Triangular	60, 65,80	
	high	Trapezoid	70, 80, 90,90	
		Triangular	70,78,90	
ALB	Low	Trapezoid	34,34,35,36	
		Triangular	34, 35,36	
	normal	Trapezoid	35, 36, 48,50	
		Triangular	35, 45,50	
	high	Trapezoid	48, 50, 60,60	
		Triangular	48, 55,60	
	A/G	Low	Trapezoid	0, 0, 1,1.2
			Triangular	0,0.6, 1.2
Normal		Trapezoid	1,1.2,1.8,2	
		Triangular	1,1.5,2	

TB	Normal	Trapezoid	0,0,5,10
		Triangular	0,5,10
	High	Trapezoid	5,10,35,35
		Triangular	5,20,40
DB	Normal	Trapezoid	0,0,1,2
		Triangular	0,1,2
	High	Trapezoid	1,2,20,20
		Triangular	1,10,20

Tableau 6. Fonction d'appartenance des nœuds flous.

Les fonctions d'appartenance vont être stockées dans un fichier de format « texte » qui contient les noms des nœuds, nombre d'états, nom de la fonction d'appartenance et les arguments de chaque nœud.

10. Inférence

Dans notre système, il existe deux types d'inférence, l'inférence classique et l'inférence floue

- **Inférence classique** : les valeurs que nous allons changer en type nominal ne prendront qu'un seul état, si la valeur appartient à un intervalle, (normal) par exemple alors les états (Low) et (high) deviendront négligeables.

Nous citons un exemple de transformations nominales des valeurs pour l'inférence classique, nous avons comme exemple les valeurs d'albumine (ALB) : 35, 45,49 qu'elles deviendraient dans des états nominaux que nous présenterons dans le tableau suivant:

Valeurs de Albumin	Intervalle	Etats
35 g/L	[34<35<36] g/L	Low
45 g/L	[36<45<48] g/L	Normal
49 g/L	[48<49<50] g/L	High

Tableau 7. Transformations nominale d'ALB pour l'inférence classique.

Après la transformation des valeurs des nœuds existant en type nominal, nous calculons l'inférence classique.

- **Inférence floue** : les valeurs que nous allons changer vers le nominal prendront un ou plusieurs états, si la valeur appartient à n'importe quel intervalle d'état elle peut appartenir à un autre état aussi avec certains pourcentages, et la somme des pourcentages des états égale à 100 %.

- Nous citons un exemple de transformations nominales des valeurs pour l'inférence floue, nous avons comme exemple les mêmes valeurs d'albumine (ALB) : 35, 45, 49 qui deviendraient dans des états nominaux que nous présenterons dans le tableau suivant:

Valeurs d'Albumin	Etats
35 g/L	Low=1.0 Normal=0.0 High=0.0
45 g/L	Low=0.0 Normal=0.90 High=0.10
49 g/L	Low=0.0 Normal=0.10 High=0.90

Tableau 8. Transformations nominale d'ALB pour l'inférence floue.

Après avoir transformé les valeurs des nœuds existants en valeurs nominales à l'aide des fonctions À partir de l'appartenance, nous calculons l'inférence floue avec Virtual évidence.

11. Conclusion

Dans ce chapitre nous avons présenté les processus de construction du réseau bayésien flou afin d'établir les prédictions et l'aide au diagnostic grâce à l'inférence floue. Dans le chapitre suivant, nous allons passer à la construction de notre système d'aide au diagnostic pour aider les médecins à diagnostiquer une maladie du foie

Chapitre 4

Implémentation

et test du

systeme

1. Introduction

Dans ce chapitre, nous présentons la mise en œuvre de notre système d'aide au diagnostic médical basé sur la conception présentée dans le chapitre précédent. Nous présentons ainsi les langages et outils utilisés pour la mise en œuvre du système et nous donnons un aperçu des résultats obtenus à l'aide de sérigraphies.

2. Outils de développement de notre système

2.1. Java

Java est un langage de programmation simple et orienté objet créé par James Gosling et Patrick Naughton, officiellement présenté avec les intégrations de 1995 de la JVM dans Netscape. En 2006, Java devient Open Source.

- Le compilateur Java génère un bytecode, c'est-à-dire un format intermédiaire, L'objectif central de Java est que les logiciels écrits dans ce langage doivent être facilement sur différents systèmes d'exploitation tels que Unix, Windows, Mac [29]. Il est fourni avec un ensemble d'outils (le JDK Java Development Kit) et un ensemble de packages: ensemble de classes. Ses différentes classes de base couvrent de nombreux domaines (entrée / sortie, interface graphique, réseau, etc.) Cette richesse de "bibliothèques standard" explique sûrement en partie le succès de Java. La langue elle-même se trouve dans le package java. Lang

2.2. NetBeans

NetBeans est un environnement de développement intégré (IDE) pour Java, placé en open source par Sun en juin 2000 sous licence CDDL (Common Development and Distribution License). En plus de Java, NetBeans prend également en charge divers autres langages, tels que Python, C, C ++, XML et HTML. Il intègre toutes les caractéristiques d'un IDE moderne (éditeur de couleurs, projets multi-langues, refactoring, éditeur graphique d'interfaces et de pages web). NetBeans est disponible sur Windows, Linux, Solaris (sur x86 et SPARC), Mac OS X et Open VMS. NetBeans se développe en Java [30].

2.4.JSMILE+

Jsmile est compatible avec jdk 7 et versions ultérieures.

Jsmile contient deux parties :

Le fichier jar et la bibliothèque native, ces derniers sont spécifiques au système d'exploitation. Le fichier jar, nommé jsmile-xyz jar, où xyz est le numéro de version, peut être utilisé comme n'importe quel pot java par exemple, ajoutez-le à votre référentiel Maven local, la bibliothèque native est utilisée par le code contenu dans le fichier jar via JNI (java interface native). le nom du fichier de bibliothèque natif dépend de la plateforme:

jsmile.dll sous windows

jsmile So sous Linux.

2.5.GeNIe

Le génie modélisé est un environnement de développement permettant la création de modèles décisionnels théoriques. Il a été créé et développé par le Decision Systems Laboratory, Université de Pittsburgh entre 1995 et 2015. Son nom et son caractère inhabituel la capitalisation vient du nom interface de réseau graphique, donné à l'interface simple et originale avec SMILE, notre bibliothèque de classe pour probabiliste graphiques et modèles de décision. GeNIe a été écrit pour les systèmes d'exploitation Windows. Bien que nous ne puissions garantir une compatibilité à 100%, nous vérifions avec chaque version que cela fonctionne sous macos (anciennement OS X) sous wine Genie sur Linux en utilisant wine.Genie nous permet de construire des modèles de toutes tailles et de complexité limitées uniquement par la capacité de votre ordinateur mémoire de fonctionnement.

Génie est un environnement de modélisation. Les modèles développés avec Genie peuvent être intégrés dans toutes les applications et sur n'importe quelle plateforme informatique, en utilisant SMILE, qui est entièrement [31].

3. Les interfaces

3.1 L'interface d'accueil

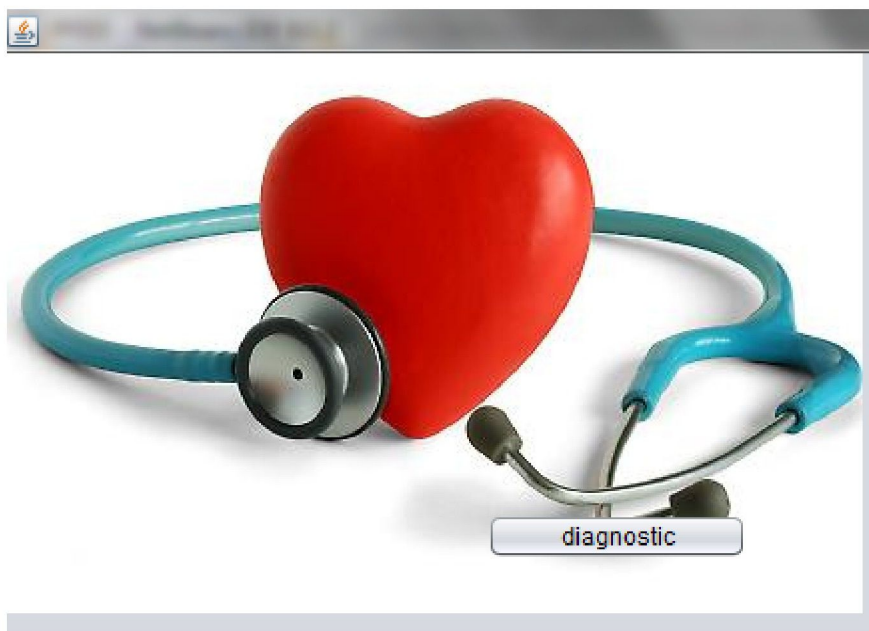


Figure 11 L'interface d'accueil

3.2. L'interface du réseau bayésien classique

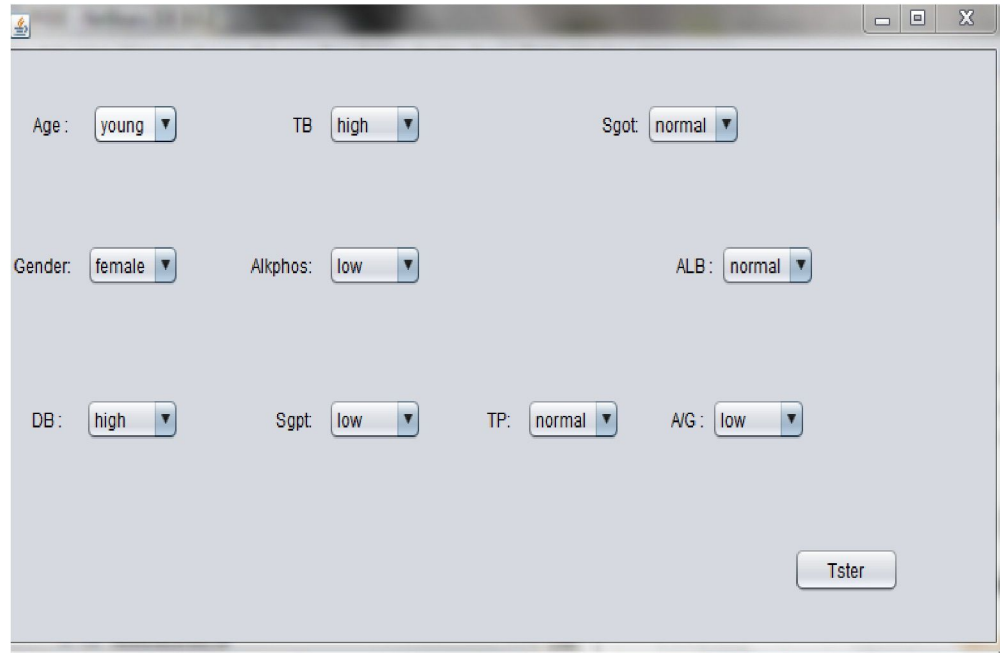


Figure 12. Interface du réseau bayésien classique

Tster : ce bouton nous permet de faire l'inférence classique de réseau bayésien classique.

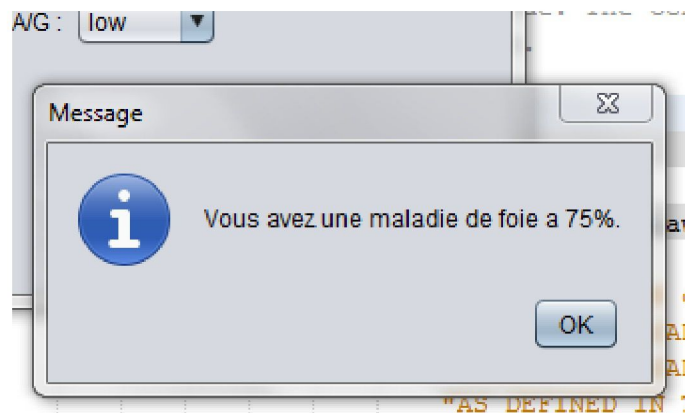
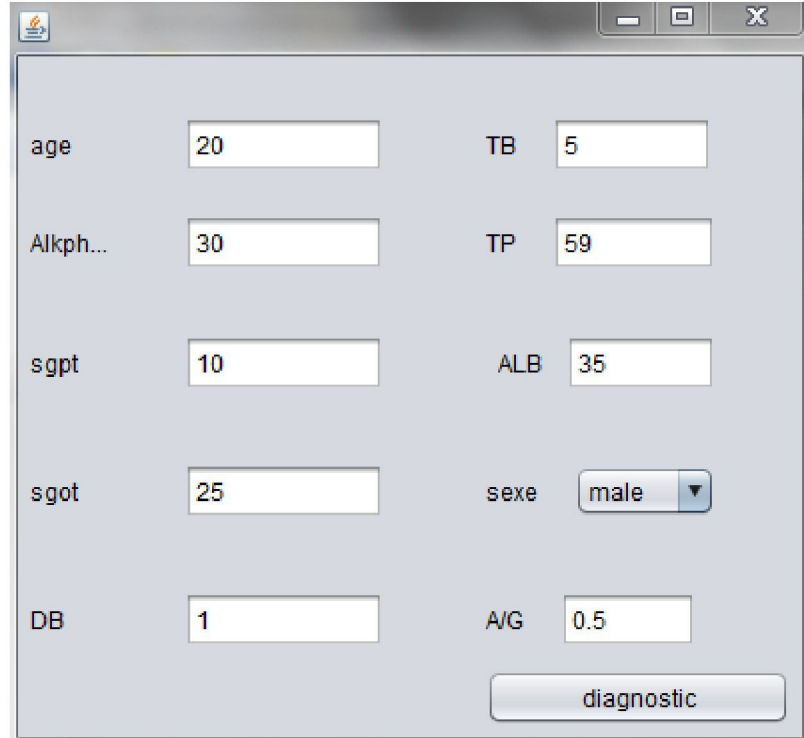


Figure 13. Inférence du réseau bayésien classique.

3.3 Interface de réseau bayésien flou



The screenshot shows a software window with a light gray background. It contains several input fields arranged in two columns. The left column has fields for 'age' (value 20), 'Alkph...' (value 30), 'sgpt' (value 10), 'sgot' (value 25), and 'DB' (value 1). The right column has fields for 'TB' (value 5), 'TP' (value 59), 'ALB' (value 35), 'sexe' (a dropdown menu showing 'male'), and 'A/G' (value 0.5). At the bottom right, there is a button labeled 'diagnostic'.

Figure 14. Interface du réseau bayésien flou.

Diagnostic : ce bouton nous permet de faire l'inférence floue du réseau bayésien flou.

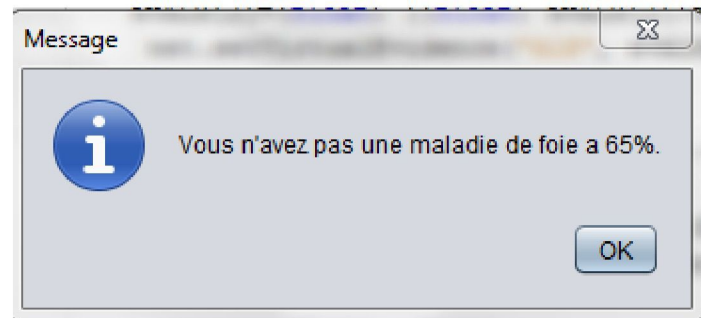


Figure 15. Interface d'inférence floue de réseau bayésien flou.

4. Test du systeme

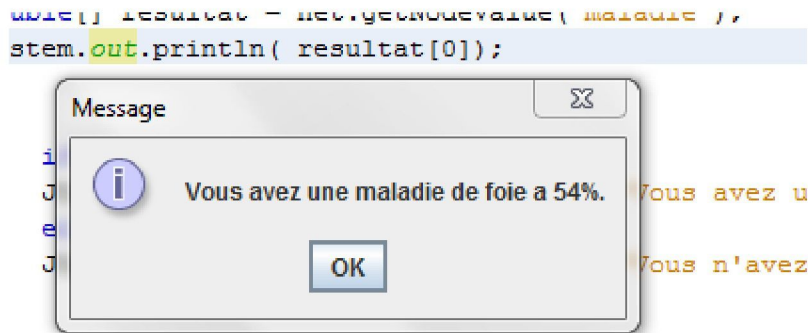


Figure 16: Résultat de test

5. La validation croisée

Est une méthode d'estimation de la fiabilité d'un modèle basée sur une technique d'échantillonnage. En fait, il existe au moins trois techniques de validation croisée: «tests et validation» ou «méthode d'exclusion», «validation croisée k-fold» et «validation croisée avec un seul retrait». Nous avons utilisé la méthode du k-fold, cette méthode nous divisons l'échantillon k fois, puis nous sélectionnons l'un des k échantillons comme ensemble de validation.

Dans notre cas $K = [2, \dots, 10]$ et les précisions du système sont calculées selon les étapes: à chaque itération, le modèle est formé sur $k - 1$ partie, et testé sur la partie restante obtient à la fin k détails, la précision du modèle est égale à leur moyenne.

5.1. Résultat de la validation croisée

Le calcul des performances a donné les résultats suivants :

Valeurs K(folds)	Performances
2	66%
3	65%
4	65 %
5	65%
6	65%
7	67%
8	67%

Tableau 9 Performance du système

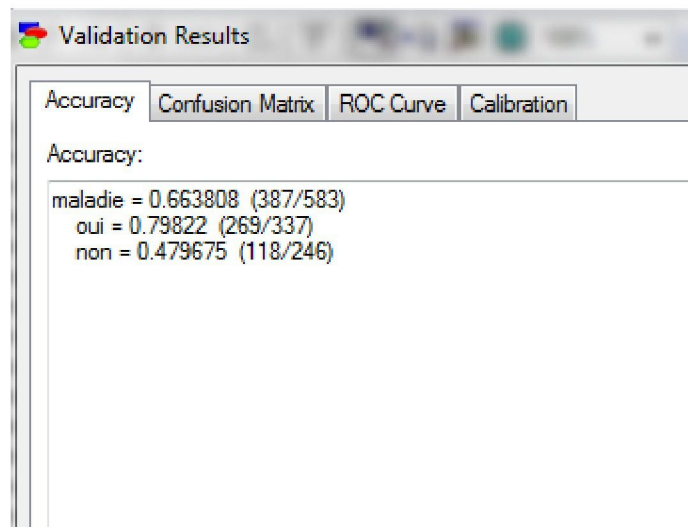


Figure 18 Exemple d'un résultat de validation (K=2)

Dans l'état "oui" où le patient est affecté par une maladie du foie, 269 résultats obtenus à partir d'une validation croisée similaire avec les patients de l'ensemble de données qui est égal à 337 "états = oui", ce qui nous donne un pourcentage de 80%.

Dans l'état "non" où le patient est saint 118 résultats obtenus à partir d'une validation croisée similaire aux états patient de l'ensemble de données qui équivaut à 246 "états = non", ce qui nous donne un pourcentage de 48%.

La somme de ces deux états nous donne 387 résultats obtenus similaires avec les patients de l'ensemble de données qui est égal à 583 dans les deux cas où les patients sont affectés (oui) ou (non), avec un pourcentage de 66%.

6. Conclusion

Nous avons présenté dans ce chapitre l'implémentation, le test et la validation de notre système en décrivant les outils et langages utilisés pour l'implémentation des différents composants du système. Nous avons principalement utilisé le langage Java et l'environnement de développement NetBeans pour le développement de notre solution. Enfin, nous avons présenté certaines des principales fenêtres de notre application.

Conclusion

Générale

Conclusion générale

Dans ce mémoire, nous avons présenté un système d'aide au diagnostic médical utilisant des réseaux bayésiens flous, il est prouvé que les réseaux bayésiens sont une approche utile et fiable pour traiter et modéliser l'incertitude. Choisir le modèle pour aider le médecin à prendre une décision efficace devient un défi, pour cela nous avons fait le meilleur entre deux approches différentes:

Modèle probabiliste « réseau bayésien standard »: car l'activité médicale accorde une importance considérable au facteur de doute, en raison de l'indisponibilité d'informations utiles et certaines au moment de la décision.

Modèle flou « réseau bayésien flou » : joue un rôle clé dans la représentation de l'imprécision. Nous avons proposé un processus pour cette modélisation qui comprend des étapes:

Premièrement: création du réseau bayésien à partir de structures d'apprentissage et de paramètres.

Deuxièmement, l'application des algorithmes EM une approche pour l'apprentissage des données manquantes (nœuds) avec un algorithme (EM) basé sur les réseaux bayésiens.

Enfin, Diagnostic: calculé le résultat avec virtuelle évidence.

Afin de réaliser notre conception, nous avons implémenté notre système, implémenté le réseau hépatique bayésien et calculé l'inférence du réseau bayésien classique et flou, ainsi, nous avons montré les différentes interfaces de notre système. La validation des résultats de notre système à l'aide de la validation croisée donne la valeur de la précision.

Références

[1] : Batifoulier, P. (1994). Incertitude sanitaire et risque probabilisable. Journal d'économie médicale, 2(1), 3-17.

[2] : Bottéro, J. (1974). *Symptômes, signes, écritures en Mésopotamie ancienne* (pp. 70-197).

Lien :<https://www.worldcat.org/title/symptomes-signes-ecritures-en-mesopotamie-ancienne-extrait-de-divination-et-rationalite/oclc/492948808?referer=di&ht=edition>

[3]: Rinaldy, A. (2018). «Christian Baudelot, Jean-Loïc Le Quellec, Valérie Louys, Isabelle Richer, Catherine Robert et al., L'Anthropologie pour tous», Les comptes rendus, 2018, mis en ligne le 19 février 2018. URL: <http://journals.openedition.org/lectures/24247>.

Lien:<http://journals.openedition.org/lectures/24247>.

[4] : Boule, M., & Vallois, H. (1932). L'homme fossile d'Asselar (Sahara). Arch. de l'Inst. de Paleont. Hum. Mem, 9, 90.

Lien :<https://www.futura-sciences.com/sante/definitions/medecine-symptome-834/>

[5] : James-Raoul, D. (2005). La poétique de la lettrine dans Le Roman de Silence. Cahiers de recherches médiévales et humanistes. Journal of medieval and humanistic studies, (12), 227-245.

[6] : Cleyet-Merle, G. (2017). *Rôle et intervention du pharmacien d'officine auprès de la personne âgée présentant des troubles locomoteurs nécessitant des dispositifs médicaux et du matériel d'aide à la vie: étude à travers des cas cliniques* (Doctoral dissertation, Université de Lorraine).

Lien:<https://www.revmed.ch/RMS/2012/RMS-349/Diagnostic-et-traitement-de-la-rupture-du-tendon-d-Achille>

[7] : François, 2009, étude du raisonnement du clinicien expérimenté et de l'étudiant : Apports de la psychologie cognitive, l'Université Paul-Sabatier de Toulouse.

[8]: Ravenstein, E. G. (1889). Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at. Journal of the royal statistical society, 52(2), 241-305.

[9]: Pearl, J. (2014). Probabilistic reasoning in intelligent systems: networks of plausible inference. Elsevier.

Lien :<https://www.elsevier.com/books/probabilistic-reasoning-in-intelligent-systems/pearl/978-0-08-051489-5>

- [10] : Jean-Jacques Boreux, Éric Parent et Jacques Bernier, *Pratique du calcul bayésien*, Springer, 2009 (ISBN 9782287996665, lire en ligne [archive]), p. 38-39
- [11] : Mrad, A. B., Maalej, M. A., Delcroix, V., Piechowiak, S., & Abid, M. *Intégration de l'Evidence Floue dans les Réseaux Bayésiens*.
- [12] : Naïm, P., Wuillemin, P. H., & Leray, P. (2008). Olivier pourret. *Anna becker, Réseaux bayésiens, Eyrolles, 3ème édition, Paris*.
- [13]: Friedman, (1998), The Bayesian structural EM algorithm, In Proceedings of The Fourteenth conference on Uncertainty in artificial intelligence (pp. 129-138), Morgan Kaufmann Publishers Inc.
- [14] : Frédéric Santos, (2015), L'algorithme EM : une courte présentation.
- [15] : Patrick Naïm, Pierre-Henri Wuillemin, Philippe Leray, Olivier Pourret et Anna Becker. (2011). *Réseaux bayésiens*, Eyrolles, 3 è éd
- [16]: Steffen L. Lauritzen and David J. Spiegelhalter. (1988). « Local Computations with Probabilities on Graphical Structures and Their Application to Expert Systems ». *Journal of the Royal Statistical Society, Series B*, 50 :157–224 1988.
- [17] :Dempster et al. (1977).
- [18]: Pearl, 1988.
- [19]: Zadeh, L. A. (1965). Fuzzy sets. *Information and control*, 8(3), 338-353.
- [20]:P. Naïm, P-H. Wuillemin, P.Leray, O. Pourret, and A. Becker. (2007). *Réseaux bayésiens*, Eyrolles, Paris.
- [21] : Ali BEN MRAD, 2012, *Les évidences incertaines dans les Réseaux Bayésiens*, Kerkennah, Tunisie.
- [22] :Vianna, V. N. (2019). *Proposta d'implantação do centro de testagem e aconselhamento no serviço universitário da UFOP*.

- [23]: Guennec-Coppens, F. L. (2012). Informations. *Journal des africanistes*, 82(1-2), 397-405.
- [24] : Balcou-Debussche, M. (2012). Construire l'éducation à la santé ensemble L'exemple du «passeport santé».
- [25] : Roth, M. (1967). Dosage fluorimétrique de la bilirubine. *Clinica Chimica Acta*, 17(3), 487-492.
- [26]: Lemire, M., Paré, G., Sicotte, C., & Harvey, C. (2008), Determinants of Internet use as a preferred source of information on personal health. *International journal of medical informatics*, 77(11), 723-734.
- [27]: Cerquides J., R.L.Mantaras. (2003). Tractable bayesian learning of tree augmented naïve bayes classifiers.
- [28] :Hugues Bersini, Introduction à Java, Université Libre de Bruxelles.
- [29] : belgacem hichem,bouazza Mohammed Reda, 2014, gestion d'une agence de voyage selon le modèle client/serveur, Université Abou Bakr Belkaid–Tlemcen.
- Lien** : <https://studylibfr.com/doc/7876547/gestion-d-une-agence.pdf>.
- [30]: Manual,18/7/2019,Structural Modeling, Inference, and Learning Engine Programmer's.
- [31]: Embarki miloud,2012, Etude comparative entre réseaux bayésien et autre méthodes de classification appliqué sur une base cardiologue,Université Abou Bakr Belkaid– Tlemcen.