

**Université Saad Dahlab Blida**

N° D'ordre : .....



Faculté des sciences

**Département d'informatique**

Mémoire Présenté par :

BEN MAMMAR Azzdine

CHERKELAINE Ali

**En vue d'obtenir le diplôme de master**

Domaine : mathématique et informatique

Filière : Informatique

Spécialité : Informatique

Option : Système informatique et réseaux

**Thème : Prédiction du diagnostic médical guidée par les réseaux bayésiens  
flous.**

***Soutenu le :***

M : Président  
M : Examineur  
M : Examineur  
M : Fareh M Promotrice

**Promotion**

2017/2018

*Dédicaces*

*A :*

*Nos chers parents*

*Nos frères et sœurs*

*Toute Notre famille*

*Tous nos amis*

*On dédie ce mémoire*

*BEN MAMMAR AZZEDINE*

*CHERKELAINE ALI*

*Remerciement*

*Nous tenons tout d'abord à remercier Dieu le tout puissant et miséricordieux, qui nous a donné la force et la patience d'accomplir ce modeste travail.*

*En second lieu, Nous tenons à remercier notre promotrice Mme : **FAREH.M** enseignante à l'université USDB, pour son aide et de nous avoir soutenues durant toute la période du travail.*

*Nous remercions le doctorant Mr **RIALI Ishak** pour nous avoir aidés et suivis durant la réalisation de ce projet*

*Nos vifs remerciements vont également aux membres du jury pour l'intérêt qu'ils ont porté à notre travail en acceptant d'examiner et de l'enrichir par leurs propositions.*

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

**BEN MAMMARE AZZEDINE**

**CHERKELAINE ALI**

**ملخص:**

يعتمد المجال الطبي على التشخيصات غير مؤكدة أو المعرفة الغير مؤكدة, و على الرغم من النجاح الذي لا جدال فيه و الذي حققته الشبكة البايزيانية الكلاسيكية, حيث يتم تطبيقها على نطاق واسع للتمثيل و الاستدلال مع المعارف, و

مع ذلك يكمن العيب الرئيسي فيها أنها لايمكنها التمثيل و الإستدلال مع المعرفة الغير مؤكدة,و من ناحية أخرى الشبكة البايزيانية الغامضة هي الأمتل في التعامل مع المعرفة الغير مؤكدة .  
-الهدف من هذا العمل هو تصميم نظام للتنبؤ و تشخيص الطبي مبني على معرفة غير مؤكدة بإستخدام الشبكة البايزيانية الغامضة .

**كلمات مفتاحية:** الشبكات البايزيانية, الشبكات البايزيانية الغامضة.

## **Résumé**

Le domaine médical est basé sur les diagnostics où les connaissances imprécises, malgré le succès indubitable des réseaux bayésiens classiques sont largement utilisées pour modéliser et faire des raisonnements avec les connaissances. Cependant, ils ne fournissent pas un support adéquat pour traiter la connaissance imprécise. Les réseaux bayésiens flous sont plus optimaux dans la représentation de l'imprécision.

L'objectif de ce travail est de concevoir un système de prédiction du diagnostic médical basé sur des connaissances imprécise en utilisant les réseaux bayésiens flous qui passe par l'apprentissage des structures et des paramètres jusqu'à la validation.

**Mots clés :** réseaux bayésiens, réseaux bayésiens flous, l'imprécision.

## **Abstract:**

The medical field is based on diagnostics where imprecise knowledge, despite the indisputable success of traditional Bayesian networks, is widely used to model and reason with knowledge. However, they do not provide adequate support to deal with imprecise knowledge. Fuzzy Bayesian networks are more optimal in the representation of imprecision.

The objective of this work is to design a system for predicting medical diagnosis based on imprecise knowledge using fuzzy Bayesian networks passed through the learning of structures and parameters until validation.

**Key words:** Bayesian networks, fuzzy bayesiens network, imprecision.

## *Sommaire*

Introduction générale	1
1. Contexte	1
2. Problématique	1
3. Objectif	2
4. Organisation du mémoire	2
Chapitre 1 : Les réseaux bayésiens	
1. Introduction	5
2. Réseaux bayésiens	5
2.2. Définition des réseaux bayésiens	6
2.3. Un exemple simple de réseau bayésien	7
2.4. L'inférence dans les réseaux bayésiens	12
2.4.1. Inférence exacte	13
2.4.2. Inférence approchée	13
2.5. La mise en œuvre du réseau Bayésien	8
2.6. Apprentissage dans les réseaux bayésien	9
2.6.1. Apprentissage de structure	9
a. Algorithmes de structure (Données complètes)	9
1) Algorithmes basés sur un score	10
2) Algorithmes basés sur l'indépendance conditionnelle	10
b. Algorithmes de structure (Données incomplètes)	10
2.6.2. Apprentissage des paramètres	10
a. A partir de données complètes	11
b. A partir de données incomplètes	11
3. Les avantages des réseaux bayesiens	13

4. Quelques domaines d'application	14
5. La théorie des ensembles flous	14
6. Réseaux bayésiens flous	16
6.1. Définition	16
6.2. Propagation de l'observation de vraisemblance : méthode de Pearl (virtual evidence)	17
6.2.3. Exemple	17
7. Conclusion	18
Chapitre 2 : Le diagnostic médical	
1. Introduction	20
2. Définitions	20
2.1. Le signe et le symptôme	20
2.1.1. Le signe positif	21
2.1.2. Le signe négatif	21
3. La maladie et le syndrome	21
4. Le diagnostic médical	22
4.1. Le raisonnement déductif	23
4.2. Le raisonnement inductif	24
4.3. Le raisonnement abductif	24
4.4. Le raisonnement par analogie	25
5. Les incertitudes médicales	25
6. Les réseaux bayésiens et raisonnement abductif	27
7. La discrétisation des variables continues	28
8. Conclusion	28
Chapitre 3 : La conception du système	29
1. Introduction	30
2. Motivation de l'étude de cas	30

3. Maladie de L'insuffisance rénale	30
4. Base des données sur l'insuffisance rénale	31
a. Les informations sur la source	31
b. Les attributs	31
c. Informations des attributs	33
d. Information générale sur la base des donnés	35
5. Architecture globale de notre système	36
6. créer le RB	38
6.1. L'apprentissage de la structure	38
6.1.1. Transformation d'une base de donnée vers nominale	38
6.1.2. Algorithmes de structure	39
a. Naive Bayes	39
a .1. Naive Bayes structure	40
b. Augmented Naive Bayes	40
b.1. Augmented Naive Bayes structure	41
c. Tree Augmented Naive Bayes	42
c.1. Tree Augmented Naive Bayes structure	42
d. PC	43
d.1. PC structure	44
e. Bayesian Search	45
e.1. Bayesian Search structure	46
f. Greedy Thick Thinning ( notre choix)	46
f.1. Algorithme Greedy Thick Thinning	47
f.2. Greedy Thick Thinning structure	47
6.1.2.1. Discussion sur algorithmes d'apprentissage des structures	48
6.2. L'apprentissage des paramètres	49
6.2. 1. L'algorithme EM	49
7. La fuzzification	52
8. Inférence	55
9. Conclusion	56

. Chapitre4 : Implémentationdu système	
1. Introduction	58
2. Outils de développement de notre système	58
3. La mise en œuvre du système	60
4. Les interfaces	60
4.1. L’interface d’accueil	60
4.2. Interface de l’ordre de causalité	61
4.3. Interface pour choisi des nœuds flous	62
4.4. Interface de création d’un structure	63
4.5. Interface de structure réseau bayésien	64
4.5. Interface pour choisi un réseau bayésien pour la prédiction	68
4.6. Interface de test	69
5. conclusion	71
Chapitre 5 : Test du système	
1. Introduction	73
2. Teste de système	73
2.1. Teste de RB classique	73
2.2. Teste de RB flou	76
3. La validation croisée	79
3.1. Résulta de validation croisée	79
3.2. Discussions de résulta :	79
4. Conclusion	80
Conclusion générale	<b>Erreur ! Signet non défini.</b>

## La liste des figures

Figure 1: Les étapes de construction d'un réseau bayésien .....	8
Figure 2: L'algorithme EM pour le maximum de vraisemblance .....	12
Figure 3: Exemple d'ensemble classique: cote de crédit .....	15
Figure 4: Exemple d'ensemble flou: cote de crédit .....	15
Figure 5: observation de vraisemblance sur D .....	17
Figure 6: schéma globale.....	37
Figure 7: Naive Bayes .....	40
Figure 8: Naive Bayes structure .....	40
Figure 9: Augmented Naive Bayes structure.....	41
Figure 10: Tree Augmented Naive Bayes .....	42
Figure 11: Tree Augmented Naive Bayes structure .....	43
Figure 12: PC.....	44
Figure 13: PC structure.....	45
Figure 14: Bayesian Search structure .....	46
Figure 15: Greedy Thick Thinning structure .....	47
Figure 16: Structure après la modification de l'expert.....	48
Figure 17: Etapes de l'exécution de l'algorithme EM.....	50
Figure 18: Paramètres de White Blood cell count après EM .....	50
Figure 19: Paramètres de hemoglobine après EM .....	51
Figure 20: Paramètres de hypertension après EM.....	51
Figure 21: Structure après l'apprentissage des paramètres .....	51
Figure 22: Interface d'accueil.....	60
Figure 23: Importer une base de donnée .....	61
Figure 24: Interface de l'ordre de causalité.....	62
Figure 25: Interface pour sélectionner des noeuds flous .....	63
Figure 26: Interface de confirmation et création d'une structure .....	63
Figure 27: Message de rappel.....	64
Figure 29: Interface pour l'apprentissage des paramètres.....	65
Figure 28: Structure du réseau bayésien.....	65
Figure 30: Message de confirmation d'apprentissage de paramètre .....	66
Figure 31: Interface de calculs d'inférence.....	66
Figure 32: Fichier Rules.texte triangulaire.....	67
Figure 33: Fichier Rules.texte trapèze.....	67
Figure 34: Message de confirmation de création d'un réseau .....	68

Figure 35: Interface pour importer un réseau bayésien .....	68
Figure 36: Interface de choisi l'état des nœuds .....	69
Figure 37: Message de repens dans le cas de la maladie égala « true » .....	70
Figure 38: Message de repens dans le cas de la maladie égala « false » .....	70
Figure 39: Importer le fichier « RBC EN.xdsl » .....	73
Figure 40: Sélectionner des états .....	74
Figure 41: Message de résultat «maladie =false» dans RB classique .....	75
Figure 42: Message de résultat «maladie =true » dans RB classique.....	75
Figure 43: Importer le fichier « RBF EN.xdsl ».....	76
Figure 44: Sélectionner et donner les valeurs des états.....	76
Figure 45: Message de résultat «maladie =true » dans RB flou .....	77
Figure 46: Message de résultat «maladie =false » dans RB flou .....	78

## La liste des tableaux

Tableau 1: La modélisation de la table de probabilité.....	7
Tableau 2: Tableau de probabilité conditionnelle .....	7
Tableau 3: Renseignement sur la source .....	31
Tableau 4: Attributs et abréviation .....	32
Tableau 5: Transformation vers nominale.....	39
Tableau 6: Exemple des données incomplètes. ....	49
Tableau 7: Fonction d'appartenance des nœuds flous .....	53
Tableau 8: Fonction d'appartenance trapezoid de Blood Pressure .....	54
Tableau 9: Fonction d'appartenance triangular de Blood Pressure .....	54
Tableau 10: Transformations nominale de rbcc pour l'inférence classique .....	55
Tableau 11: Transformations nominale de rbcc pour l'inférence flou.....	56
Tableau 12: Précision de système dans RB et RB flou .....	79

### Introduction générale

#### 1. Contexte

Dans le domaine médical, les médecins sont régulièrement amenés à prendre plusieurs décisions, comme la prévention des pathologies, la prescription d'un examen, etc...

Certaines décisions sont prises facilement, lorsque le diagnostic est facile, le traitement choisi est efficace et les risques sont nuls, alors que dans d'autres cas, la bonne décision à prendre n'est pas évidente, par exemple lorsque le médecin se trouve dans une situation où le diagnostic est particulièrement douteux, parce que, généralement le domaine de la santé donne une importance considérable au facteur d'imprécision. Cette dernière tient à plusieurs raisons : certaines connaissances sont d'ordre statistique et sont associées par nature à un risque d'erreur, mais d'autres connaissances sont incomplètes, par défaut d'exploration, la complexité et l'imprécision liées aux maladies contribuent à de grandes variations dans la pratique clinique à des erreurs et à un non respect des guidelines.

Plusieurs approches existent dans la littérature médicale pour aider le médecin à prendre des meilleures décisions, parmi ces méthodes existe les réseaux bayésiens flous. Un RBF est une extension de réseau bayésien classique, permettant de modéliser des connaissances imprécises et floues simultanément. Il permet également de faire des raisonnements probabilistes basé sur des observations imprécise et/ou vague.

La maladie laquelle on va parler est difficile à diagnostiquer, c'est pourquoi il est complexe de mettre en évidence les causes de son incertitude complexe.

L'hypertension artérielle ou le diabète sont des épidémies silencieuses et progressent d'une façon alarmante parce qu'elles touchent des millions d'êtres humains autour du globe. Ces derniers sont souvent la conséquence de l'insuffisance rénale chronique dite «terminale» qui se développe sans signe d'alerte. Pour les cas les plus graves, les seules solutions sont la dialyse ou la greffe d'organe.

Les maladies rénales touchent près de 753 millions de personnes dans le monde, l'insuffisance rénale chronique est un problème de santé publique au niveau mondial. Dans les pays développés, cette affection survient essentiellement chez les sujets âgés, mais en Afrique, elle s'installe plutôt chez les sujets jeunes actifs.

#### 2. Problématique

L'incertitude 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'incertitude comprennent la variabilité biologique des patients, les préjugés des patients et des médecins, l'erreur dans l'interprétation des tests, les valeurs et opinions divergentes des patients et des médecins et l'incertitude entourant la prise de décision qui dut souvent à la complexité du diagnostic, vu la nature incertaine de la connaissances médicale, ce qui contribue à de grandes variations dans la pratique clinique, à des erreurs cliniques et à un mauvais respect des guidelines.

Cette incertitude peut se manifester par exemple, lors de la probabilité qu'un patient est atteint par l'insuffisance rénale, cela conduit à poser les questions suivantes

1. Comment modéliser ces connaissances incertaines dans le domaine médical ?
2. Comment construire un réseau bayésien flou pour traiter ces connaissances incertaines qui servent à inférer le diagnostic médical ?
3. La mise en œuvre du système de prédiction commence par l'identification d'un ensemble des signes pertinents pour produire un diagnostic basé sur ces derniers, mais comment faire ce système et quels sont les outils nécessaires ?

### 3. Objectif

Malgré l'omniprésence de l'incertitude dans le domaine médical, les médecins font des jugements et des diagnostics dans ces circonstances basées sur leur savoir et savoir-faire.

Dans ce contexte, l'objectif de ce travail est de mettre en œuvre un système d'aide à la décision pour le diagnostic médical basé sur des observations sur les patients (biologique, tests....etc.). C'est pourquoi nous avons choisi d'utiliser les réseaux bayésiens flous (RBF) qui sont des techniques très performantes en matière d'aide à la décision et qui sont issues d'une combinaison entre la théorie bayésienne des probabilités conditionnelles, la théorie des graphes et des ensembles flous.

Ces techniques ont prouvés leurs efficacités dans plusieurs domaines, tels que le domaine d'aide à la décision et le domaine de modélisation de systèmes complexes.

L'objectif est de concevoir et d'implémenter un système d'aide au diagnostic médical guidé par les RBF.

### 4. Organisation du mémoire

**Chapitre 1- Les réseaux bayésiens :** Ce chapitre a pour objectif de donner une vision sur les réseaux bayésiens standards et flous. Tout au long de ce chapitre, nous éclaircissons la notion des réseaux bayésiens standards et flous et les différentes méthodes d'apprentissage et d'inférence.

**Chapitre 2 - Le diagnostic médical :** Ce chapitre contient les définitions des signes, des symptômes, de diagnostic médical et l'incertitude médicale. Ce chapitre pour objectif de donner une vision sur les différents types de raisonnements.

**Chapitre 3 - La conception du système :** Nous allons voir dans ce chapitre les processus de la construction des réseaux bayésiens flous. Cela sera introduit à travers un diagramme qui représente le système global de la modélisation et qui sera détaillé par la suite dans ce chapitre. Il constitue un point de départ à l'implémentation.

**Chapitre 4 - Implémentation du système:** Ce chapitre est le résultat de la conception pour réaliser le système d'aide à la décision. La fin de ce chapitre constitue un point de départ pour effectuer un test sur notre application.

**Chapitre 5 - Test du système:** expose un cas d'étude pour bien éclaircir notre approche. Nous prenons le domaine médical et plus précisément Maladie de L'insuffisance rénale comme un champ d'étude.

La conclusion de ce mémoire synthétise les principales contributions de notre travail.

# **Chapitre 1 :**

# **Les réseaux**

# **bayésiens**

### 1. Introduction

La représentation des connaissances et le raisonnement à partir de ces représentations a donné naissance à de nombreux modèles. Les modèles graphiques probabilistes, et plus précisément les réseaux bayésiens (RB), initiés par Judea Pearl dans les années 1980.

La théorie des ensembles flous un pas vers un rapprochement entre la précision, elle a donc pour objet d'étude, la représentation des connaissances imprécises.

Différents travaux de recherche ont récemment permis l'utilisation de la théorie des ensembles flous dans le cadre des réseaux bayésiens parmi ces méthodes le réseau bayésien flou.

### 2. Réseaux bayésiens

Le raisonnement probabiliste et inférentiel des experts nous conduit naturellement à l'utilisation des réseaux bayésiens afin de pouvoir représenter leur connaissance et automatiser leur raisonnement.

Plusieurs autres techniques de représentation pourraient être utilisées (logique floue, système expert, arbre de décision...), mais la préférence donnée aux réseaux Bayésiens tient plus particulièrement à leur polyvalence pour :

- **L'extraction de connaissance probabiliste** : c'est-à-dire trouver quelles variables sont corrélées, dépendantes ou conditionnellement indépendantes.
- **Le diagnostic** : l'évaluation de  $P(\text{causes}|\text{symptômes})$ .
- **La prédiction** : l'évaluation de  $P(\text{symptômes}|\text{causes})$ .
- **La classification** : le calcul de  $\max_{\text{classe}} P(\text{classe}|\text{observations})$ [1].

L'intérêt des réseaux bayésiens réside aussi dans leur capacité à tenir compte dans le même modèle de plusieurs types d'informations (informations subjectives d'experts et celles issues du traitement statistique de données).

Enfin, la représentation graphique des réseaux bayésiens est intuitive et compréhensible par un non spécialiste. Cela explique leur utilisation de plus en plus grande dans le domaine médical et industriel où ils montrent des résultats encourageants. Par exemple, nous pouvons citer le système de diagnostic interactif des pannes des imprimantes [2].

Thomas bayes (1702-1761) est né à Londres en Angleterre a développé un théorème qui repose sur la propagation de l'information au sein du réseau, c'est-à-dire les calculs de probabilités a posteriori de certaines variables à partir d'un certain nombre d'observations sur d'autres variables.

$$P(B / A) = \frac{P(A / B) \cdot P(B)}{P(A)}$$

Cette règle, permet de faire un raisonnement dans les deux sens, le calcul de la probabilité de B sachant A mais aussi de A sachant B. Dans un sens nous cherchons à expliquer une cause dans l'autre nous quantifions une conséquence.

### 2.2. Définition des réseaux bayésiens

La théorie des réseaux bayésiens résulte d'une fusion entre la théorie des probabilités et la théorie des graphes [3]. La définition classique d'un réseau bayésien c'est comme un graphe acyclique dirigé. Il est formé d'un ensemble de variables et d'un ensemble d'arcs entre les variables. Chaque variable correspond à un nœud du réseau.

Un réseau bayésien  $B = \{G, P\}$  est défini par :

- Un graphe dirigé sans circuit  $G = (X, E)$  où  $X$  est l'ensemble des nœuds (ou sommets) et où  $E$  est l'ensemble des arcs.
- Un espace probabilisé  $(\Omega, P)$ .

Un ensemble de variables aléatoires  $X = \{X_1 \dots X_n\}$  associées aux nœuds du graphe et définie sur  $(\Omega, P)$  telles que

$$P\{X_1 \dots X_n\} = \prod_{i=1}^n P(x_i / Pa(x_i)) \quad (1)$$

ou  $Pa(X_1)$  est l'ensemble des parents du nœud  $X_1$  dans  $G$ .

Les réseaux bayésiens sont des modèles graphiques probabilistes représentant la connaissance et permettant de calculer des probabilités conditionnelles en s'appuyant sur le théorème de Bayes et apportant des solutions à différentes sortes de problématiques. Ils décrivent la distribution des probabilités associées à un ensemble de variables dont certains sont directement dépendants et d'autres sont indépendantes conditionnellement.

Un réseau bayésien est représenté par un graphe orienté sans circuit (GOSC) où les nœuds représentent les variables auxquelles sont associées des tables de probabilités conditionnelles.

Le graphe d'un réseau bayésien représente d'une façon visuelle les relations (dépendances et indépendances) entre les variables du modèle et les distributions de probabilités permettent d'enrichir cette structure graphique par une quantification de ces relations, afin d'apporter une meilleure modélisation à l'aspect incertain qui relie les différents variables.

### 2.3. Un exemple simple 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 alors un coup d'œil à la pelouse de son voisin, M. Watson, et s'aperçoit qu'elle est humide. Il en déduit alors qu'il a probablement plu, et il décide de partir au travail sans vérifier [10].

Dans ce texte, monsieur Holmes fait un raisonnement qui peut être décrit à l'aide d'un Réseau bayésien.

L'herbe de mon jardin est mouillée (J) si et seulement si : il a plu cette nuit (P) ou j'ai oublié de débrancher mon arroseur (A). Ce qui se traduit par la table de probabilité ci-dessous :

	A=vrai		A=faux	
	P=vrai	P=faux	P=vrai	P=faux
J=vrai	1	1	1	0
J=faux	0	0	0	1

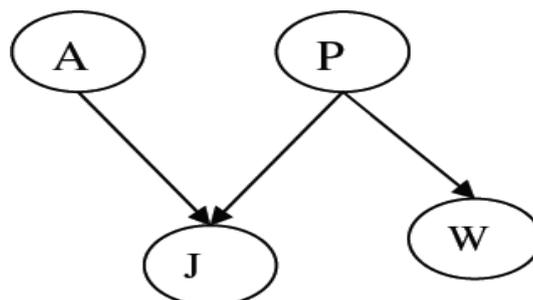
**Tableau 1: La modélisation de la table de probabilité**

L'herbe du jardin de mon voisin Watson est humide (W) si et seulement s'il a plu cette nuit. La table de probabilité associée :

	P=vrai	P=faux
W=vrai	1	0
W=faux	0	1

**Tableau 2: Tableau de probabilité conditionnelle**

Cela conduit au graphe causal ci-dessous.



**Figure 1: Graphe bayésienne**

**A** : J'ai oublié de débrancher mon arroseur

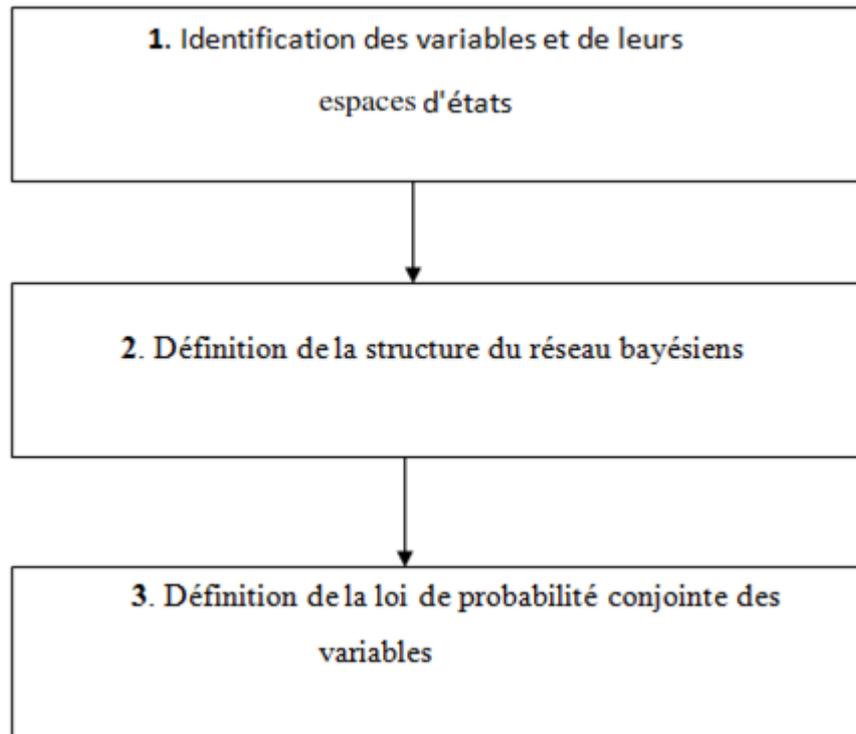
**P** : Il a plu pendant cette nuit

**J** : L'herbe de mon jardin est humide

**W** : L'herbe du jardin de mon voisin Watson est humide

### 2.4. La mise en œuvre du réseau Bayésien

La construction d'un réseau bayésien se décompose en trois étapes distinctes présentées au dessous :



**Figure 2: Les étapes de construction d'un réseau bayésien [10]**

La première étape dite **qualitative** : consiste à la définition de l'ensemble des variables du système, avec précision de l'espace d'états de chaque variable.

La deuxième est l'**étape probabiliste** : elle introduit l'idée de distribution jointe définie sur les variables à généré la base d'observations et choisir une structure de graphe qui sera compatible avec les variables.

La troisième est l'**étape quantitative** : elle consiste en l'évaluation numérique des distributions de probabilités conditionnelles.

### 2.5. Apprentissage dans les réseaux bayésien

Nous avons montré précédemment qu'un réseau bayésien est constitué à la fois d'un graphe et d'un ensemble de probabilités conditionnelles.

L'apprentissage d'un réseau bayésien doit répondre aux deux questions suivantes :

- Comment estimer les lois de probabilités conditionnelles ?
- Comment trouver la structure du réseau bayésien ?

Le problème de l'apprentissage se divise en deux parties :

- a. L'apprentissage des paramètres, où nous supposons que la structure du réseau a été fixée, et où il faudra estimer les probabilités conditionnelles de chaque nœud du réseau.
- b. L'apprentissage de la structure, où le but est de trouver le meilleur graphe représentant la tâche à résoudre.

Il existe différentes techniques pour modéliser et traiter le problème de l'apprentissage. Ces techniques peuvent se partager en deux grandes familles :

- apprentissage à partir de données, complètes ou non, par des approches statistiques ou bayésiennes.
- acquisition de connaissances avec un expert du domaine [11].

#### 2.5.1. Apprentissage de structure

Lorsque la structure du réseau n'est pas fournie à priori par un expert, La première idée est de trouver la meilleure structure d'un réseau bayésien. C'est un problème NP-complet à cause du nombre exponentiel des structures possibles, [12] a montré que le nombre de structures différentes  $r(n)$  pour un réseau bayésien possédant  $n$  noeuds est :

$$r(n) = \sum_{i=1}^n (-1)^{i+1} \binom{n}{i} 2^{i(n-i)} r(n-i) = n^{20(n)}$$

Avec :  $r(1) = 1$ ,  $r(2) = 3$ ,  $r(3) = 25$ ,  $r(5) = 29281$  et  $r(10) = 4, 2 \times 10^{18}$ .

Beaucoup de travaux existent dans la littérature pour apprendre cette structure à partir des données complètes ou données incomplètes, nous commençons par quelques algorithmes basés sur des données complètes:

##### a. Algorithmes de structure (Données complètes)

Nous trouverons deux familles d'algorithmes, les algorithmes basées sur les scores et

d'autres sur les indépendances conditionnelles :

### 1) Algorithmes basés sur un score

Cette première famille consiste à parcourir tous les graphes possibles, associer un score à chaque graphe, puis choisir le graphe ayant le score le plus élevé. Toutefois, cette méthode n'est pas simple, principalement à cause de la taille super exponentielle de l'espace de recherche en fonction du nombre de variables.

Plusieurs méthodes ont été proposées pour résoudre ce problème telle que : L'arbre de poids maximal, GS (Greedy Search), Algorithm K2.

### 2) Algorithmes basés sur l'indépendance conditionnelle

La deuxième famille consiste à déterminer dans un premier temps un graphe non orienté en tenant compte les différentes indépendances conditionnelles qui existent entre les variables de ce graphe, puis à orienter ce graphe pour obtenir un réseau bayésien. Ces algorithmes sont peu efficaces dans le cas de problèmes de grande taille, car la détermination des ces indépendances devient difficile avec un nombre important de variables.

Parmi les algorithmes basés sur l'indépendance conditionnelle on cite: Algorithmes PC (recherche de causalité), Algorithme IC (causalité inductive).

#### b. Algorithmes de structure (Données incomplètes)

Dans le cas où les données ne sont pas complètes (critère : Apprentissage de structures données incomplètes), la première idée est de manipuler que les données complètes, mais le risque d'avoir une petite quantité qui peut être insuffisante pour l'apprentissage.

Pour faire l'apprentissage de structure d'un réseau bayésien à partir de données Incomplètes, il est indispensable de résoudre les deux problèmes rencontrés, l'un lors de l'apprentissage de paramètre d'une structure à partir des données incomplètes, l'autre lors de l'apprentissage de structure à partir de données incomplètes.

Un algorithme SEM (Structural EM) a été proposé. Cet algorithme se base sur l'algorithme EM pour faire de l'apprentissage de structure [11].

#### 2.5.2. Apprentissage des paramètres

L'apprentissage des paramètres d'un réseau bayésien se fait à partir de données relatives au problème à modéliser. Toutefois, ces données peuvent être complètes ou incomplètes. Les

algorithmes d'apprentissage des paramètres ne sont pas les mêmes dans ces deux cas.

### a. A partir de données complètes

Lorsque la structure du réseau bayésien est connue et les données à traiter sont complètes.

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'apparition de l'événement dans la base de données. Cette approche, appelée maximum de vraisemblance (MV)[11].

### b. A partir de données incomplètes

Dans les applications pratiques, les bases de données sont très souvent incomplètes. 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, etc [17].

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

- **Algorithme EM**

L'algorithme EM — pour Expectation-Maximisation — est un algorithme itératif du à Dempster, Laird et Rubin (1977). Il s'agit d'une méthode d'estimation paramétrique s'inscrivant dans le cadre général du maximum de vraisemblance. Lorsque les seules données dont on dispose ne permettent pas l'estimation des paramètres, et/ou que l'expression de la vraisemblance est analytiquement impossible à maximiser, l'algorithme EM peut être une solution. De manière grossière et vague, il vise à fournir un estimateur lorsque cette impossibilité provient de la présence de données cachées ou manquantes ou plutôt, lorsque la connaissance de ces données rendrait possible l'estimation des paramètres. L'algorithme EM tire son nom du fait qu'à chaque itération il opère deux étapes distinctes

- la phase « Expectation », souvent désignée comme « l'étape E », procède comme son nom le laisse supposer à l'estimation des données inconnues, sachant les données observées et la valeur des paramètres déterminée à l'itération précédente.

- la phase « Maximisation », ou « étape M », procède donc à la maximisation de 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 [13].

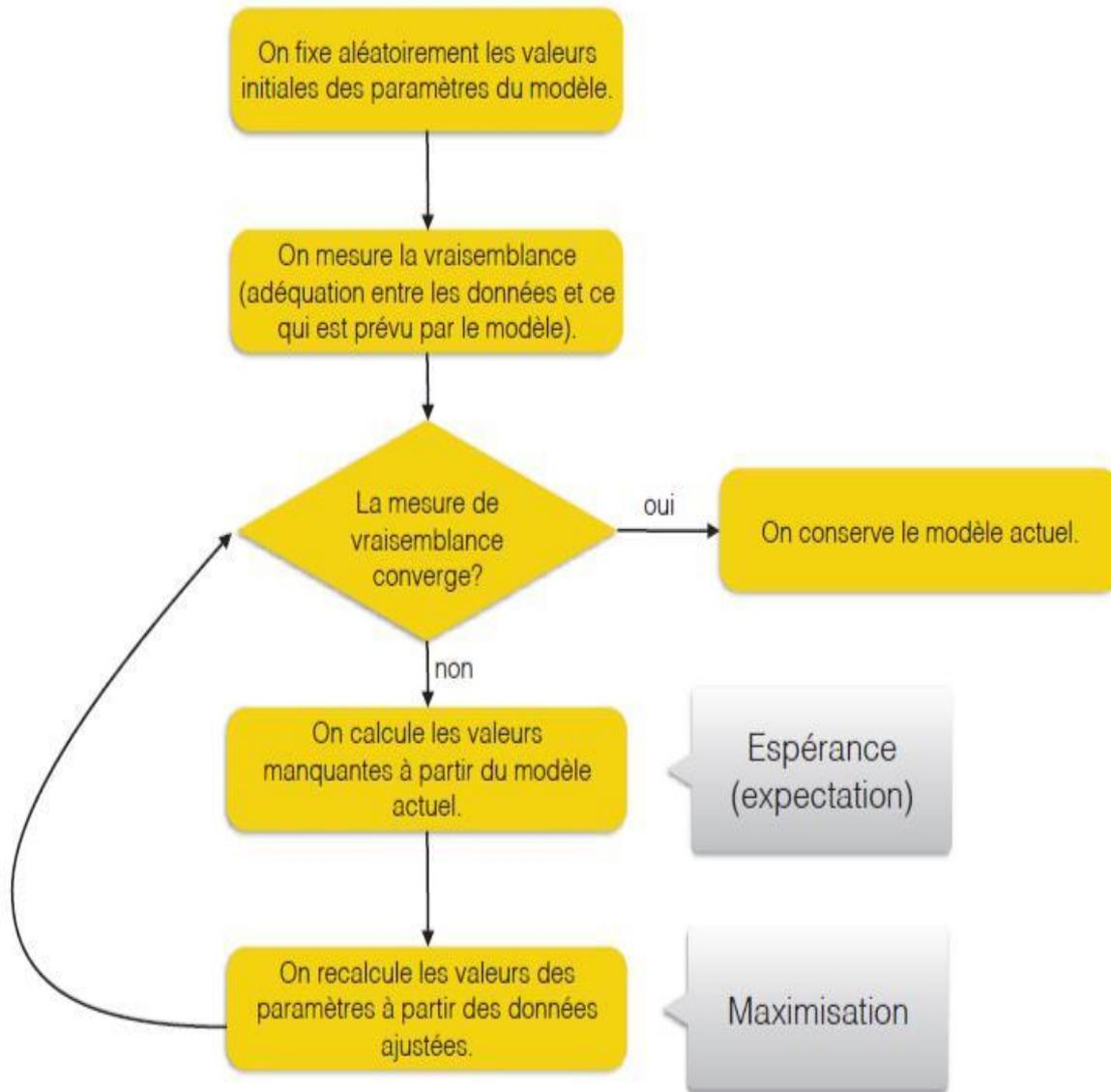


Figure 3: L'algorithme EM pour le maximum de vraisemblance

### 2.6. L'inférence dans les réseaux bayésiens

L'inférence dans un réseau bayésien concerne le calcul de la probabilité de n'importe quelle variable ou sous ensemble de variables à partir des autres variables observées. Il s'agit donc de déterminer les probabilités conditionnelles d'événement relié par des relations d'influence.

Les algorithmes d'inférence dans les réseaux bayésiens se répartissent en deux groupes : algorithmes d'inférence exacte et algorithmes d'inférence approchée. Les algorithmes d'inférence exacte exploitent les indépendances conditionnelles contenues dans le réseau pour calculer les probabilités a posteriori exactes [4], [5].

Concernant la deuxième catégorie d'algorithmes, les méthodes utilisées donnent des estimations approchées des probabilités à posteriori [13], [7].

### 2.6.1. Inférence exacte

La tâche de base de tout système d'inférence probabiliste consiste à calculer la distribution de probabilités a posteriori d'un ensemble de variables de requête, étant donné un événement observé autrement dit, une affectation de valeurs à un ensemble de variables d'observation.  $X$  représente la variable de requête,  $E$  représente l'ensemble de variable d'observation  $E_1, \dots, E_m$ , et  $e$  est un événement observé particulier,  $Y$  les variables non observées  $Y_1, \dots, Y_n$  (parfois nommées variables cachées). D'où l'ensemble complet de variables  $X = \{X\} \cup E \cup Y$ .

Une requête de type demande la distribution de probabilité à posteriori :  $P(X|e)$  [8].

### 2.6.2. Inférence approchée

Etant donné que l'inférence exacte est impraticable dans de grands réseaux multiplement connectés, il est essentiel d'envisager des méthodes d'inférence approchée. Exemple l'algorithme d'échantillonnage aléatoire, également nommé algorithme de Monte-Carlo qui fournit des réponses approchées dont l'exactitude dépend du nombre d'échantillons générés. Ces dernières années, les algorithmes de Monte-Carlo ont été largement utilisés en informatique pour estimer des quantités difficiles à calculer exactement [9].

## 3. Les avantages des réseaux bayésiens

Un réseau bayésien peut être avantageusement substitué à différents modèles d'évaluation, de prévision, de diagnostic, d'aide à la décision ou de data mining. Pour les avantages des réseaux bayésien on peut citer :

- La possibilité de rassembler et de fusionner des connaissances de diverses natures dans un même modèle : données de retour d'expérience, expertise (exprimée sous forme de règles logiques, d'équations, ou de probabilités subjectives), observations.
- Facile à utiliser: un réseau bayésien est un modèle graphique, compréhensible et manipulable par un non-spécialiste.
- La polyvalence : on peut se servir d'un même modèle pour évaluer, prévoir, diagnostiquer, optimiser des décisions.

### 4. Quelques domaines d'application

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

- **Santé** : Les premières applications des réseaux bayésiens ont été développées dans le domaine du diagnostic médical. Les réseaux bayésiens sont particulièrement adaptés à ce domaine parce qu'ils offrent la possibilité d'intégrer des sources de connaissances hétérogènes (expertise humaine et données statiques), et surtout parce que leur capacité à traiter des requêtes complexes peuvent constituer une aide véritable et interactive pour le praticien.

- Dans le domaine de la santé, une application intéressante des algorithmes issus des réseaux bayésiens a permis d'améliorer considérablement la recherche de la localisation de certains gènes, dans le cadre du projet Human Genom [16].

- **Informatique** : Les réseaux bayésiens sont probablement l'une des technologies les plus adaptées pour construire l'intelligence des agents. Ils assurent en effet ses différentes propriétés :

- **L'autonomie** : Représentée par la capacité des réseaux bayésiens de fournir des décisions en présence d'incertitude, ou en l'absence de certaines informations.

- **La motivation** : peut être représentée par certains types d'inférences, ou par un système de planification.

- **La réactivité** : c'est le principe même de l'inférence dans les réseaux bayésiens.

- **L'adaptation à l'environnement** : elle est rendue possible par les capacités d'apprentissage incrémental des réseaux bayésiens.

- **Gestion de connaissance** : 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 [16].

### 5. La théorie des ensembles flous

En théorie classique des ensembles, un objet est un élément, ou un non-élément, d'un ensemble. Toutefois, en réalité, due à l'insuffisance des connaissances ou de l'imprécision des données, il n'est pas toujours évident de déterminer si un objet appartient ou non à un ensemble. Par contre, les ensembles flous interprètent l'incertitude d'une manière approximative. Au plan conceptuel, la théorie des ensembles flous permet à un objet d'appartenir à plusieurs ensembles exclusifs dans le cadre du raisonnement. À chaque

ensemble correspond un degré de certitude selon lequel un objet appartient à un ensemble flou. Prenons l'exemple des cotes de crédit. Supposons qu'il existe trois niveaux de cote - faible, moyen, élevé - qui peuvent être considérés comme trois ensembles. Selon la théorie classique des ensembles, l'ensemble complet se compose de ces trois ensembles exclusifs. Dès que la cote de crédit est connue, son niveau est établi. La figure 3 présente un exemple d'ensemble classique de cotes de crédit. Si la cote s'établit à 3,5, il est certain à 100 % que la cote de crédit est élevée.

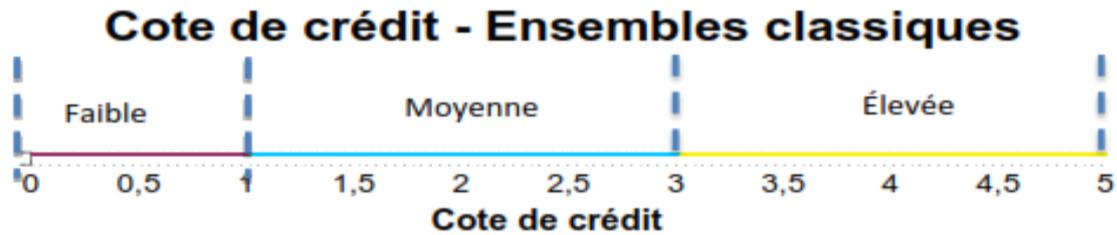


Figure 4: Exemple d'ensemble classique: cote de crédit [18].

La figure 4 présente un exemple d'ensemble flou pour les cotes de crédit. Chaque ensemble possède sa propre **fonction d'appartenance**, qui détermine le degré de certitude qu'un élément appartient à l'ensemble. Par exemple, si l'on suppose une cote de crédit de 3,5, il est certain à 60 % que la cote est élevée, et à 22 % sûr que la cote est moyenne. Il est faux de croire que la cote est faible. Dans la théorie de logique floue, le degré de certitude de tous les ensembles ne totalise pas nécessairement a un pour un objet précis.

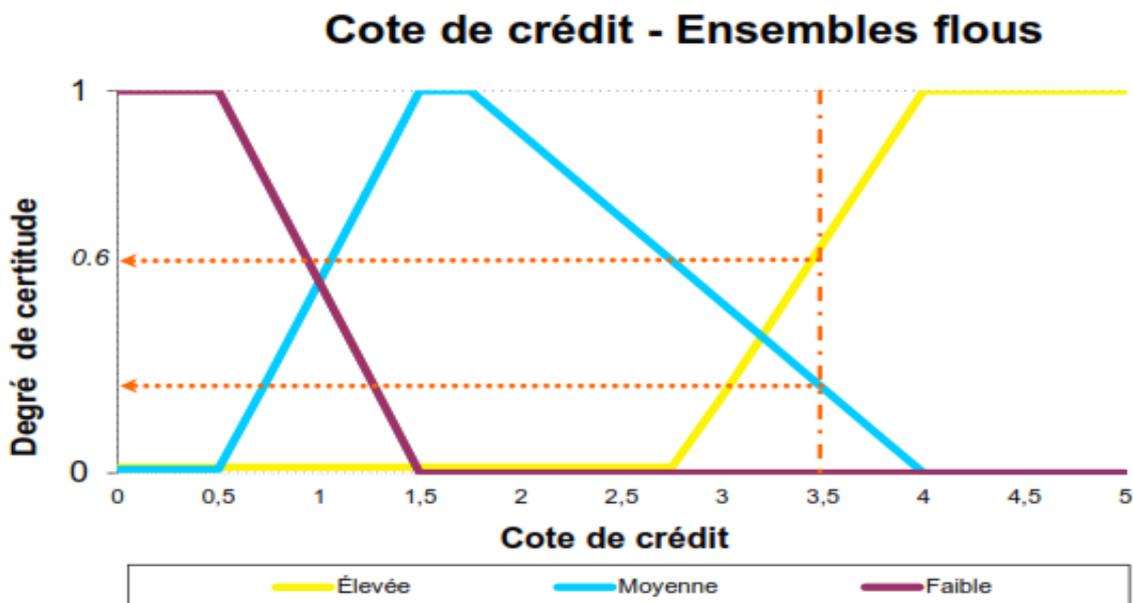


Figure 5: Exemple d'ensemble flou: cote de crédit [18].

- Dans cet exemple, les fonctions d'appartenance pour les trois ensembles sont précisées ci-après.

$$\mu^{\text{Élevée}}(x) = \begin{cases} 0 & x \leq 2.75 \\ \frac{x-2.75}{1.25} & 2.75 < x \leq 4 \\ 1 & x > 4 \end{cases}$$

$$\mu^{\text{Moyenne}}(x) = \begin{cases} 0 & x \leq 0.5 \\ \frac{x-0.5}{1} & 0.5 < x \leq 1.5 \\ 1 & 1.5 < x \leq 1.75 \\ \frac{4-x}{2.25} & 1.75 < x \leq 4 \\ 0 & x > 4 \end{cases}$$

$$\mu^{\text{Faible}}(x) = \begin{cases} 1 & x \leq 0.5 \\ \frac{1.5-x}{1} & 0.5 < x \leq 1.5 \\ 0 & x > 1.5 \end{cases}$$

- Les ensembles flous présentent une caractéristique incontournable : il n'existe pas de règles parfaites au sujet de la définition des fonctions d'appartenance. La forme mathématique de la fonction et ses paramètres dépendent de la contribution des experts.

## 6. Réseaux bayésiens flous

Différents travaux de recherche ont récemment permis l'utilisation de la théorie des ensembles flous dans le cadre des réseaux bayésiens. Différentes approches ont été proposées, mais à ce jour aucune ne fait consensus. Fogelberg et al. [14] ont cherché à généraliser les réseaux bayésiens aux réseaux bayésiens flous qui remplacent les états discrets par états flous dans tout le réseau. Ceci contribue au renforcement de la robustesse du système.

- Plus précisément le réseau bayésien flou est une extension de réseau bayésien standard telle qu'il fait la combinaison avec la théorie des ensembles flous.

### 6.1. Définition

Les réseaux bayésiens sont des outils très puissants et sont utilisés dans plusieurs situations et différents domaines.

- Des ensembles flous peuvent représenter des données dans des termes linguistiques qui aident à améliorer la compréhensibilité. De plus, la fonction d'adhésion floue fournit un cadre pour représenter les degrés d'adhésion dans un ensemble.
- Combine ces deux idées peut être conceptuellement difficile parce que la signification d'une valeur d'adhésion floue et une probabilité est très différente, pourtant ils sont représentés de la même façon dans le champ  $[0,1]$ .

## 6.2. Propagation de l'observation de vraisemblance : méthode de Pearl (virtual evidence)

La première présentation de la notion d'observation de vraisemblance se trouve dans [15].

-Les observations de vraisemblance concernent les cas où l'observation d'une variable est incertaine. Il existe une incertitude sur l'observation elle-même, par exemple parce que l'information est fournie par une source qui n'est pas parfaitement fiable ou précise. Pour spécifier cette observation, il est nécessaire d'avoir une évaluation, au moins partielle, du lien entre les différentes valeurs réelles possibles de la variable et la valeur observée.

-La méthode de Pearl pour propager une observation incertaine sur X consiste à ajouter un nœud "virtuel" O au réseau bayésien pour transformer l'observation incertaine sur X en une observation classique sur O. Le nouveau nœud est ajouté comme fils du nœud X. Il comporte deux états que l'on peut noter o et  $\bar{o}$ , où la valeur o désigne l'observation. Ainsi, X représente le nœud observé par le biais d'un observateur ou un capteur imprécis et O représente l'observation, c'est-à-dire le "biais" introduit par l'observateur.

-Dans ce réseau bayésien augmenté, l'observation incertaine sur X se traduit simplement l'instanciation  $O = o$ , et peut être propagée à l'aide des algorithmes classiques d'inférence.

### 6.2.3. Exemple

Considérons un réseau bayésien contenant une variable D représentant la valeur d'un dé à six faces.

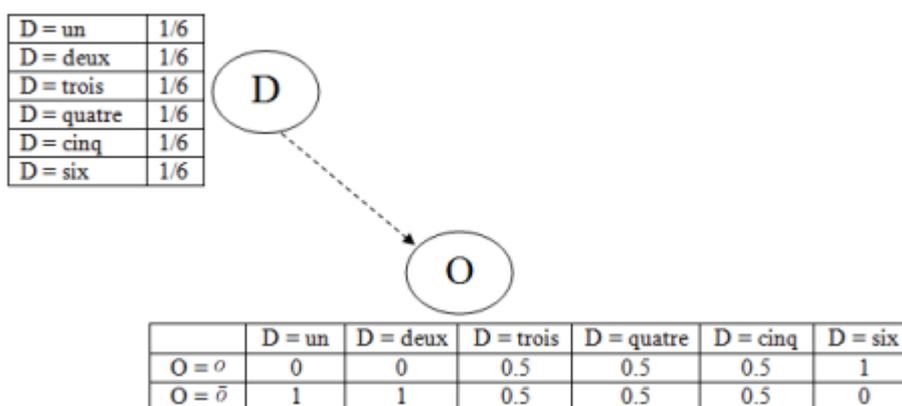


Figure 6: observation de vraisemblance sur D [6]

L'observateur est un jeune enfant qui confond encore certains chiffres, et il déclare "c'est un six". Connaissant le niveau d'acquisition des chiffres de l'enfant, on estime qu'il y a deux

fois plus de chances que l'enfant déclare avoir vu un six si le dé est effectivement tombé sur la face numéro six que s'il est tombé sur la face trois, quatre ou cinq ; et on estime que l'enfant n'affirmera pas cela si le dé est tombé sur la face un ou deux car il connaît ces chiffres. Cette information (déclaration de l'enfant + notre connaissance de l'enfant) indique entre autre que  $P(O \mid D = \text{six}) = 2 \times P(O \mid D = \text{quatre})$ . La figure 2.4 illustre l'ajout temporaire du nœud O pour prendre en compte cette observation incertaine sur la variable D. La table de probabilité conditionnelle associée au nœud est donnée par la vraisemblance de l'observation sachant X :  $P(O = o \mid X)$ .

### 7. Conclusion

Dans ce chapitre nous avons expliqué la notion des réseaux bayésiens qui est une technologie très utile pour informatiser le domaine médical, nous avons discuté ensuite tout ce qui concerne les réseaux bayésiens flous

- Dans le prochain chapitre, nous allons expliquer comment faire un diagnostic médical

# **Chapitre 2 :**

# **Le diagnostic**

# **médical**

### 1. Introduction

Etablir un diagnostic médical fiable requiert l'identification de la maladie d'un patient basé sur l'observation de ses signes et symptômes et de faire un bon raisonnement. Ces prédictions gèrent les connaissances incertaines qui doivent être présentées dans un graphe de RB (réseau bayésien).

-Parmi les données que nous avons reçues à partir d'un réseau bayésien il y a certaines données floues représentées dans l'incertitude médicale, nous allons voir toutes ces notions et les expliquer dans ce chapitre

### 2. Définitions

#### 2.1. Le signe et le symptôme

En médecine, un « signe » ou un « ensemble de signes » peut être la manifestation objective d'une pathologie, relevée par le médecin [19]. Un signe peut apparaître à des niveaux d'observation différents (par exemple clinique, para-clinique ou complémentaire), impliquant des techniques de recueil plus ou moins rapides et/ou coûteuses [20]. Les signes peuvent être scindés comme suit :

- ❖ Les signes cliniques sont des signes recueillis sans instrument lourd, par exemple le stéthoscope, le thermomètre, le tensiomètre, etc. Les signes cliniques peuvent être scindés en :
  - o Signes généraux (par exemple, la température, la fatigue, etc.).
  - o Signes focaux (délimité à une zone).
  - o Signes fonctionnels ou symptômes. Ce sont des signes récupérés par l'interrogatoire, non vérifiable par un autre signe clinique, on peut citer la douleur, ou la dyspnée
  - o Signes physiques. Ce sont ceux qui peuvent se vérifier à l'examen clinique. Par exemple, une rougeur, un bruit anormal au stéthoscope, etc.
  
- ❖ Les signes para-cliniques. Ceux qui sont observables via la radiographie, le scanner, etc.

- Un symptôme représente une des manifestations subjectives d'une maladie ou d'un processus pathologique, tel qu'exprimé par le patient [19]. Les symptômes peuvent être multiples pour une pathologie donnée.

- Une autre catégorisation des signes et symptômes existe et s'appuie sur la relation entre ces derniers et le diagnostic que veut établir un médecin. Cette catégorisation est fondée sur le postulat que chaque signe pris séparément n'a pas de sens en soi, et c'est plutôt sur un ensemble de signes que sera établi le diagnostic. Cette catégorisation est la suivante : signes ou symptômes positifs et signes ou symptômes négatifs.

### 2.1.1. Le signe positif

Un signe est dit positif si c'est sa présence qui a une signification, par exemple une quantité d'immunoglobuline de type A supérieure à 3g/dL est un signe positif pour le myélome multiple [21].

### 2.1.2. Le signe négatif

Werner Ceusters [22], [23] définit les « signes négatifs » ou « signes absents » comme étant tous les signes qui sont exprimés avec des négations, par exemple « pas de douleur lombaire, absence de fièvre etc. ». C'est aussi, dans certaines situations un signe dont c'est l'absence qui a une signification, par exemple, l'absence de CRAB (C-hypercalcémie, R- insuffisance rénale, A- anémie, B- lésion osseuse) est caractéristique de la gammopathie monoclonale à signification indéterminée (MGUS) [21].

## 3. La maladie et le syndrome

Nous présentons dans cette partie deux définitions de la maladie, celle de l' « Ontology for General Medical Science » (OGMS) [24], celle de Émile Littré et celle de Anne Fagot-Largeault [20].

- En effet, selon l'OGMS [24] : « Une maladie est une disposition à subir un processus pathologique causé par un ou plusieurs désordres sous-jacents ». En d'autres termes, la maladie est une disposition réalisée dans des processus pathologiques. Nous utilisons le terme « réalisation » pour faire référence au processus par lequel une disposition est réalisée. Chaque disposition dans l'organisme a un site physique. Le site physique d'une maladie est une certaine combinaison de composants physiques dans l'organisme. Quel que soit l'organisme, il y a une large variété de dispositions, certaines sont associées à la bonne santé, d'autres à la maladie. Par exemple, le SIDA est une disposition du patient sain à subir une persistance et une duplication pathogène (processus pathologique) suivis des infections opportunistes survenant à cause de la faiblesse du système immunitaire (site physique).

Une disposition est une fonction d'un organisme en vertu de laquelle elle peut initier des processus spécifiques quand certaines conditions sont satisfaites. Par exemple, la disposition du corps humain à avoir besoin d'eau en cas de déshydratation, la disposition du corps humain à avoir des hausses de température en cas d'infection. Le processus pathologique est un processus physique qui est une manifestation d'un désordre. Un désordre est une combinaison causale des composants physiques qui sont cliniquement anormaux [24]. Selon Émile Littré [19]: « La maladie est une altération organique ou fonctionnelle de la santé considérée dans son ensemble comme une entité définissable ».

### 4. Le diagnostic médical

Le diagnostic médical est une démarche par laquelle le médecin va déterminer l'affection dont souffre le patient, à partir des symptômes et des signes que ce dernier présente, et à l'aide d'éventuelles investigations complémentaires [25],[19]. Cette démarche repose sur la recherche des causes (étiologie) et des effets (signes et symptômes) de l'affection [19]. Randolph Miller [26] définit le diagnostic médical comme : « l'acte d'associer le nom d'une ou plusieurs maladies ou syndromes à des manifestations observées (antécédents, symptômes, signes) sur un patient ».

- C'est un processus qui résulte de la confrontation d'un problème réel à l'expérience acquise et à un corpus de connaissances théoriques. Il se déroule en trois étapes : (a) l'enquête interrogative, (b) l'examen clinique proprement dit et (c) les examens complémentaires [27].

- ❖ Premièrement, le médecin prend connaissance des symptômes se manifestant chez un patient. A 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 médecin « réalise » une mise en correspondance entre les informations obtenues au cours des trois étapes précédentes avec les connaissances qu'il possède de par sa formation et son expérience. Si au terme des étapes précédentes, le taux de confiance d'une certaine hypothèse s'accroît au point de dissiper le doute sur la maladie à laquelle est confronté le médecin, ce dernier peut alors formuler son diagnostic final et prescrire le traitement adéquat au patient. Si

le cas reste ambigu après les trois étapes indiquées, le médecin cherche alors une autre source d'informations qui puisse apporter une quantité d'informations supplémentaires permettant d'éliminer l'ambiguïté. Souvent, il demande un examen complémentaire qui peut être sous forme d'analyses biologiques, d'imagerie médicale, etc. Il acquiert des informations supplémentaires qui viennent compléter la quantité d'informations dont il dispose déjà, et qui lui permet de confirmer ou d'infirmer la ou les hypothèses qu'il a déjà faites. Si le médecin n'arrive toujours pas à établir un diagnostic, une dernière étape consiste à ce qu'il ait recours à l'étude d'une base de cas similaires traités par le passé (son expérience) afin d'établir une correspondance avec le cas actuel auquel il est confronté en s'appuyant sur toutes les informations dont il dispose. Il utilise alors les cas les plus similaires (leurs solutions) afin d'en extraire des informations susceptibles de l'aider à trouver une solution à son cas.

- Le processus du diagnostic médical peut relever d'un raisonnement déductif, inductif, abductif, par analogie, etc. [28],[29],[30].

### 4.1. Le raisonnement déductif

Le raisonnement déductif va du général au particulier ou du principe à la conséquence et permet de déduire des conclusions valides à partir d'un ensemble de prémisses, c'est une approche descendante [31] [32]. Par exemple :

- ❖ Toutes les rhinopharyngites virales ont une évolution spontanément favorable (règle),
- ❖ Ce patient a une rhinopharyngite virale (cas).
  - Ce patient aura une évolution spontanément favorable (résultat).

- Ce mode de raisonnement peut être appliqué via un chaînage avant, un chaînage arrière ou un chaînage mixte suivant les règles du « modus ponens » et du « modus tollens »

#### Modus ponens

Si A est vrai		A
Et A implique C	ou	<u>A→C</u>
Alors C est vrai		C

### Modus tollens

Si $\neg C$ est vrai		$\neg C$
Et A implique C	ou	<u><math>A \rightarrow C</math></u>
Alors $\neg C$ est vrai		$\neg C$

- Le « chaînage avant » est un mode de raisonnement qui part des faits et règles de la base de connaissances, et tentent de s'approcher des conclusions recherchées par le problème.
- Le « chaînage arrière » quant à lui part des hypothèses recherchées par le problème, et tente par l'intermédiaire des règles, de « remonter » à des faits connus.
- Le « chaînage mixte » est une combinaison du chaînage avant et du chaînage arrière.

### 4.2. Le raisonnement inductif

- Le raisonnement inductif montre une proposition générale à partir d'observations particulières. Il est utilisé dans les processus d'apprentissage et est ascendant, c'est-à-dire qu'à partir d'exemples, on produit des règles qui décrivent ces exemples. Si ces règles sont bonnes elles s'appliquent non seulement à ces exemples particuliers mais aussi d'une manière générale à d'autres cas [31]; [32]. Par exemple :

- ❖ Un groupe de patients a une rhinopharyngite virale (cas).
- ❖ Tous les patients ont guéri spontanément (résultat).
- ❖ Toutes les rhinopharyngites virales ont une évolution spontanément favorable (règle).

### 4.3. Le raisonnement abductif

Le raisonnement abductif est une forme d'inférence qui part des données (prémises) décrivant quelque chose vers les hypothèses (conclusion) permettant d'expliquer ces données. Par exemple, étant donné une proposition  $a$ , on essaie de savoir ce qui peut l'avoir induite. Pour cela, on regarde s'il y a une implication de la forme  $b \rightarrow a$  et, si c'est le cas, on établit que  $b$  est la « cause » de  $a$ .

- Dans le cadre de la médecine, selon Douglas Walton [33], Charniak et McDermott [34], le diagnostic médical est une forme de raisonnement abductif car il permet d'inférer les causes d'un phénomène, d'un dysfonctionnement, d'un signe, d'un symptôme etc.

- Soit D une collection de données (faits, observations, signes et symptômes), soit H une hypothèse de D (si H est vrai, il explique D) et Aucune autre hypothèse ne peut mieux expliquer D que H. Alors, l'hypothèse H est probablement vrai.

- Le raisonnement abductif génère des hypothèses permettant d'expliquer la présence des signes ou des symptômes et il donne la possibilité de générer des hypothèses même si on n'a pas observé tous (de manière simultanée) les signes ou symptômes caractérisant ces hypothèses [29]. Mais, il ne se limite pas seulement à la génération d'hypothèse, il permet aussi d'évaluer ces hypothèses. C'est un raisonnement incrémental qui pour chaque signe ou symptôme observé peut confirmer ou infirmer une hypothèse précédemment générée [34].

#### 4.4. Le raisonnement par analogie

Le raisonnement par analogie est le procédé qui consiste à affecter une ou plusieurs propriétés à un objet (la cible) sachant qu'un objet de référence (la base) l'a ou les possède déjà. Ce transfert peut nécessiter ou induire l'apparition d'autres propriétés sur la cible.

- L'analogie exploite des relations de correspondances et des relations de dépendances afin de transposer des connaissances d'un univers à un autre. L'analogie se fonde sur la notion de similitude de termes dans un certain univers, du type : A est à B, ce que C est à D. Le raisonnement à partir de cas en est un sous-type [35].

### 5. Les incertitudes médicales

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 (prévention primaire ou passive), étape de dépistage de la maladie (prévention secondaire ou active), étape de diagnostic et, enfin étape de soins curatifs.

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

- L'incertitude porte sur l'occurrence de la maladie (incertitude événementielle). Or, cette occurrence n'est pas toujours probabilisable, comme le suggère Faugère [41], en rappelant, entre autres, que la maladie n'est pas indépendante des facteurs de risque. L'incertitude est alors en partie endogène : le patient pourra, par son comportement, modifier le risque sans modifier nécessairement le jugement qu'il lui porte. On pense, par exemple que le "régime

méditerranéen" réduit le risque cardio-vasculaire. Or ce régime est suivi inintentionnellement. Toute modification de régime alimentaire ne s'accompagne donc pas nécessairement d'une révision du jugement sur l'occurrence de la maladie. Ceci ne signifie pas que le risque n'est pas probabilisable, mais que le jugement de probabilité est faussé par le manque d'information : le jugement a lui-même une dimension aléatoire.

- L'incertitude à laquelle le patient est confronté est souvent jugée secondaire puisque celui-ci s'en remet au praticien. L'un des rôles du médecin consiste en effet, à "lever le doute" du patient sur son état de santé. L'incertitude est alors transférée du patient au praticien.

- Le médecin hérite des doutes de ses patients, il est donc confronté à la même incertitude que ce dernier mais aborde le problème de façon radicalement différente : il effectue une expertise de la situation, ce que le patient était dans l'incapacité cognitive de réaliser. Le jugement qu'il forme est "éclairé" par un faisceau d'informations, fourni par le diagnostic, que le médecin combine avec un savoir et un savoir-faire. Là où le patient ne pouvait que mesurer une incertitude, le médecin mesure le risque. L'expertise est alors, selon l'expression de Salais, "une compétence à réagir à l'aléa".

- 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 soin :

- La première est celle de la justesse du diagnostic. Le rôle du diagnostic consiste à dévoiler toute l'information nécessaire à son 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) et aiguillées par le patient. Or, le patient n'est pas l'expert, il peut donc omettre, 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. A l'inverse, le patient peut influencer le diagnostic du médecin en attirant son attention sur une information erronée [36] La réponse à cette incertitude structurelle du diagnostic consiste, pour le médecin, à multiplier les examens complémentaires au détriment de la durée et du coût d'élaboration de ce diagnostic, à multiplier les analyses cliniques au détriment de l'écoute du malade [37], [38].

- La 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 fréquentistes. 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 [39]. 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 de 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 est 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 raisons 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 selon Schmidt. L'environnement dans lequel le médecin prend sa décision est donc à information incomplète (incomplétude des sources d'information : diagnostic et études épidémiologiques) et à information imparfaite (asymétrie d'information).

### **6. Les réseaux bayésiens et raisonnement abductif**

Pour rappel, le diagnostic médical est abductif [33]. Le raisonnement abductif permet de produire la meilleure hypothèse explicative d'un fait. Alor, il consiste à produire l'hypothèse de maladie la plus significative pouvant justifier la survenue d'un signe ou d'un symptôme. En effet, un moteur d'inférences mettant en œuvre un raisonnement abductif exploite une base de connaissances contenant les triplets (maladie, signe ou symptôme, relation) et, à partir des observations (signes et symptômes) effectuées, il extrait les maladies ou hypothèses de

maladies correspondantes en fonction des situations matérialisées par les relations (est une cause, est une manifestation, est un site anatomique, exclut, implique etc.).

### 7. La discrétisation des variables continues

La discrétisation d'une variable continue nécessite de déterminer le nombre d'intervalles et l'ensemble des seuils. Ces choix sont parfois un compromis difficile. En effet, dans le cas d'un réseau bayésien discret, le nombre d'états des variables, ainsi que le nombre des parents d'un nœud détermine la taille de la table de probabilités conditionnelles de ce nœud, et donc la taille du modèle. Or, la taille du modèle est un facteur important pour l'apprentissage et l'inférence dans le réseau bayésien. A l'opposé, une discrétisation grossière peut conduire à une perte de précision. La solution proposée dans [40], pour propager des observations sur les variables continues dans un réseau bayésien discret consiste à choisir une discrétisation avec un petit nombre d'intervalles, et à propager des observations probabilistes fixes sur ces états pour représenter des observations certaines sur les variables continues. Ainsi, les observations certaines sur les variables continues seront remplacées par des observations probabilistes sur les mêmes variables après discrétisation. Les observations probabilistes sont obtenues par discrétisation floue.

### 8. Conclusion

Le diagnostic médical relève du raisonnement scientifique, mais il possède des difficultés spécifiques telles que l'incertitude et l'imprécision. Chaque étape du diagnostic médical peut donner lieu à des erreurs qui, dans le domaine de la santé peuvent être lourdes de conséquences. C'est pour cela que nous avons essayé de résoudre ce problème par les observations probabilistes qui sont obtenues par discrétisation floue.

# **Chapitre 3 :**

# **La conception**

# **du système**

### 1. Introduction

Le diagnostic médical est une tâche qui permet aux médecins de faire des prédictions sur une maladie selon les signes qui apparaissent sur un patient. Ces prédictions gèrent les connaissances incertaines qui doivent être introduites dans un réseau bayésien. La construction de ce dernier se fait à l'aide d'une certaine méthode qui se déroule à travers plusieurs étapes que nous allons présenter dans ce chapitre.

### 2. Motivation de l'étude de cas

L'insuffisance rénale chronique a touché 753 millions de personnes dans le monde en 2016, dont 417 millions de femmes et 336 millions d'hommes [42]. En 2015, elle a entraîné 1,2 million de décès, contre 409 000 en 1990[43]. Les causes qui contribuent au plus grand nombre de décès sont l'hypertension artérielle à 550 000, suivie par le diabète à 418 000 et la glomérulonéphrite à 238 000.

- L'application de notre système ne se limite pas à ce cas uniquement. En effet, les incertitudes qui existent dans le domaine médical sont multiples et plus particulièrement les maladies dont le diagnostic se fait difficilement.

### 3. Maladie de L'insuffisance rénale

L'insuffisance rénale est la réduction ou l'impossibilité que présente le rein, à assurer la filtration et l'élimination des déchets du sang, C'est aussi la réduction de la capacité des reins, à contrôler l'équilibre du corps en eau et en sels minéraux.

- La maladie rénale chronique comprend des affections qui endommagent les reins et diminuent leur capacité à rester en bonne santé en accomplissant les tâches énumérées. Si la maladie rénale s'aggrave, les déchets peuvent atteindre des niveaux élevés dans le sang et rendre les gens malades. Elle peut développer des complications telles que l'hypertension artérielle, l'anémie (faible numération globulaire), la faiblesse des os, une mauvaise santé nutritionnelle et des lésions nerveuses. En outre, la maladie rénale augmente le risque d'avoir une maladie cardiaque et des vaisseaux sanguins. Ces problèmes peuvent se produire lentement sur une longue période de temps. La maladie rénale chronique peut être causée par le diabète, l'hypertension artérielle et d'autres troubles. La détection précoce et le traitement peuvent souvent empêcher la maladie rénale chronique d'empirer. Lorsque la maladie rénale progresse, il peut éventuellement conduire à une insuffisance rénale, ce qui nécessite une dialyse ou une greffe de rein pour maintenir la vie.

### 4. Base des données sur l'insuffisance rénale

La base des données est sur les patients qui probablement avaient une insuffisance rénale, nous représentons les informations à propos de notre BD :

- Titre: Early stage of Indians Chronic Kidney Disease(CKD).

#### a. Les informations sur la source

Les informations sur la source Sont listées dans le tableau suivant :

Date	Juillet 2015
Source	Dr.P.Soundarapandian.M.D.,D.M (Senior Consultant Nephrologist), Apollo Hospitals, Managiri, Madura Main Road, Karaikudi, Tamilnadu, India.
Créateur	L.Jerlin Rubini(Research Scholar) Alagappa University, EmailId :jel.jerlin@gmail.com ContactNo : +91-9597231281
Guidée par	Dr.P.Eswaran Assistant Professor, Department of Computer Science and Engineering, Alagappa University, Karaikudi, Tamilnadu, India. Emailid:eswaranperumal@gmail.com

**Tableau 3: Renseignement sur la source**

#### b. Les attributs

Nous présentons la liste des attributs qui existe dans La base de donnée de l'insuffisance rénale et leur abréviation dans le tableau suivant :

Attributs	Abréviations
Age	age
blood pressure	bp
specific gravity	sg
albumin	al
sugar	su
red blood cells	rbc
pus cell	pc
pus cell clumps	pcc
bacteria	ba
blood glucose random	bgr
blood urea	bu
serum creatinine	sc
sodium	sod
potassium	pot
hemoglobin	hemo
packed cell volume	pcv
white blood cell count	wc
red blood cell count	rc
hypertension	htn
diabetes mellitus	dm
coronary artery disease	cad
appetite	appet
pedal edema	pe
anemia	ane
class	class

**Tableau 4: Attributs et abréviation**

- Nombre d'instances : 399 (250 chronic kidney disease) et 149( not chronic kidney disease)
- Nombre d'attributs : 24 + class = 25 (11 numérique ,14 nominal)

### c. Informations des attributs

La description de tous les attributs qui ont été utilisées

- **Age (numérique)**
- **Blood Pressure(numérique) in mm/Hg** : Lorsque notre cœur bat, il serre et pousse le sang à travers nos artères vers le reste de notre corps. Cette force crée une pression sur ces vaisseaux sanguins, et c'est notre tension artérielle systolique.
- **Specific Gravity(nominal) - (1.005,1.010,1.015,1.020,1.025)** : La gravité de l'insuffisance rénale chronique est estimée par la clairance de la créatinine (ClCr) après dosage du taux de créatinine dans le sang du patient.
- **Albumin (nominal) - (0,1,2,3,4,5)** : Les albumines (du latin albus, blanc) sont des protéines solubles dans l'eau pure, moins dans l'eau salée, elles sont composées d'environ 580 acides aminés et ne contiennent pas de glucides.
- **Sugar (nominal) - (0,1,2,3,4,5)** : Le sucre est le nom général des hydrates de carbone sucrés qui se dissolvent dans l'eau.
- **Red Blood Cells (nominal) - (normal, abnormal)** : Les globules rouges (hématies) sont les cellules les plus nombreuses du sang (près de 5 millions par mm<sup>3</sup>), dont le rôle est de transporter l'oxygène entre les poumons et les organes.
- **Pus Cell (nominal) - (normal, abnormal)** : Ce sont des leucocytes (globules blancs), spécifiquement des neutrophiles, qui attaquent les bactéries et les champignons. Certains vivent, mais le pus est principalement constitué de cellules mortes.
- **Pus Cell clumps (nominal) - (present, notpresent)** : Clumped est le processus dans lequel les substances se rassemblent dans des collections inégales ou des touffes. C'est souvent une conséquence indésirable lorsqu'une variation des niveaux d'humidité se produit.
- **Bacteria(nominal) - (present, notpresent)** : Le terme bactérie est un nom vernaculaire qui désigne certains organismes vivants microscopiques et procaryotes présents dans tous les milieux.
- **Blood Glucose Random (numérique) in mgs/dl** : Une analyse de sang réalisée de manière aléatoire en vue de surveiller le niveau de glucose dans le sang.
- **Blood Urea(numérique) in mgs/dl** : L'urée est un déchet azoté issu de la dégradation hépatique des protéines. À elle seule, cette molécule constitue l'un des principaux modes d'élimination des déchets de l'organisme. À ce titre, le dosage du taux d'urée dans le sang, également appelé "urémie", est souvent préconisé pour évaluer l'efficacité de la fonction rénale. En effet, l'urée est généralement filtrée par les reins

avant d'être évacuée par voie naturelle (urines). La mesure de l'urémie est obtenue par l'analyse d'un prélèvement sanguin réalisé par ponction veineuse. Il arrive parfois que le dosage sanguin soit accompagné d'un dosage de l'urée dans les urines pratiqué sur une durée de 24 heures. L'association des deux examens permet, en effet, au praticien de bénéficier d'une vision plus globale de la capacité rénale du patient.

- **Serum Creatinine(numérique) in mgs/dl** : La créatinine est issue de la dégradation de la créatine, elle-même synthétisée par le foie et stockée dans les muscles où elle joue un rôle important dans la production d'énergie. L'utilisation de créatine par les muscles produit des déchets, le plus notable étant la créatinine. Celle-ci est transportée par le sang, filtrée par les reins et éliminée dans les urines. Ainsi, comparer le niveau de créatinine dans le sang à la quantité de créatinine évacuée par les urines permet d'évaluer l'activité des reins.
- **Sodium(numérique) in mEq/L** : Le sodium est un minéral qui joue un rôle important dans l'état d'hydratation de l'organisme. Il est présent dans le sang et dans le liquide extracellulaire dans lequel baignent les cellules. Le sodium aide également à maintenir l'équilibre acido-basique et est essentiel dans la transmission des influx nerveux ainsi que la contraction musculaire. Cependant, en excès il peut avoir des conséquences délétères. C'est pourquoi les recommandations actuelles visent à limiter la consommation de sodium.
- **Potassium(numérique) in mEq/L** : Le potassium joue un rôle majeur dans notre organisme. Sa baisse ou son augmentation dans le sang peuvent avoir des répercussions importantes sur la santé. Le potassium agit en collaboration avec le sodium. Tous deux sont chargés de maintenir l'équilibre acido-basique du corps et des fluides. Le potassium, seul, permet la transmission d'impulsions nerveuses, la contraction musculaire, le bon fonctionnement de la fonction rénale. Ce minéral occupe ainsi une place essentielle au sein de notre organisme.
- **Hemoglobin(numérique) in gms** : L'hémoglobine, couramment symbolisée par Hb, parfois Hgb, est une métalloprotéine contenant du fer, présente essentiellement dans le sang des vertébrés au sein de leurs globules rouges, ainsi que dans les tissus de certains invertébrés. Elle a pour fonction de transporter l'oxygène O<sub>2</sub> depuis l'appareil respiratoire (poumons, branchies) vers le reste de l'organisme.
- **Packed Cell Volume (numérique)** : L'hématocrite (Ht ou HCT), également connu sous plusieurs autres noms, est le pourcentage volumique (vol%) de globules rouges dans le sang.

- **White Blood Cell Count (numérique) in cells/cumm** : Un nombre de globules blancs est un test qui mesure le nombre de globules blancs dans notre corps.
- **Red Blood Cell Count (numérique) in millions/cmm** : Un nombre de globules rouges est un test qui mesure le nombre des globules rouges dans notre corps.
- **Hypertension(nominal) - (yes,no)** : La tension artérielle, ou pression artérielle, mesure la force exercée par le sang sur les parois des artères.
- **Diabetes Mellitus(nominal) - (yes,no)** : Le diabète sucré (ou diabète par abus de langage) est une maladie liée à une défaillance des mécanismes biologiques de régulation de la glycémie (concentration de glucose dans le sang) menant à une hyperglycémie.

Cette maladie se manifeste par des symptômes propres au diabète (syndrome polyuro-polydipsique) et par des lésions d'organes tels la rétine, les reins, les artères coronaires, etc., dues à la toxicité du glucose.

- **Coronary Artery Disease(nominal) - (yes,no)** : La coronaropathie, également connue sous le nom de cardiopathie ischémique (CIH), fait référence à un groupe de maladies comprenant l'angor stable, l'angor instable, l'infarctus du myocarde et la mort subite d'origine cardiaque. C'est dans le groupe des maladies cardiovasculaires dont c'est le type le plus commun.
- **Appetite(nominal) - (good,poor)** : L'appétit se distingue particulièrement du désir en ce qu'il n'est pas constant, mais périodique, et qu'apaisé pour un temps il renaît après des intervalles déterminés.
- **Pedal Edema(nominal) - (yes, no)** : L'œdème de la pédale est l'accumulation de liquide dans jambes les pieds et inférieures.
- **Anemia(nominal) - (yes, no)** : L'anémie est une condition dans laquelle vous n'avez pas assez de globules rouges sains pour transporter l'oxygène adéquat aux tissus du corps. Avoir de l'anémie peut vous rendre fatigué et faible.
- **Class (nominal) class - (ckd, notckd)** : La class est le résultat de la maladie.

### d. Information générale sur la base des données

- Valeurs d'attribut manquantes : Oui (désigné par "espace vide").
- Classe Distribution: (2 classes).

Classe : ckd (chronic kidney disease), notckd (not chronic kidney disease) et le nombre d'instance en ordre est 250, 150 instances.

### 5. Architecture globale de notre système

L'objectif de notre travail est de transformer un ensemble des données à un graphe d'un réseau bayésien afin de pouvoir réduire l'incertitude grâce au réseau bayésien flou pour calculer et prendre la décision la plus appropriée.

- Le diagramme dans la figure suivante donne une vision globale de notre travail et les différents processus de transformation.

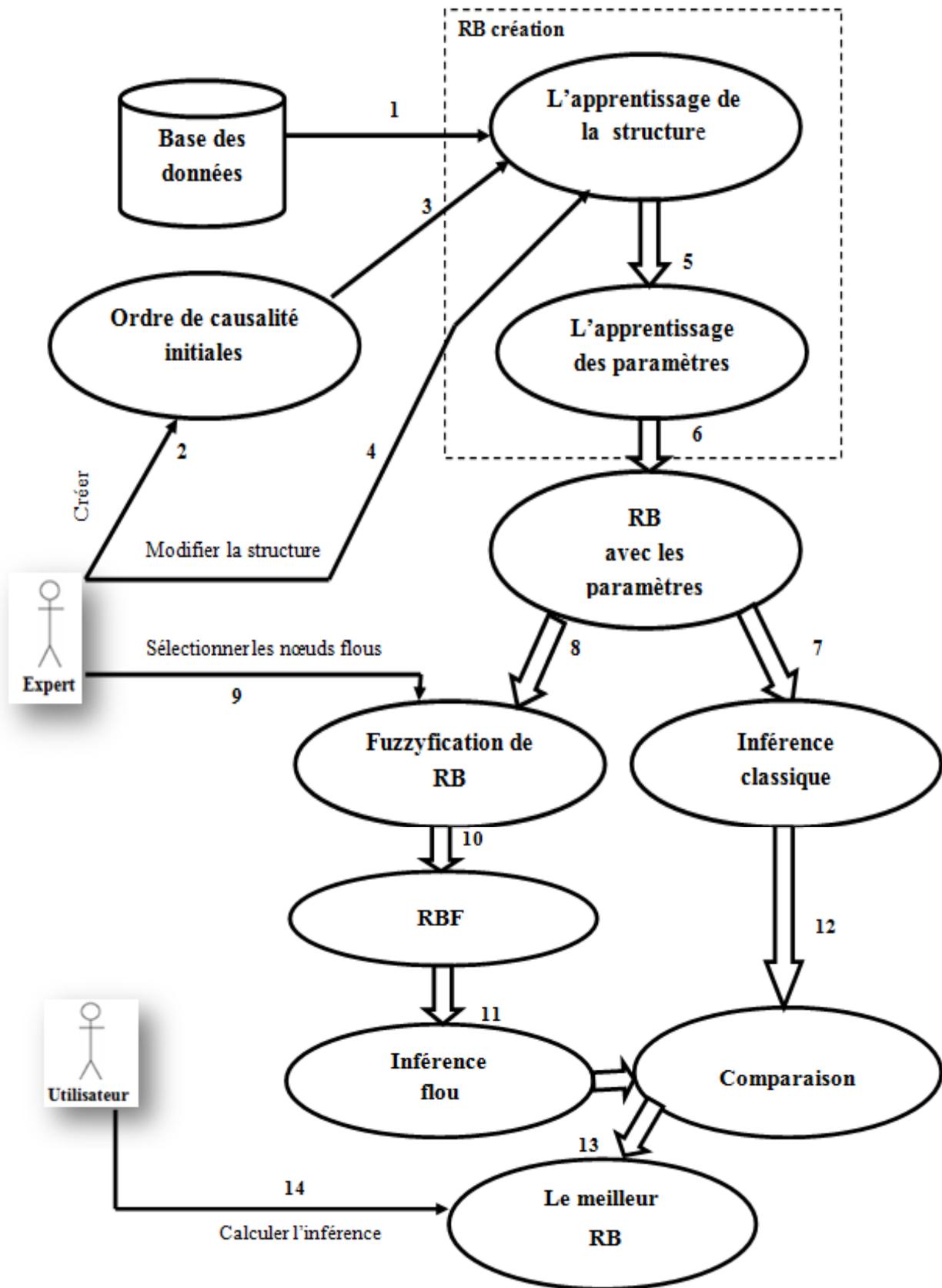


Figure 7: schéma globale

### 6. créer le RB

Créer le RB est une étape essentielle pour pouvoir calculer la probabilité de la maladie.

#### 6.1. L'apprentissage de la structure

L'objectif de l'apprentissage de la structure est de trouver une structure du graphe à partir des données disponibles et qui représente le mieux un problème.

- L'apprentissage de la structure d'un réseau bayésien à partir d'un ensemble de données est un NP-difficile, cela devient de plus en plus difficile comme la taille de l'ensemble de données d'apprentissage diminue. Les approches algorithmiques conçues pour résoudre ce problème.

##### 6.1.1. Transformation d'une base de donnée vers nominale

Cette conversion permet de changer les valeurs numériques en valeurs nominales en utilisant ces normes que nous avons reçues de l'expert du domaine médicale afin pouvoir la présenter dans un tableau.

Blood Pressure	Min=50 mm Hg Max=180 mm Hg	low normal high	[0 - 60[ mm Hg [60 - 85] mm Hg ]85 - 180] mm Hg
Blood Glucose	Min=70 mgs/dl Max=490 mgs/dl	low normal high	[0 - 63[ mgs/dl [63 - 140] mgs/dl ]140 - 490] mgs/dl
Blood Urea	Min =10 mgs/dl Max =391 mgs/dl	low normal high	[0 - 15[ mgs/dl [15 - 49] mgs/dl ]49 - 391] mgs/dl
Serum Creatinine	Min =10 mgs/dl Max =391 mgs/dl	low normal high	[0 - 56,6[ $\mu$ mol/L [56,6 - 112,3] $\mu$ mol/L ]112,3 - 391] $\mu$ mol/L
Sodium	Min=76 MEQ/L Max=163 MEQ/L	low normal high	[ 0 - 135[ MEQ/L [135 - 145] MEQ/L ]145 - 163 ]MEQ/L
		low	[0 - 3.5[ MEQ/L

Potassium	Min=2.5 MEQ/L Max=47 MEQ/L	normal high	[3.5 - 5.2] MEQ/L ]5.2 – 47] MEQ/L
Hemoglobin	Min=3.1 gms Max=17.8 gms	low normal high	[0 - 11.7[ gms [11.7 - 15] gms ]15.6 - 17.8] gms
Packed Cell Volume	Max=9% Min=54%	low normal high	[0 - 29[% [29 - 46] % ]46 – 54] %
WhiteBlood Cell Count	Min=2200/mm <sup>3</sup> . Max=26400/mm <sup>3</sup>	low Normal high	[0 - 4 000[/mm <sup>3</sup> . [4 000-10 000]/mm <sup>3</sup> ]10 000 -26400] /mm <sup>3</sup>
Red Blood Cell Count	Min=2.1 millions /mm <sup>3</sup> Max=8 millions /mm <sup>3</sup>	low normal high	[0 – 4[ millions/mm <sup>3</sup> [4 – 6] millions/mm <sup>3</sup> ]6 - 8] millions/mm <sup>3</sup>
Age	Min=2 ans Max=90 ans	Young Adulte	[1 - 20[ ans [20 - 90] ans

**Tableau 5: Transformation vers nominale**

- Le Min et le Max représenter le maximum et le minimum dans la base de données de l'insuffisance rénale notre cas d'étude.

### 6.1.2. Algorithmes de structure

Nous allons montrer les différents algorithmes que nous avons appliqués, afin d'obtenir une structure de réseau bayésien. (Naive Bayes, TAN, Augmented Naive Bayes, PC, Bayesian Search , Greedy Thick Thinning).

#### a. Naive Bayes

Les réseaux bayésiens naïfs sont largement utilisés pour résoudre des problèmes de classification [44], [45]. Cet algorithme d'apprentissage crée une structure de graphe Naive Bayes dans laquelle une seule variable de classe pointe vers toutes les autres variables.

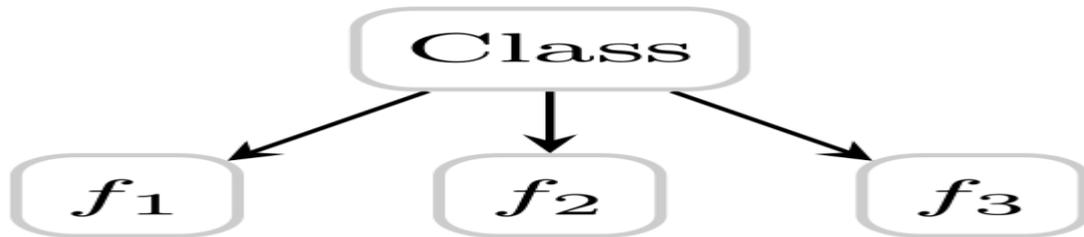


Figure 8: Naive Bayes

**a .1. Naive Bayes structure**

Exemple de structure après que nous avons utilisés Naive Bayes algorithme sur la base de donnée de l'insuffisance rénale.

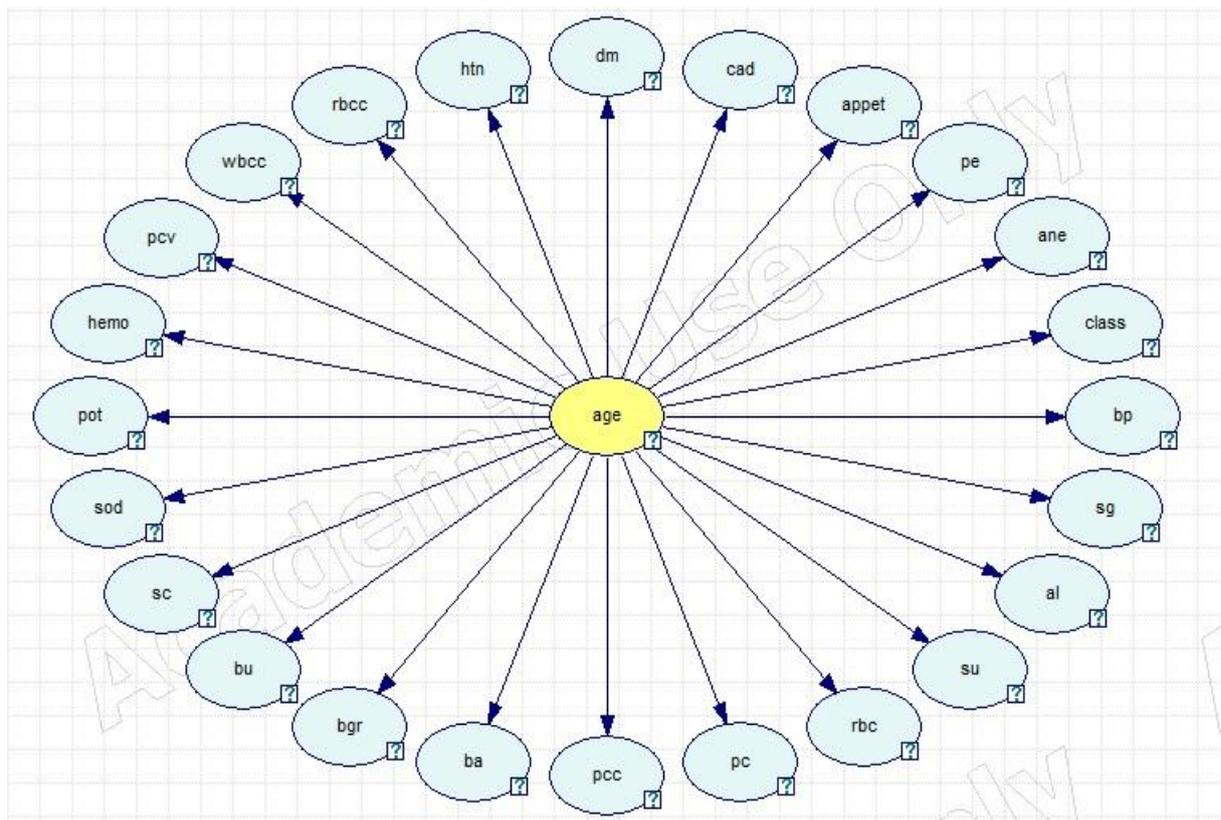


Figure 9: Naive Bayes structure

**b. Augmented Naive Bayes**

L'algorithme d'apprentissage de structure de Augmented Naive Bayes (ANB) est une méthode d'apprentissage de structure semi-naïve basée sur l'approche de la recherche bayésienne, décrite et soigneusement évaluée dans [46]. L'algorithme ABN commence par une structure Naive Bayes (c'est-à-dire, dans laquelle la variable de classe est le seul parent de toutes les variables d'entité restantes) et ajoute des connexions entre les variables d'entité pour

conditionner la dépendance de classe. Il n'y a pas de limite au nombre de connexions supplémentaires entrant dans chaque variable d'entité, sauf si elle est imposée par l'un des paramètres de l'algorithme (Max Parent Count). Veuillez noter que la structure de Naive Bayes suppose que les caractéristiques sont indépendantes de la variable de classe, ce qui conduit à des inexactitudes lorsqu'elles ne sont pas indépendantes. L'algorithme ANB est simple et a été trouvé pour fonctionner de façon fiable mieux que Naive Baye.

### b.1. Augmented Naive Bayes structure

Exemple de structure après que nous avons appliqué Augmented Naive Bayes algorithm sur la base de donnée de l'insuffisance rénale.

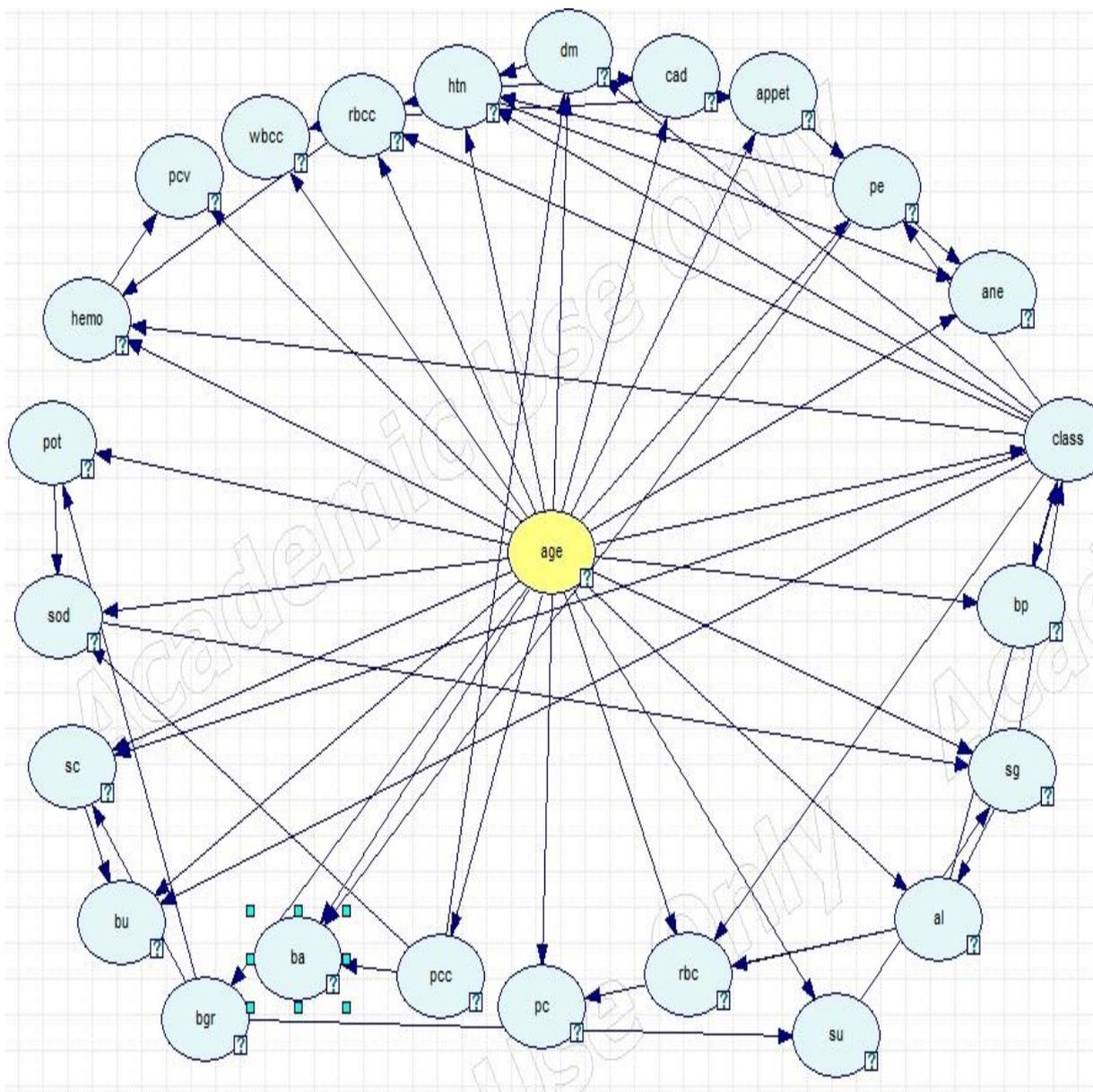


Figure 10: Augmented Naive Bayes structure

### c. Tree Augmented Naive Bayes

Cet algorithme d'apprentissage crée une structure de graphe Tree Augmented Naive Bayes (TAN) dans laquelle une variable de classe unique n'a pas de parents et toutes les autres variables ont la classe en tant que parent et au plus un autre attribut en tant que parent.

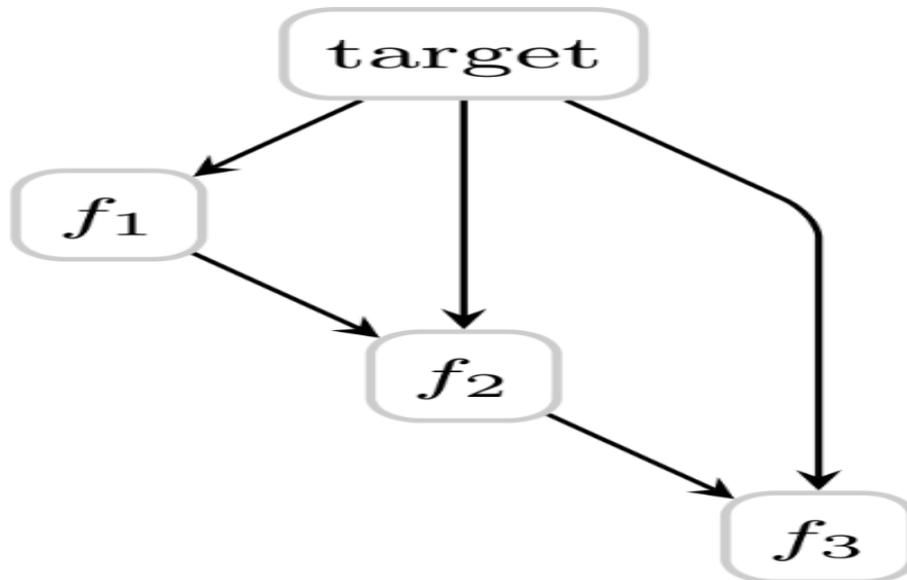


Figure 11: Tree Augmented Naive Bayes

L'algorithme d'apprentissage de structure TAN (Tree Augmented Naive Bayes) est une méthode d'apprentissage de structure semi-naïve basée sur l'approche bayésienne de la recherche. L'algorithme TAN commence par une structure Naive Bayes (c'est-à-dire, dans laquelle la variable de classe est le seul parent de toutes les variables d'entité restantes) et ajoute des connexions entre les variables d'entité. L'algorithme impose la limite d'un seul parent supplémentaire de chaque variable d'entité (en plus de la variable de classe, qui est un parent de chaque variable d'entité). Veuillez noter que la structure de Naive Bayes suppose que les caractéristiques sont indépendantes de la variable de classe, ce qui conduit à des inexactitudes lorsqu'elles ne sont pas indépendantes. L'algorithme TAN est simple et a été trouvé pour fonctionner de façon fiable mieux que Naive Bayes.

#### c.1. Tree Augmented Naive Bayes structure

Exemple de structure après que nous avons utilisés Augmented Naive Bayes algorithme sur la base de donnée de l'insuffisance rénale.

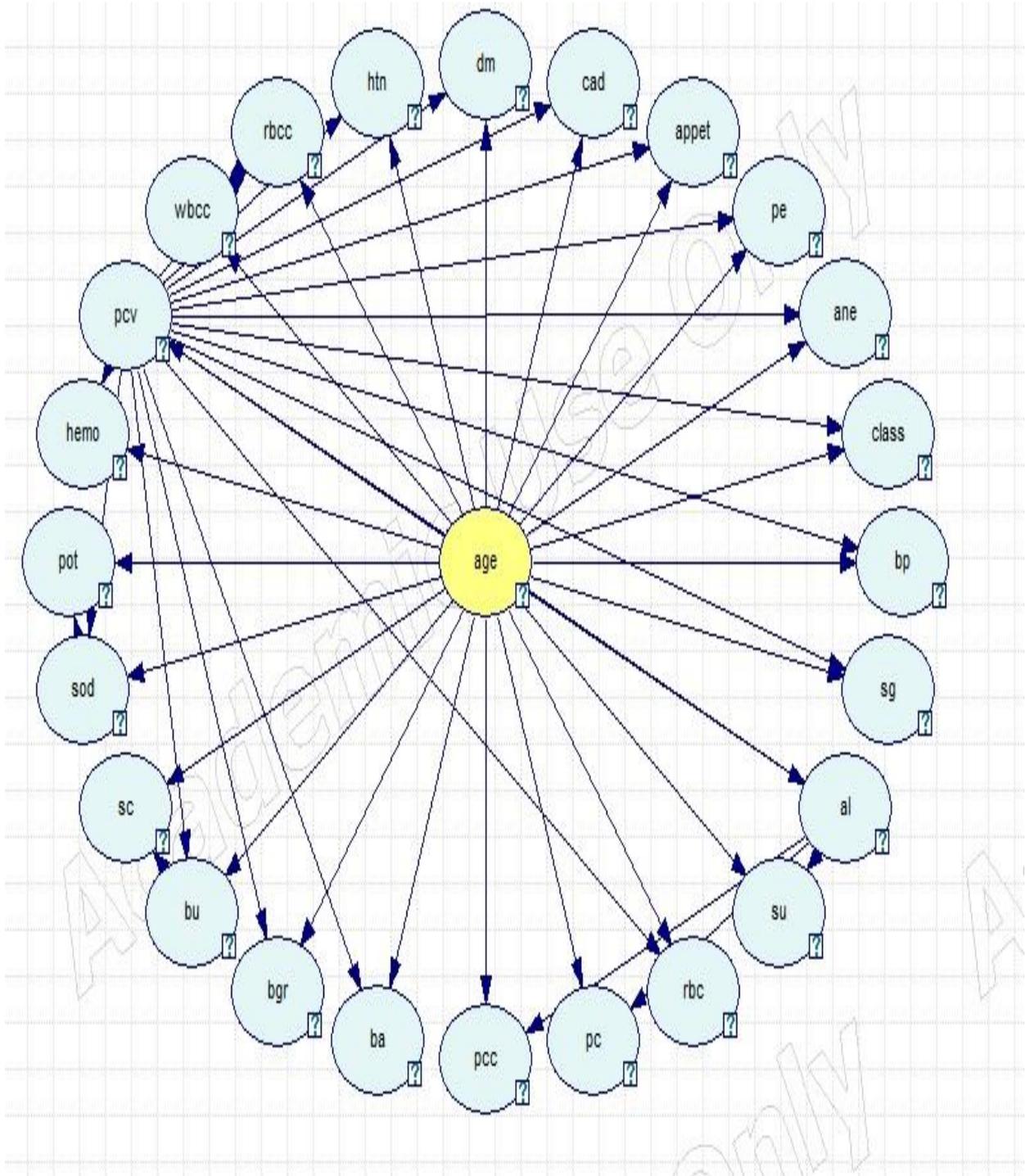
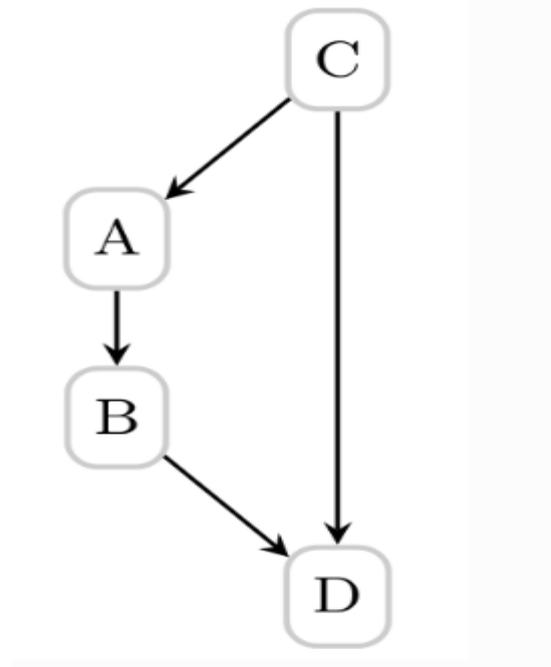


Figure 12: Tree Augmented Naive Bayes structure

**d. PC**

L'algorithme d'apprentissage de structure de PC est l'un des algorithmes les plus anciens et les plus populaires, introduit par [47]. Cet algorithme d'apprentissage utilise la procédure PC. C'est un algorithme d'apprentissage de structure de graphe à usage général, ce qui signifie qu'il tentera de rechercher l'intégralité de l'espace des graphes pour le meilleur graphe.



**Figure 13: PC**

- Commencez avec un graphique non orienté complet sur toutes les  $n$  variables, avec des arêtes entre tous les nœuds.
- Pour chaque paire de variables  $X$  et  $Y$ , et chaque ensemble d'autres variables  $S$ , voir si  $X$  et  $Y$  sont conditionnellement indépendants étant donné  $S$ . Si oui, enlever le bord entre  $X$  et  $Y$ .
- Trouvez des collisionneurs en vérifiant la dépendance conditionnelle; orienter les bords des collisionneurs.
- Essayez d'orienter les bords non orientés par cohérence avec les bords déjà orientés; faites ceci de manière récursive jusqu'à ce que plus aucun bord ne puisse être orienté.

### **d.1. PC structure**

L'algorithme PC est le seul algorithme d'apprentissage de structure permettant des données continues. Les données doivent répondre raisonnablement à l'hypothèse qu'elles proviennent d'une distribution normale multivariée. Pour vérifier cette hypothèse, vérifiez que les histogrammes de chaque variable sont proches de la distribution normale et que les diagrammes de dispersion de chaque paire de variables montrent des relations approximativement linéaires. Voortman et Druzdzel [48], ont vérifié expérimentalement que l'algorithme PC est assez robuste à l'hypothèse de normalité multi-variée.

- Exemple de structure après que nous avons essayés PC algorithme sur la base de donnée de l'insuffisance rénale.

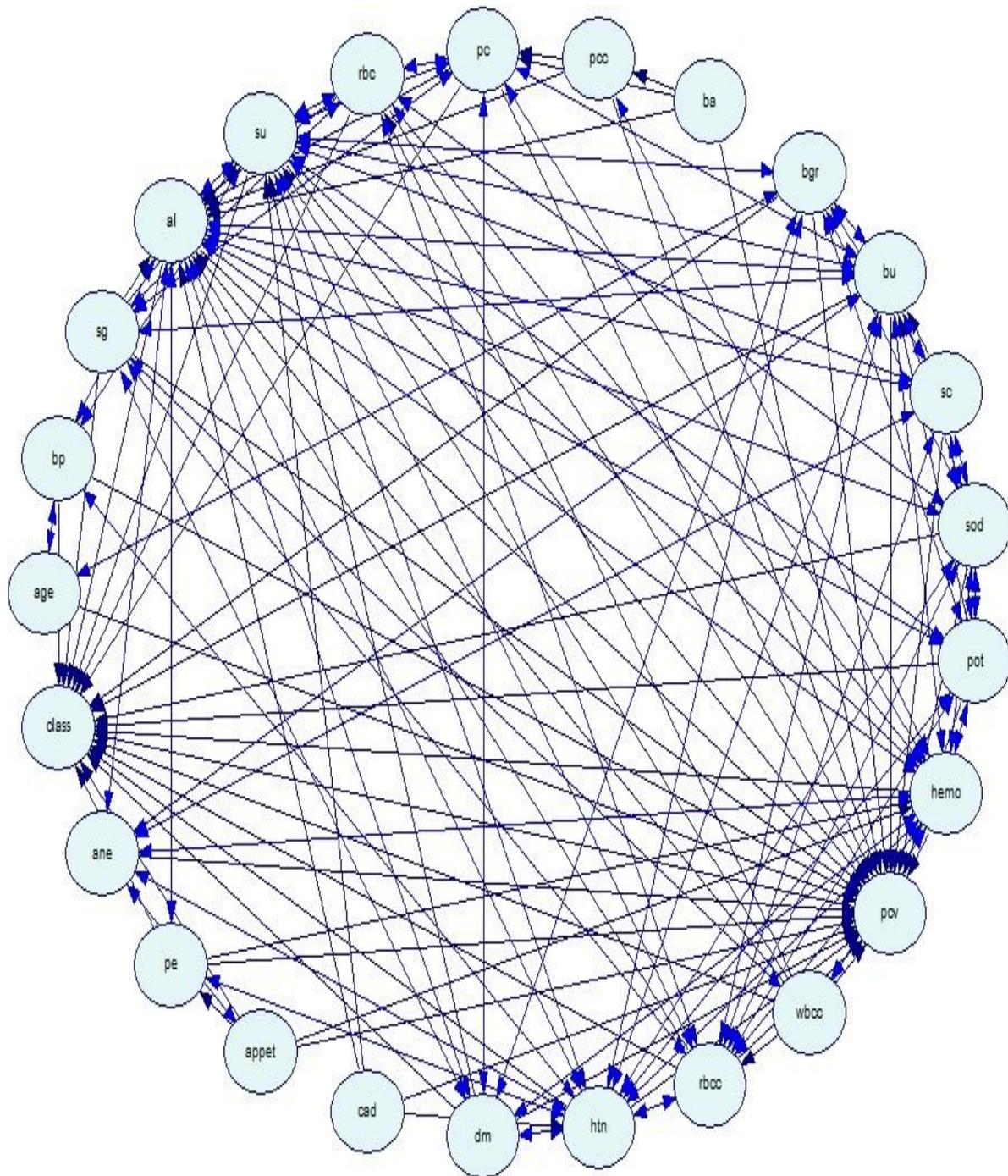


Figure 14: PC structure

### e. Bayesian Search

Cet algorithme d'apprentissage utilise la procédure de recherche bayésienne. C'est un algorithme d'apprentissage de structure de graphe à usage général, ce qui signifie qu'il tentera de rechercher l'intégralité de l'espace des graphes pour le meilleur graphe.

- L'algorithme exécute une recherche de graphe partiellement orientée sur les classes d'équivalence de Markov au lieu de chercher directement dans l'espace des DAG (qui est sur-exponentiel dans  $n$ ).

### e.1. Bayesian Search structure

Exemple de structure après que nous avons essayés Bayesian Search algorithm sur la base de donnée de l'insuffisance rénale.

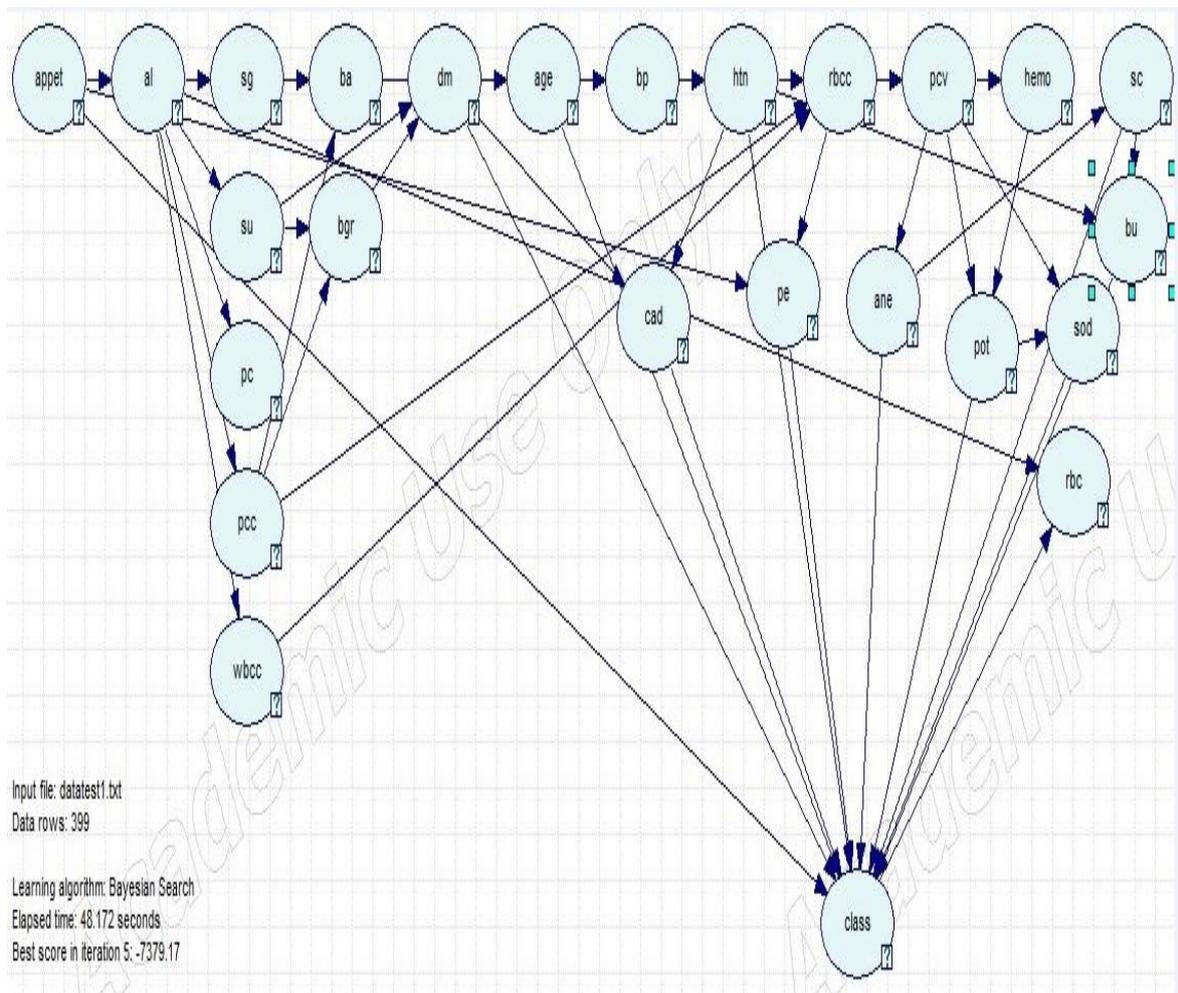


Figure 15: Bayesian Search structure

### f. Greedy Thick Thinning (notre choix)

Cet algorithme d'apprentissage utilise la procédure Greedy Thick Thinning. C'est un algorithme d'apprentissage de structure de graphe à usage général, ce qui signifie qu'il tentera de rechercher l'intégralité de l'espace des graphes pour le meilleur graphe.

- Greedy Thick Thinning [51], qui optimise une structure existante en modifiant la structure et en notant le résultat, a été réalisée. En partant d'un Directed Acyclic Graph (DAG) entièrement connecté et en supprimant ensuite les arcs entre les nœuds en fonction des tests d'indépendance conditionnelle [52], l'algorithme Greedy Thick Thinning est capable d'isoler le meilleur réseau de notation. L'une des fonctions de notation les plus courantes est la métrique bayésienne [53], qui mesure la probabilité d'observer les données en fonction de la

structure du réseau, c'est-à-dire que le meilleur réseau en termes de métrique bayésienne basé sur les données reçu.

### f.1. Algorithme Greedy Thick Thinning

L'algorithme d'apprentissage de la structure Greedy Thick Thinning (GTT) qui s'est basé sur l'approche Bayesian Search et a été décrit dans [49]. GTT commence par un graphe vide et ajoute à plusieurs reprises l'arc (sans créer de cycle) qui augmente au maximum la vraisemblance marginale  $P(D | S)$  jusqu'à ce qu'aucune addition d'arc n'aboutisse à une augmentation positive (c'est la phase d'épaississement). Ensuite, il supprime à plusieurs reprises les arcs jusqu'à ce que la suppression de l'arc n'entraîne une augmentation positive de  $P(D | S)$  (c'est la phase d'amincissement).

### f.2. Greedy Thick Thinning structure

Notre structure après que nous avons appliqué Greedy Thick Thinning algorithme sur la base de donnée de l'insuffisance rénale.

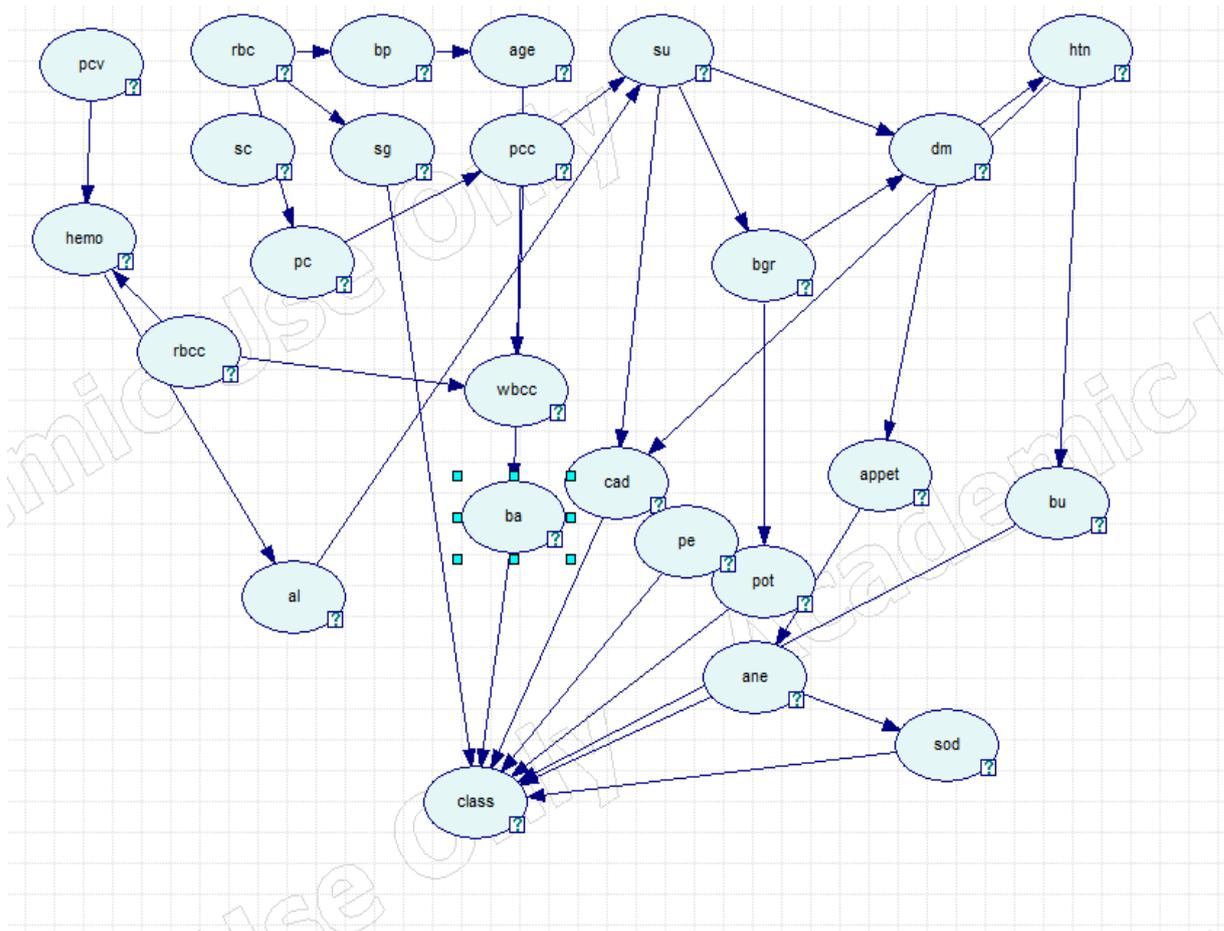


Figure 16: Greedy Thick Thinning structure

### 6.1.2.1. Discussion sur algorithmes d'apprentissage des structures

Les structures que nous avons réussi à créer grâce aux algorithmes d'apprentissage de structure ne se sont pas arrangées correctement, en d'autres termes, les relations entre les attributs ne sont pas forcément correctes après la consultation de l'expert médical.

- Les algorithmes Tree Augmented Naive Bayes, Augmented Naive Bayes et Naive Bayes, nous ne pouvons pas donner l'ordre initial à eux, alors si un nœud A influence sûrement sur un autre nœud B les algorithmes que nous avons cités ne **permettent pas de diriger A vers B**. Par contre l'algorithme de Greedy Thick Thinning, Bayesian Search et PC nous pouvons donner l'ordre initial à eux alors **nous pouvons diriger les nœuds entre eux**.

- Nous avons choisi l'algorithme Greedy Thick Thinning parce que nous pouvons donner l'ordre initial de la structure, et par conséquent la structure optimisée mieux adaptée que la structure Bayesian Search et la structure de PC selon l'expert du domaine. Cependant, la complexité de l'algorithme Greedy Thick Thinning que nous avons choisi d'utiliser est NP-complet et le temps d'estimation de la structure est long et indéterminable.

### 6.1.2.2. Structure finale

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

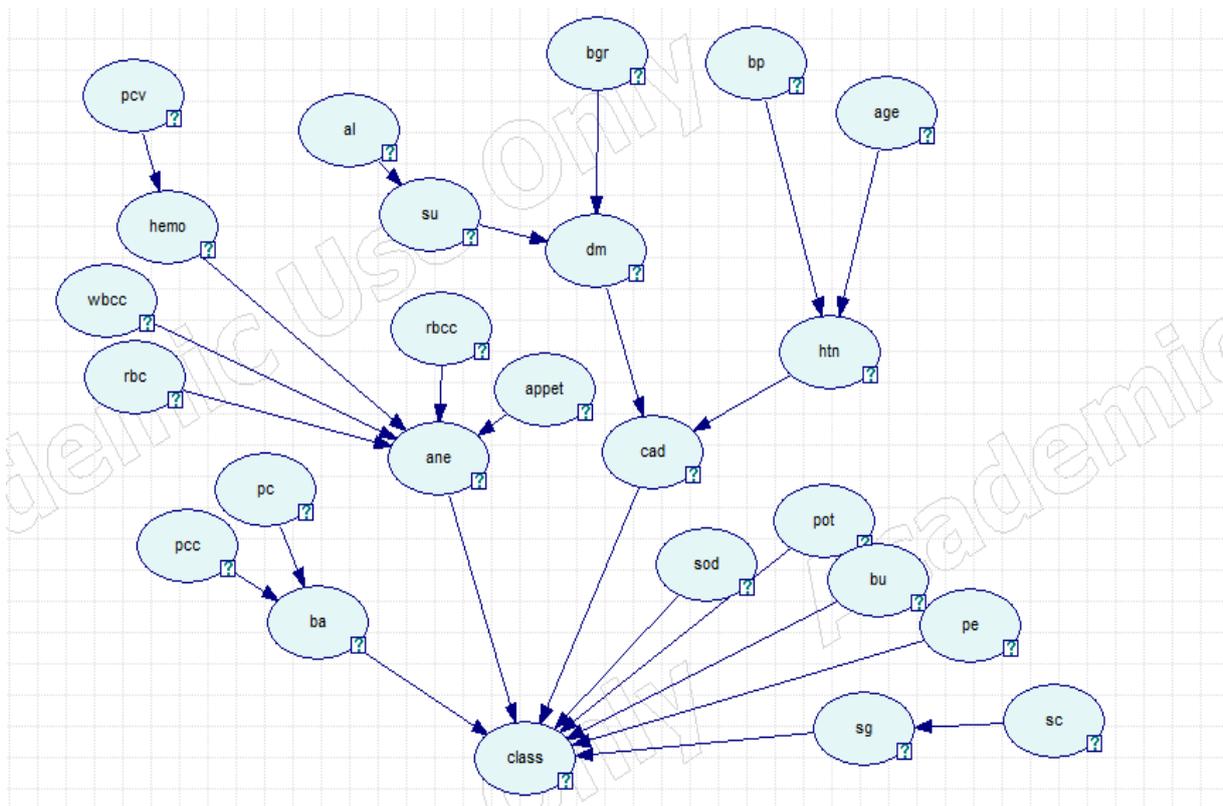


Figure 17: Structure après la modification de l'expert

## 6.2. L'apprentissage des paramètres

L'apprentissage des paramètres pour une structure particulière consiste à estimer les distributions de probabilités a priori ou les paramètres des lois de probabilités à partir des données disponibles. Dans le cas où l'on dispose de données complètes, on peut utiliser le maximum de vraisemblance qui utilise la fréquence d'apparition d'un événement dans les données. Au contraire, si les données sont incomplètes, l'algorithme Expectation-Maximisation peut être utilisé. Cet algorithme itératif part du modèle, infère le modèle pour calculer la distribution de probabilité. Puis, sur la base de cette distribution, l'algorithme construit un modèle meilleur. Le processus se répète jusqu'à obtenir le modèle le plus vraisemblable.

### 6.2.1. L'algorithme EM

L'apprentissage EM est une des techniques d'apprentissage des variables totalement "caché" (aussi connu comme «latentes»), qui sont, des variables pour lesquelles on n'a pas d'observations, mais on soupçonne qu'elle existe et peut être utile pour la modélisation du monde étudié. Ces exemples de données manquantes sont présentés au tableau.

	age	bp	sg	al	su	rbc	pc	pcc	ba	bgr	bu	sc	sod
▶	adult	Normal	D	S1	S0		normal	no	no	normal	normal	low	
	young	low	D	S4	S0		normal	no	no		normal	low	
	adult	Normal	B	S2	S3	normal	normal	no	no	high	high	low	
	adult	Normal	A	S4	S0	normal	abnormal	yes	no	normal	high	low	low
	adult	Normal	B	S2	S0	normal	normal	no	no	normal	normal	low	
	adult	high	C	S3	S0			no	no	normal	normal	low	normal
	adult	Normal	B	S0	S0		normal	no	no	normal	high	high	low
	adult		C	S2	S4	normal	abnormal	no	no	high	normal	low	
	adult	high	C	S3	S0	normal	abnormal	yes	no	normal	high	low	
	adult	high	D	S2	S0	abnormal	abnormal	yes	no	normal	high	normal	low
	adult	Normal	B	S2	S4		abnormal	yes	no	high	high	low	
	adult	Normal	B	S3	S0	abnormal	abnormal	yes	no	high	high	low	low
	adult	Normal	C	S3	S1		normal	yes	no	high	high	low	normal
	adult	Normal						no	no	normal	high	low	normal
	adult	Normal	B	S3	S2	normal	abnormal	yes	yes	high	high	low	low
	adult	Normal	C	S3	S0		normal	no	no	normal	high	normal	normal
	adult	Normal	C	S2	S0		normal	no	no	normal	normal	low	normal
	adult	Normal						no	no	normal	high	low	normal
	adult	high	E	S0	S3		normal	no	no	high	normal	low	normal

Tableau 6: Exemple des données incomplètes.

- L'algorithme EM est une méthode itérative efficace pour calculer l'estimation de vraisemblance maximale (MLE) en présence de données manquantes. Le but du MLE est d'estimer les paramètres du modèle pour lesquelles les données observées sont les plus

vraisemblables tout en tenant compte de l'existence de données manquantes (Figure 2 ,chapitre 1).

- Une fois qu'on a la structure du réseau bayésien faite par l'algorithme Greedy Thick Thinning et la base ses données, l'algorithme EM va s'occuper des calculs des paramètres de notre RB pour qu'elle soit prêt pour les inférences.

-La figure suivante montre les étapes de l'exécution de l'algorithme EM.

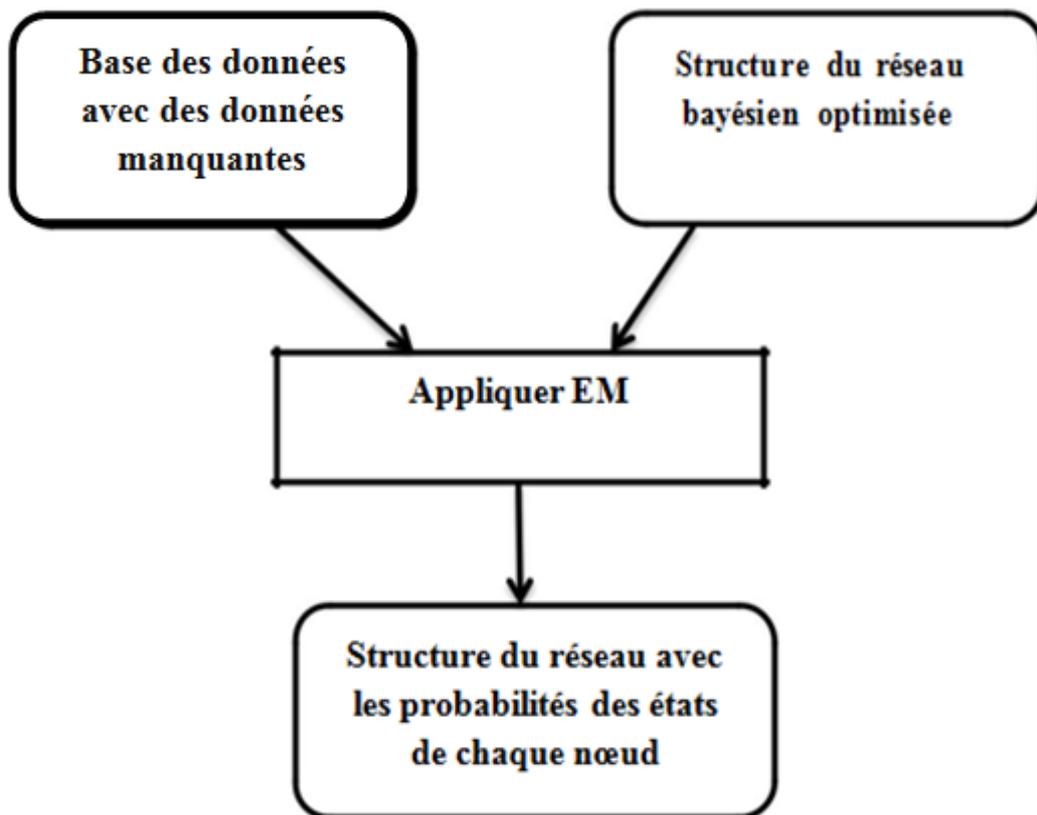


Figure 18: Etapes de l'exécution de l'algorithme EM

- Après avoir utilisé l'algorithme EM les résultats obtenu sont les suivant :

-Dans le cas où les nœuds sont sans parents nous avons le nœud white Blood cell count

▶	high	0.22395296
	low	0.013876428
	normal	0.76217061

Figure 19: Paramètres de White Blood cell count après EM

### Chapitre 3 : La conception du système

-Dans le cas où les nœuds ont un parent, nous avons le nœud hemoglobin son parent est packed cell volume

	pcv	low	high	normal
▶ high		0.0053907...	0.59062976	0.20472407
low		0.98896042	0.004464022	0.37872237
normal		0.0056487...	0.40490622	0.41655356

Figure 20: Paramètres de hemoglobin après EM

-Dans le cas où les nœuds ont deux parents, nous avons le nœud hypertension ses parents sont blood pressure et age

bp	high		low		Normal	
age	adult	young	adult	young	adult	young
▶ no	0.30084056	0.74984507	0.50150348	0.87753326	0.7052605	0.90945859
yes	0.69915944	0.25015493	0.49849652	0.12246674	0.2947395	0.090541412

Figure 21: Paramètres de hypertension après EM

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

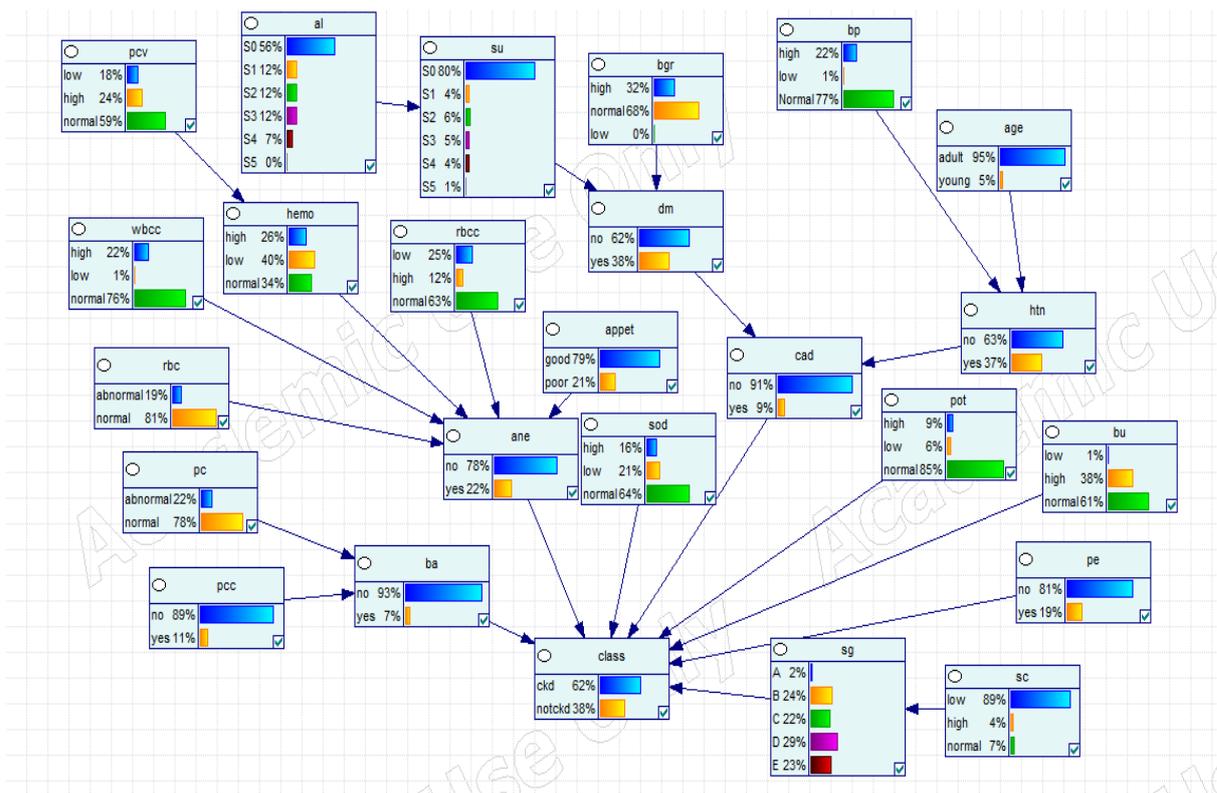


Figure 22: Structure après l'apprentissage des paramètres

## Chapitre 3 : La conception du système

L'apprentissage des paramètres va modifier le pourcentage des états des nœuds dans notre structure lors ce que nous appliquons l'algorithme EM.

### 7. La fuzzification

L'étape de fuzzification à pour but de transformer une donnée numérique en variable linguistique. Pour cela, le concepteur du système flou doit créer des fonctions d'appartenances. Une fonction d'appartenance est une fonction qui permet de calculer le degré d'appartenance d'une donnée numérique associé à une variable linguistique.

- Bien évidemment, on peut utiliser plusieurs variables linguistiques pour caractériser un seul type de données. Ici, nous choisirons trois variables linguistiques pour qualifier le degré des nœuds flous de la base de données : low, normal et high. Sauf age sons variables linguistiques est young et adulte. Pour cela, il faut créer une fonction d'appartenance pour chaque variable.

- Le tableau suivant montre les arguments et les type de chaque état qui appartienne à des nœuds flous que nous avons reçu a coté de l'expert :

Nœud flou	états	Fonction d'appartenance	
		type	Arguments
Age	young	Trapezoid	0, 0, 20, 25
		Triangular	0, 12.5, 25
	adulte	Trapezoid	20, 25, 100, 100
		Triangular	20, 25, 100
bp	low	Trapezoid	0, 0, 50 ,60
		Triangular	0, 30, 60
	normal	Trapezoid	50, 60, 80, 90
		Triangular	50, 70, 90
	high	Trapezoid	80, 90, 200, 200
		Triangular	80, 130, 200
bgr	low	Trapezoid	0, 0, 50, 60
		Triangular	0, 30, 60
	normal	Trapezoid	50, 60, 130, 150
		Triangular	50, 100, 150
	high	Trapezoid	130, 150, 700, 700
		Triangular	130, 150, 700
bu	low	Trapezoid	0, 0, 15, 20
		Triangular	0, 10, 20
	normal	Trapezoid	15, 20, 45, 55
		Triangular	15, 35, 55
	high	Trapezoid	45, 55, 500, 500
		Triangular	45, 55, 500
	low	Trapezoid	0, 0, 6, 7
		Triangular	0, 3.5 ,7

### Chapitre 3 : La conception du système

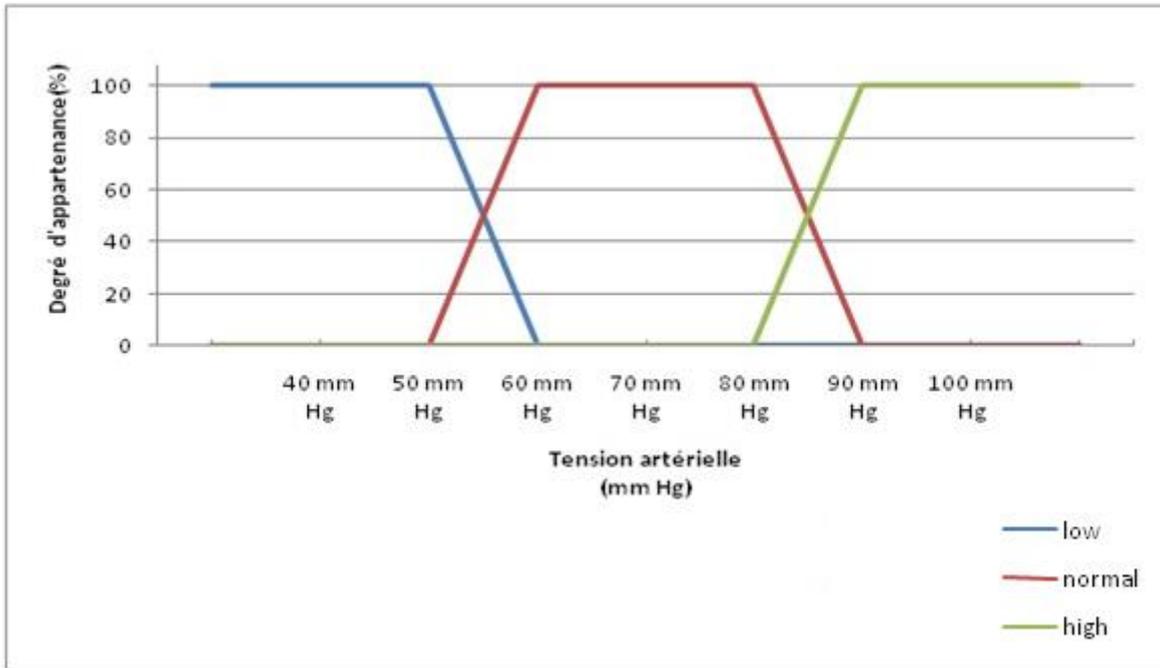
sc	normal	Trapezoid	6, 7, 12, 14
		Triangular	6, 10, 14
	high	Trapezoid	12, 14, 100, 100
		Triangular	12, 14, 100
sod	low	Trapezoid	0, 0, 125, 135
		Triangular	0, 65, 135
	normal	Trapezoid	125, 135, 145, 155
		Triangular	125, 140, 155
	high	Trapezoid	145, 155, 500, 500
		Triangular	145, 155, 500
pot	low	Trapezoid	0, 0, 3, 3.9
		Triangular	0, 2, 3.9
	normal	Trapezoid	3, 3.9, 5, 5.7
		Triangular	3, 4.3, 5.7
	high	Trapezoid	5, 5.7, 100, 100
		Triangular	5, 5.7, 100
hemo	low	Trapezoid	0, 0, 11, 11.5
		Triangular	0, 5.5, 11.5
	normal	Trapezoid	11, 11.5, 15, 15.5
		Triangular	11, 13.5, 15.5
	high	Trapezoid	15, 15.5, 100, 100
		Triangular	15, 16.5, 100
pcv	low	Trapezoid	0, 0, 20, 30
		Triangular	0, 15, 30
	normal	Trapezoid	20, 30, 45, 50
		Triangular	20, 35, 50
	high	Trapezoid	45, 50, 100, 100
		Triangular	45, 55, 100
wbcc	low	Trapezoid	0, 0, 3000, 4500
		Triangular	0, 2200, 4500
	normal	Trapezoid	3000, 4500, 9500, 10500
		Triangular	3000, 6500, 10500
	high	Trapezoid	9500,10500,100000,100000
		Triangular	9500, 11500, 100000
rbcc	low	Trapezoid	0, 0, 3, 4.5
		Triangular	0, 2.2, 4.5
	normal	Trapezoid	3, 4.5, 5.9, 6.3
		Triangular	3, 4.5, 6.3
	high	Trapezoid	5.9, 6.3, 100, 100
		Triangular	5.9, 6.5, 100

**Tableau 7: Fonction d'appartenance des nœuds flous**

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

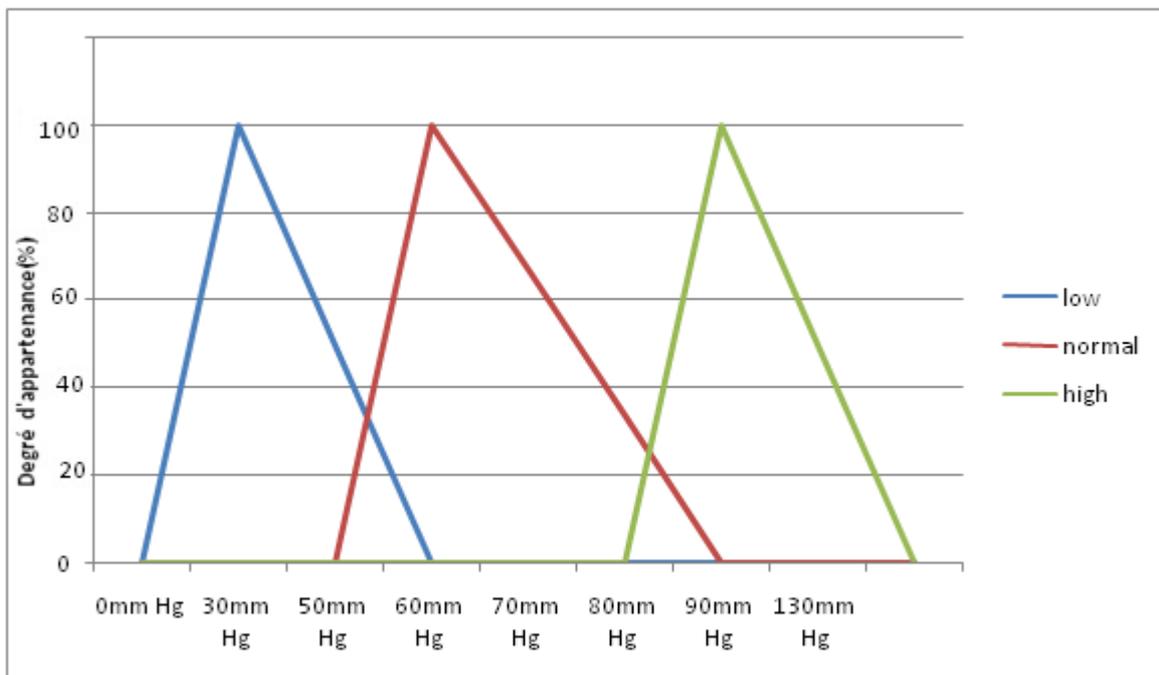
### Chapitre 3 : La conception du système

- Nous avons crée un graphe de fonctions d'appartenance de type trapezoid de Blood Pressure (tension artérielle) dans la figure suivant.



**Tableau 8: Fonction d'appartenance trapezoid de Blood Pressure**

- Nous avons crée un graphe de fonctions d'appartenance triangular de type de Blood Pressure (tension artérielle) dans la figure suivant



**Tableau 9: Fonction d'appartenance triangular de Blood Pressure**

Dans la fonction triangular, la somme de degré d'appartenance entre les états ne pas toujours 100%, pour résoudre ce problème nous faisons une normalisation qui consiste que la somme de degré d'appartenance entre les états que nous avons trouvé est 100% et nous recalculons le degré d'appartenance de chaque états

- Cette première étape de fuzzification va donc traduire les données numériques de BD en différentes variables linguistiques

### 8. Inférence

Dans notre système il existe deux types d'inférence, inférence classique et inférence flou :

- **Inférence classique** : Les valeurs que nous allons changer vers le nominale vont prendre un seul état seulement, si la valeur appartient a un intervalle, (normal) par exemple alors les états (low) et (high) deviendront carrément négligeables.

Nous citons un exemple de transformations nominale des valeurs pour l'inférence classique, nous avons comme exemple les valeurs de red blood cell count (rbcc) : 4.9, 6.1, 6.2 qui ils deviendraient en nominale comme des états que nous allons présenter dans le tableau suivant :

Valeurs de red blood cell count	Intervalle	Etats
4,9 millions/mm <sup>3</sup>	[4< 4,9 <6] millions/mm <sup>3</sup>	normal
6,1 millions/mm <sup>3</sup>	[6< 6,1 <8] millions/mm <sup>3</sup>	high
6,2 millions/mm <sup>3</sup>	[6< 6,2 <8] millions/mm <sup>3</sup>	high

**Tableau 10: Transformations nominale de rbcc pour l'inférence classique**

Après la transformation des valeurs des nœuds existants vers nominale nous calculons l'inférence classique.

- **Inférence flou** : Les valeurs que nous allons changer vers le nominale vont prendre un ou plusieurs états, si la valeur appartient a un intervalle d'état quelconque il peut appartenir a un autre état aussi avec certaine pourcentage, et la somme de pourcentage des états égala 100%

- Nous citons un exemple de transformations nominale des valeurs pour l'inférence flou, nous avons comme exemple les même valeurs de red blood cell count (rbcc) : 4.9, 6.1, 6.2 qui ils

deviendraient en nominale comme des états que nous allons présenter dans le tableau suivant :

Valeurs de red blood cell count	Etats
4,9 millions/mm <sup>3</sup>	low=0.0 medium=1.0 eleve=0.0
6,1 millions/mm <sup>3</sup>	low= 0.0 medium= 0.5000002384185793 eleve = 0.4999997615814207
6,2 millions/mm <sup>3</sup>	low = 0.0 medium = 0.2500004768371581 eleve = 0.7499995231628419

**Tableau 11: Transformations nominale de rbcc pour l'inférence flou**

- Après la transformation des valeurs des nœuds existants vers nominale à l'aide des fonctions d'appartenances nous calculons l'inférence floue avec Virtual Evidence.

### 9. Conclusion

Dans ce chapitre nous avons présenté notre architecture du système ainsi que tous les processus de construction de réseau bayésien flou, **en suivant eux passé par la création de RB et fuzzification pour arriver finalement à faire les prédictions de la maladie grâce a l'inférence flou.**

- Dans le chapitre suivant, nous passerons à la réalisation de notre système de prédiction qui doit aider les médecins à diagnostiquer l'insuffisance rénale.

# **Chapitre4 :**

# **Implémentation**

# **du système**

## 1. Introduction

Nous allons présenter dans ce chapitre, les outils de développement que nous avons utilisé pour la réalisation de notre application, cette dernière était programmée en Java, sous Netbeans 8.2 en utilisant la bibliothèque jSMILE.

- Implémenter c'est effectuer l'ensemble des opérations qui permettent de définir un projet et de réaliser, dans ce chapitre nous allons décrire de façon visuelle l'implémentation de notre système, en effectuant des captures d'écran des différentes interfaces du système.

## 2. Outils de développement de notre système

- **Java**

Java est un langage de programmation orienté objet créé par James Gosling et Patrick Naughton, employés de Sun Microsystems, avec le soutien de Bill Joy, présenté officiellement le 23 mai 1995 au SunWorld.

- La société Sun a été ensuite rachetée en 2009 par la société Oracle qui détient et maintient désormais Java.

- La particularité et l'objectif central de Java est que les logiciels écrits dans ce langage doivent être très facilement portables sur plusieurs systèmes d'exploitation tels que Unix, Windows, Mac OS ou GNU/Linux, avec peu ou pas de modifications, mais qui ont l'inconvénient d'être plus lourd à l'exécution (en mémoire et en temps processeur) à cause de sa machine virtuelle. Pour cela, divers plateformes et frameworks associés visent à guider, sinon garantir, cette portabilité des applications développées en Java.



- **NetBeans**

Est un environnement de Développement intégré (EDI), placé en open Source par Sun en juin 2000 sous licence CDDL (Common Development and Distribution License) et GPLv2. En plus de Java, NetBeans permet également de supporter différents autres langages, comme C, C++, JavaScript, XML, Groovy, PHP et HTML de façon native ainsi que bien d'autres (comme Python ou Ruby) par l'ajout de greffons. Il comprend toutes les caractéristique d'un



IDE moderne (éditeur en couleur, projets multi-langage, refactoring, éditeur graphique d'interfaces et de pages Web). Conçu en Java, NetBeans est disponible sous Windows, Linux, Solaris (sur x86 et SPARC), Mac OS X ou sous une version indépendante des systèmes d'exploitation (requérant une machine virtuelle Java). Un environnement Java Développment Kit JDK est requis pour les développements en Java. NetBeans constitue par ailleurs une plateforme qui permet le développement d'applications spécifiques (bibliothèque Swing (Java)). L'IDE NetBeans s'appuie sur cette plateforme.

- **SMILE**

Est l'interface du programmeur d'application de GeNIe (API) et pratiquement toutes les opérations élémentaires effectuées avec GeNIe se traduisent par des appels aux méthodes SMILE.

SMILE est une bibliothèque logicielle C++ qui implémente des modèles graphiques probabilistes et décisionnels, tels que les réseaux bayésiens, les diagrammes d'influence et les modèles d'équation structurelle. Ses classes individuelles, définies dans l'API SMILE (Application Programming Interface), vous permettent de créer, éditer, sauvegarder et charger des modèles graphiques, et de les utiliser pour le raisonnement probabiliste et la prise de décision dans l'incertitude. SMILE supporte directement la méthodologie orientée objet.

-SMILE peut être intégré dans des programmes qui utilisent des modèles probabilistes graphiques comme moteurs de raisonnement. Les modèles développés en GeNIe peuvent être équipés d'une interface utilisateur qui utilise SMILE comme moteur backend. SMILE est publié sous la forme d'une bibliothèque de liens dynamiques (DLL). Il existe également plusieurs enveloppes SMILE, comme SMILE.NET (.NET interface), SMILEX (Active X), jSMILE (Java interface), etc.

- **jSMILE**

jSMILE est compatible avec JDK 7 et les versions plus récentes.

jSMILE contient deux parties : le fichier jar et la bibliothèque native, cette dernière est spécifique à votre système d'exploitation. Le fichier jar, nommé jsmile-x.y.z.jar où x.y.z est le numéro de version, peut être utilisé comme n'importe quel jar Java. Vous pouvez l'ajouter à votre dépôt Maven local, par exemple.

- **GeNIe**

GeNIe Modeler est une interface utilisateur graphique (GUI) pour SMILE Engine et permet la construction de modèles interactifs et l'apprentissage. Il est écrit pour l'environnement Windows mais peut également être utilisé sous Mac OS et Linux sous les émulateurs



Windows. Il a été testé de manière approfondie sur le terrain depuis 1998, a été largement accepté par les universités et l'industrie et compte des milliers d'utilisateurs dans le monde entier. Un critère de conception important pour tous les produits a été dès le début qu'ils devraient permettre une liberté de modélisation complète. Les modèles GeNIe ne plient pas la réalité aux outils de modélisation disponibles. Tout ce que le domaine exige, peut être modélisé dans notre logiciel. Parce qu'il n'existe pas d'algorithmes exacts pour certains types de modèles, ce logiciel est équipé d'une suite d'algorithmes d'échantillonnage stochastique approximatifs, capables de résoudre tous les modèles créés par les utilisateurs.

### 3. La mise en œuvre du système

Après l'effectuation des étapes de la création de réseau bayésien flou, nous avons eu un fichier avec une extension xdsl, qui porte la structure et paramètres de RB. Nous donnons un exemple d'un patient probablement touché par l'insuffisance rénale par remplir toutes Les états du paramètre dans l'interface de test sauf le nœud « classe » qui est un nœud qui sert à faire la décision.

## 4. Les interfaces

### 4.1. L'interface d'accueil

La première interface à afficher est la suivante :

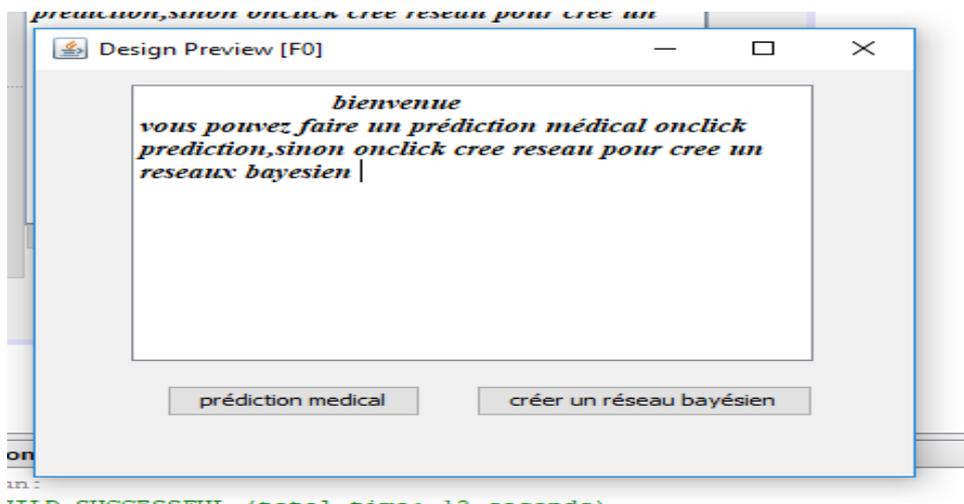


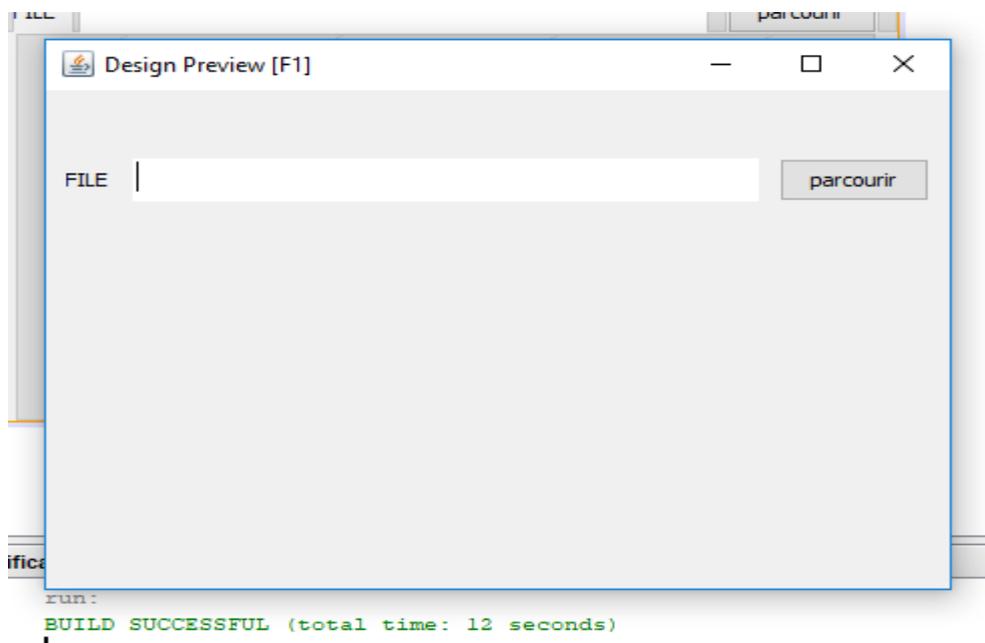
Figure 23: Interface d'accueil

- Cette interface nous permet de créer un réseau bayésien ou bien faire une prédiction de la maladie.

- **Créer un réseau bayésien** : va nous conduire vers la création de réseau bayésien ce parti conservé à l'expert.

- **prédiction médicale** : permet l'utilisateur à faire les prédictions possible sur la maladie est généralement pour l'utilisateur.

- Après un clic sur le bouton « **créer un réseau bayésien** » qui est dans l'interface précédente, cette interface va s'afficher. Puisque notre système est générique, on doit commencer par le choix du fichier csv qui contient la base de données et qui correspond notre cas d'étude.



**Figure 24: Importer une base de donnée**

- **Parcourir** : nous permet de ramener notre base de données nominale d'une maladie quelconque à partir de notre espace de stockage et l'ouvrir.

#### 4.2. Interface de l'ordre de causalité

Cette interface nous permet de donner l'ordre de causalité aux nœuds de base de données.

- Les zones de sélection à gauche nous permettent d'influencer les zones qui sont à droite, au début cela va nous donner 5 nœuds qui influencent les autres 5 nœuds, si cela ne suffit pas, on clique sur le bouton add.

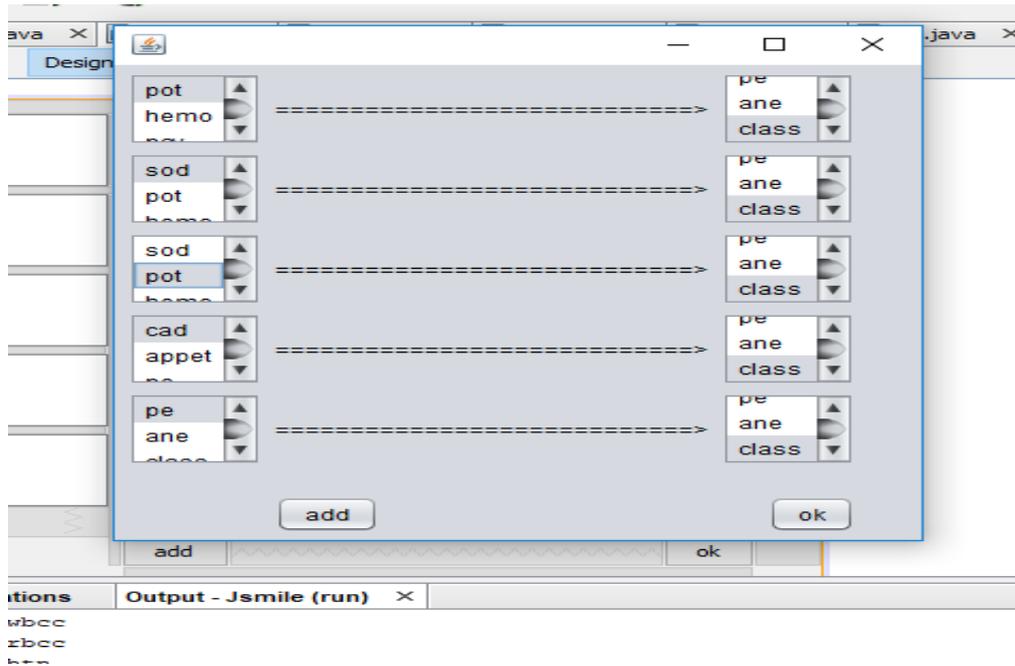
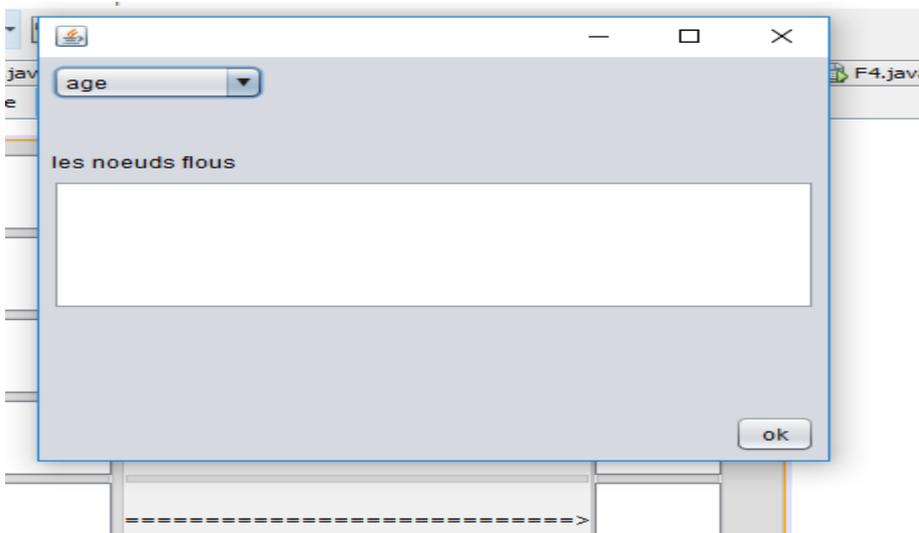


Figure 25: Interface de l'ordre de causalité

- **Add** : va nous agrandir l'interface que le contenu devient 12 nœuds qui influence autres 12 nœuds, et si nous cliquons sur add de nouveau, ca va devenir 24 qui influencent sur des autres 24.
- **Ok** : si nous cliquons sur le bouton ok , nous allons confirmer l'ordre de causalité que nous avons sélectionné, et va nous diriger vers une autre interface .

### 4.3. Interface pour choisi des nœuds flous

Cette interface nous permet de sélectionner les nœuds flous.



### Figure 26: Interface pour sélectionner des nœuds flous

- Quand nous cliquons sur la flèche une liste de tous les nœuds va sortir, après nous allons sélectionner les nœuds flous que nous voulons, les nœuds que nous avons sélectionné vont apparaître dans le champ qui est à gauche.

- **ok** : va nous valider les nœuds flous que nous avons sélectionné et nous dirige vers l'interface de confirmation.

- Dans le cas ou nous ne sélectionnons aucun nœud et les champs des nœuds flous restent vide, et nous cliquons sur ok, nous allons avoir un réseau bayésien classique.

#### 4.4. Interface de création d'un structure

Nous allons faire entrer le nom du réseau bayésien que nous voulons créer dans le champ ou le nom de notre réseau est écrit.

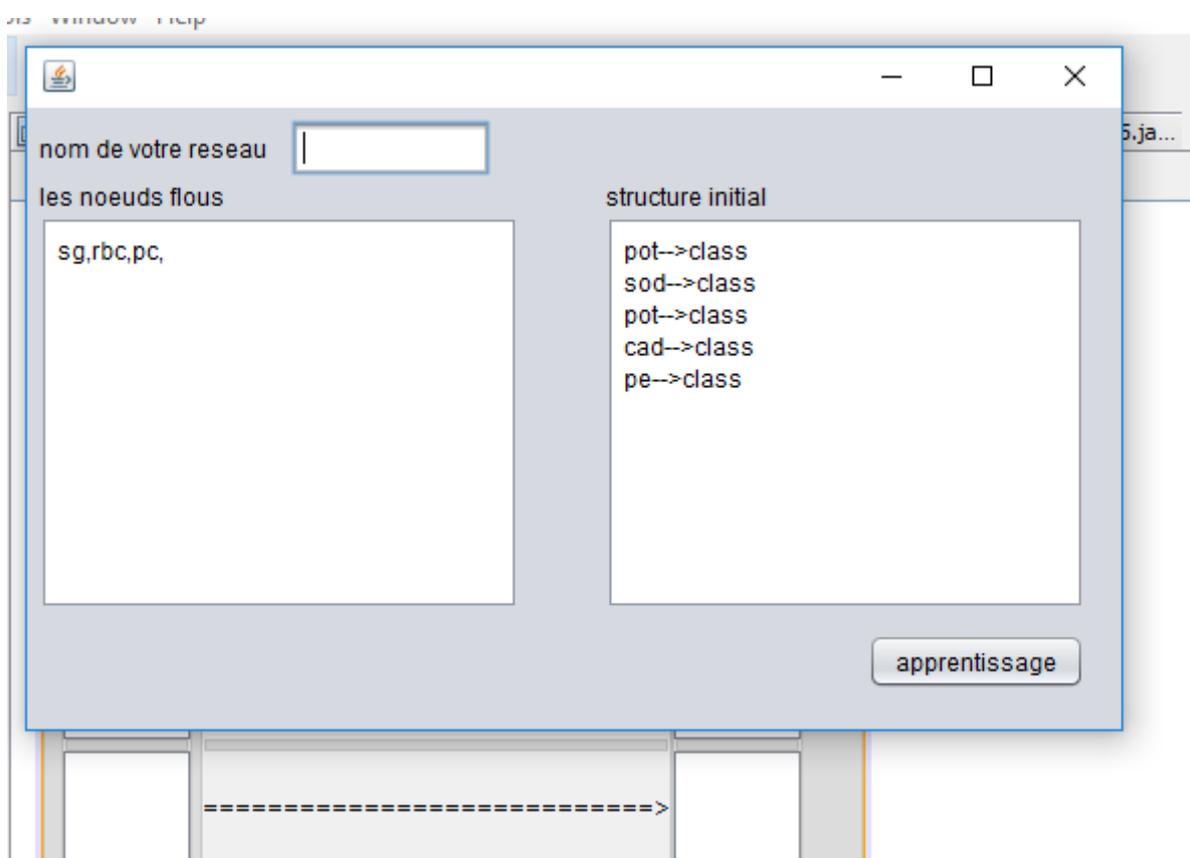


Figure 27: Interface de confirmation et création d'une structure

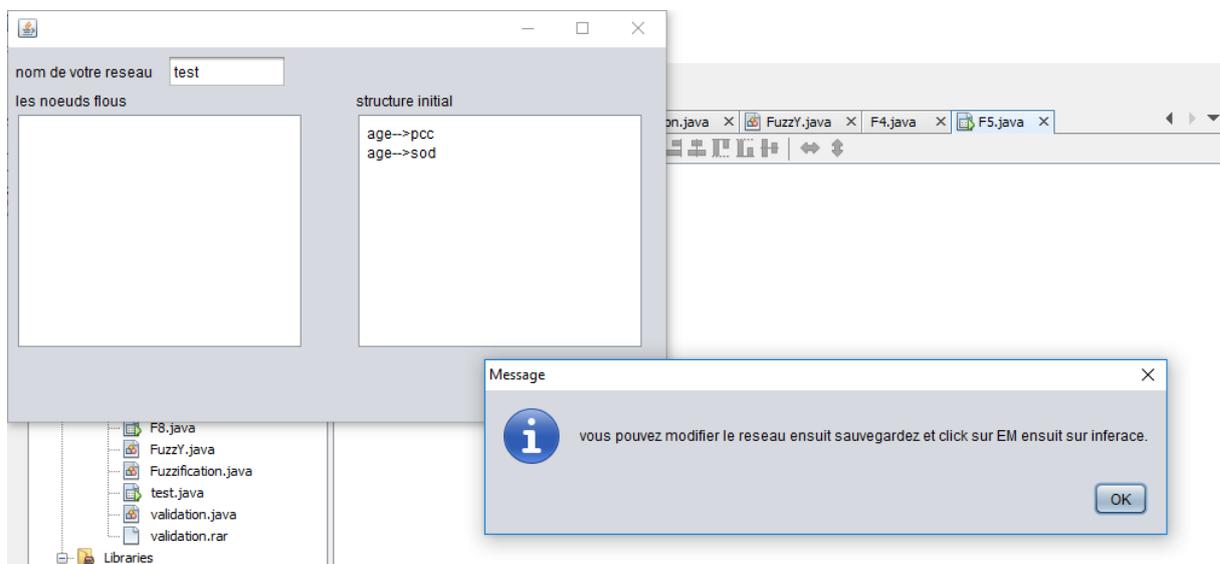
- Cette interface nous permet de confirmer nos nœuds flous et l'ordre de causalité.

- Le champ qui est à gauche « les nœuds flous », sont les nœuds flous que nous avons sélectionné au paravent, nous avons la possibilité de supprimer ou ajouter les nœuds flous en écrivant leur noms.

- Le champ qui est à droite « structure initial » qui est l'ordre de causalité qu'on a créé au paravent, il est possible de supprimer l'ordre de causalité lorsque nous effaçons ce nom du tableau, ou l'ajouter en écrivant sur le champ en respectant le format qui est apparu au dessus de ce paragraphe (les nœuds a gauche influence les nœuds qui sont an droite).

• **Apprentissage** : ce bouton nous permet de faire l'apprentissage de la structure et a qui nous respectant l'ordre de causalité que nous lui a donné.

- Après que nous avons cliqué sure « apprentissage » un message de rappel va apparaître (vous pouvez modifier le réseau ensuite sauvegarder et click sur EM ensuite sur inférence).



**Figure 28: Message de rappel**

- Après nous cliquons sur « ok » un fichier xsdl format va être créé et notre structure va apparaître dans le genie\_academic

#### 4.5. Interface de structure réseau bayésien

La figure suivant nous fais apparaître notre structure dans le genie\_academic, nous pouvons modifier les causalités entre les nœuds dans genie\_academic selon l'expert (ajouter ou supprimer entre les arcs).

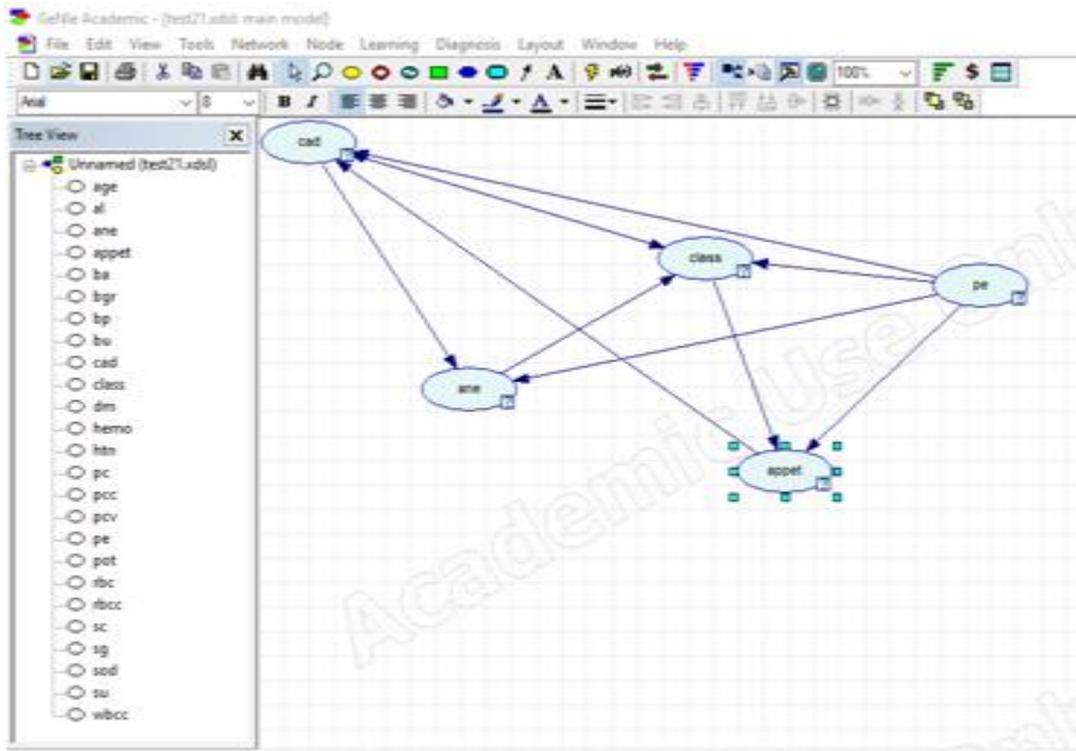


Figure 29: Structure du réseau bayésien

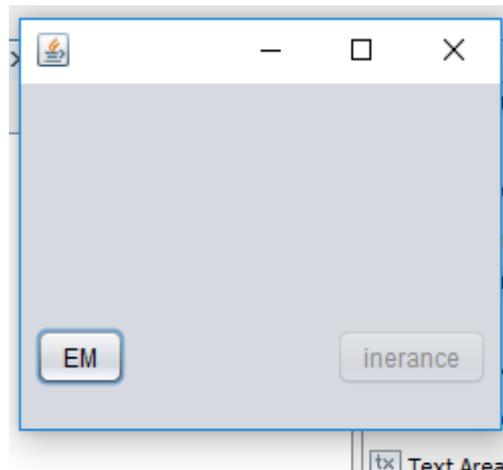
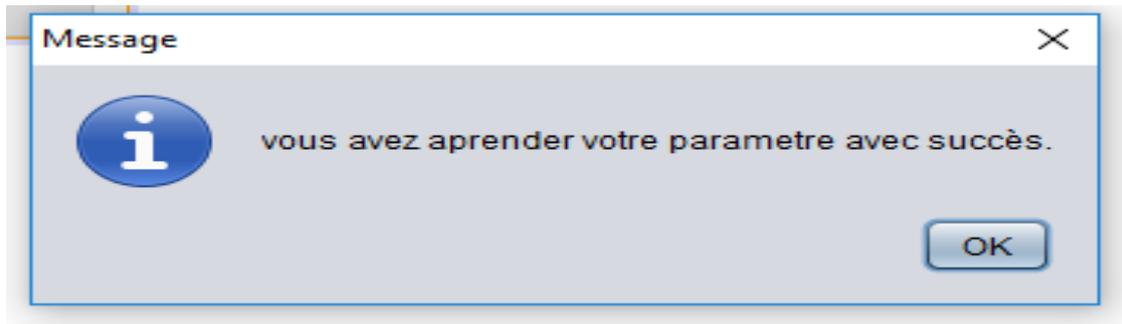


Figure 30: Interface pour l'apprentissage des paramètres

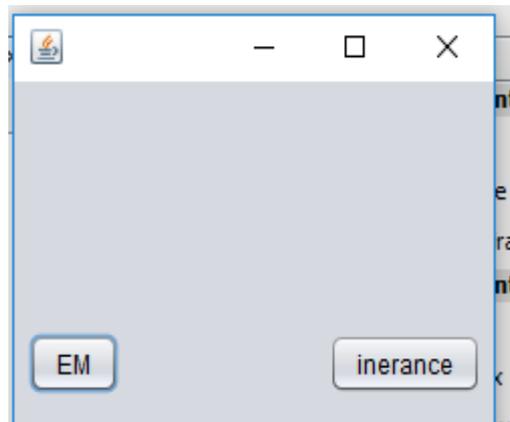
• **EM** : le bouton « EM » permet de faire l'apprentissage des paramètres de la structure que nous avons créé. Après avoir appliqué l'algorithme EM, nous arriverons à calculer l'inférence floue ou bien classique, **mais nous ne pouvons pas le calculer avant.**

- Après que nous cliquons sur «EM » un message de confirmation va apparaître



**Figure 31: Message de confirmation d'apprentissage de paramètre**

- Après que nous cliquons sur « ok » nous pouvons faire l'inférence



**Figure 32: Interface de calculassions d'inférence**

•**Inférence** : quand nous cliquons sur « inférence », si il existe un fichier Rules.text et la base de données non transformer dans le même endroit de stockage de notre réseau alors l'inférence va être effectué, en cas ou l'un de ces fichier n'existe pas il faut le mettre dans le même endroit de stockage de notre réseau pour pouvoir faire l'inférence.

- Le fichier Rules.text contient les noms des nœuds flous dans la première ligne, dans la deuxième nous avons le nombre d'états de chaque nœud, la ligne suivante contient le nom de fonction d'appartenance, les valeurs qui reste son des arguments de chaque nœud comme les figures suivante :

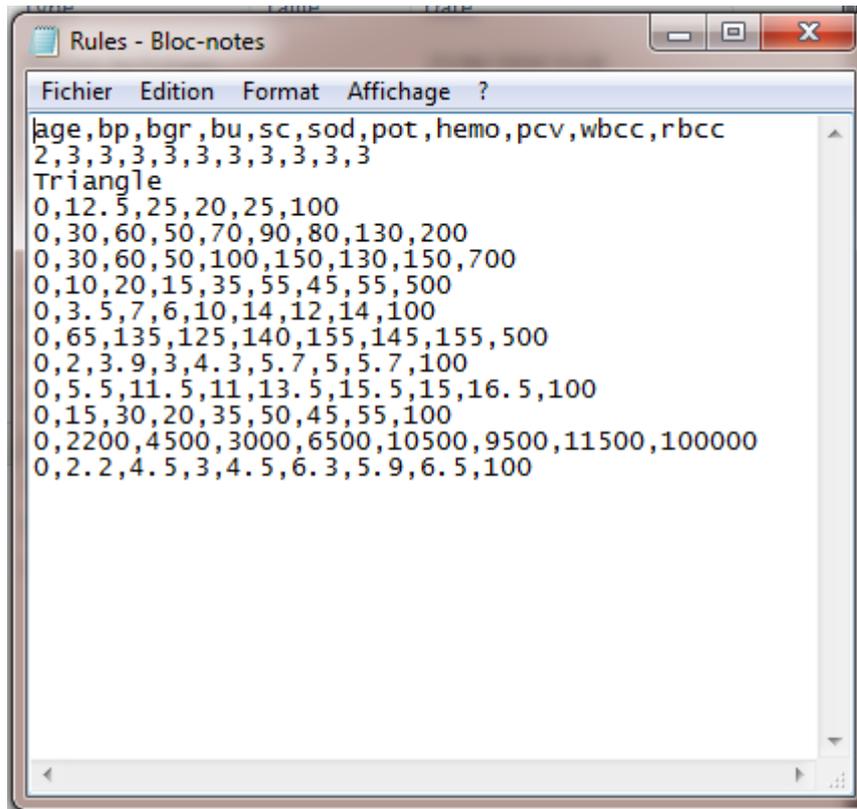


Figure 33: Fichier Rules.texte triangulaire

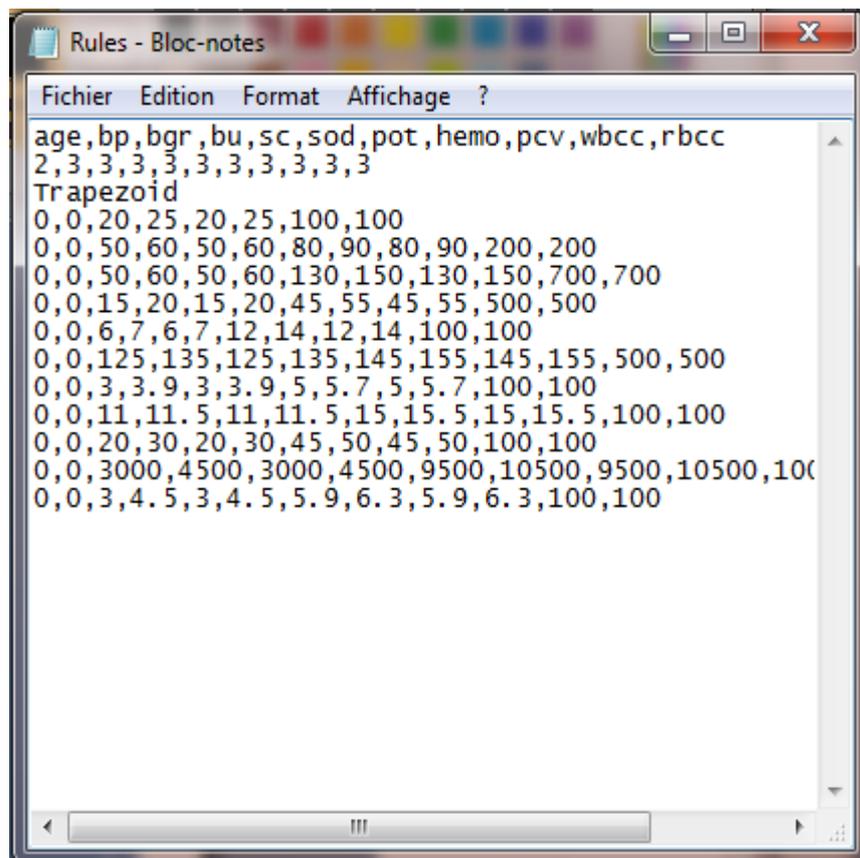


Figure 34: Fichier Rules.texte trapèze

Après l'inférence un message de confirmation va apparaître que notre réseau a été créé avec succès

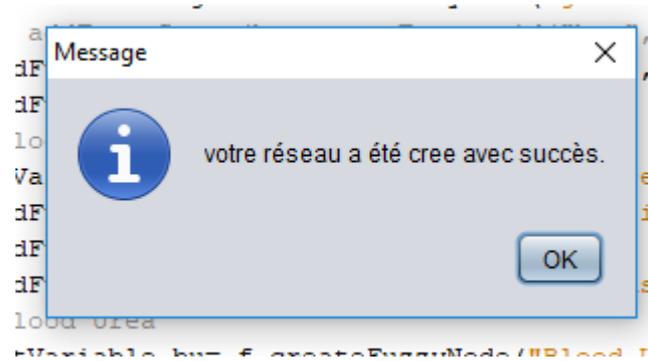


Figure 35: Message de confirmation de création d'un réseau

#### 4.5. Interface pour choisi un réseau bayésien pour la prédiction

Quand nous cliquons sur « prédiction » qui est dans l'interface d'accueil, une interface va apparaître dans lequel il ya un champ ou on peu choisir le RB

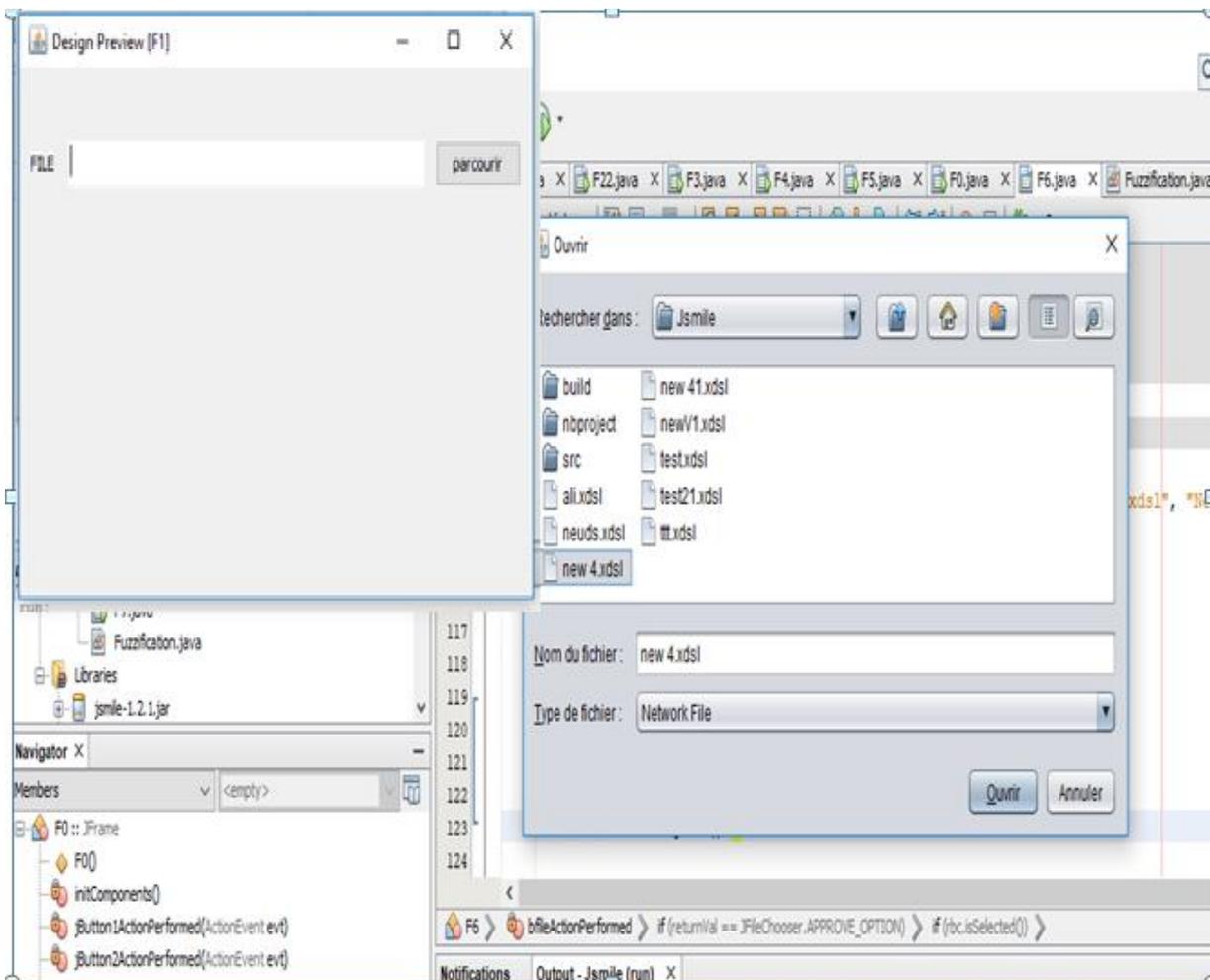


Figure 36: Interface pour importer un réseau bayésien

- **Parcourir** : nous permet de ramener un réseau nominal d'une maladie quelconque à partir de notre espace de stockage et l'ouvrir.

#### 4.6. Interface de test

Age	<input type="text"/>	Sodium	<input type="text"/>
Blood Pressure	<input type="text"/>	Potassium	<input type="text"/>
Specific Gravity	A <input type="text"/>	Hemoglobin	<input type="text"/>
Albumin	S0 <input type="text"/>	Packed Cell Volume	<input type="text"/>
Sugar	S0 <input type="text"/>	White Blood Cell Count	<input type="text"/>
Red Blood Cells	normal <input type="text"/>	Red Blood Cell Count	<input type="text"/>
Pus Cell	normal <input type="text"/>	Hypertension	yes <input type="text"/>
Pus Cell clumps	yes <input type="text"/>	Diabetes Mellitus	yes <input type="text"/>
Bacteria	yes <input type="text"/>	Coronary Artery Disease	yes <input type="text"/>
Blood Glucose R	<input type="text"/>	Appetite	good <input type="text"/>
Blood Urea	<input type="text"/>	Pedal Edema	yes <input type="text"/>
Serum Creatinine	<input type="text"/>	Anemia	yes <input type="text"/>

**prédiction médical**

Figure 37: Interface de choisi l'état des nœuds

- **Prédiction médical** : Après la sélection des états et le remplissage les champs vide des nœuds avec des valeurs, nous cliquons sur cette fonction, un message va apparaitre contient un résultat de patient possède une anomalie d'un certain pourcentage ou bien un message qu'il n'a pas, a certain pourcentage.

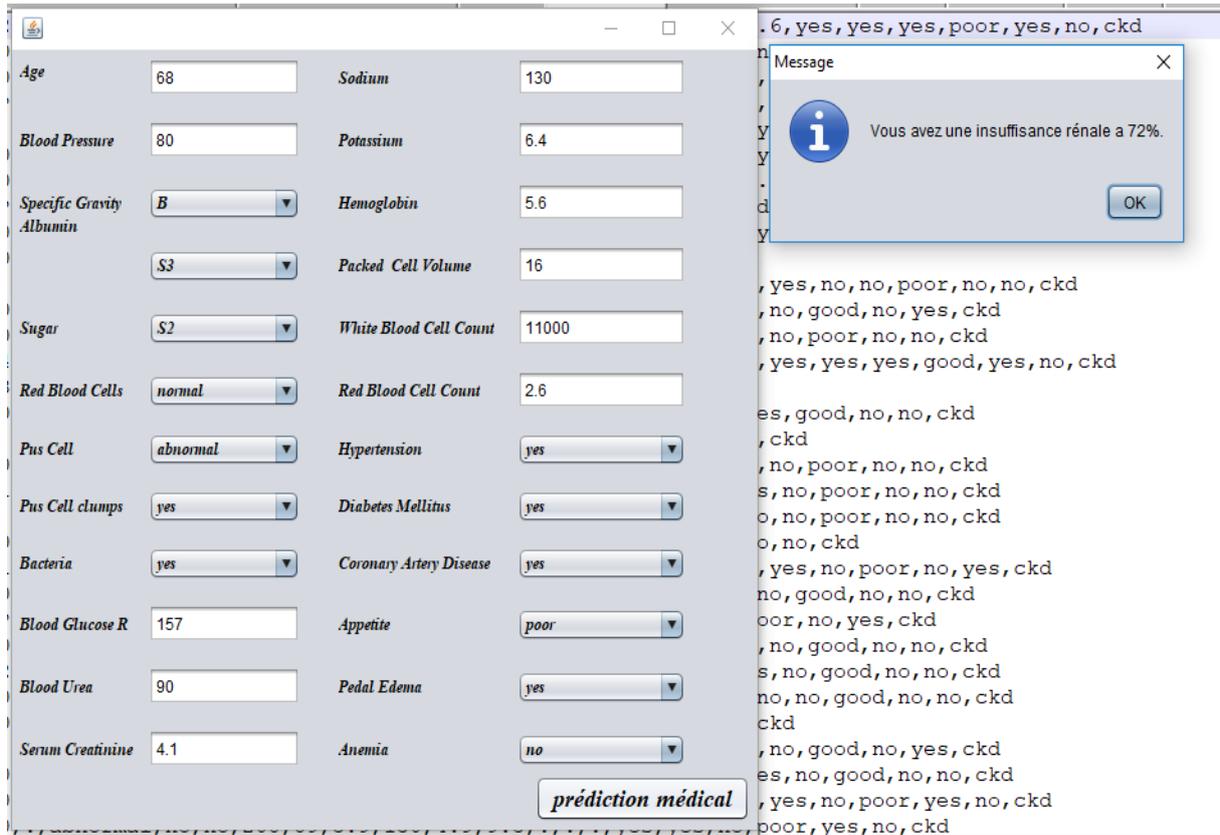


Figure 38: Message de repens dans le cas de la maladie égala « true »

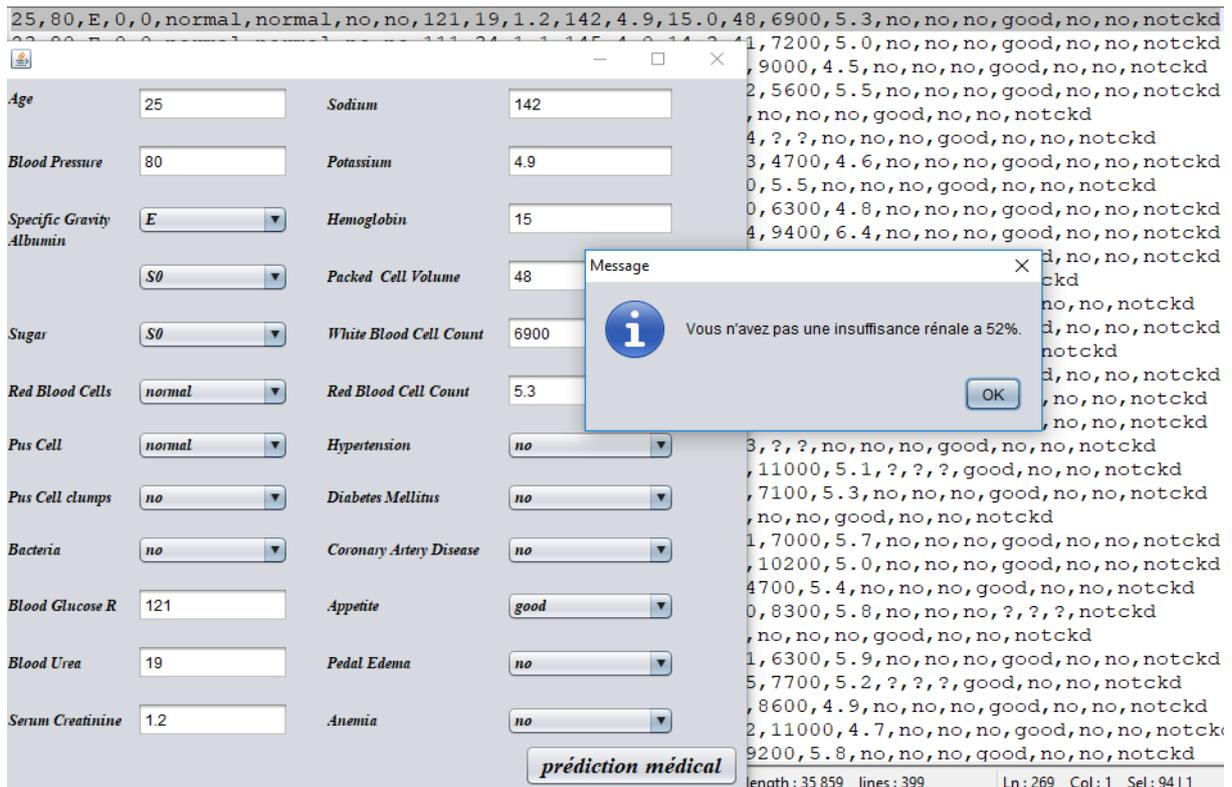


Figure 39: Message de repens dans le cas de la maladie égala « false »

## **5. conclusion**

Nous avons présenté dans ce chapitre les détails techniques liés à la mise en œuvre de notre application, nous avons commencé par présenter les différentes technologies utilisées pour la réalisation de l'application. Ensuite nous avons présenté les différentes interfaces de notre système

# **Chapitre 5 :**

# **Test du système**

### 1. Introduction

Dans cette dernière partie nous allons effectuer un test sur notre système implémenté.

-La méthode que nous avons utilisée pour tester notre système c'est d'effectuer les étapes pour créer un réseau bayésien a partir d'une base de données et faire l'inférence sur un ensemble des patients enregistrées dans une base de donnée, ou bien d'un nouveau patient entrée avec des symptômes compatibles avec notre système.

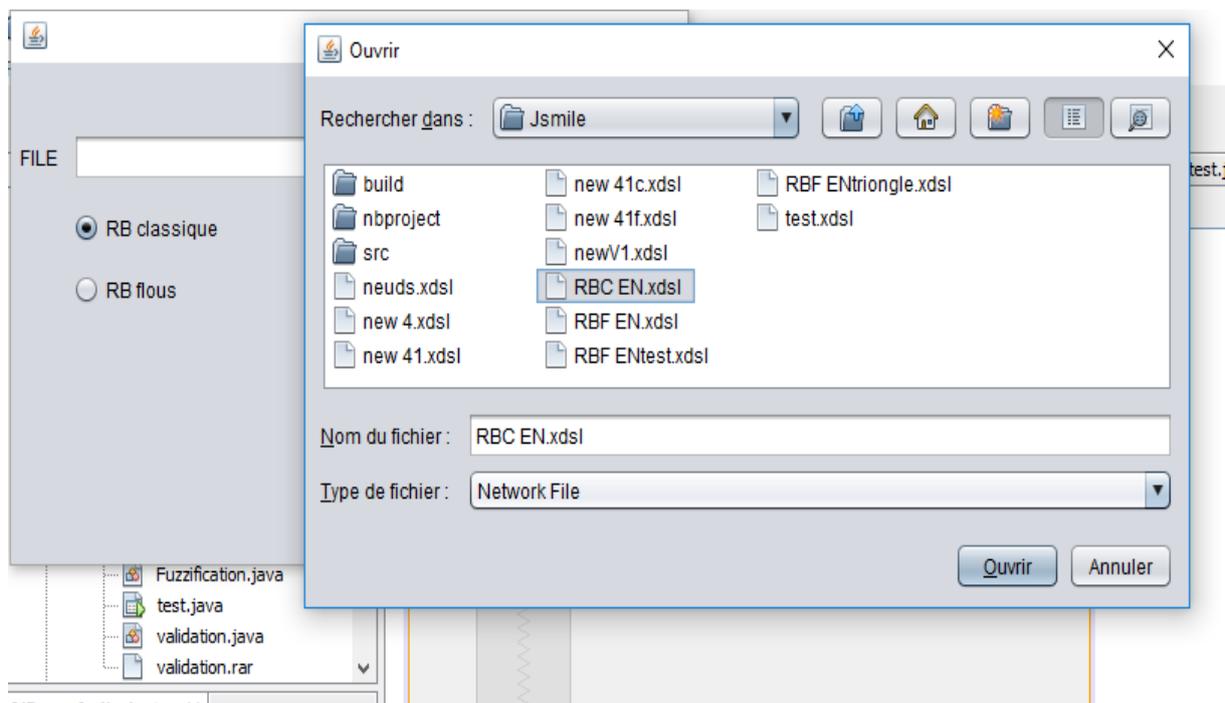
- Nous validons notre résultat par la méthode validation croisée qui nous indique la précision de notre système.

### 2. Teste de système

#### 2.1. Teste de RB classique

Après la création de RB classique, nous obtenons un fichier xdsl, que nous avons nommé RBC EN.xdsl dans la création.

- Nous ouvrons notre application et nous choisissons prédiction médicale sur l'interface d'accueil ensuite nous couchons sur RB classique et nous importons notre fichier RBC EN.xdsl comme dans la figure suivant :



**Figure 40: Importer le fichier « RBC EN.xdsl »**

-Ensuite nous sélectionnons les états des paramètres comme nous le faisons dans la figure suivant :

The image shows a software window titled "prédiction médicale" with a grid of 24 dropdown menus. Each menu is labeled with a medical parameter and has a selected value. The parameters and their values are:

Age	adult	Sodium	normal
Blood Pressure	Normal	Potassium	normal
Specific Gravity	D	Hemoglobin	high
Albumin	S0	Packed Cell Volume	normal
Sugar	S0	White Blood Cell Count	normal
Red Blood Cells	normal	Red Blood Cell Count	normal
Pus Cell	normal	Hypertension	no
Pus Cell clumps	no	Diabetes Mellitus	no
Bacteria	no	Coronary Artery Disease	no
Blood Glucose R	normal	Appetite	good
Blood Urea	normal	Pedal Edema	no
Serum Creatinine	low	Anemia	no

At the bottom right of the window is a button labeled "prédiction médicale".

Figure 41: Sélectionner des états

-Après nous cliquons sur prédiction médicale, un message va apparaitre indique le résultat de notre teste, représenter comme un pourcentage soit touché ou non par la maladie comme les figures suivantes **que nous avons choisi deux instance différent comme exemple, le premier ne pas touché par la maladie et deuxième oui :**

## Chapitre 5 : Test du système

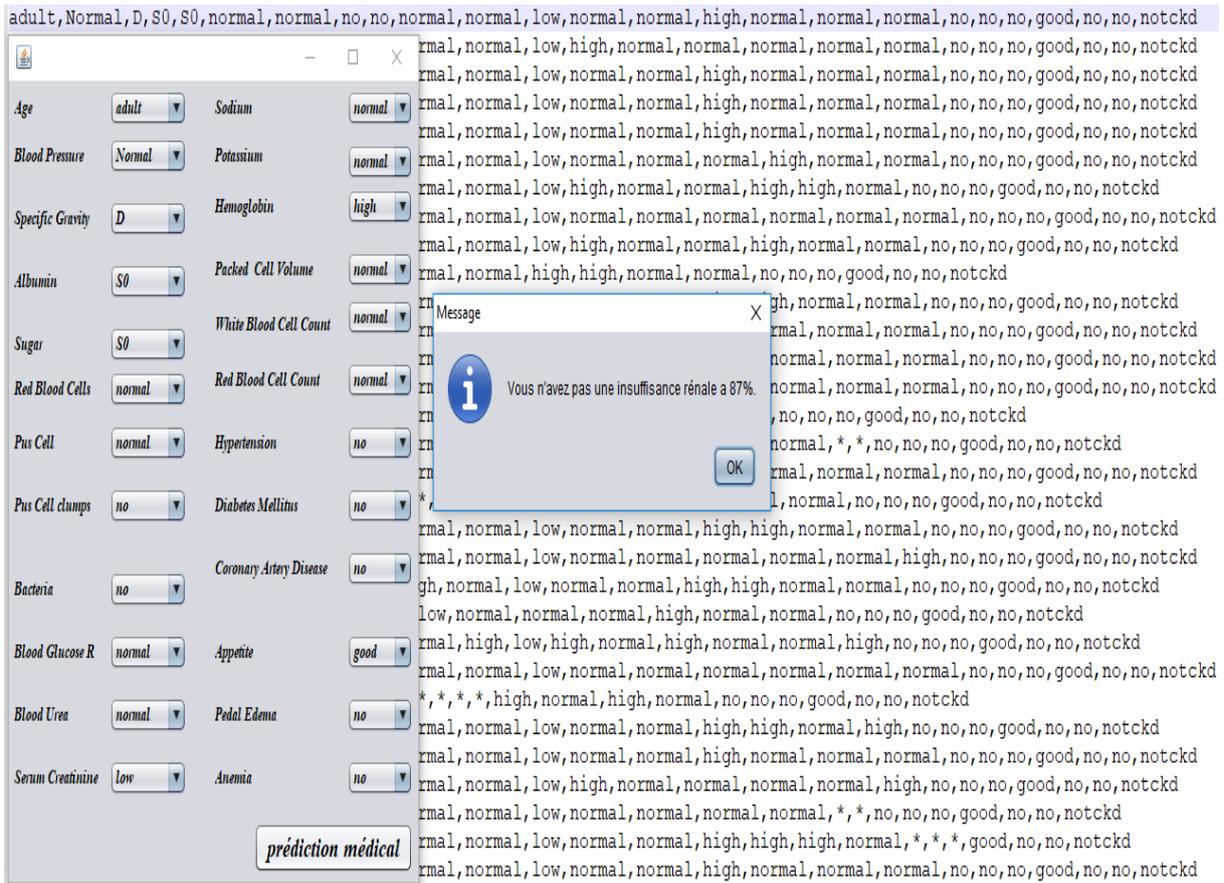


Figure 42: Message de résultat «maladie =false» dans RB classique

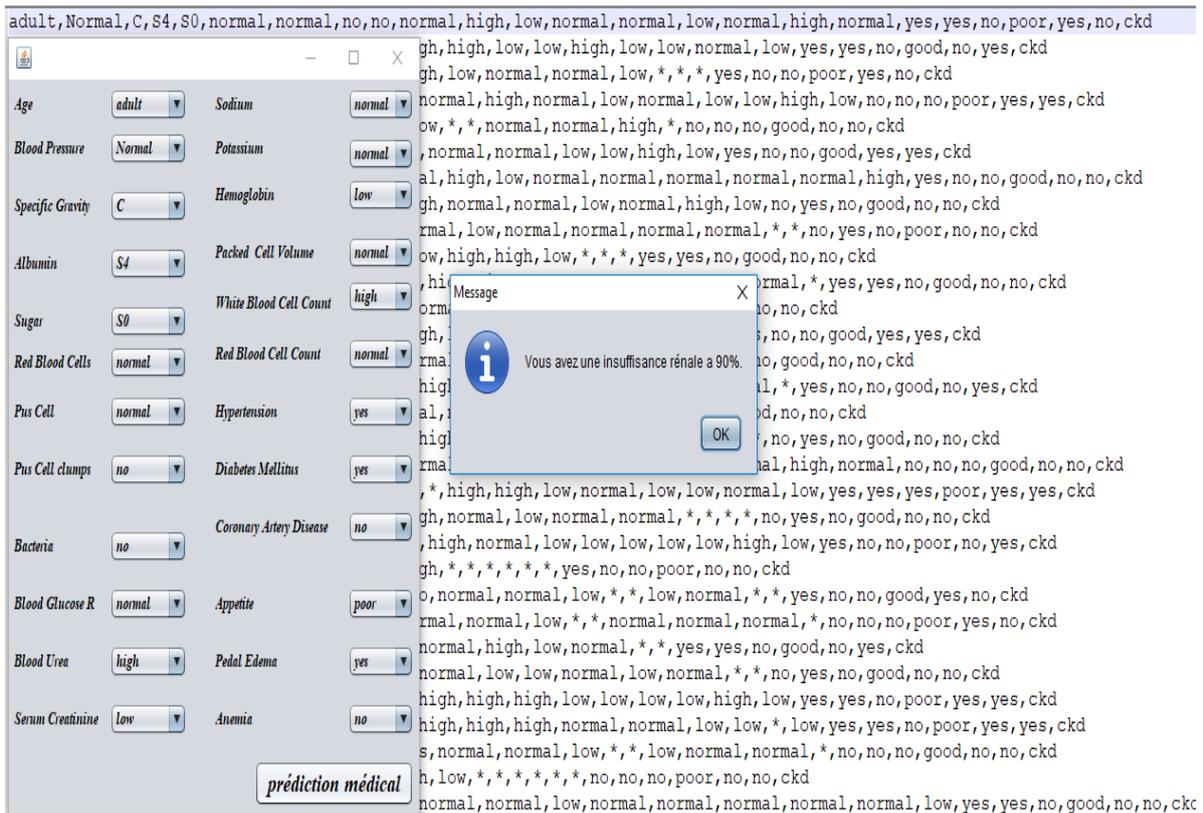


Figure 43: Message de résultat «maladie =true » dans RB classique

### 2.2. Teste de RB flou

-Après la création de RB flou, nous obtenons un fichier xdsl, que nous avons nommé RBF EN.xdsl dans la création.

- Nous ouvrons notre application et nous choisissons prédiction médicale sur l'interface d'accueil ensuite nous cochons sur RB flous et nous importons notre fichier RBF EN.xdsl comme dans la figure suivant :

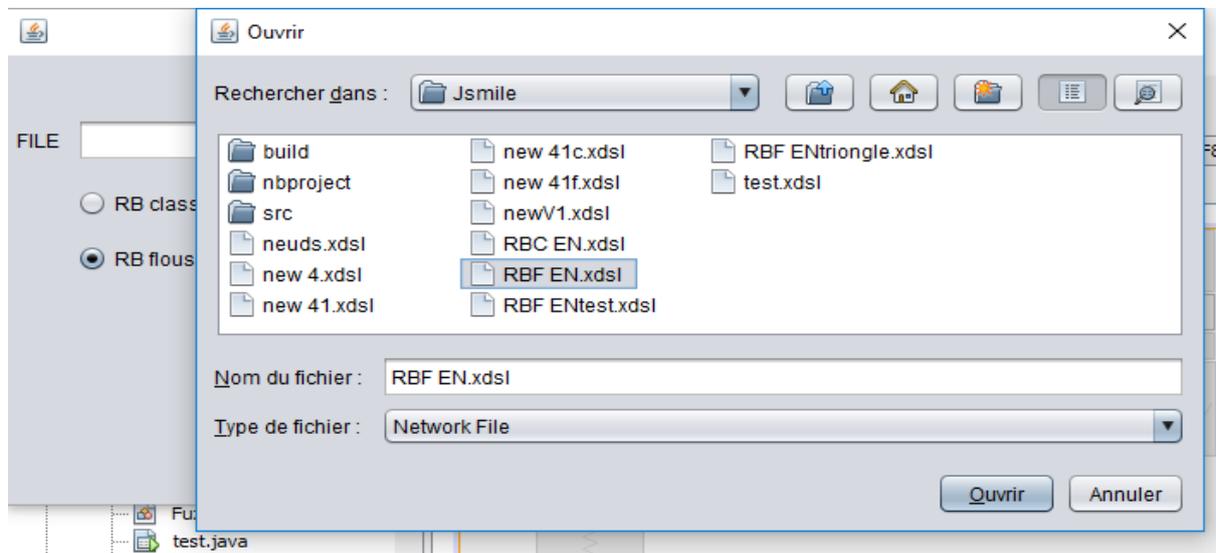
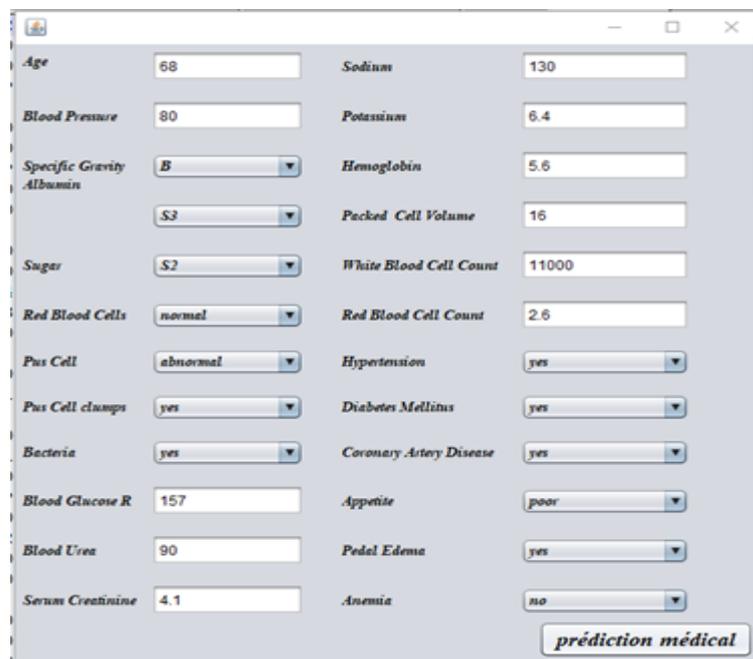


Figure 44: Importer le fichier « RBF EN.xdsl »

- Ensuite nous sélectionnons les états et nous donnons les valeurs des paramètres comme nous le faisons dans la figure suivant :

A screenshot of a medical prediction form. The form contains various input fields and dropdown menus for medical parameters. The parameters and their values are as follows:

Age	68	Sodium	130
Blood Pressure	80	Potassium	6.4
Specific Gravity	B	Hemoglobin	5.6
Albumin	S3	Packed Cell Volume	16
Sugar	S2	White Blood Cell Count	11000
Red Blood Cells	normal	Red Blood Cell Count	2.6
Pus Cell	abnormal	Hypertension	yes
Pus Cell clumps	yes	Diabetes Mellitus	yes
Bacteria	yes	Coronary Artery Disease	yes
Blood Glucose R	157	Appetite	poor
Blood Urea	90	Pedal Edema	yes
Serum Creatinine	4.1	Anemia	no

A "prédiction médical" button is located at the bottom right of the form.

Figure 45: Sélectionner et donner les valeurs des états

## Chapitre 5 : Test du système

Après nous cliquons sur prédiction médicale, un message va apparaître indique le résultat de notre teste représenter comme un pourcentage, soit toucher ou non par la maladie comme dans les figures suivantes que nous avons choisi deux instance différent comme exemple le premier toucher par la maladie et deuxième non :

:

The screenshot shows a web-based medical prediction interface. On the left, there is a form with various input fields for patient data. At the bottom of the form is a button labeled "prédiction médicale". On the right, a "Message" dialog box is displayed, containing an information icon and the text "Vous avez une insuffisance rénale a 72%." with an "OK" button.

Parameter	Value	Parameter	Value
Age	68	Sodium	130
Blood Pressure	80	Potassium	6.4
Specific Gravity	B	Hemoglobin	5.6
Albumin	S3	Packed Cell Volume	16
Sugar	S2	White Blood Cell Count	11000
Red Blood Cells	normal	Red Blood Cell Count	2.6
Pus Cell	abnormal	Hypertension	yes
Pus Cell clumps	yes	Diabetes Mellitus	yes
Bacteria	yes	Coronary Artery Disease	yes
Blood Glucose R	157	Appetite	poor
Blood Urea	90	Pedal Edema	yes
Serum Creatinine	4.1	Anemia	no

prédiction médicale

Message

Vous avez une insuffisance rénale a 72%.

OK

Figure 46: Message de résultat «maladie =true » dans RB flou

25, 80, E, 0, 0, normal, normal, no, no, 121, 19, 1.2, 142, 4.9, 15.0, 48, 6900, 5.3, no, no, no, good, no, no, notckd

1, 7200, 5.0, no, no, no, good, no, no, notckd

9000, 4.5, no, no, no, good, no, no, notckd

2, 5600, 5.5, no, no, no, good, no, no, notckd

, no, no, no, good, no, no, notckd

4, ?, ?, no, no, no, good, no, no, notckd

3, 4700, 4.6, no, no, no, good, no, no, notckd

0, 5.5, no, no, no, good, no, no, notckd

0, 6300, 4.8, no, no, no, good, no, no, notckd

4, 9400, 6.4, no, no, no, good, no, no, notckd

d, no, no, notckd

ckd

no, no, notckd

d, no, no, notckd

notckd

d, no, no, notckd

, no, no, notckd

, no, no, notckd

3, ?, ?, no, no, no, good, no, no, notckd

, 11000, 5.1, ?, ?, ?, good, no, no, notckd

, 7100, 5.3, no, no, no, good, no, no, notckd

, no, no, good, no, no, notckd

1, 7000, 5.7, no, no, no, good, no, no, notckd

, 10200, 5.0, no, no, no, good, no, no, notckd

4700, 5.4, no, no, no, good, no, no, notckd

0, 8300, 5.8, no, no, no, ?, ?, ?, notckd

, no, no, no, good, no, no, notckd

1, 6300, 5.9, no, no, no, good, no, no, notckd

5, 7700, 5.2, ?, ?, ?, good, no, no, notckd

, 8600, 4.9, no, no, no, good, no, no, notckd

2, 11000, 4.7, no, no, no, good, no, no, notckd

9200, 5.8, no, no, no, good, no, no, notckd

length : 35 859 lines : 399 Ln: 269 Col: 1 Sel: 94 | 1

Age 25 Sodium 142

Blood Pressure 80 Potassium 4.9

Specific Gravity E Hemoglobin 15

Albumin S0 Packed Cell Volume 48

Sugar S0 White Blood Cell Count 6900

Red Blood Cells normal Red Blood Cell Count 5.3

Pus Cell normal Hypertension no

Pus Cell clumps no Diabetes Mellitus no

Bacteria no Coronary Artery Disease no

Blood Glucose R 121 Appetite good

Blood Urea 19 Pedal Edema no

Serum Creatinine 1.2 Anemia no

prédiction médicale

Message

Vous n'avez pas une insuffisance rénale à 52%

OK

Figure 47: Message de résultat «maladie =false » dans RB flu

### 3. La validation croisée

Elle va nous permettre d'utiliser l'intégralité de notre base de données pour l'entraînement et pour la validation.

- On découpe la base de données en k parties (folds en anglais) à peu près égales. Tour à tour, chacune des k parties est utilisée comme données de test. Le reste (autrement dit, l'union des k-1 autres parties) est utilisé pour l'entraînement.

- À la fin, chaque instance (ou observation) a servi 1 fois dans les données de test, (k-1) fois dans les données d'entraînement. Nous avons donc 1 prédiction par instance de nos données initiale, et aucune de ces prédictions n'a été faite avec les données d'entraînement qui contiennent cette instance. Nous n'avons pas violé le principe de ne pas valider sur les données d'entraînement.

#### 3.1. Résultats de validation croisée

Nous avons donné plusieurs valeurs à K, pour confirmer notre résultat.

- Après la validation croisée sur RB classique, RB flous avec une fonction d'appartenance triangle et RB flous avec une fonction d'appartenance trapézoïde, nous présentons les résultats dans le tableau suivant :

K(folds)	RB classique	RB Flous fonction d'appartenance Triangle	RB Flous fonction d'appartenance Trapézoïde
K=2	88.9%	95.9%	96.2%
K=3	89.2%	96.2%	96.4%
K=4	88.7%	95.9%	96.2%
K=5	88.4%	95.9%	96.2%
K=6	88.7%	96.4%	96.7%
K=7	89.2%	97.2%	97.5%
K=8	88.2%	95.9%	96.2%
K=9	88.7%	96.9%	97.2%
K=10	87.9%	95.7%	95.9%
K=11	88.7%	97.4%	97.7%

Tableau 12: Précision de système dans RB et RB flou

#### 3.2. Discussions de résultats :

Dans toutes les valeurs que nous avons données à « K », la précision de RB flous soit triangulaire ou trapézoïde sont élevées à la précision de RB classique. et nous avons remarqué que quand nous utilisons la fonction trapézoïde dans RB flou, les résultats de précision seront un peu augmentés par rapport aux résultats de RB flous avec une fonction d'appartenance triangle.

Après ces résultats nous avons modifié notre application pour faire la prédiction médicale seulement avec RB flou dans l'implémentation.

### 4. Conclusion

Ce chapitre a été consacré aux tests de notre système, Pour tester ce qui est considéré comme un objectif de l'application.

-Faire le teste d'une Maladie d'insuffisance rénale avec des symptômes compatible avec notre structure ou bien de tester les maladies qui sont enregistrées dans la base de données avec RB et RB flous, après nous pouvons faire une validation sur les deux et choisir le meilleure d'entre eux pour l'implémentation.

# **Conclusion générale**

### Conclusion générale

Les réseaux bayésiens sont une approche puissante pour permettre le partage des connaissances et de soutenir l'interopérabilité entre les personnes et les systèmes informatiques. Cependant, les réseaux bayésiens traditionnelles (formelles) ne fournissent pas un soutien adéquat à l'incertitude, une caractéristique fondamentale d'un environnement complexe dans un monde réel ouvert.

Pour faire face à tels problème important. Ce mémoire présente Les réseaux bayésiens flous qui porte les procédures générales pour la modélisation de l'incertitude.

Le domaine médical est basé sur les diagnostics. Mais les réseaux bayésiens classiques (traditionnelles) n'intègre pas de connaissances incertaines, alors que le réseau bayésien flous joue un rôle important dans la représentation de l'incertitude.

Nous avons proposé pour cette modélisation un processus qui porte des étapes :

- **L'importation de base de donnée** : sélectionner quelle base de donnée qu'on veut traiter.
- **L'ordre de causalité et les nœuds flous** : sélectionner L'ordre de causalité et les nœuds flous.
- **Application de l'algorithme Greedy Thick Thinning** : nous avons appliqué l'algorithme Greedy Thick Thinning afin d'optimiser la structure de notre réseau bayésien, cette structure va être un paramètre d'entrée de l'algorithme EM.
- **Application de l'algorithme EM**: une approche pour l'apprentissage des données manquantes (les nœuds) avec un algorithme (EM) basé sur les réseaux bayésiens qui arrive à faire une prédiction efficace sur les données manquantes.
- **Calculer l'évidence** : calculer le résultat avec virtuel évidence.

Nous avons implémenté notre système qui est générique et nous l'avons testé avec une base se donnée d'un domaine médicale, ainsi nous avons montré les différentes interfaces nécessaires pour la construction de réseaux bayésiens, et les interfaces nécessaires pour l'application des algorithmes que nous avons utilisé pour l'apprentissage de la structure et l'apprentissage des paramètres manquant EM afin pouvoir calculer le résultat (evidence).





## Référence

[1] Rajai Mourid El Dajani. (2002). Modélisation cognitive et prédictive du comportement dynamique de l'activité électrique du cœur par approche neuronale, thèse de doctorat, L'institut national des sciences appliquées de Lyon. .

[2] Ouahib GUENOUNOU. (2009). Méthodologie de conception de contrôleurs intelligents par l'approche génétique- application à un bioprocédé, thèse de doctorat, université de Toulouse.

[3] D. Rumelhart, G. Hinton. (1986). Learning internal representations by error propagation, in *Parallel Data Processing*, vol. 1. The M.I.T. Press, Cambridge,

[4] Pearl, J. (1988). Morgan Kaufmann series in representation and reasoning. Probabilistic reasoning in intelligent systems: Networks of plausible inference.

[5] Lauritzen, S. L, & Spiegelhalter, D. J. (1990). Local computations with probabilities on graphical structures and their application to expert systems. *Readings in uncertain reasoning*, 415-448.

[6] Ben Mrad, A. (2015) Observations probabilistes dans les réseaux bayésiens. Intelligence artificielle Université de Valenciennes et du Hainaut-Cambresis.

[7] Shachter, R. D., & Peot, M. A. (2013). Simulation approaches to general probabilistic inference on belief networks. arXiv preprint arXiv:1304.1526.

[8] Fung, R., & Del Favero, B. (1994, July). Backward simulation in Bayesian networks. In *Proceedings of the Tenth international conference on Uncertainty in artificial intelligence* (pp. 227-234). Morgan Kaufmann Publishers Inc.

[9] Intelligence, A., & Approach, A. M. (2010). Stuart Russell and Peter Norvig.

[10] Embarki,M . (2012). Etude comparative entre réseaux bayésien et autre méthodes de classification appliqué sur une base cardiologie. Mémoire de master. Université de tlemcen.

[11] Patrick Naïm, Pierre-Henri Wuillemin, Philippe Leray, Olivier Pourret et Anna Becker. (2011). Réseaux bayésiens, Eyrolles, 3<sup>e</sup> éd.

[12]: Robinson R. W. (1977). Counting unlabeled acyclic digraphs. In C. H. C. Little, Ed., *Combinatorial Mathematics V*, volume 622 of *Lecture Notes in Mathematics*, p. 28–43, Berlin: Springer.

[13]: Frédéric Santos. (2015). L'algorithme EM : une courte présentation.

- [14] Christopher Fogelberg, Vasile Palade, and Phil Assheton. (2008). Belief propagation in fuzzy bayesian networks. In 1st International Workshop on Combinations of Intelligent Methods and Applications (CIMA) at ECAI'08, p. 19–24.
- [15] Judea Pearl. (1988). Probabilistic reasoning in intelligent systems : networks of plausible inference. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.
- [16] P. Naïm, P-H. Wuillemin, P.Leray, O. Pourret, and A. Becker. (2007). Réseaux bayésiens, Eyrolles, Paris.
- [17] 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.
- [18] Kailan Shang, & Zakir Hossen. (2013). Application de la logique floue à l'évaluation des risques et à la prise de décisions. Casualty Actuarial Society, Institut canadien des actuaires, Society of Actuaries.
- [19] Bertaud-Gounot, V. (2008). Mémoire de thèse d'université : Proposition d'un modèle constructiviste de base de connaissances fondé sur une ontologie des signes permettant les raisonnements abductif et déductif. Université de Rennes 1. Retrieved from <http://www.theses.fr/2007REN1B105>
- [20] Fagot-Largeault, A. (2010). Le concept de maladie sous-jacent aux tentatives d'informatisation du diagnostic médical. In Presses Universitaires de France - PUF (20 janvier 2010) (Ed.), Histoire et philosophie (1st ed., Vol. 1, pp. 89–110). Paris.
- [21] Rajkumar, V. S. (2011). Multiple myeloma: 2011 update on diagnosis, risk-stratification, and management. American journal of hematology, 86(1), 57–65.
- [22] Ceusters, W., Elkin, P., & Smith, B. (2006). Referent tracking: the problem of negative findings. Studies in health technology and informatics, 124(1), 741–6
- [23] Ceusters, W., Elkin, P., & Smith, B. (2007). Negative findings in electronic health records and biomedical ontologies: a realist approach. International journal of medical informatics, 76(3), 326–33.
- [24] Scheuermann, R. H., Ceusters, W., & Smith, B. (2009). Toward an Ontological Treatment of Disease and Diagnosis Department of Pathology and Division of Biomedical Informatics , University of Texas. In AMIA 2009 summit on translational bioinformatics (pp. 116–120).
- [25] Arocha, J., Wang, D., & Patel, V. (2005). Identifying reasoning strategies in medical decision making: a methodological guide. Journal of biomedical informatics, 38(2), 154–71.

[26] Miller, R. A. (2009). Computer-assisted diagnostic decision support: history, challenges, and possible paths forward. *Advances in health sciences education : theory and practice*, 14(1), 89–106.

[27] Masquelet, A. C. (2006a). Le raisonnement diagnostique. In P. U. de F. PUF (Ed.), *Le raisonnement médical* (1st ed., pp. 17–43). Paris: PUF, Presses Universitaires de France.

[28] Patel, V., Arocha, J., & ZHANG, J. (2004). Thinking and Reasoning in Medicine Citation : What is Medical Reasoning? Early Research on Medical Problem Solving and Reasoning, 1–34.

[29] Pottier, P., & Planchon, B. (2011). Description of the mental processes occurring during clinical reasoning. *La Revue de médecine interne / fondée ... par la Société nationale française de médecine interne*, 32(6), 383–90.

[30] Seroussi, B., Le Beux, P., & Venot, A. (2013). L'aide au diagnostic médical. In Springer Verlag France (Ed.), *Informatique médicale, e santé. Fondements et applications*

[31] Degoulet, P., & Fieschi, M. (1994a). Raisonnement médical et décision. In Masson (Ed.), *Informatique médicale* (1st ed., pp. 47–63). Paris: MASSON.

[32] Masquelet, A. C. (2006b). Les modes fondamentaux d'inférence. In P. U. de F.- PUF (Ed.), *Le raisonnement médical* (1st ed., pp. 75–93). Paris: PUF, Presses Universitaires de France.

[33] Walton, D. (2004). Abductive inference in medical diagnosis. In D. Walton (Ed.), *Abductive reasoning* (1st ed., pp. 143–157). Alabama: University Alabama Press.

[34] Josephson, J., Josephson, S., & Tanner, M. (1996). Conceptual analysis of abduction. In J. R. Josephson & S. G. Josephson (Eds.), *Abductive Inference - Computation, Philosophy, Technology* (1st ed., pp. 1–29). Cambridge: Cambridge University Press.

[35] Aamodt, A., & Plaza, E. (1994). Case-Based Reasoning : Foundational Issues ,Methodological Variations , and System Approaches. *AI Communications*, 7(1), 39–59.

[36] Abecassis, P. et Batifoulier, P., (1998), “Comment penser l'incertitude médicale à l'aide des probabilités ?”, Actes du Colloque de l'Association d'Econométrie Appliquée, Econométrie de la santé XI- Maîtrise de la complexité en santé, Lyon, 9-10 juillet.

[37] Barnes, B. A., Wennberg, J. E. et Zubkoff, M., (1982), “Professional Uncertainty and the problem of Supplier-Induced Demand.”, *Social Science and Medecine*, 16 : 811-824.

[38] Batifoulier, P., (1994), “Incertainitude sanitaire et risque probabilisable.”, *Journal d'Economie Médicale*, 2, 1(février) : 3-17.

[39] Popper, K., (1990), *A word of Prependencies*, Thoemmes, Bristol.

[40] E. Di Tomaso and J. F. Baldwin. An approach to hybrid probabilistic models. *International Journal of Approximate Reasoning*, 47(2) :202– 218, 2008.

[41] Faugère, J.-P., (1982), “L'incertitude en économie de la santé : du risque maladie à l'incertitude de la pratique médicale.”, in P.M. Clair et alii (eds), *Economie et incertitudes*, Economica, Paris.

[42] Bikbov B, Perico N, Remuzzi G (23 May 2018). "Disparities in Chronic Kidney Disease Prevalence among Males and Females in 195 Countries: Analysis of the Global Burden of Disease 2016 Study". *Nephron*.

[43] GBD 2015 Mortality and Causes of Death, Collaborators. (8 October 2016). "Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980-2015: a systematic analysis for the Global Burden of Disease Study 2015". *Lancet*. 388 (10053): 1459–1544.

[44] Cerquides J., R.L.Mantaras. (2003). Tractable bayesian learning of tree augmented naïve bayes classifiers,

[45] Smail L. (2004). Algorithmique pour les réseaux Bayésiens et leurs extensions. Thèse de doctorat,

[46] Friedman, Nir, Dan Geiger & Moises Goldszmidt (1997). Bayesian network classifiers. *Machine Learning*, 29, 131–163

[47] Spirtes, Peter, Clark Glymour & Richard Scheines (1993). *Causation, Prediction, and Search*. Springer Verlag Lectures in Statistics.

[48] Voortman, Mark & Marek J. Druzdzel (2008). Insensitivity of constraint-based causal discovery algorithms to violations of the assumption of multivariate normality. In *Recent Advances in Artificial Intelligence: Proceedings of the Twenty First International Florida Artificial Intelligence Research Society Conference (FLAIRS-2008)*, David Wilson, H. Chad Lane (eds), pages 690-695, Menlo Park, CA: AAAI Press.

[49] heng, Jie, David A. Bell & Weiru Liu (1997). An Algorithm for Bayesian Belief Network Construction from Data. *Proceedings of AI & Statistics*, pages 83-90.

[51] Dash D, M.J Druzdzel. (2002). Robust independence testing for constraint-based learning of causal structure; *Proceedings of the Nineteenth conference on Uncertainty in Artificial Intelligence*; Morgan Kaufmann Publishers Inc;

[52] Spirtes P, Glamour C, Scheines R. New York: Springer. (1993). *Causation, Prediction, and Search*.

[53] Heckerman D, Geiger D, Chickering D.M. (1995). Learning Bayesian networks: The combination of knowledge and statistical data. *Machine learning*.



