



REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université Mohamed Khider – BISKRA
Faculté des Sciences Exactes, des Sciences de la Nature et de la Vie
Département d'informatique

N° d'ordre :/M2/2020

Mémoire

présenté pour obtenir le diplôme de master académique en

Informatique

Parcours : **SIOD**

Modèle d'aide au diagnostic des tumeurs cérébrales à l'aide des réseaux bayésiens

Par :

ZOUZAL KHEDIJA

Soutenu le, devant le jury composé de :

.....

.....

Président

BAHI NAIMA

MCB

Rapporteur

.....

.....

Examineur

Dédicaces

Je dédie ce modeste travail à :

Mon père et ma Mère

Toute ma famille

Et tous mes amis.

Remerciements

Mes remerciements vont d'abord au tout-puissant ALLAH, pour m'avoir fourni courage et patience surtout, pendant la période du virus corona, et pour avoir guidé mes pas vers le chemin de la connaissance.

Je tiens à remercier mon encadreur BAHI NAÏMA pour ses conseils, ses orientations et son aide si Précieuse.

Je remercie également les membres de jury d'avoir accepté d'évaluer ce travail.

Enfin, je présente mes remerciements à toutes les personnes qui m'ont encouragé à construire et élaborer ce mémoire.

Sans oublier de remercier tous ceux qui ont participé directement ou indirectement à la réalisation de ce travail.

Résumé

La tumeur cérébrale est devenue une tumeur très forte ces dernières années avec le développement de nouveaux types et sa détection est devenue plus difficile en raison de leur similitude.

Grâce aux progrès de la science, le travail du cerveau est de mieux en mieux connu. On sait maintenant qu'ils fonctionnent grâce à des milliards de neurones, connectés les uns aux autres, et on sait aussi qu'ils sont organisés en plusieurs régions, gérant chacune des fonctions différentes, cette connaissance du cerveau a permis de comprendre les maladies les meilleures et les plus difficiles.

Notre travail commence par trouver un modèle pour aider à diagnostiquer cette tumeur à l'aide des réseaux bayésiens, et en raison de l'expansion des réseaux bayésiens, nous avons choisi certaines techniques qui nous aideront à trouver le modèle approprié pour diagnostiquer une tumeur cérébrale. C'est pourquoi nous apporterons des solutions efficaces et intelligentes dans de nombreux domaines en général et dans le domaine de la santé en particulier.

Mots clé

Apprentissage, Apprentissage par paramètres, Apprentissages par structures, Réseau bayésien, Structure fixée, Réseau bayésien naïve, Algorithme K2, Validation croisée, La précision.

Table des matières

Dédicaces	i
Remerciements	ii
Résumé	iii
Table des matières	iv
Listes des figures	vii
Listes des tableaux	ix
Listes des algorithmes	x
Introduction générale	1
Chapitre 01 : Tumeur cérébrale	3
1. Introduction	4
2. Définition.....	4
3. Familles de tumeur	4
3.1.Tumeur bénigne	4
3.2.Tumeur maligne	4
3.3.Tumeur bénigne Vs tumeur maligne	5
4. Tumeur primitive (précancéreuse)	5
5. Causes.....	6
6. Type des tumeurs	6
7. Tumeur cérébrale.....	9
7.1.Signes	10
7.2.Diagnostic	10
7.3.Variété de tumeur cérébrale.....	11
8. Conclusion.....	15
Chapitre 02 : Réseaux bayésiens	16
1. Introduction	17
2. Définition.....	17
3. Domaine d'application	17
4. Définition formelle.....	18
5. Taches de réseau bayésien	20
5.1.Inférer des variables non observées	20
5.2.Apprentissage des paramètres.....	21

5.2.1. Estimation vraisemblance maximum	22
5.2.2. Estimation maximum à posteriori	23
5.3.Apprentissage des structures	24
5.3.1. Approche basée-scores	25
5.3.2. Approche basée-contraintes	27
6. Apprentissage des paramètres Vs Apprentissage de la structure	28
7. Conclusion.....	28
Chapitre 03 : Conception	29
1. Introduction	30
2. Schéma général.....	30
3. Présentation du modèle.....	31
4. Modélisation de la structure du réseau bayésien.....	31
4.1.Choix de la structure	33
4.1.1. Changement de l'orientation des arcs	33
4.1.2. Elimination des variables.....	34
4.1.3. Modélisation par réseau bayésien naïf	34
4.1.4. Modélisation par algorithme de recherche k2	34
5. Modélisation des paramètres.....	34
5.1.Description de la base.....	35
5.2.Format structure de réseau	37
6. Les algorithmes appliqués	39
6.1.Algorithme principal	39
6.2.Les méthodes appliquées	41
6.2.1. La structure fixée.....	41
6.2.2. Réseau bayésien naïve	42
6.2.3. L'algorithme k2	43
6.3.Les classes utilisées	44
7. Conclusion.....	44
Chapitre 04 : Implémentation	45
1. Introduction	46
2. Outils de Programmation	46
2.1. Langage de programmation.....	46
2.2. Outil de Développement.....	46
2.3. Les Packages	46
3. Implémentation.....	46

3.1. Résultats	47
3.2. Comparaisons des techniques	50
4. Discussion des résultats	51
5. Conclusion	51
Conclusion générale	52
Bibliographie	53
Sitographie	54

Liste des figures :

Figure 1-1 : Tumeur maligne dans la bouche.....	6
Figure 1-2 : Tumeurs dans l'estomac.....	7
Figure 1-3 : Tumeurs dans le poumon.....	7
Figure 1-4 : Tumeurs dans les os et cartilage.	8
Figure 1-5 : Mésothéliome de la plèvre.	8
Figure 1-6 : Tumeur de cerveau.	9
Figure 1-7 : Structure du cerveau.....	9
Figure 1-8 : Variété de tumeur cérébrale.	11
Figure 1-9 : Evolution d'un Astrocytome.	11
Figure 1-10 : Un Gliome de l'hémisphères.	12
Figure 1-11 : Un Oligodendrogliome.	12
Figure 1-12 : Un Lymphome de cerveau.	13
Figure 1-13 : Métastase au cerveau d'autre cancers.....	13
Figure 1-14 : Un Ependymomes intracrâniens.	14
Figure 1-15 : Médulloblastome de L'Enfant.	14
Figure 2-1 : Exemple 1 indépendance conditionnel entre les variables.	19
Figure 2-2 : Exemple 2 indépendance conditionnel entre les variables.	20
Figure 2.3 : NP Complet.	21
Figure 3-1 : Schéma général de travail.	30
Figure 3-2 : Etapes nécessaire à la modélisation de l'approche utilisée.....	31
Figure 3-3 : Structure du réseau bayésien déduite à partir du raisonnement effectué par le domaine médical.	32
Figure 3-4 : Structure de notre réseau bayésien en changeant l'orientation des arcs.	33
Figure 3-5 : Structure de notre réseau bayésien en éliminant des variables.	34
Figure 3-6 : Base de données de la tumeur cérébrale.....	35
Figure 3-7 : Première partie de la base de données format ARFF.....	36
Figure 3-8 : Deuxième partie de la base de données format ARFF.....	36

Liste des figures

Figure 3-9 : Troisième partie de la base de données format ARFF.....	37
Figure 3-10 : Une partie de BIF fichier.....	38
Figure 3-11 : Exemple d'une base de données simple.	42
Figure 3-12 : Etape deuxième de principe réseau bayésien naïve.....	42
Figure 4-1 : Logo NetBeans.	47
Figure 4-2 : Résultat d'une itération de l'application.	48
Figure 4-3 : Graphe décrivent la précision par rapport au nombre de validation croisée avec la structure initiale.	49
Figure 4-4 : Graphe décrivent la précision par rapport au nombre de validation croisée avec la structure de changement les arcs.	49
Figure 4-5 : Graphe décrivent la précision par rapport au nombre de validation croisée avec la structure de suppression les nœuds.....	50
Figure 4-6 : Graphe décrivent la précision par rapport au nombre de validation croisée avec réseau bayésien naïve.	50
Figure 4-7 : Graphe décrivent la précision par rapport au nombre de validation croisée avec réseau bayésien appliquer algorithme de recherche k2.	51
Figure 4-8 : Comparaison des techniques appliquées avec les bases de données utilisées.....	51

Liste des tableaux :

Tableau 1-1 : Les tumeurs et leurs caractéristiques.....	5
Tableau 2-1 : Liste de variables avec leurs paramètres.....	21
Tableau 2-2 : Liste des paramètres et de leurs distributions a priori.....	24
Tableau 2-3 : Comparaison entre apprentissage de la structure et les paramètres.	28
Tableau 3-1 : Signification des nœuds du réseau bayésien.	32
Tableau 4-1 : Temps d'exécution en fonction des itérations exécutées.	52

Liste des algorithmes :

Algorithme 3-1 : Algorithme principal.	40
Algorithme 3-2: Algorithme K2.	43
Algorithme 3-3 : Classe FromFile et ses méthodes.	44
Algorithme 3-4 : Classe main et ses méthodes.	44

L'étude des tumeurs cérébrales possède une grande importance dans cette dernière décennie. Ce domaine médical prend son importance de la gravité des conséquences de ces tumeurs. Ces dernières mettent la vie en danger. Le nombre de variétés de ces tumeurs rend leur diagnostic assez difficile. Plusieurs caractéristiques entrent en jeu pour décider le type exact de la tumeur. Certaines de ces caractéristiques sont extraites à partir des images par résonance magnétique. D'autres caractéristiques sont extraites des données cliniques du patient. Pour prendre la bonne décision, le processus de diagnostic des tumeurs cérébrales nécessite l'intervention de plusieurs avis de médecins. Notre travail porte sur la modélisation de l'expertise relative au diagnostic d'une famille des tumeurs cérébrales qui est celle des tumeurs de type intra-axial sustentorielle. Les tumeurs de cette famille se caractérisent par la forte ressemblance de leurs caractéristiques ce qui rend leur diagnostic assez difficile. Nous limitons notre champ d'études aux sept tumeurs suivantes : Astrocytome, Gliome, Oligodendrogliome, lymphome, Métastase, Ependymome, Médulloblastome. En observant plusieurs rapports médicaux, nous avons constaté que les informations qui entrent dans le diagnostic de ces tumeurs cérébrales se répartissent en deux catégories. La première catégorie rassemble les données extraites du dossier médical du patient comme par exemple l'âge, le sexe et l'historique médical. La deuxième catégorie est constituée d'un ensemble de données extraites à partir des images IRM. La difficulté du diagnostic de cette famille de tumeurs est due aux valeurs communes pour certaines caractéristiques. L'utilisation de ces caractéristiques dans le cadre d'une consultation médicale est insuffisante. L'expertise du médecin est nécessaire afin d'avoir une bonne précision dans ce type de diagnostic. Ce dernier intervient dans la majorité de cas étudiés pour lever l'incertitude d'identification du type de la tumeur. Son intervention se base sur les situations traitées auparavant. Chaque situation donne de nouvelles connaissances sur une tumeur donnée et met à jour l'expertise du médecin concernant son incertitude sur la fréquence d'apparition d'une tumeur par rapport à une autre.

Un modèle graphique est une famille de distribution de probabilités définie en termes de graphe orienté ou non. Il est constitué de nœuds qui représentent des variables aléatoires et des arcs représentant les relations de dépendances entre ces variables. On distingue deux formes de modèles graphiques probabilistes : les modèles orientés basés respectivement sur les graphes acycliques orientés connus sous le nom de réseaux bayésiens et les graphes non orientés comme les champs de Markov.

Les modèles graphiques probabilistes semblent être l'outil d'intelligence artificielle le plus approprié à la modélisation des systèmes d'aide au diagnostic médical. Par nature, ces systèmes intègrent un certain degré d'incertitude et expriment de façon lisible le phénomène de dépendance. Les réseaux bayésiens offrent une représentation graphique explicite, intuitive et compréhensible par les non spécialistes. Ils donnent aussi la possibilité de rassembler et de fusionner de connaissances de diverses natures dans un même modèle. Pour ces raisons, le formalisme des réseaux bayésiens a été retenu pour la modélisation du problème de diagnostic des tumeurs cérébrales.

Dans le cadre d'un modèle d'aide au diagnostic des tumeurs cérébrales, les réseaux bayésiens permettent d'utiliser différents types de caractéristiques dans un même mo-

dèle. Ces caractéristiques sont modélisées par les nœuds du réseau bayésien. En plus, les arcs du réseau bayésien représentent les indépendances entre ces différents nœuds. L'incertitude de l'expert est exprimée par un ensemble de tables de probabilités conditionnelles qui permettent de quantifier le taux de certitude d'un nœud en tenant compte de ses parents. Les tables de probabilités conditionnelles CPTs sont remplies au passé par l'expert du domaine. Cette tâche devient de plus en plus difficile pour la modélisation de problème de grande taille. Ce travail se fait aujourd'hui par un apprentissage automatique à partir d'une base de données collectée par l'ensemble de cas étudiés par l'expert du domaine. Ce processus connu sous le nom de « apprentissage de paramètres dans le réseau bayésien » permet de remplir les CPTs même en présence de données manquantes dans la base d'apprentissage.

Notre objectif est de construire un modèle qui pourra aider le diagnostic des tumeurs cérébrales, grâce à une application qui sera testée dans la réalité, ainsi le médecin pourra rapidement diagnostiquer son patient.

Ce travail, qui vise à assurer de trouver le modèle à l'aide des réseaux bayésien, il est organisé en quatre chapitres : Le chapitre 1 présente les généralités de la tumeur, y compris ses caractéristiques et la tumeur qu'on a étudié. Le chapitre 2 examine d'abord la définition de réseau bayésien. Ensuite, les tâches de réseaux bayésien (le principe de réseau bayésien). Dans le chapitre 3, une conception du système pour les différentes techniques de réseaux bayésien avec ses résultats obtenus dans le chapitre 4. Une conclusion générale est présentée à la fin du manuscrit.

Chapitre 01

1. Introduction

Le monde a changé son mode de vie avec l'alimentation transformée, la sédentarité, le stress et le rythme de vie non équilibré. Depuis de nombreuses années beaucoup de nouvelles maladies ont menacés la vie de l'homme comme le sida, les virus tel que Ébola, grippe aviaire, ..., etc. Mais le cancer ou la tumeur demeure l'une des maladies les plus mortelles pour l'homme dans le monde. Les statistiques affirment que le nombre de décès avoisine les 200 morts/jour. Selon l'OMS la tumeur est la deuxième cause de décès dans le monde a fait 8,8 millions de morts en 2015 Près d'un décès sur 6 dans le monde est dû à la tumeur. [13]

Ce chapitre parle sur ce qu'est une tumeur, sa famille, ses causes, ses types, on va se concentrer sur les tumeurs cérébrales, quelles sont ses signes et les types sur lesquelles on va travailler.

2. Définition

Une tumeur est une masse plus ou moins volumineuse due à une multiplication normale ou anormale des cellules [1]. La multiplication normale définit une tumeur bénigne ou une tumeur non maligne, non cancéreuse. La multiplication anormale définit une tumeur maligne ou un cancer.

3. Familles de tumeur

3.1.Tumeur bénigne

Les tumeurs bénignes : résultent d'une simple anormalité, mais ne mettent pas en danger la vie du sujet et possèdent les caractéristiques suivantes [2] :

- Rester à un seul endroit et ne se propage pas à d'autres parties du corps ;
- Ne réapparaître habituellement pas après avoir été enlevée ;
- Avoir tendance à posséder une forme régulière et lisse et être recouvertes d'une capsule ;
- Pouvoir être facilement déplacées dans le tissu.

3.2.Tumeur maligne

Les tumeurs malignes : elles sont cancéreuses et donc dangereuses. Le cancer peut prendre naissance dans n'importe quelle partie du corps. Quand les cellules cancéreuses forment une masse, on dit que la tumeur est cancéreuse. La tumeur est cancéreuse quand [2]:

- Elle envahit les tissus voisins ;
- Certaines cellules peuvent s'en détacher et circuler dans le sang ou le système lymphatique pour se propager aux ganglions lymphatiques et à des parties du corps éloignées.

Un cancer qui se propage à partir du premier lieu où il a pris naissance (tumeur primitive) jusqu'à une nouvelle partie du corps est appelé cancer métastatique.

Quand les cellules cancéreuses se propagent et forment une nouvelle tumeur, on parle de métastases.

3.3. Tumeur bénigne Vs tumeur maligne

Contrairement à ce que l'on croit communément, le terme de tumeur n'est pas automatiquement associé à celui de cancer.

En revanche, dans le cadre d'un cancer, on peut classer les tumeurs selon leur apparence et leur comportement dans l'organisme, et dans le cas des nodules, des kystes ou des polypes, il convient de distinguer deux types de tumeurs :

- les tumeurs bénignes ;
- les tumeurs malignes.

Ces deux catégories de tumeurs s'opposent les unes aux autres, voici le tableau suivant.

Caractéristiques	Tumeurs bénignes	Tumeurs malignes
Volume	Restreint.	Important et qui augmente.
Limites	Bien délimitées.	Mal délimitées.
Action sur les tissus voisins	Compriment et refoulent les tissus avec lesquels elles entrent en contact sans les envahir.	Infiltrant et envahissent les tissus voisins avec lesquels elles entrent en contact.
Métastases	Aucune.	Tendance à métastaser.
Évolution	Pas de conséquences sévères sur l'état de santé du patient.	Grave et souvent récidivante.

Tableau 1-1 : Les tumeurs et leurs caractéristiques [3].

4. Tumeur primitive (précancéreuse)

Les cellules précancéreuses sont anormales et elles peuvent se transformer en cancer si on ne les traite pas. Certaines de ces cellules présentent de légers changements susceptibles de disparaître sans traitement. Mais il y a des cellules précancéreuses qui transmettent des changements génétiques et qui deviennent ainsi graduellement de plus en plus anormales en se divisant jusqu'à ce qu'elles se transforment en cellules cancéreuses. Cela peut prendre beaucoup de temps avant qu'un état précancéreux devienne un cancer.

Les changements précancéreux peuvent varier de légers à graves. Il existe différentes façons de décrire les changements précancéreux selon qu'ils sont légers ou graves, par exemple :

Hyperplasie signifie que les cellules anormales se divisent et augmentent en nombre plus rapidement qu'à la normale.

Atypie signifie que les cellules sont légèrement anormales, ou atypiques.

Métaplasie signifie qu'un changement a été subi par les types de cellules qu'on trouve normalement dans cette région du corps.

Dysplasie signifie que les cellules sont anormales, qu'il y a plus de cellules qu'à l'habitude, que les cellules se développent plus vite que normalement et qu'elles ne sont pas disposées comme des cellules normales.

Le **carcinome in situ** est le type le plus grave de changement précancéreux. Les cellules sont très anormales, mais elles n'ont pas envahi le tissu voisin.

5. Causes

Les causes d'apparition d'une tumeur sont multiples et souvent mal identifiées. Toutefois, dans le cas des tumeurs malignes, certains facteurs connus, externes comme internes, peuvent entrer en ligne de compte : l'âge, le mode de vie (chaque année, 20 % des nouveaux cas de cancers sont liés à la consommation de tabac et 8 % à la consommation d'alcool), les antécédents familiaux, l'environnement...etc.

6. Type des tumeurs

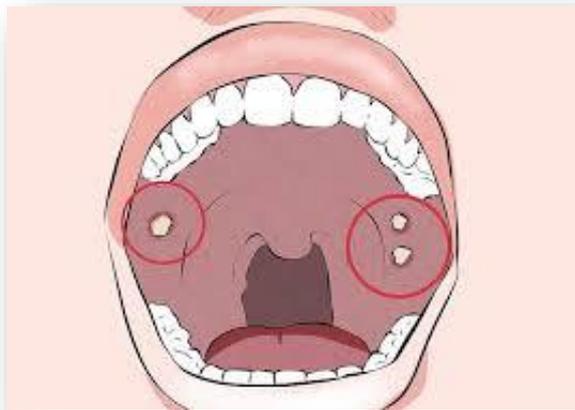


Figure 1-1 : Tumeur maligne dans la bouche.

➤ **Les tumeurs de la cavité buccale :** peuvent se développer à différents niveaux de la cavité buccale (Le plancher de la bouche, La langue, Les amygdales, Le palais, Les joues, Les gencives, Les lèvres.) ce sont à majorité des carcinomes épidermoïdes [4].

Le papillome et l'adénome pléomorphe sont des tumeurs bénignes qui se trouvent au niveau du pharynx ou sur les lèvres.



Figure 1-2 : Tumeurs dans l'estomac.

➤ **Les tumeurs de l'estomac :** L'estomac est un organe creux, situé entre l'œsophage et le duodénum. Dans 9 cas sur 10, le cancer primaire de l'estomac est un adénocarcinome, développé aux dépens des glandes de la muqueuse gastrique. [5]

Il y a aussi des polypes hyperplasiques et des glandes fundiques qui peuvent être réséqués par chirurgie.



Figure 1-3 : Tumeurs dans le poumon.

➤ **Les tumeurs du poumon :** Une tumeur cancéreuse du poumon peut envahir les tissus voisins et les détruire. Elle peut également se propager (métastases) à d'autres parties du corps comme le cancer du poumon à petites cellules et celui du poumon non à petites cellules [6].

Une tumeur bénigne est très rare ; on peut citer l'hamartome pulmonaire, l'adénome ou le papillome.



Figure 1-4 : Tumeurs dans les os et cartilage.

➤ **Les tumeurs des os et du cartilage :** sont bénignes à 90 % pour l'enfant, l'adolescent, l'adulte jusqu'à 60 ans ; l'ostéochondrome est la plus connue. Après 60 ans, elles deviennent malignes comme les métastases osseuses (myélomes), chondrosarcome...etc. [7].



Figure 1-5 : Mésothélium de la plèvre.

➤ **Les tumeurs du mésothélium :** Le mésothélium est une couche de tissu qui recouvre nos organes. Différentes tumeurs bénignes peuvent se développer sur ce tissu : tumeur fibreuse de la plèvre, mésothéliome multi kystique, mésothéliome adénomatoïde... [12].

Dans certains cas les cellules mésothéliales se modifient et engendrent un mésothéliome malin qui est très rare, c'est le mésothéliome de la plèvre ou pleural [7].

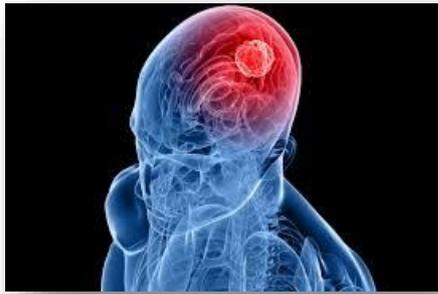


Figure 1-6 : Tumeur de cerveau.

➤ **Les tumeurs de cerveau :** On appelle tumeur du cerveau toutes les tumeurs qui se développent à l'intérieur du crâne. Elles peuvent se développer dans n'importe quelle zone du cerveau : les hémisphères, le cervelet, le tronc cérébral, l'hypophyse, etc.

7. Tumeur cérébrale

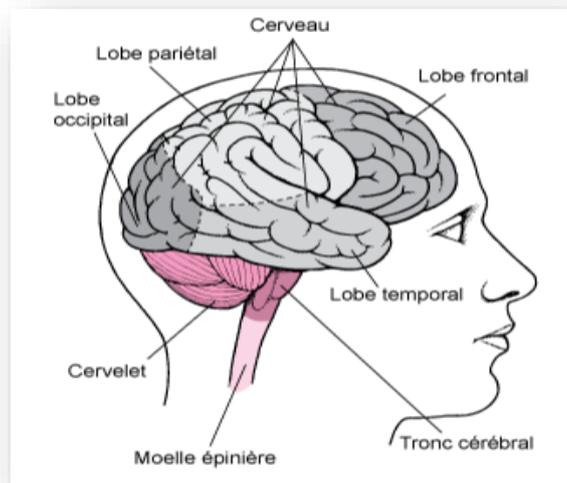


Figure 1-7 : Structure du cerveau.

Une tumeur du cerveau est une masse résultant de la multiplication anarchique de cellules anormales. On ne sait pas, à l'heure actuelle, quelle est l'influence exacte de l'environnement du patient et de son bagage héréditaire sur le développement de la tumeur. Très rarement, elle apparaît en relation avec une maladie transmise génétiquement[8].

Certaines tumeurs cérébrales sont peu agressives, c'est-à-dire qu'elles sont constituées de cellules qui croissent lentement. Ces tumeurs dites bénignes peuvent

être simplement surveillées radiologiquement ou être extirpées chirurgicalement. D'autres tumeurs sont malignes, c'est-à-dire qu'elles sont constituées de cellules se divisant relativement vite. Ces tumeurs croissent donc rapidement et peuvent envahir et endommager des zones importantes du cerveau. Elles peuvent être traitées par la chirurgie, la radiothérapie, la chimiothérapie ou une combinaison de ces différentes techniques.

7.1. Signes

Une tumeur cérébrale peut se manifester de diverses façons, en fonction de son emplacement, de son volume ou de l'œdème qui y est associé[8].

Le ou la patient(e) peut subir un ou plusieurs des symptômes suivants :

- Maux de tête
- Crises d'épilepsie
- Vertiges et étourdissements
- Déficits moteurs ou sensitifs
- Nausées matinales ou vomissements
- Troubles auditifs ou visuels.

7.2. Diagnostics

Le médecin neuro-oncologue décide, en fonction de la situation de chaque patient(e), quels examens doivent être effectués. Il pourra recommander les mesures suivantes[8]:

- Un examen clinique
- Des examens électrophysiologiques menés par le Service de neurologie, comme un électroencéphalogramme (EEG) pour analyser l'activité électrique du cerveau du ou de la patient(e)
- Des examens radiologiques menés par le Service de radiodiagnostic et radiologie interventionnelle, comme le scanner à rayon X, l'imagerie par résonance magnétique (IRM), une spectroscopie par résonance magnétique ou une tomographie par émission de positrons (TEP)
- une intervention chirurgicale menée au sein du service, comme une biopsie pour déterminer le stade de la tumeur et son comportement.

Il y a différents stades (degrés d'extension) d'une tumeur [9] :

- **Stade 1** : la tumeur est unique et de petite taille.
- **Stade 2** : la tumeur est plus volumineuse.
- **Stade 3** : la tumeur envahit les ganglions lymphatiques ou les tissus avoisinants.
- **Stade 4** : présence de métastases dans d'autres organes à distance de la tumeur d'origine.

7.3.Variété de tumeur cérébrale

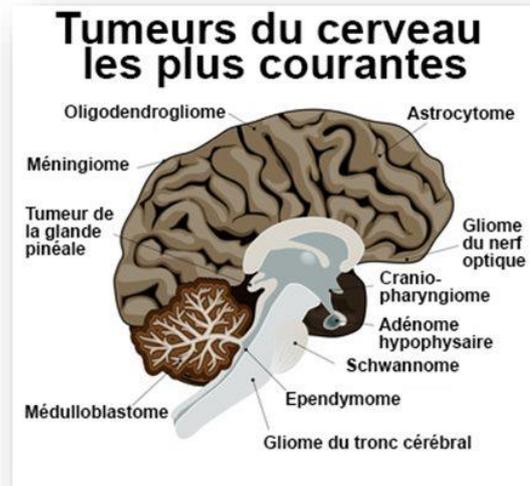


Figure 1-8 : Variété de tumeur cérébrale

Il existe une dizaine de tumeurs cérébrales, on nomme habituellement les tumeurs selon le type de cellule où elles se développent. Voici quelques types de tumeur cérébrale connue :

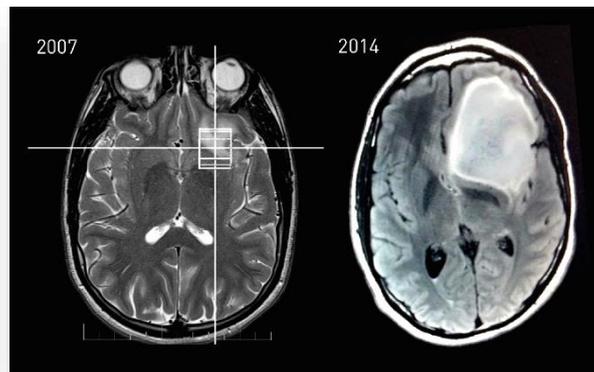


Figure 1-9 : Evolution d'un Astrocytome.

- **Astrocytome** : Tumeur maligne développée au dépend des hémisphères cérébraux (plutôt chez l'adulte) et le cervelet (plutôt chez l'enfant).

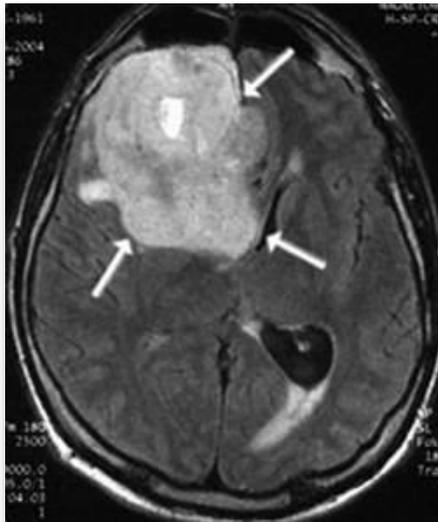


Figure 1-10 : Un Gliome de l'hémisphères.

- **Gliome** : Ils peuvent se situer n'importe où dans les principales régions du cerveau (hémisphères cérébraux, cervelet, tronc cérébral) ou dans la moelle épinière (60% des tumeurs primitives). Ils sont bénins ou malins.

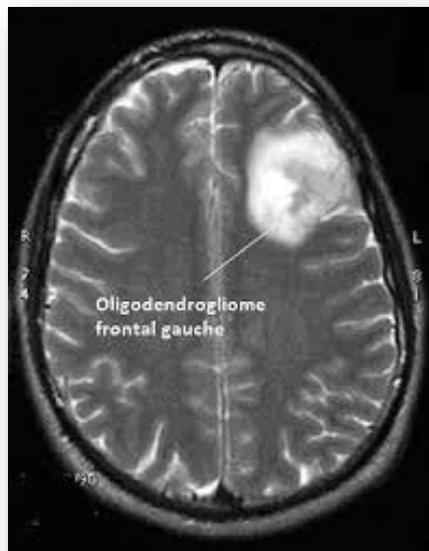


Figure 1-11 : Un Oligodendrogliome.

- **Oligodendrogliome** : est une tumeur primaire du système nerveux central (SNC). Cela signifie qu'il commence dans le cerveau ou la moelle épinière [10].

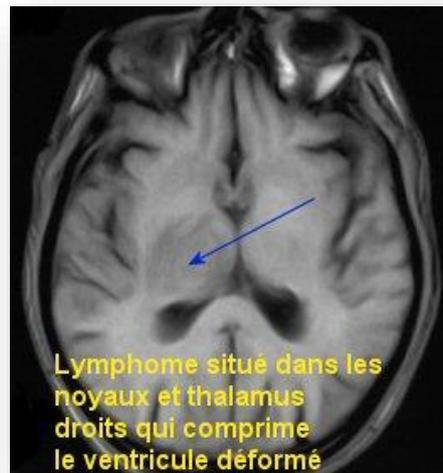


Figure 1-12 : Un Lymphome de cerveau.

- **Lymphome** : un type de lymphome primitif du SNC, ont leur origine dans le tissu nerveux et sont habituellement des tumeurs liées à « la prolifération de lymphocytes B [13].

- **Les lymphocytes B** : sont le support de l'immunité humorale adaptative dont les effecteurs terminaux sont les anticorps ou Immuno- globulines (Ig) qui ont pour fonction de « neutraliser/détruire » les éléments étrangers et pathogènes [14].

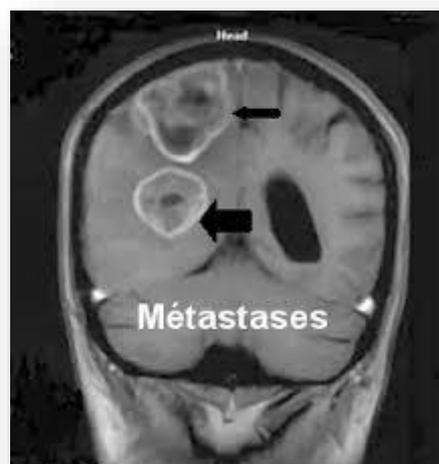


Figure 1-13 : Métastase au cerveau d'autre cancers.

- **Métastase** : Ce sont les tumeurs cérébrales les plus fréquentes. Elles sont malignes ou secondaires à une tumeur d'une autre région du corps (poumon, sein, rein, peau). Des cellules tumorales migrent vers le cerveau en empruntant le plus souvent la voie sanguine. Les métastases les plus fréquentes sont celles du cancer du poumon et du sein.

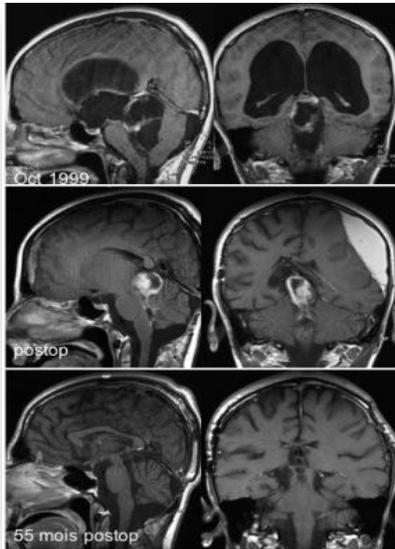


Figure 1-14 : Un Ependymomes intracrâniens .

- **Ependymome** : la tumeur intramédullaire la plus fréquente chez l'adulte (mais il représente seulement 10-12% des tumeurs pédiatriques du système nerveux central), et peut être bénin ou anaplasique. L'Ependymome est issu des cellules épendymaires des ventricules cérébraux, des inclusions corticales et du canal central de la moelle épinière. Dans la majorité des cas il s'agit d'une tumeur bénigne.

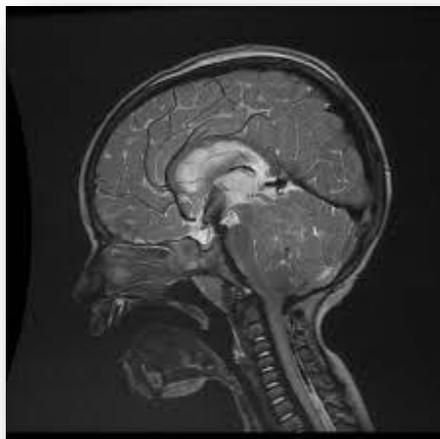


Figure 1-15: Médulloblastome de L'Enfant.

- **Médulloblastome** : Un type de tumeur primitive neuroectodermale (TPN) située dans le cervelet et le quatrième ventricule (un espace rempli de liquide dans l'encéphale).

- **Neuroectodermale** : Groupe de tumeurs malignes du système nerveux qui présentent des cellules primitives avec des éléments de différenciation neuronale et / ou gliale. L'utilisation de ce terme est limitée par certains auteurs aux tumeurs du système nerveux central et d'autres dont les néoplasmes d'origine similaire survivent de façon extra crânienne.

8. Conclusion

Aujourd'hui avec les progrès scientifique et technologique des nouvelles technologies et l'IA beaucoup de travaux et de recherches tentent d'apporter des solutions basées sur les technologies de l'IA pour le diagnostic médical utilisant la méthode du "Machine Learning", ou apprentissage automatique, qui permet à des machines d'effectuer des tâches complexes pour lesquelles elles ont été entraînées, comme pouvant effectivement détecter des maladies allant de cancers à des pathologies ophtalmiques avec autant d'exactitude que des professionnels de santé.

Dans ce chapitre, nous avons présenté une étude sur la tumeur, quelles est ça famille, quelles sont les causes, ses types. Plus le diagnostic est pronostiqué plus le patient a de chance de guérir. Dans notre prochain chapitre on parlera de la méthode qui nous aide à trouver le meilleur modèle d'aide au diagnostic de tumeur cérébrale.

Chapitre 02

1. Introduction

Depuis l'antiquité, le sujet des machines préoccupe les esprits. Ce concept est la base de pensées pour ce qui devient l'IA, ainsi qu'une de ses sous-branches : l'apprentissage automatique. Les dernières années, il consiste à prédire ou reconnaître automatiquement une forme. Cette branche a conquis tous les domaines de la vie face à l'évolution, parmi celle-ci il y a les réseaux bayésiens.

2. Définition

En informatique et en mathématique, un réseau bayésien est un modèle graphique probabiliste représentant un ensemble de variables aléatoires sous la forme d'un graphe orienté acyclique. Intuitivement, un réseau bayésien est à la fois :

1. un modèle de représentation des connaissances ;
2. Une « machine à calculer » des probabilités conditionnelles
3. une base pour des systèmes d'aide à la décision

Ces modèles sont capables d'effectuer des raisonnements probabilistes à partir de données incomplètes alors que peu de méthodes sont actuellement capables d'utiliser les bases d'exemples incomplètes pour leur apprentissage [19].

3. Domaine d'application

Les réseaux bayésiens sont un formalisme de raisonnement probabiliste de plus en plus utilisé pour des tâches aussi diverses que :

- La fouille de données : cette tâche a plusieurs exemples comme dans le domaine aéronautique, une interruption opérationnelle est un retard au départ (décollage) de plus de quinze minutes, une annulation ou une interruption de vol suite à un problème technique (panne ou dysfonctionnement).

De ce fait, lors du lancement de nouveaux projets avions, les ingénieurs doivent fournir dès la phase

de conception une prédiction la plus réaliste possible de la fréquence des interruptions opérationnelles, qui sera mesurée lors de la future exploitation commerciale des avions. Ces prédictions initient, guident et valident les choix de conception.

Pour effectuer cette activité, les ingénieurs utilisent un outil informatique implémentant un modèle mathématique stochastique intégrant les paramètres dont les impacts sur la fréquence des interruptions opérationnelles sont connus. Cet outil est calibré et paramétré par le retour d'expérience obtenu à partir d'avions, de systèmes ou d'équipements en service comparables.

Les besoins de recherche portent sur l'amélioration des modèles de calcul utilisés par cet outil de prédiction. Ainsi, la fouille des données en service est intéressante car elle permet de découvrir de nouveaux facteurs qui pourraient être intégrés à ces modèles pour améliorer la prédiction de la fréquence des interruptions opérationnelles [26].

- La fouille de texte est l'héritière directe de la fouille de données et qui implique la recherche d'information textuelle[20].

- La robotique devient la programmation bayésien robotique qu'elle a proposée un cadre simple et générique pour la programmation des robots en présence d'incertitude et d'incomplétude. Au-delà d'applications robotiques, le formalisme de PBR autorise une reformulation et une comparaison de nombreux modèles probabilistes standards comme les Réseaux Bayésiens.

Tout système robotique est confronté à cette difficulté centrale : comment utiliser un modèle incomplet par rapport à son environnement pour percevoir, inférer, décider et agir efficacement ?

Pour cela, le raisonnement rationnel à partir d'informations incomplètes est un défi pour les systèmes artificiels. Le but de l'inférence et de l'apprentissage bayésien est précisément d'aborder ce problème avec une théorie formelle et bien établie [21].

- En médecine, une même combinaison de symptômes peut être observée pour différentes maladies. Il n'y a donc pas toujours de règles strictes pour obtenir un diagnostic. Pour un système complexe, un expert humain est capable de porter un jugement même lorsque toutes les données nécessaires ne sont pas observées. Un système expert ne peut pas faire cela alors qu'un modèle probabiliste le permet.

De plus, les réseaux bayésiens sont plus facilement adaptés et mis à jour en fonction du contexte que les systèmes à base de règles [20].

Ils sont surtout utilisés aussi pour l'analyse de risques, la détection des spams...etc.

4. Définition formelle

Pour définir un réseau bayésien on va exprimer la loi de probabilité jointe sur l'ensemble des variables aléatoires modélisées dans le réseau .Un réseau bayésien est un graphe orienté acyclique $G = (V, E)$ avec V l'ensemble des nœuds du réseaux et E l'ensemble des arcs. À chaque nœud x appartenant à V du graphe est associé la distribution de probabilité conditionnelle suivante :

$$P(x|pa(x)) \dots\dots\dots \text{équation (1)}$$

Où $pa(x)$ représente les parents immédiats de x dans V . Selon **équation (1)**, L'ensemble V est donc un ensemble de variables aléatoires discrètes. Chaque nœud de V est conditionnellement indépendant de ses non-descendants, étant donné ses parents immédiats. Il est ainsi possible de factoriser les distributions de probabilité conditionnelles sur l'ensemble des variables en en faisant le produit :

$$P(V) = \prod_{x \in V} P(x | pa(x)) \dots\dots\dots \text{équation (2)}$$

Et pour un réseau bayésien, il faut évaluer si le graphe orienté acyclique $G = (V, E)$ vérifie la propriété de Markov globale étant donné une loi de probabilité P sur l'ensemble des variables V . Soit X, Y, S trois sous-ensemble de V , tels que X et Y sont d-séparés par S , noté $(X|S|Y)$: la propriété de Markov globale exprime l'indépendance conditionnelle entre X et Y , *c'est-à-dire* que X est indépendant de Y conditionnellement à S . la propriété dans un graphe orienté s'énonce comme suit :

$$\forall X, Y, S \subset V \text{ disjoints}, (X|S|Y) \implies X \perp\!\!\!\perp Y \mid S$$

Il convient de définir la d-séparation, pour « séparation dirigée » ou « séparation orientée », dans un graphe orienté. Les sous-ensembles X et Y sont d-séparés par S si et seulement si toute chaîne de nœuds de X à Y est « bloquée ou ouverte » par S. Pour trois nœuds x appartenant à X, y appartenant à Y et s appartenant à S.

✓ une chaîne est bloquée: soit la chaîne contient une séquence « $x \rightarrow z \leftarrow y$ » où z n'appartient pas à S et aucun descendant de z appartient à S .

✓ une chaîne est ouverte : soit la chaîne contient une séquence « $x \rightarrow s \rightarrow y$ » ou « $x \leftarrow s \rightarrow y$ » ou « $x \leftarrow s \leftarrow y$ »

L'indépendance conditionnelle, notée $X \perp\!\!\!\perp Y \mid S$, exprime que X est conditionnellement indépendant de Y étant donné S. Formellement :

$$X \perp\!\!\!\perp Y \mid S \text{ si et seulement si } \left\{ \begin{array}{l} P(X \mid Y, S) = P(X \mid S) \\ \text{Et} \\ P(Y \mid X, S) = P(Y \mid S) \end{array} \right.$$

Ou $X \perp\!\!\!\perp Y \mid S$ si et seulement si $P(X, Y \mid S) = P(X \mid S) \times P(Y \mid S)$

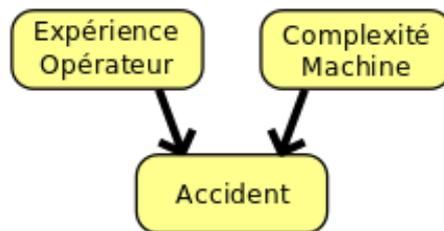


Figure 2-1 : Exemple 1 indépendance conditionnel entre les variables.

Exemple 1:

La probabilité jointe est égale à :

$$P(\text{ExpérienceOpérateur}) \times P(\text{ComplexiteMachine}) \times P(\text{Accident} \mid \text{ExpérienceOpérateur}, \text{ComplexiteMachine})$$

Pour l'exemple de figure 2-1, on peut écrire que l'expérience de l'opérateur est indépendante de la complexité de la machine, mais qu'elle est conditionnellement dépendante de la complexité de la machine étant donné le risque d'accident [24].

Exemple 2 : Soit un réseau bayésien à 5 variables

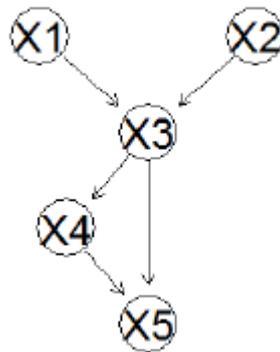


Figure 2-2 : Exemple 2 indépendance conditionnel entre les variables.

En utilisant le théorème de Bayes on a :

$$P(X1,X2,X3,X4,X5) = P(X5|X1 ,X2,X3,X4)P(X1,X2,X3,X4)$$

Or d'après l'hypothèse de Markov le nœud X5 n'est dépendant que de ses parents, donc:

$$P(X1,X2,X3,X4,X5) = P(X5|X3 ,X4)P(X1,X2,X3,X4)$$

de proche en proche on arrive à la décomposition suivante :

$$P(X1,X2,X3,X4,X5) = P(X5|X3,X4)P(X4|X3)P(X3|X2 ,X1)P(X2)P(X1)$$

5. Taches de réseau bayésien

Un graphe orienté acyclique $G = (V, E)$ est un réseau bayésien si et seulement s'il vérifie la propriété de Markov globale orientée étant donné une loi de probabilité P sur l'ensemble des variables V .

Les réseaux bayésiens effectuent trois tâches d'inférence principales :

5.1. Inférer des variables non observées

L'inférence, ou la mise à jour des croyances, consiste à calculer la probabilité à posteriori au niveau de certaines variables en prenant en compte les informations sur d'autres variables appelées observations. Les réseaux bayésiens permettent de mettre à jour les distributions de probabilités au niveau des variables : on commence par fixer des observations, c'est à dire affecter des valeurs aux variables d'observation, puis on utilise un algorithme d'inférence pour calculer les probabilités a posteriori des autres variables [25].

Bien que le problème d'inférence dépende de la complexité du réseau (plus la topologie du réseau est simple, plus l'inférence est facile), il s'agit d'un problème NP-difficile (voir figure 2-3).

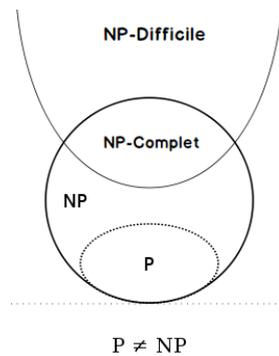


Figure 2-3 : NP Complet.

À la suite de ces résultats, deux grandes catégories d'algorithmes d'inférence viennent naturellement: les algorithmes d'inférence exacte, qui calculent les probabilités à posteriori généralement en temps exponentiel, et les algorithmes d'inférence approchée qui fournissent plutôt une approximation des distributions a posteriori, mais avec une complexité computationnelle moindre.

5.2.Apprentissage des paramètres

On a présenté de nombreuses distributions différentes pour les variables aléatoires, et toutes ces distributions ayant des paramètres: les nombres fournis en entrée lorsqu'on définit une variable aléatoire. Alors lorsqu'on travaille avec des variables aléatoires, soit on donne explicitement les valeurs des paramètres, ou on pourrait deviner les valeurs en comprenant le processus qui génère les variables aléatoires[25]. Ces idées sont critiques pour l'intelligence artificielle. Presque tous les algorithmes d'apprentissage automatique modernes travaillent comme ceci:

- (1) Spécifier un modèle probabiliste ayant des paramètres.
- (2) Apprendre la valeur de ces paramètres de données.

Etant donné un modèle, les paramètres sont les nombres qui donnent la distribution réelle. Dans le cas la variable aléatoire de Bernoulli , le paramètre unique est la valeur p . Dans le cas d'une variable aléatoire Uniforme, les paramètres sont les valeurs a et b qui définissent les valeurs min et max.

Dans la suite, on utilise la notation θ pour être un vecteur de tous les paramètres.

Distribution	Paramètres
Bernoulli(p)	$\theta = p$
Poisson(λ)	$\theta = \lambda$
Uniform(a,b)	$\theta = (a,b)$
Normal(μ,σ^2)	$\theta = (\mu,\sigma^2)$
$Y=mX + b$	$\theta = (m,b)$

Tableau 2-1 : Liste de variables avec leurs paramètres

Dans le monde réel, on ne connaît souvent pas les «vrais» paramètres, mais on doit observer des données. Ensuite, on explorera comment utiliser les données pour estimer les paramètres du modèle.

Pour estimer la valeur des paramètres, il y a deux principales approches: estimation de vraisemblance maximale (EVM) et maximum a posteriori (MAP). Les deux approches supposent que vos données sont des échantillons (IID): X_1, X_2, \dots, X_n où tous les X_i sont indépendants et ont la même distribution.

5.2.1. Estimation vraisemblance maximum

Le premier algorithme d'estimation de paramètres s'appelle estimation de vraisemblance maximale. L'idée principale de cet algorithme est de sélectionner les paramètres (θ) qui rendent les données observées les plus utiles[25].

Les données que l'on va utiliser pour estimer les paramètres vont être n indépendantes et échantillons distribués de manière identique (IID): X_1, X_2, \dots, X_n .

On suppose que les données étaient distribuées de manière identique. Cela signifie qu'elles doivent avoir soit la même fonction de masse de probabilité (si les données sont discrètes), soit la même densité de probabilité fonction (si les données sont continues). Pour simplifier la conversation sur l'estimation des paramètres, on va utiliser la notation $f(X | \theta)$. La nouvelle notation est intéressante de deux manières.

- Premièrement, on a maintenant inclus une condition sur θ qui est notre façon d'indiquer que la probabilité de valeurs différentes de X dépend des valeurs de nos paramètres.
- Deuxièmement, on va utiliser le même symbole f pour les distributions discrètes et continues.

Que signifie vraisemblance et en quoi la «vraisemblance» diffère-t-elle de la «probabilité»? Dans le cas de distributions discrètes, la vraisemblance est synonyme de la probabilité conjointe de vos données. Dans le cas de distribution continue, la vraisemblance fait référence à la densité de probabilité conjointe de vos données.

Vu que chaque point de données est indépendant, la vraisemblance de toutes les données est le produit de la vraisemblance de chaque point de données. Mathématiquement, la vraisemblance que nos données soient données avec les paramètres θ est la suivante:

$$L(\theta) = \prod_{i=1}^n f(X_i | \theta) \dots\dots\dots \text{équation (3)}$$

Pour différentes valeurs des paramètres, la probabilité de nos données sera différente. Si on a des paramètres corrects, les données seront beaucoup plus probables que si on a des paramètres incorrects. C'est pour laquelle on écrit la vraisemblance en fonction de nos paramètres (θ).

Dans l'estimation du maximum de vraisemblance, l'objectif est de choisir les valeurs des paramètres (θ) qui maximise la fonction de vraisemblance de la section précédente. On va utiliser la notation $\hat{\theta}$ pour représenter le meilleur choix de valeurs pour les paramètres. Formellement, l'estimation du maximum de vraisemblance suppose que:

$$\hat{\theta} = \arg \max L(\theta) \dots\dots\dots \text{équation (4)}$$

«arg max» est l'abréviation de l'argument du maximum. L'« arg max » d'une fonction est la valeur du domaine dans lequel la fonction est maximisée. Cela s'applique aux domaines de n'importe quelle dimension.

Une propriété intéressante de « arg max » est que puisque « log » est une fonction monotone, l'« arg max » d'une fonction est le même que l'« arg max » de la fonction log ! C'est bien parce que les logs simplifient les calculs.

Si on trouve l'« arg max » du log de vraisemblance, il sera égal à l'arg max de la vraisemblance. Par conséquent, pour l'estimation du maximum de vraisemblance, on écrit d'abord la fonction de log vraisemblance (LL)

$$LL(\theta) = \log L(\theta) = \log \prod_{i=1}^n f(X_i|\theta) = \sum_{i=1}^n \log f(X_i|\theta) \dots\dots\dots \text{équation (5)}$$

Pour utiliser un estimateur de vraisemblance maximale, on écrit d'abord le log vraisemblance des paramètres données, puis on choisit la valeur des paramètres.

5.2.2. Estimation maximum à posteriori

Le paradigme de l'estimation maximum à posteriori est de choisir la valeur pour des paramètres. À première vue, cela ressemble à l'estimation du maximum de vraisemblance; Cependant, l'estimation du maximum de vraisemblance choisit la valeur des paramètres qui donne les meilleurs données ou qui correspond aux meilleurs données [25]. Formellement, pour les variables aléatoires (IID) : X_1, \dots, X_n

$$\theta_{MAP} = \arg \max_{\theta} f(\theta | X_1, \dots, X_n) \dots\dots\dots \text{équation (6)}$$

Dans l'équation (6) ci-dessus, on essaie de calculer la probabilité conditionnelle des variables aléatoires non distinguées. Etant donné les variables aléatoires qui ont été calculées. On pense alors au théorème de Bayes! On développe la fonction f utilisant la version continue du théorème de Bayes:

$$\begin{aligned} \theta_{MAP} &= \arg \max_{\theta} f(\theta | X_1, \dots, X_n) \\ &= \arg \max_{\theta} \frac{f(X_1, \dots, X_n | \theta) g(\theta)}{h(X_1, \dots, X_n)} \quad \text{par théorème de Bayes} \end{aligned}$$

On note que f, g et h sont toutes des densités de probabilité. On a utilisé différents symboles pour rendre clair les différentes fonctions qu'elles peuvent avoir. On va maintenant tirer parti de deux observations.

- Premièrement, les données sont supposées être IID pour pouvoir décomposer la densité des données en θ .
- Deuxièmement, le dénominateur est une constante par rapport à θ . En tant que tel, sa valeur n'affecte pas l'arg max, donc on peut s'en passer. Mathématiquement:

$$\begin{aligned} \theta_{MAP} &= \arg \max_{\theta} \frac{\prod_{i=1}^n f(X_i|\theta) g(\theta)}{h(X_1, \dots, X_n)} \\ &= \arg \max_{\theta} \prod_{i=1}^n f(X_i|\theta) g(\theta) \dots\dots\dots \text{équation (7)} \end{aligned}$$

Comme auparavant, il sera plus pratique de trouver le arg max du log de la fonction θ_{MAP} , qui nous donne la forme finale pour l'estimation maximum à posteriori des paramètres.

$$\theta_{MAP} = \arg \max_{\theta} (\log(g(\theta)) + \sum_{i=1}^n \log(f(X_i|\theta))) \dots\dots\dots \text{équation (8)}$$

Afin de se préparer à l'estimation maximum à posteriori on a besoin de mettre à jour nos distributions. Par exemple, si on utilise une distribution de Poisson, quel est le type adéquat de variable aléatoire pour prédire λ ?

Voici une liste des différents paramètres et des distributions les plus souvent utilisées pour leurs a priori:

Paramètre	Distribution
Bernoulli p	Beta
Binomial p	Beta
Poisson λ	Gamma
Exponentiel λ	Gamma
Multinomial p_i	Dirichlet
Normal μ	Normal
Normal σ^2	Inverse Gamma

Tableau 2-2 : Liste des paramètres et de leurs distribution a priori

Les distributions utilisées pour représenter la croyance «priori» sur une variable aléatoire auront souvent leurs propres paramètres. Par exemple, une distribution bêta est définie à l'aide de deux paramètres (a, b). Doit-on utiliser l'estimation de paramètres pour évaluer a et b aussi? Non, ces paramètres s'appellent «Hyperparamètres». C'est un terme qu'on réserve aux paramètres d'un modèle qu'on corrige avant d'exécuter l'estimation de paramètre. Avant d'exécuter l'estimation maximum a posteriori, on donne des valeurs pour (a, b).

La distribution de Dirichlet généralise bêta de la même manière que multinomial généralise Bernoulli. La variable aléatoire X de Dirichlet est paramétrée comme $X \sim \text{Dir}(a_1, a_2, \dots, a_m)$.

La distribution gamma (k, θ) est utilisé par le paramètre λ de Poisson (elle est aussi utilisé pour le λ dans l'exponentielle).

Les hyperparamètres peuvent être interprétés comme: on a vu k événements imaginaires totaux au cours de θ imaginaire périodes. Après avoir observé n événements au cours des t périodes suivantes, la distribution postérieure est Gamma ($k + n, \theta + t$).

5.3. Apprentissage des structures

La tâche d'apprentissage de la structure pour les réseaux bayésiens consiste à apprendre la structure du graphe acyclique dirigé (GAD) à partir de données. Les données peuvent être continues ou discrètes. Il existe deux approches principales pour l'apprentissage de la structure: l'approche basée-scores et l'approche basée-contraintes [11].

5.3.1. Approche basée-scores

L'approche basée sur les scores consiste d'abord à définir un critère permettant d'évaluer l'adéquation du réseau bayésien aux données, puis à rechercher dans l'espace des GAD une structure avec un score maximal [17]. Cette approche comprend deux parties: la définition de la métrique du score et l'algorithme de recherche [11].

5.3.1.1. La métrique du score

Les métriques de score pour une structure G ou un graphe et données D qui sont les paramètres de graphe associé, [11] peuvent être généralement définies comme :

$$Score(G:D) = LL(G:D) - \phi(|D|)\|G\| \dots \dots \dots \text{équation (9)}$$

Selon **équation (9)**, LL(G:D) se réfère à la log-vraisemblance des données sous la structure graphique G .

|D| est le nombre d'échantillons, \|G\| est le nombre de paramètres dans le graphique G. Pour calculer $\phi(|D|)$, il y a de nombreuses fonctions de scores tels que l'AIC, le BIC, le BD...etc.

- Pour la fonction AIC (critère d'information d' Akaike) $\phi(|D|)=1$.
- Pour la fonction BIC(le critère d'information bayésien) $\phi(|D|)=\log(|D|)/2$.
- Pour le score BD (Bayésien Dirichlet), si on définit en premier la probabilité de données D conditionnel à la structure du graphe G comme :

$$P(D|G) = \int P(D|G, \Theta_G)P(\Theta_G|G)d\Theta_G \dots \dots \dots \text{équation (10)}$$

Selon **équation (10)**, P(D|G,Θ_G) est la probabilité des données compte tenu de la structure et des paramètres du réseau, et P(Θ_G|G) est la probabilité a priori des paramètres. Lorsque la probabilité a priori est spécifiée sous forme de distribution de Dirichlet,

$$P(D|\Theta_G) = \prod_i \prod_{\pi_i} \left[\frac{\Gamma \sum_j N'_{i,\pi_i,j}}{\Gamma(\sum_j N'_{i,\pi_i,j} + N'_{i,\pi_i,j})} \prod_j \frac{\Gamma(\sum_j N'_{i,\pi_i,j} + N'_{i,\pi_i,j})}{\Gamma(N'_{i,\pi_i,j})\pi_i} \right] \dots \dots \dots \text{équation (11)}$$

Selon **équation (11)**, π_i fait référence à la configuration parente de la variable i et $N_{i,\pi_i,j}$ est le nombre de variable i prendre de la valeur j avec configuration parent π_i . N' représente les comptes dans le priori respectivement.

Avec un priori pour la structure du graphe P(Θ_G) (par exemple, un uniforme), le score BD est défini comme :

$$\log P(D|\Theta_G) + \log (P(\Theta_G))$$

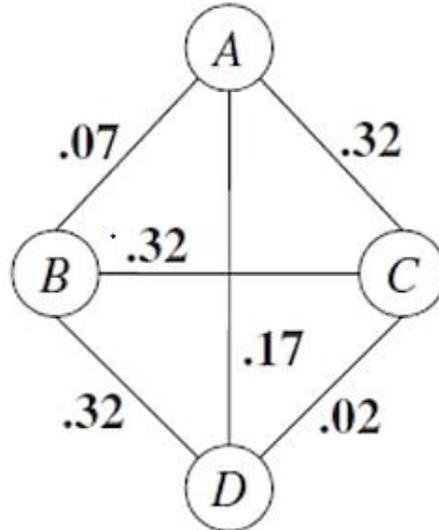
Algorithme de Chow-Liu

L'algorithme de Chow-Liu est un type spécifique d'approche basée sur les scores. Il utilise la structure arborescente de maximum de vraisemblance dans laquelle chaque nœud a au plus un parent. On note que le score est simplement le maximum de vraisemblance.

L'algorithme comporte trois étapes :

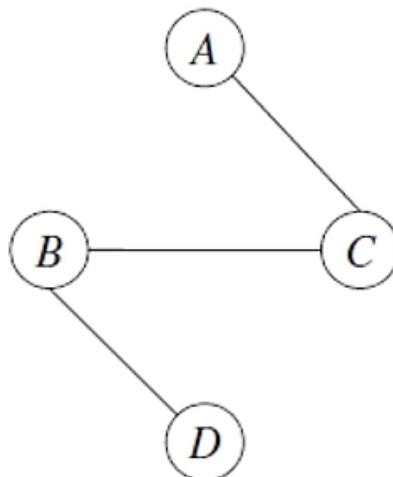
1) Calculer les informations communes pour toutes les paires de variables X,U, et former le graphe d'information commune où les variables X,U a un poids MI(X,U) entre elles:

$$MI(X, U) = \sum_{x,u} \hat{p}(x, u) \log \left[\frac{\hat{p}(x, u)}{\hat{p}(x)\hat{p}(u)} \right] \dots\dots\dots \text{équation (12)}$$



$$\hat{p}(x, u) = \frac{\text{Count}(x, u)}{\# \text{ data points}} \dots\dots\dots \text{équation (13)}$$

2) Trouver l'arbre d'interpolation de poids maximal: l'arbre de poids maximal qui relie tous les sommets d'un graphique. Ceci peut être trouvé en utilisant des algorithmes de Kruskal ou de Prim.



3) Choisir n'importe quel nœud la variable racine et assigner les directions qui partent de ce nœud (les flèches s'éloignent de lui). Cette étape transforme l'arbre non dirigé résultant en un arbre dirigé.



Maintenant qu'on a décrit l'algorithme, on explique pourquoi cela fonctionne. Il se trouve que le score vraisemblance se décompose en informations mutuelles et en termes d'entropie :

$$\log p(D|\theta^{ML}, G) = |D| \sum_i MI_{\hat{p}}(X_i, X_{pa(i)}) - |D| \sum_i H_{\hat{p}}(X_i) \dots\dots \text{équation (14)}$$

On cherche un graphique G qui maximise cette log-vraisemblance. Les entropies étant indépendantes de l'ordre de dépendance dans l'arborescence, les seuls termes qui changent avec le choix de G sont les termes d'information commune. Donc on veut:

$$\arg \max \log P(D|\theta^{ML}(G), G) = \arg \max \sum_i MI(X_i, X_{pa(i)})$$

Maintenant si on suppose $G=(V,E)$ est un arbre où chaque nœud a au plus un parent, on obtient :

$$\arg \max \log P(D|\theta^{ML}(G), G) = \arg \max \sum_{i,j \in E} MI(X_i, X_{pa(i)}) \dots\dots \text{équation (15)}$$

5.3.1.2. Algorithmes de recherche

Le choix le plus courant pour les algorithmes de recherche est la recherche locale et la recherche gloutonne [17][11].

Pour l'algorithme de recherche locale, il commence par un graphique vide ou un graphique complet. À chaque étape, il tente de modifier la structure du graphique en ajoutant une arête, en supprimant une arête ou en inversant une arête. (Bien entendu, l'opération doit conserver la propriété acyclique.) Si le score augmente, il adopte la tentative et effectue la modification, sinon il tente une autre tentative.

Pour la recherche gloutonne on suppose d'abord un ordre topologique du graphe. Pour chaque variable, on limite son ensemble parent aux variables de rang supérieur. Lors de la recherche d'un ensemble parent pour chaque variable, il faut adopter une approche gloutonne en ajoutant le parent qui augmente le plus le score jusqu'à ce qu'aucune amélioration ne soit possible.

5.3.2. Approche basée-contraintes

Le cas basé sur les contraintes utilise le test d'indépendance pour identifier un ensemble de contrainte de côté pour le graphique, puis recherche le meilleur GAD qui satisfait les contraintes. Par exemple, on pourrait distinguer la structure en V et la structure en fourche en effectuant un test d'indépendance pour les deux variables de côté conditionnellement à la variable du milieu. Cette approche fonctionne bien avec

d'autres connaissances (expertes) préalables de la structure, mais nécessite de nombreux échantillons de données pour garantir la puissance de test [11].

6. Apprentissage des paramètres Vs Apprentissage de la structure

Ce tableau définit la différence entre l'apprentissage de la structure et l'apprentissage des paramètres et que chacune d'entre elles doit prendre son rôle: l'apprentissage de structure doit être le premier suivi par l'apprentissage des paramètres ; il faut connaître la structure pour continuer passer à l'étape suivante qui est l'apprentissage des paramètres.

Apprentissage des paramètres	Apprentissage de la structure
<ul style="list-style-type: none">- Estimer les tables de probabilités d'un réseau- Peuvent modifier leur connaissance des probabilités antérieures des variables ou des relations statistiques entre les variables considérées comme potentiellement dépendantes	<ul style="list-style-type: none">- Estimer la structure d'un réseau- Apprendre de nouvelles relations entre des variables qui ne sont pas considérées comme potentiellement dépendantes.

Tableau 2-3 : Comparaison entre apprentissage de la structure et les paramètres.

7. Conclusion

Les réseaux bayésiens est une partie de l'apprentissage automatique qui nous présente des techniques variés et efficace pour trouver le meilleur modèle pour l'aide de décision.

Dans notre travail, on s'intéresse à l'application de ce type dans le domaine médical et spécialement le diagnostic pour la tumeur cérébrale. Cette solution va aider les oncologues et les pathologistes à prendre une décision fiable concernant le type de la tumeur. Cette étape est très importante pour le traitement.

Chapitre 03

1. Introduction

Notre objectif est de présenter une solution adaptée basée sur les réseaux bayésiens pour l'aide au diagnostic des tumeurs cérébrales. Cette opération est nécessaire pour pouvoir prescrire le bon traitement.

Nous commençons ce chapitre par une présentation du schéma général du modèle. Ensuite, on va proposer des structures ou des méthodes puis on va modéliser ces paramètres de la structure et enfin expliquer le principe de chaque partie de travail.

2. Schéma général

La figure 3-1 représente un schéma de notre travail pour l'aide au diagnostic de tumeur cérébrale.

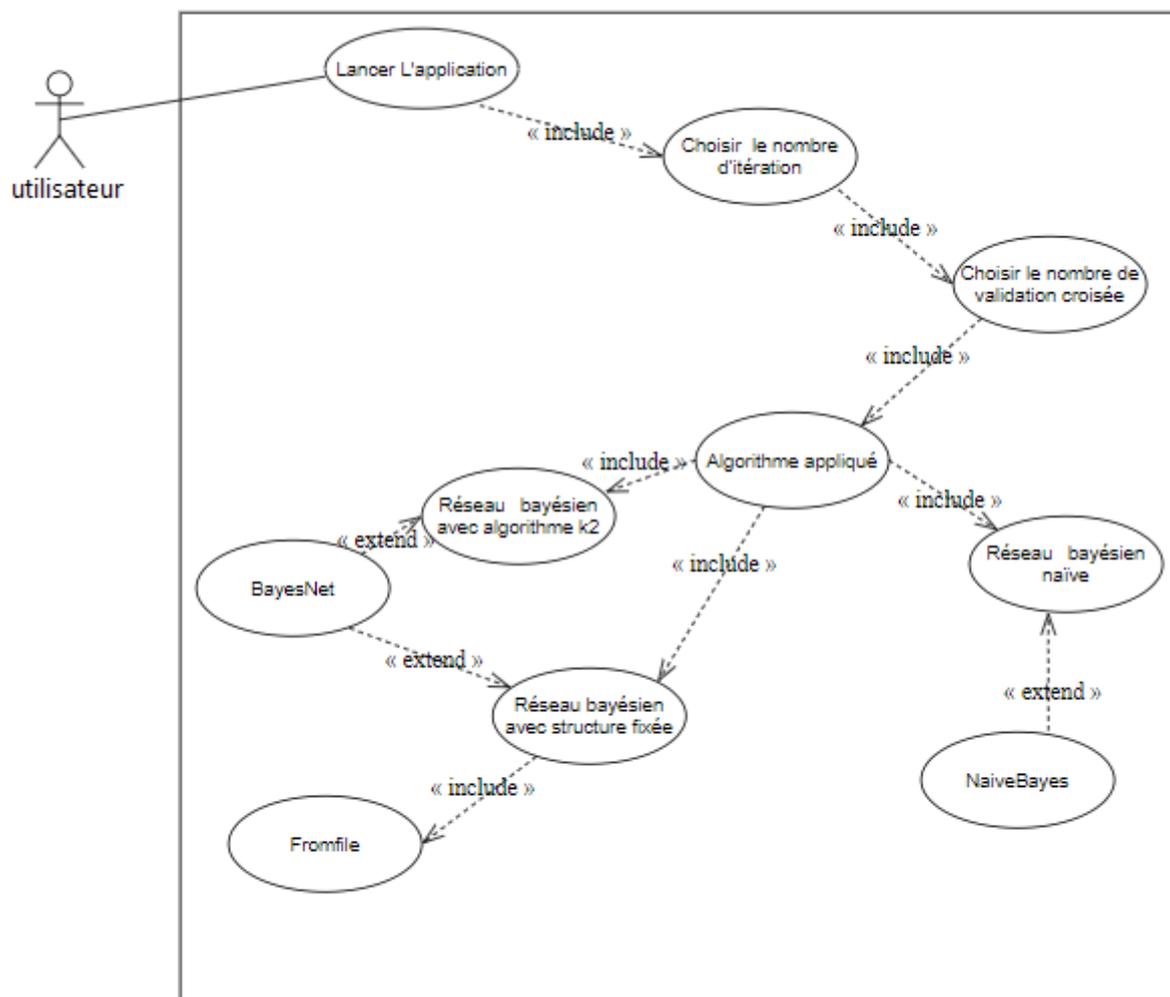


Figure 3-1 : Schéma général de travail

On a fait un système pour que l'utilisateur trouve le bon modèle d'aide au diagnostic de tumeur cérébrale et ce schéma résume comment faire le travail.

Au début on lance l'application puis on choisit des itérations à suivre, en choisissant un nombre de validation croisée. Chaque itération applique plusieurs techniques. Pour chaque technique on utilise une entité mère sauf l'entité réseau bayésien avec structure fixée qui a besoin d'une autre entité appelé Fromfile.

3. Présentation du modèle

La figure 3-2, représente les étapes de base nécessaires à la conception et le développement d'un modèle d'aide au diagnostic des tumeurs cérébrales en utilisant les réseaux bayésiens. Ces étapes sont : la modélisation de la structure, la modélisation des paramètres par apprentissage et l'inférence.

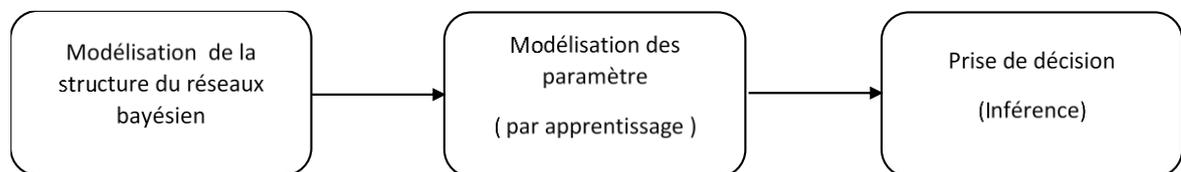


Figure 3-2 : Etapes nécessaires à la modélisation de l'approche utilisée

La modélisation de la structure du réseau bayésien est une phase assez complexe. Cette difficulté est due au grand nombre de caractéristiques qui entrent en jeu dans le diagnostic des tumeurs cérébrales.

Dans certains cas, la structure du réseau Bayésien est fournie à priori par un expert. Une fois la structure du réseau bayésien fixée, on passe à définir l'ensemble des paramètres. Cette définition peut être accomplie soit par l'expert du domaine soit par apprentissage de paramètres de façon automatique et donc à déterminer les probabilités conditionnelles de chaque variable qui se trouve dans le réseau. Les données disposées peuvent être complètes ou incomplètes, discrètes ou continues [26]. Pour chaque cas, l'algorithme d'apprentissage des paramètres diffère (Voir Chapitre 2 Section 5.2. Apprentissage des paramètres Page 20).

La troisième phase est une étape où on valide la structure et les paramètres du réseau bayésien. Elle permet d'injecter des évidences dans la meilleure modélisation obtenue afin de prendre la décision la plus adéquate (Voir Chapitre 2 Section 5.1. Inférer des variables non observées page 19).

4. Modélisation de la structure du réseau bayésien

La modélisation de la structure du réseau bayésien est très importante. Elle permet de donner les connaissances utilisées par les non spécialistes sur cette problématique de manière simple et compréhensible. La structure est directement déduite à partir du raisonnement effectué par le domaine médical comprenant seize nœuds répartis sur quatre niveaux. Les caractéristiques qui sont mises en jeu pour la prise de décision dans le cadre de diagnostic des tumeurs cérébrales sont représentées au niveau 1. Le deuxième et le troisième niveau sont des nœuds intermédiaires de décision.

Le quatrième niveau est composé d'un seul nœud qui indique le type de la tumeur cérébrale (nœud de décision). La structure de ce réseau est donnée par la figure 3-3.

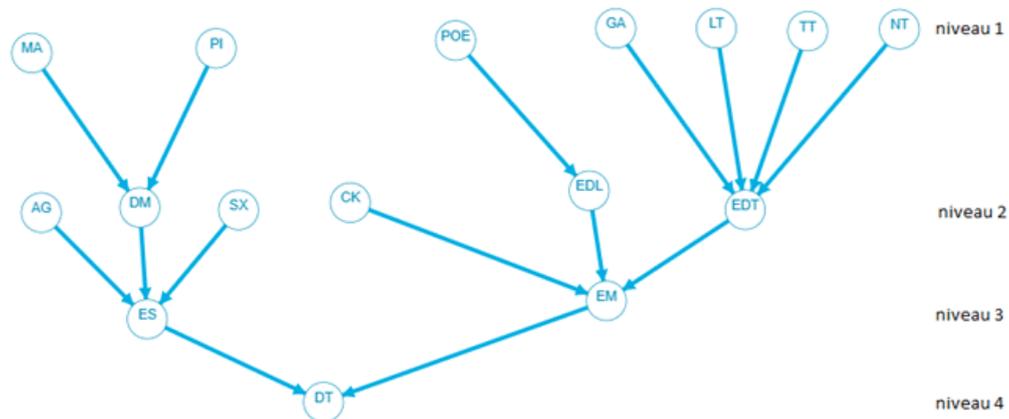


Figure 3-3 : Structure du réseau bayésien déduite à partir du raisonnement effectué par le domaine médical.

Nom du nœud	Signification	Nom du nœud	Signification
MA	Maladies Auxiliaires	DM	Dossier Médical
PI	Première Infection	SX	Sexe
Poe	Présence Œdème	CK	Composante Kystique
GA	Grade	EDL	Etat Données Liquides
LT	Location	EDT	Etat Données Tumeur
TT	Taille Tumeur	ES	Etat clinique
NT	Nombre Tumeur	EM	Etat radiologique
AG	Age	DT	Décision Tumeur

Tableau 3-1 : Signification des nœuds du réseau bayésien.

4.1. Choix de la structure

Dans la suite, on va décrire les contraintes de la structure qui augmentent la précision dans le diagnostic des tumeurs cérébrales:

- Changement de l'orientation des arcs en conservant le même nombre de variables,
- Elimination des variables,
- Modélisation par réseau bayésien naïf,
- Modélisation par algorithme de recherche k2.

4.1.1. Changement de l'orientation des arcs

La stratégie consiste à définir les caractéristiques prioritaires dans le diagnostic et de les orienter directement au nœud de décision « DT ». On signale à cet effet, que plusieurs critères entrent dans l'aide la décision des tumeurs cérébrales. On cite dans un premier plan le grade, la localisation de la tumeur et la composante kystique

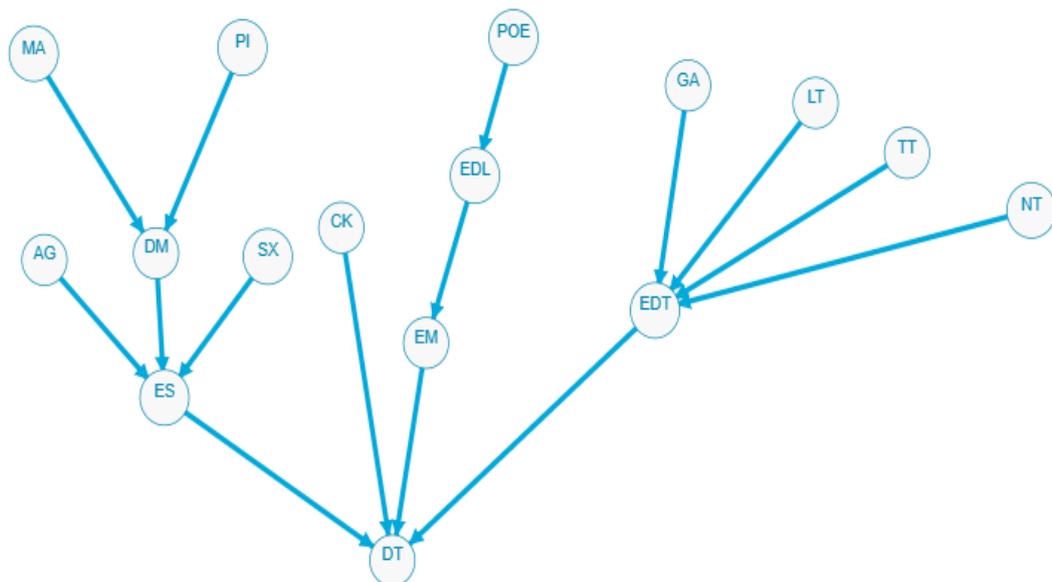


Figure 3-4 : Structure de notre réseau bayésien en changeant l'orientation des arcs.

en plus de l'œdème c'est un choix facultatif. La figure 3-4 illustre le changement de la direction d'arc de deux nœuds « CK » et « EDT » vers le nœud décision au lieu d'être dirigée vers le nœud « EM ». Ces nœuds regroupent la majorité des caractéristiques qui entrent dans le diagnostic des tumeurs cérébrales.

En outre, les nœuds « GA » et « LT » sont les deux caractéristiques similaires dans les tumeurs étudiées. En effet, ces nœuds jouent un rôle important dans la prise de décision.

4.1.2. Elimination des variables

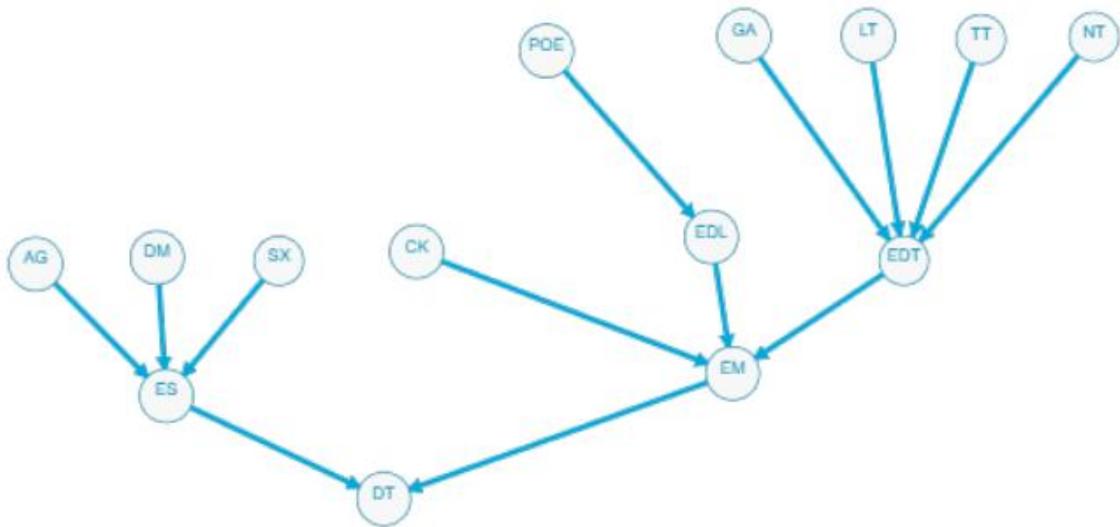


Figure 3-5 : Structure de notre réseau bayésien en éliminant des variables

Le principe consiste à éliminer les variables qui ne sont pas prioritaires dans le diagnostic ou celles qui possèdent beaucoup des valeurs manquantes. En supprimant les nœuds « MA » et « PI » la structure obtenue est la même que la structure initiale mais en éliminant les nœuds « MA » et « PI ». Ces deux caractéristiques n'entrent pas dans le diagnostic de la majorité des tumeurs.

4.1.3. Modélisation par réseau bayésien naïf

Tout d'abord le réseau bayésien naïf utilise le théorème de Bayes avec l'hypothèse de l'indépendance conditionnelle des caractéristiques de la tumeur cérébrale conditionnellement aux caractéristiques du diagnostic.

4.1.4. Modélisation par algorithme de recherche k2

La double nature d'un réseau bayésien fait de l'apprentissage d'un réseau bayésien en deux étapes une division naturelle : apprendre d'abord une structure de réseau, puis apprendre les tables de probabilité [20].

L'idée de l'algorithme K2 est de maximiser la probabilité de la structure sachant les données dans l'espace de graphe $G = (V, E)$ respectant cet ordre d'énumération. Cette technique utilise originellement le score bayésien (on a utilisé la fonction AIC) et teste l'ajout de parents de la manière suivante : le premier nœud ne peut pas posséder de parents et pour les nœuds suivants, l'ensemble des parents choisis est le sous-ensemble de nœuds qui augmente le plus le score parmi l'ensemble des nœuds qui le précèdent dans l'ordre d'énumération.

5. Modélisation des paramètres

Les sources d'incertitude dans le cadre d'un diagnostic médical sont présentes dans la majorité de modèles d'intelligence artificielle. L'incertitude peut être provoquée par l'absence d'un historique clinique complet et détaillé pour les patients. En plus, ces derniers ne se rappellent pas parfois de tous les symptômes pendant l'évaluation de leurs maladies.

Pendant la phase d'interrogation, l'information donnée par le patient peut être mal exprimée et il peut même mentir au médecin. En médecine, beaucoup de données sont imprécises en particulier celles qui proviennent des images IRM et de tests de laboratoire. Ces causes conduisent à une base d'apprentissage souvent incomplète qui nécessite des algorithmes appropriés pour travailler dessus.

On va utiliser réseau bayésien avec simple estimation de paramètre. Une fois la structure du réseau apprise, on doit choisir comment apprendre les tables de probabilité en sélectionnant une classe. Une méthode simple est utilisée pour produire des estimations directes des probabilités conditionnelles, c'est-à-dire

$$P(x_i = k | pa(x_i) = j) = \frac{N_{ijk} + N'_{ijk}}{N_{ij} + N'_{ij}}$$

où N'_{ijk} est le paramètre alpha qui peut être défini et est 0,5 par défaut. Avec alpha = 0, nous obtenons des estimations du maximum de vraisemblance.

5.1. Description de la base

Cette section a pour objectif de présenter les paramètres qu'on va modéliser. Nous possédons une base de données sur les tumeurs cérébrales sous le format ARFF (Attribute Relation File Format) qui se compose de 87 patients (enregistrements):

```
@relation tumeur_cerebrale

@attribute AG {enfant,adult,jeune_adult}
@attribute SX {h,f}
@attribute DM {incomplet,complet}
@attribute MA {non,oui}
@attribute PI {dd,c,pd,dm,mp,ma,mo,v,pe,mt}
@attribute CK {n,e}
@attribute EDL {non,oui}
@attribute POE {p,n}
@attribute ES {n,p}
@attribute LT {'fosse posterieure','tissu nerveux',melanome,'tissu de soutien glie',rein,colorectal,poumon,astrocytes,oligodendrocytes,'cellules epitheliales'}
@attribute NT numeric
@attribute GA {haut,faible}
@attribute TT {grand,petit,moyen}
@attribute EM {p,n}
@attribute EDT {n,p}
@attribute DT {medulloblastom,lymphome,metastase,gliome,autre,astrocytome,oligodendrogliomes,ependymomes}

@data
enfant,h,incomplet,non,dd,n,non,p,n,'fosse posterieure',7,?,grand,p,n,medulloblastom
adult,f,incomplet,?,c,n,oui,n,n,'tissu nerveux',3,?,petit,n,n,lymphome
enfant,h,complet,oui,pd,n,oui,p,p,melanome,7,haut,grand,n,p,metastase
adult,h,incomplet,oui,dm,n,?,p,n,'tissu nerveux',6,haut,petit,n,n,lymphome
adult,h,incomplet,?,mp,n,?,n,p,'tissu de soutien glie',6,faible,grand,p,p,gliome
adult,h,complet,non,ma,e,non,p,n,'tissu de soutien glie',1,haut,grand,p,n,gliome
enfant,f,complet,?,dm,n,oui,p,p,rein,7,faible,petit,n,n,autre
adult,h,incomplet,oui,ma,e,?,p,p,melanome,6,haut,grand,p,n,autre
enfant,f,complet,non,c,e,non,p,n,melanome,2,haut,moyen,n,n,metastase
adult,h,incomplet,non,mo,e,?,n,n,colorectal,2,haut,petit,n,p,metastase
adult,h,incomplet,non,mp,e,non,p,p,'tissu de soutien glie',4,faible,moyen,p,p,gliome
adult,f,complet,non,c,e,oui,p,n,'tissu de soutien glie',4,haut,grand,n,n,gliome
```

Figure 3-6 : Base de données de la tumeur cérébrale

Cette base de données contient 3 parties, à savoir:

La figure 3-7 définit la première caractéristique de fichier ARFF et qui est le nom de l'ensemble des données.

```
@relation tumeur_cerebrale
```

Figure 3-7 : Première partie de la base de données format ARFF

La figure 3-8 est la deuxième caractéristique, elle définit les attributs de la base de données et leurs types.

```
@attribute AG {enfant,adult,jeune_adult}
@attribute SX {h,f}
@attribute DM {incomplet,complet}
@attribute MA {non,oui}
@attribute PI {dd,c,pd,dm,mp,ma,mo,v,pe,mt}
@attribute CK {n,e}
@attribute EDL {non,oui}
@attribute POE {p,n}
@attribute ES {n,p}
@attribute LT {'fosse posterieure','tissu nerveux',melanome,'tissu de soutien glie',rein,colorectal,poumon,astrocytes,oligodendrocytes,'cellules epitheliales'}
@attribute NT numeric
@attribute GA {haut,faible}
@attribute TT {grand,petit,moyen}
@attribute EM {p,n}
@attribute EDT {n,p}
@attribute DT {medulloblastom,lymphome,metastase,gliome,autre,astrocytome,oligodendroglomes,ependymomes}
```

Figure 3-8 : Deuxième partie de la base de données format ARFF

Voici la description des attributs de la base de données :

AG : définit l'âge de patient soit un enfant, un adulte ou un jeune adulte.

SX : Sexe du patient (femme, homme).

DM : le dossier médical de patient soit complet ou non.

MA : les maladies auxiliaires est ce qu'il y a ou non.

PI : la première infection de maladie (vertiges, modification de parole, modification d'audition, perte de poids ...etc.)

CK : est-ce que le patient a des composantes kystiques ou non.

EDL : l'état donné liquide est que l'œdème liquide ou non (oui, non)

POE : la présence de l'œdème positif ou négatif.

ES : l'état clinique de résultat d'existence d'œdème soit liquide ou non (négatif, positif)

LT : la location de la tumeur ; chaque type de la tumeur à sa place exacte.

NT : la nombre des tumeurs qui existe dans la corp de patient.

GA : le grade ou le stade de tumeur (faible, haut).

TT : la taille de tumeurs soit petite, moyen ou grand.

EM : l'état radiologique de la tumeur, est ce qu'il y a une tumeur (positif ou négatif)

EDT : l'état donné de la tumeur définit les détails principaux pour diagnostiquer une tumeur

DT : décision de la tumeur est le type de la tumeur qu'a le patient.

```
@data
enfant,h,incomplet,non,dd,n,non,p,n,'fosse posterieure',7,?,grand,p,n,medulloblastom
adult,f,incomplet,?,c,n,oui,n,n,'tissu nerveux',3,?,petit,n,n,lymphome
enfant,h,complet,oui,pd,n,oui,p,p,melanome,7,haut,grand,n,p,metastase
adult,h,incomplet,oui,dm,n,?,p,n,'tissu nerveux',6,haut,petit,n,n,lymphome
adult,h,incomplet,?,mp,n,?,n,p,'tissu de soutien glie',6,faible,grand,p,p,gliome
adult,h,complet,non,ma,e,non,p,n,'tissu de soutien glie',1,haut,grand,p,n,gliome
enfant,f,complet,?,dm,n,oui,p,p,rein,7,faible,petit,n,n,autre
adult,h,incomplet,oui,ma,e,?,p,p,melanome,6,haut,grand,p,n,autre
enfant,f,complet,non,c,e,non,p,n,melanome,2,haut,moyen,n,n,metastase
adult,h,incomplet,non,mo,e,?,n,n,colorectal,2,haut,petit,n,p,metastase
adult,h,incomplet,non,mp,e,non,p,p,'tissu de soutien glie',4,faible,moyen,p,p,gliome
adult,f,complet,non,c,e,oui,p,n,'tissu de soutien glie',4,haut,grand,n,n,gliome
adult,h,complet,?,mp,e,non,p,p,'tissu de soutien glie',3,haut,moyen,p,p,gliome
enfant,h,incomplet,oui,v,e,non,p,p,poumon,2,?,moyen,p,n,metastase
enfant,h,incomplet,non,c,e,oui,p,n,astrocytes,4,haut,grand,p,p,astrocytome
adult,h,complet,?,dd,n,non,p,p,oligodendrocytes,1,faible,petit,n,p,oligodendrogliomes
adult,f,incomplet,oui,pe,e,?,n,n,oligodendrocytes,4,faible,grand,p,p,oligodendrogliomes
jeune_adult,f,complet,oui,dm,e,?,n,p,'cellules epitheliales',3,faible,grand,p,n,ependymomes
jeune_adult,f,incomplet,oui,pe,e,oui,p,p,'cellules epitheliales',6,haut,petit,p,n,ependymomes
jeune_adult,f,incomplet,non,pe,n,non,p,p,'cellules epitheliales',3,faible,grand,p,n,ependymomes
adult,f,complet,oui,dm,e,non,n,n,oligodendrocytes,3,faible,moyen,p,p,oligodendrogliomes
adult,f,complet,oui,c,n,non,n,p,oligodendrocytes,5,haut,grand,n,n,oligodendrogliomes
enfant,f,complet,non,mo,n,oui,n,n,astrocytes,2,haut,grand,n,p,astrocytome
jeune_adult,f,incomplet,non,pe,e,oui,n,n,'cellules epitheliales',5,faible,petit,n,p,ependymomes
jeune_adult,f,complet,oui,pe,e,non,n,n,'cellules epitheliales',1,haut,moyen,n,n,ependymomes
enfant,h,complet,oui,mt,n,oui,p,n,astrocytes,2,faible,petit,n,p,astrocytome
enfant,h,complet,oui,dd,e,non,p,p,poumon,2,faible,moyen,p,n,metastase
adult,f,incomplet,oui,pd,e,non,n,n,colorectal,1,haut,petit,n,n,metastase
```

Figure 3-9 : Troisième partie de la base de données format ARFF

La figure 3-9 représente la dernière caractéristique de la base de données et qui sont les instances. On écrit @data pour commencer cette partie.

5.2. Format structure de réseau

On a travaillé aussi avec un fichier de format xml appelé XML BIF(XML à base BayesNets Interchange Format) pour définir les structures de réseau utilisé dans le travail, voir la figure 3-10.

```

<?xml version="1.0"?>
<!-- DTD for the XMLBIF 0.3 format -->
<!DOCTYPE BIF [
  <!ELEMENT BIF ( NETWORK )*>
    <!ATTLIST BIF VERSION CDATA #REQUIRED>
    <!ELEMENT NETWORK ( NAME, ( PROPERTY | VARIABLE | DEFINITION )* )>
    <!ELEMENT NAME (#PCDATA)>
    <!ELEMENT VARIABLE ( NAME, ( OUTCOME | PROPERTY )* )>
    <!ATTLIST VARIABLE TYPE (nature|decision|utility) "nature">
    <!ELEMENT OUTCOME (#PCDATA)>
    <!ELEMENT DEFINITION ( FOR | GIVEN | TABLE | PROPERTY )*>
    <!ELEMENT FOR (#PCDATA)>
    <!ELEMENT GIVEN (#PCDATA)>
    <!ELEMENT TABLE (#PCDATA)>
    <!ELEMENT PROPERTY (#PCDATA)>
]
]>

<BIF VERSION="0.3">
<NETWORK>
<NAME>tumeur_cerebrale</NAME>
<VARIABLE TYPE="nature">
<NAME>AG</NAME>
<OUTCOME>enfant</OUTCOME>
<OUTCOME>adult</OUTCOME>
<OUTCOME>jeune_adult</OUTCOME>
<PROPERTY>position = (81,210)</PROPERTY>
</VARIABLE>
<VARIABLE TYPE="nature">
<NAME>SX</NAME>
<OUTCOME>h</OUTCOME>
<OUTCOME>f</OUTCOME>
<PROPERTY>position = (225,209)</PROPERTY>
</VARIABLE>

```

Figure 3-10 : Une partie de BIF fichier

Le BIF fichier contient une balise XML initiale qui affiche des informations de base sur le document:

```
<? xml version = "1.0"?>
```

Il faut incorporer toute la spécification du document en haut du document via une balise ! DOCTYPE, cette balise contient toute la définition de type de document (DTD) pour XMLBIF (BIF fichier).

```

<! DOCTYPE BIF [
  <! ELEMENT BIF (NETWORK) *>
    <! ATTLIST BIF VERSION CDATA #REQUIRED>
    <! ELEMENT NETWORK (NOM, (PROPERTY | VARIABLE | DEFINITION)
*)>
  <! NOM DE L'ÉLÉMENT (#PCDATA)>
  <! ÉLÉMENT VARIABLE (NOM, (RÉSULTAT | PROPRIÉTÉ) *)>
    <! ATTLIST VARIABLE TYPE (nature | décision | utilité)
"nature">
  <! RÉSULTAT DE L'ÉLÉMENT (#PCDATA)>
  <! DÉFINITION DE L'ÉLÉMENT (POUR | DONNÉ | TABLE | PROPRIÉTÉ)
*>

```

```
<! ELEMENT POUR (#PCDATA)>
<! ÉLÉMENT DONNÉ (#PCDATA)>
<! TABLEAU DES ÉLÉMENTS (#PCDATA)>
<! PROPRIÉTÉ DE L'ÉLÉMENT (#PCDATA)>
]>
```

La DTD affiche toutes les balises utilisées par XMLBIF. Le réseau représenté dans un fichier doit être entouré de balises BIF.

Chaque réseau est entouré de balises NETWORK. Un réseau est défini par une balise NAME et une séquence de balises PROPERTY, VARIABLE et DEFINITION.

```
<VERSION BIF = "0.3">
<NETWORK>
<NAME> tumeur cérébrale </NAME>
<PROPERTY> information </PROPERTY>
<VARIABLE> un nœud dans le réseau et leur information </VARIABLE>
...
</NETWORK>
</BIF>
```

Une propriété n'est qu'une chaîne de texte arbitraire à associer à un bloc. Les variables sont définies par leurs noms, types et propriétés :

```
<TYPE VARIABLE = "nature">
  <NAME> dossier_médicale </NAME>
  <OUTCOME> complet </OUTCOME>
  <OUTCOME> non_complet </OUTCOME>
  <PROPERTY> position = (73, 165) </PROPERTY>
</VARIABLE>
```

Les probabilités conditionnelles sont spécifiées à l'intérieur de la balise DEFINITION.

```
<DEFINITION>
  <FOR> un nœud dans le réseau </FOR>
  <GIVEN> le parent </GIVEN>
  <TABLE> leur table de probabilité </TABLE>
</DEFINITION>
```

6. Les algorithmes appliqués

6.1. Algorithme principal

L'algorithme 3-1 présente notre programme principal qui a des techniques requises avec leurs précisions dans chaque itération et un nombre de validation croisée différent.

Algorithme principal

Var

rb: Réseau bayésien;
 db: Bases de données;
 nombre_itération, nombre_validation_croisée, Choix_Réseau, Choix_Structure: entier;
 bif1, bif2, bif3 : Fichier XML; // Dans lequel la structure du réseau est sauvegardée
 p: réel ; // Pour retourner la précision.
 Liste_CrossValidation: Liste des réel ;

Debut

lire(db);
 lire (nombre_itération);
 Liste_CrossValidation←Nil ;

Pour i = 1 à nombre_itération **faire**

Debut

lire (nombre_validation_croisée);
Pour j = 1 à nombre_validation_croisée **faire**
Debut
 Ecrire('un menu à l'utilisateur pour choisir d'appliquer l'une de ces méthode');
 lire (Choix_Réseau); // Choix de la méthode du réseau bayésien à exécuter
cas (Choix_Réseau)
 'Bayésien_Structure': // Choix de la structure du réseau bayésien

Debut

Lire(rb);
 lire (Choix_Structure); // Choix de la structure du réseau bayésien à utiliser

cas (Choix_Structure')

'Structure_Initiale': //1ière méthode: réseau bayésien avec structure initiale

Debut

// Lire la structure du fichier XML en choisissant celui de la structure fixée bif1

Parcourir_BIFXML(BIFXML, 'bif1') ;

Fin ;

'Structure_Changement_Arcs': //2eme méthode : réseau bayésien avec changement des arcs

Debut

// Lire la structure du fichier XML en choisissant la structure fixée après suppression des arcs: bif2

Parcourir_BIFXML(BIFXML, 'bif2') ;

Fin ;

'Structure_Supprimer_noeuds':

// 3eme méthode : réseau bayésien avec la suppression des nœuds

Debut

// Lire la structure du fichier XML en choisissant la structure fixée après suppression des arcs: bif3

Parcourir_BIFXML(BIFXML, 'bif3');

Fin ;

buildstructure(rb,BIFXML); // pour combiner le réseau bayésien avec la structure

```
FinCas;  
  ‘Bayésien_Naif’: // faire un réseau bayésien naïve  
    lire(rb); // lire un réseau bayésien naïve  
  ‘Bayésien_K2’: // faire un réseau bayésien avec algorithme k2  
    lire(rb); // cette partie utilise algorithme k2 qui est appliqué par défaut  
quand on lit le réseau.  
  Default: Sortir;  
  buildclassier(rb,db); //pour  
  // il divise la base de données en 2 parties et calcule la précision( un ensemble de  
  nombre_validation_croisée(partie test) et le reste partie entrain )  
  valeur←CrossValidationModel(rb,db,nombre_validation_croisée);  
  Ajouter(valeur,Liste_CrossValidation);  
  FinPour  
  p←calculeprécision(Liste_CrossValidation); //calculer la moyenne de précision  
  ecrire(p);  
FinPour  
Fin.
```

Algorithme 3-1 : Algorithme principal.

6.2. Les méthodes appliquées

Dans les réseaux bayésiens, comme on a dit auparavant, il faut connaître la structure, puis étudier les paramètres.

Chaque étape a des algorithmes associés. Dans notre travail, nous avons utilisé pour l'apprentissage de structure 3 choix, et pour l'apprentissage des paramètres nous avons choisi la vraisemblance maximale :

6.2.1. La structure fixée

L'étape d'apprentissage de la structure peut être ignorée en sélectionnant une structure de réseau fixe. Il existe deux méthodes pour obtenir une structure fixe : il suffit d'en faire un réseau Bayes naïf, ou de la lire à partir d'un fichier au format XML BIF, on a choisi la lecture à partir de BIF fichier.

6.2.2. Réseau bayésien naïve

La méthode naïve de Bayes est applicable uniquement en cas de vérification de l'indépendance entre les attributs. Aussi, les valeurs des attributs numériques doivent avoir une distribution normale. Cette méthode reste une méthode simple et moins coûteuse en temps de calcul. Voici un exemple qui explique ce principe.

La figure 3-11 représente un exemple d'une base de données

Outlook	Temperature	Humidity	Windy	Class
sunny	hot	high	false	N
sunny	hot	high	true	N
overcast	hot	high	false	P
rain	mild	high	false	P
rain	cool	normal	false	P
rain	cool	normal	true	N
overcast	cool	normal	true	P
sunny	mild	high	false	N
sunny	cool	normal	false	P
rain	mild	normal	false	P
sunny	mild	normal	true	P
overcast	mild	high	true	P
overcast	hot	normal	false	P
rain	mild	high	true	N

Figure 3-11 : Exemple d'une base de données simple.

La première étape consiste à convertir l'ensemble de données en un tableau de fréquences : $P(p)=9/14$ et $P(n)=5/14$. Où dans l'étape 2, un tableau de vraisemblance est créé en trouvant les probabilités par rapport à la classe .

outlook	
$P(\text{sunny} p) = 2/9$	$P(\text{sunny} n) = 3/5$
$P(\text{overcast} p) = 4/9$	$P(\text{overcast} n) = 0$
$P(\text{rain} p) = 3/9$	$P(\text{rain} n) = 2/5$
temperature	
$P(\text{hot} p) = 2/9$	$P(\text{hot} n) = 2/5$
$P(\text{mild} p) = 4/9$	$P(\text{mild} n) = 2/5$
$P(\text{cool} p) = 3/9$	$P(\text{cool} n) = 1/5$
humidity	
$P(\text{high} p) = 3/9$	$P(\text{high} n) = 4/5$
$P(\text{normal} p) = 6/9$	$P(\text{normal} n) = 2/5$
windy	
$P(\text{true} p) = 3/9$	$P(\text{true} n) = 3/5$
$P(\text{false} p) = 6/9$	$P(\text{false} n) = 2/5$

Figure 3-12 : Etape deuxième de principe réseau bayésien naïve.

Maintenant, nous passons à l'étape 3. Nous utilisons l'équation bayésienne naïve (Voir Chapitre 2 Section 4 Equation 2 Page 17) pour calculer la probabilité postérieure pour chaque classe. La classe avec la probabilité postérieure la plus élevée est le résultat de la prédiction.

Si on a un nouvel exemple $X = \text{"rain, hot, high, false"}$.

Donc la probabilité est calculée comme suite:

$$P(X|p)P(p) = P(\text{rain}|p).P(\text{hot}|p).P(\text{high}|p).P(p) = 3/9.2/9.3/9.6/9.9/14 = 0.010582$$

$$P(X|n)P(n) = P(\text{rain}|n).P(\text{hot}|n).P(\text{high}|n).P(n) = 2/5.2/5.4/5.2/5.5/14 = 0.018286$$

X est classifié dans la classe n.

6.2.3.L'algorithmme k2

Cette partie explique l'algorithmme k2, pour mieux comprendre (Voir Chapitre 3 Section 4.1.4 page 34).

Procédure K2

Var
 A: ensemble de n nœuds;
 D: base de données;

Debut
pour $i = 1$ à n **faire**
 parent \leftarrow Nil;
 //|D| est le nombre d'échantillons;
 //||A|| est le nombre de paramètres dans le graphique G;
 $\phi(|D|) \leftarrow 1$; //la fonction AIC;
 $P_{\text{prec}} \leftarrow LL(A: D) - \phi(|D|)||A||$
 trouve \leftarrow vrai;
Tantque trouve and $|parent| < 1$
Debut
 // noeuds(xi) sont l'ensemble noeuds qui précède xi et parent qui maximize
 U {z}
 z \leftarrow noeuds(xi) - parent ;
 $P_{\text{nouv}} \leftarrow LL(A: D) - \phi(|D|)||A|| \cup \{z\}$
 si $P_{\text{nouv}} > P_{\text{prec}}$ **Alors**
 Debut
 $P_{\text{nouv}} \leftarrow P_{\text{prec}}$;
 parent \leftarrow parent U {z};
 sinon trouve \leftarrow faux
 FinSi
FinTantque
FinPour
Fin;

Algorithme 3-2: Algorithme K2.

6.3. Les classes utilisées

Cette section a pour but de présenter les classes utilisées:

Classe Fromfile pour lire le XMLBIF, modifier, combiner la structure avec le réseau (buildstructure), et afficher le placement de fichier.

```
Classe FromFile  
Debut  
                                // Attributs  
Var BIFXML: string;  
                                // Méthodes  
setBiffile(Var BIFXML: string); // Modifier le fichier XML.  
Buildstructure(Var rb: réseau bayésien, bif :Fichier XML ) ; // Lire et combiner la  
structure avec le réseau.  
getBiffile():string; //Afficher l'emplacement de fichier XML.  
Fin;
```

Algorithme 3-3: Classe FromFile et ses méthodes.

Classe main pour appeler les méthodes et calculer leurs précisions.

```
Classe main  
Debut  
                                // Attributs  
Var rb: Réseau bayésien;  
    db: Bases de données;  
                                // Méthodes  
Fonction Calculeprécision(^Liste:réel): reel; // Fonction retourne la précision  
readDataFile(nom_fichier : string); // Lire la base de données  
//Fonction pour diviser la base de données par un nombre spécifié  
Fonction CrossValidationModel(rb;réseau bayésien ,db;base de données,  
nombre_validation : entier): réel;  
BuildClassifier(db;base de données,rb: réseau bayésien);// Génère un classifica-  
teur pour aider au calcul de la précision.  
Fin;
```

Algorithme 3-4 : Classe main et ses méthodes.

7. Conclusion

Dans ce chapitre, nous avons présenté la conception d'une application d'aide au diagnostic des tumeurs cérébrales. Nous avons utilisé trois techniques à savoir un réseau bayésien avec une structure fixée (3 structures), un réseau bayésien avec algorithme de recherche k2 et un algorithme bayésien naïve.

Dans le prochain chapitre, nous allons détailler la partie de l'implémentation et discuter les résultats obtenus.

Chapitre 04

1. Introduction

Dans ce chapitre nous présentons notre application et les outils de programmation utilisés pour son implémentation. Des résultats obtenus sont aussi présentés.

2. Outils de Programmation

2.1. Langage de programmation

Dans notre application nous avons utilisé le langage de programmation Java, ce choix est basé sur le fait que java est un bon et simple langage de programmation mais aussi pour Weka est que celle-ci est elle-même créée avec java.

2.2. Outil de Développement

NetBeans est un environnement de développement intégré (EDI ou IDE – Integrated Development Environment) et il est pleinement fonctionnel pour les développeurs d'application Java.



Figure 4-1 : Logo NetBeans.

2.3. Les Packages

On a utilisé :

- Java Weka.jar qui est un fichier qui contient plusieurs classes prédéfinis, spécialement pour le machine Learning, il permet de faire tous les type de méthodes d'apprentissage tel que :la classification, la régression, le clustering,et contient un multitude d'algorithmes comme les arbres de décision ,SVM, Réseaux de Neurones Artificiels,...,etc.
- Java Swing : package pour les interfaces graphiques de l'application avec une multitude de composant graphique.

3. Implémentation

Cette application est destinée pour un usage médical où les utilisateurs peuvent être des médecins ou du personnel médical.

3.1. Résultats

Dans cette section, nous présentons certains résultats obtenus. Noter que Cross validation ou la validation croisée est une technique utilisée pour évaluer la manière dont les résultats de l'analyse statistique se généralisent à un ensemble de données indépendant. La validation croisée est largement utilisée dans les contextes où la cible est la prédiction et où il est nécessaire d'estimer la précision des performances d'un modèle prédictif.

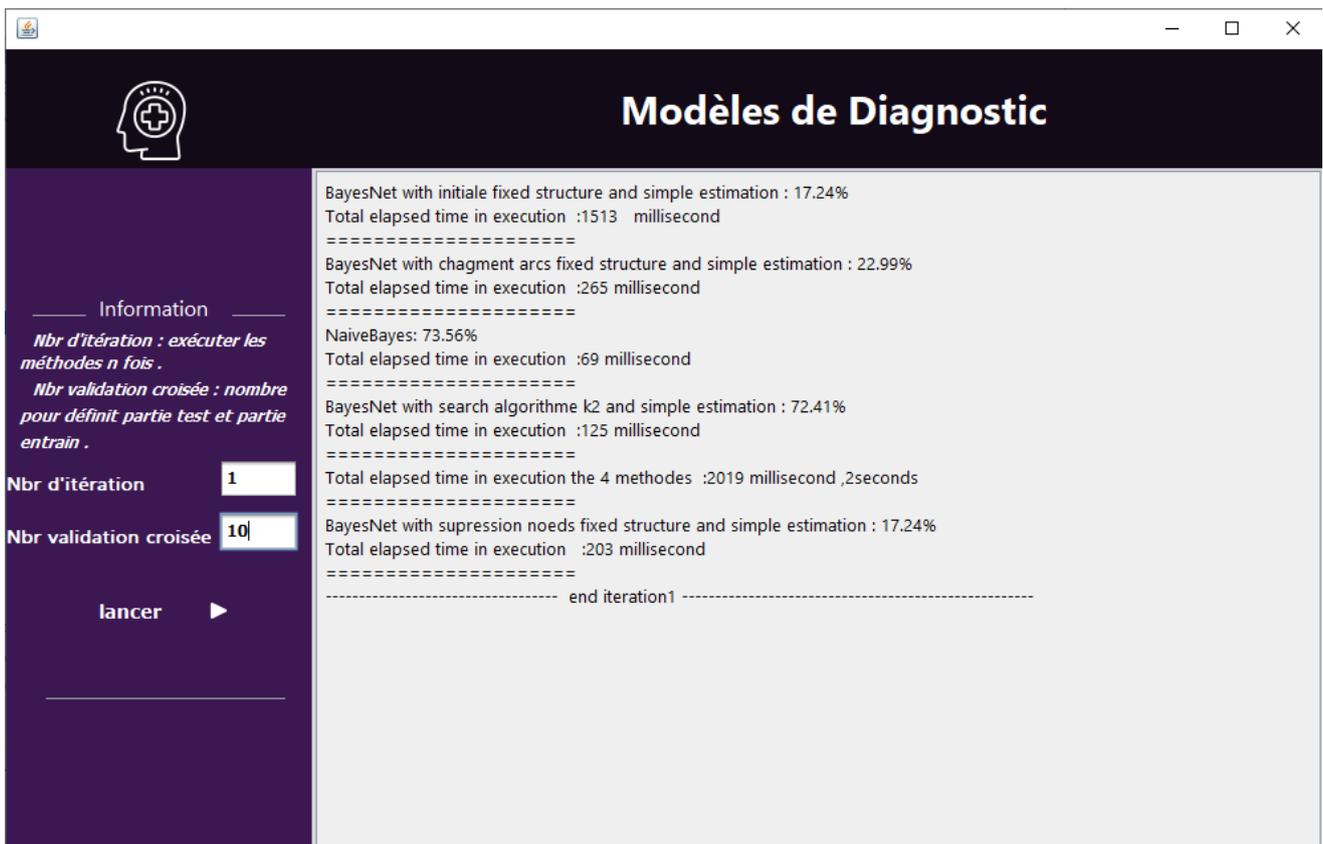


Figure 4-2 : Résultats d'une itération de l'application.

D'abord, nous appliquons le réseau bayésien avec une structure fixée en gardant la structure initiale. Ce graphe représente la précision du modèle avec un changement du nombre de partitionnement (validation croisée) la précision maximale est de 21,83 avec 4 ou 6 partitionnements.

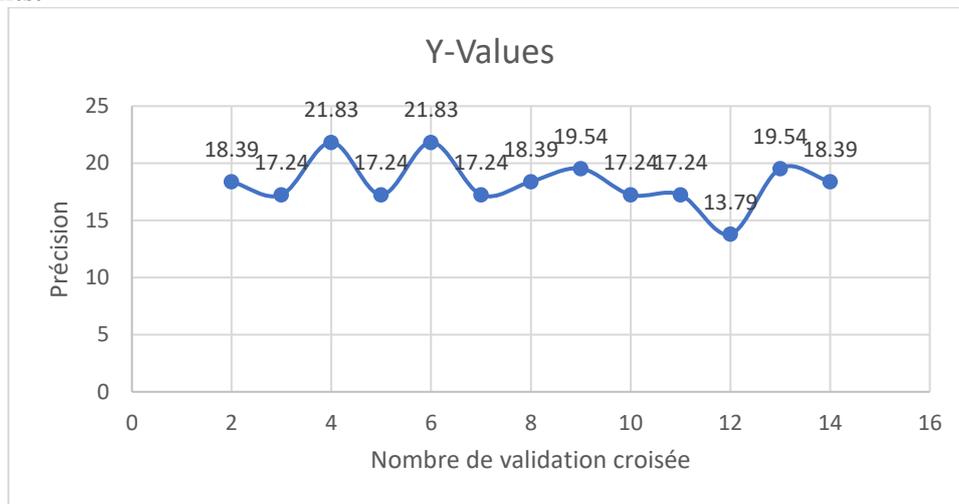


Figure 4-3 : Graphe décrivant la précision par rapport au nombre de validation croisée avec la structure initiale.

Maintenant et en appliquant le réseau bayésien avec une structure fixée qui a subit des changements d'arcs. Ce graphe représente la précision du modèle avec un changement du nombre de partitionnement, la précision maximale est de 27,58 avec 7 partitionnements.

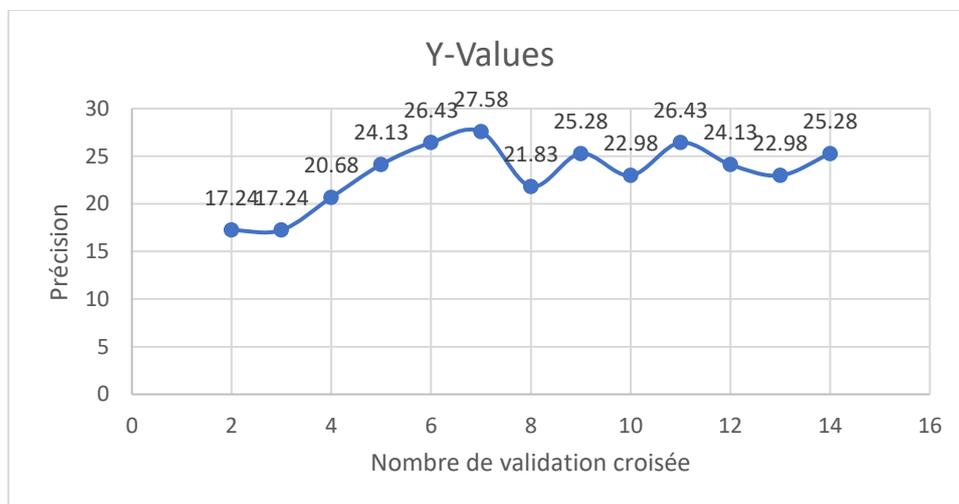


Figure 4-4 : Graphe décrivant la précision par rapport au nombre validation croisée avec la structure de changement les arcs

Avec des suppressions des nœuds, nous obtenons le graphe suivant. La précision maximale est de 21,83 avec 4,6 partitionnements.

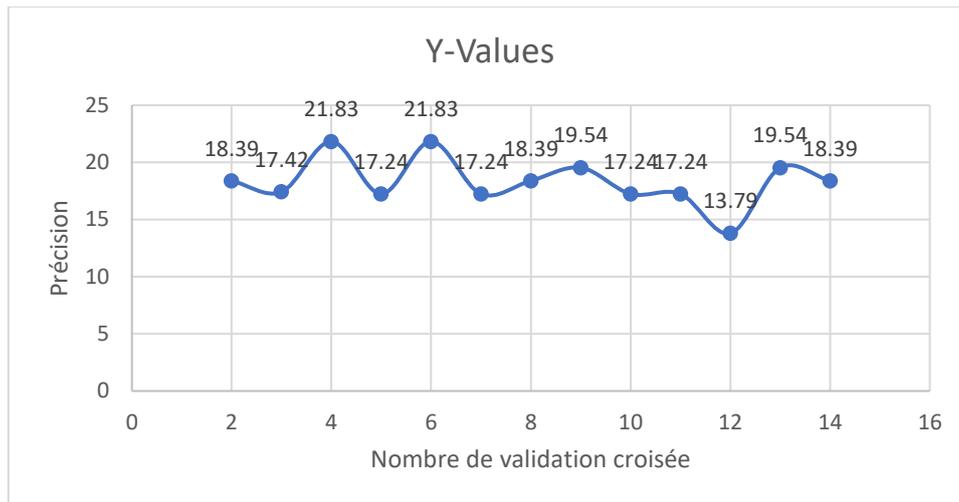


Figure 4-5 : Graphe décrivant la précision par rapport au nombre de validation croisée avec la structure de suppression les nœuds

Si nous appliquons le réseau bayésien naïf sur les deux bases de données, nous obtenons le graphe suivant. Il représente la précision du modèle avec un changement du nombre de partitionnements, la précision maximale est de 74.71 avec 8 partitionnements.

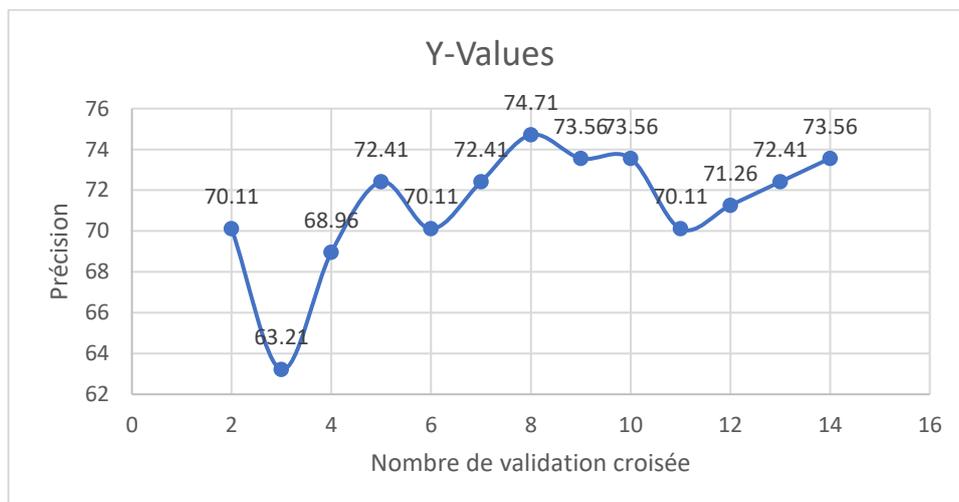


Figure 4-6 : Graphe décrivant la précision par rapport au nombre de validation croisée avec réseau bayésien naïve.

Le réseau bayésien avec k2 donne le graphe suivant où la précision maximale est de 75.86 avec 2 ou 11 partitionnements.

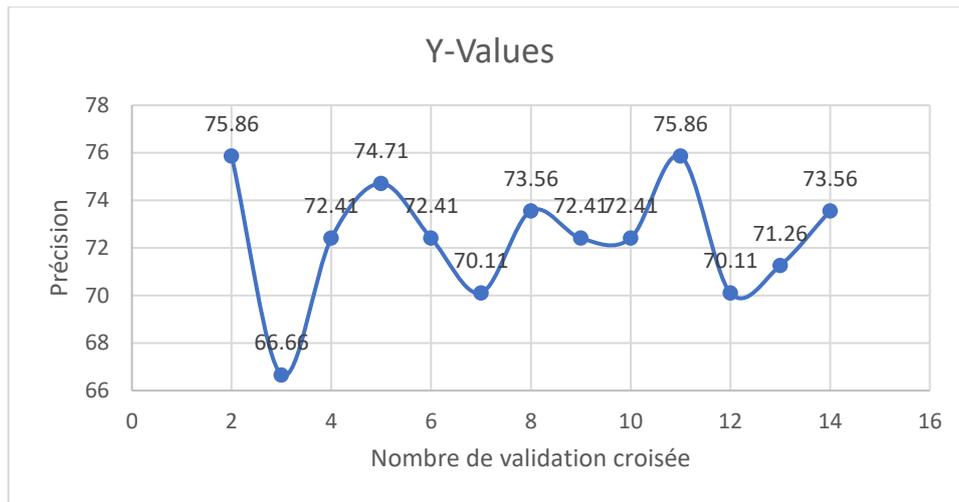


Figure 4-7 : Graphe décrivent la précision par rapport au nombre de validation croisée avec réseau bayésien appliquer algorithmme de recherche k2 .

3.2. Comparaisons des techniques

Cette partie nous offre la différence entre les techniques appliquées (Voir Figure 4-8 et Tableau 4-1) en termes du temps d'exécution et la précision. Ces derniers sont calculés pour différentes valeurs du validations croisées.

Pour le Tableau 4-1 le temps d'exécution régresse par rapport à chaque itération.

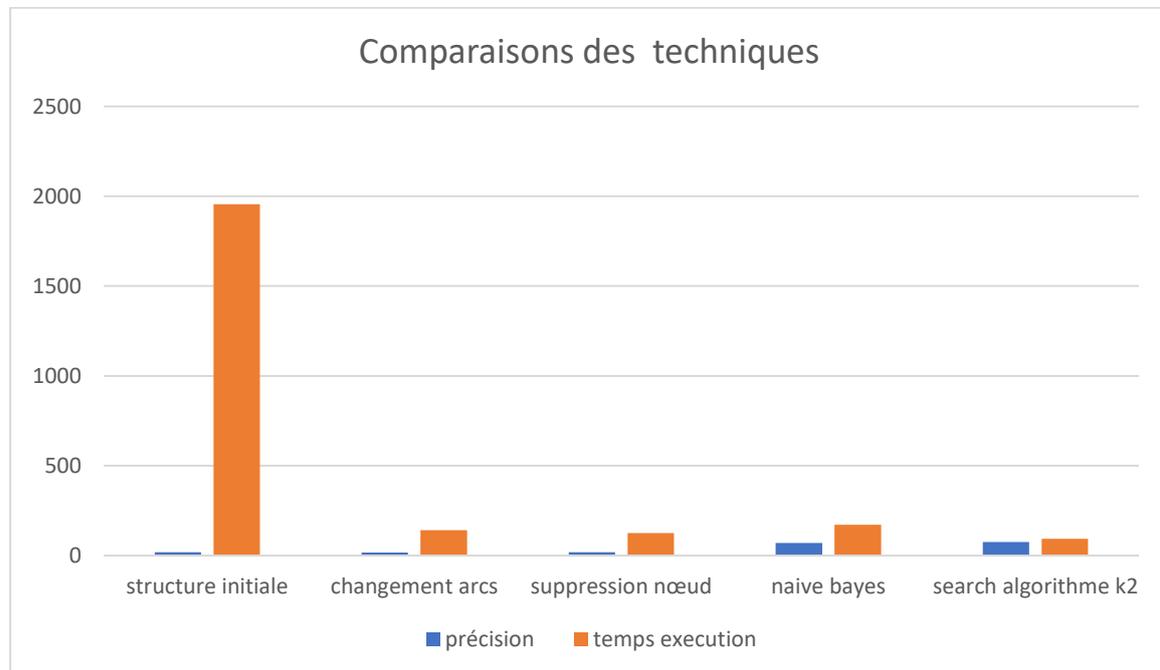


Figure 4-8 : Comparaison des techniques appliquées avec les bases de données utilisées.

Ce tableau définit le temps d'exécution par rapport 20 itérations de différent nombre de validation croisée.

Méthode	Itération	Temps d'exécution
Application de toutes les méthodes	1	2905
	5	432
	10	333
	20	346
Application de la méthode de structure fixée par la suppression des nœuds	1	169
	5	187
	10	125
	20	130
Application de méthode réseau bayésien avec algorithme k2	1	143
	5	141
	10	62
	20	56

Tableau 4-1 : Temps d'exécution en fonction des itérations exécutées.

4. Discussion des résultats

Dans cette partie, nous discuterons des résultats obtenus et comme nous l'avons remarqué, chaque technique a une certaine précision. Et cela varie en fonction du nombre de validation croisée à chaque itération. Le but de l'itération est de nous aider à démontrer la précision dans les plus brefs délais afin d'indiquer la différence pour chaque technique utilisée et le temps qui s'écoule.

Les graphiques précédents représentent chaque technique, avec leur précision, avec un nombre différent de validation croisée, nous ont montré la différence et quelles sont les techniques qui nous aident à trouver le modèle approprié pour le diagnostic de tumeur cérébrale. Nous avons également remarqué que le temps d'exécution varie selon chaque technique, par exemple lorsque le nombre de validation croisée est égal à 2, la précision des techniques du réseau bayésien avec structure fixée était moins précise que le réseau bayésien naïve, et le temps d'exécution est moindre quant au réseau bayésien avec l'algorithme k2, c'était le plus court par rapport aux autres comme représenté dans la Figure 4-8.

Comme nous l'avons déjà dit. L'itération nous aide à mettre en œuvre les mêmes techniques avec différents nombres de validation croisée et des temps d'exécution différents

Le temps d'exécution varie du début à la fin, et tant que la base de données est volumineuse, ce temps d'exécution est long.

5. Conclusion

L'implémentation de notre application est une conviviale pour trouver un modèle l'aide au diagnostic de la tumeur cérébrale. Les résultats ont montré que le réseaux bayésien naïve et réseau bayésien avec algorithme de recherche k2 donnent les meilleurs résultats car on a obtenu un meilleur taux de précision par rapport aux autres.

L'apprentissage automatique fait référence au développement, à l'analyse et à l'implémentation de méthodes qui permettent à une machine d'évoluer grâce à un processus d'apprentissage, et ainsi de remplir des tâches qu'il est difficile ou impossible de remplir par des moyens algorithmiques plus classiques.

L'un de ces algorithmes est les réseaux bayésiens, qui ont aussi une base que l'on suit au début, il faut connaître la structure et dans notre travail on a plusieurs choix, ensuite en apprenant les paramètres avec la méthode maximum vraisemblance qui est une technique d'apprentissage des paramètres et son idée de base est de représenter la vraisemblance sur les données par rapport aux paramètres du modèle, puis de trouver les valeurs des paramètres afin que la vraisemblance soit maximisée.

Ce mémoire a pour objectif de développer une approche intelligente pour l'aide au diagnostic des tumeurs cérébrales basée sur les réseaux bayésiens. Le but principal est de satisfaire les besoins des oncologues, tout en proposant une application ergonomique qui aidera les médecins au diagnostic.

Lors de la réalisation de cette étude, nous avons commencé d'abord par la définition de la tumeur, ses causes, ses types, son développement, Ensuite, nous avons présenté les réseaux bayésiens sur lesquels nous avons basé notre conception. Un ensemble des résultats sont présentés pour valider l'implémentation de notre approche.

Ce mémoire constitue une base de travail à partir de laquelle de nouvelles activités de recherche peuvent être lancées afin d'améliorer le travail présenté. Par exemple on pourra utiliser d'autres techniques des réseaux bayésiens pour trouver le meilleur modèle d'aide au diagnostic des tumeurs cérébrales et on pourra aussi améliorer le travail pour diagnostiquer un cas et fournir précisément le type de tumeur contracté.

Sitographie

- [1] https://www.doctissimo.fr/html/sante/encyclopedie/sa_1099_tumeurs_cere.htm consulté le 28-02-2020
- [2] <https://www.cancer.ca/fr-ca/cancer-information/cancer-101/what-is-cancer/types-of-tumours/?region=qc> consulté le 03-04-2020
- [3] <https://cancer.ooreka.fr/comprendre/tumeur> consulté le 14-02-2020
- [4] <https://www.sante-sur-le-net.com/maladies/cancer/cancers-bouche/> consulté le 15-02-2020
- [5] https://www.ligue-cancer.net/article/26133_cancer-de-lestomac consulté le 25-03-2020
- [6] <https://www.cancer.ca/fr-ca/cancer-information/cancer-type/lung/lung-cancer/cancerous-tumours/?region=on> consulté le 28-02-2020
- [7] <https://www.cancer.ca/fr-ca/cancer-information/cancertype/mesothelioma/mesothelioma/?region=on> consulté le 13-07-2020
- [8] <https://www.chuv.ch/fr/neurochirurgie/nch-home/patients-et-familles/maladies-traitees/les-tumeurs-cerebrales/> consulté le 15-07-2020
- [9] <https://www.cancer.be/le-cancer/comment-se-forme-une-tumeur> 20-08-2020
- [10] <https://www.cancer.gov/rare-brain-spine-tumor/tumors/oligodendroglioma> consulté le 15-02-2020
- [11] <https://ermongroup.github.io/cs228-notes/learning/structure/> consulté le 10-02-2020
- [12] <https://www.doctissimo.fr/sante/diaporamas/tumeur-benigne/mesothelium> consulté le 30-07-2020
- [13] <https://www.cancer.ca/fr-ca/cancer-information/cancer-type/lung/lung-cancer/?region=on> consulté le 02-08-2020

Bibliographie

- [14] Ali Ben Mrad . Observations probabilistes dans les réseaux bayésiens, Université de Sfax École Nationale d'Ingénieurs de Sfax ,20 juin 2015.
- [15] A. Louail . Les lymphocytes B. conférence, Faculté de médecine d'Oran, 2013-2014.
- [16] Clément Fauré, Sylvie Delprat, Alain Mille, Jean-François Boulicaut. Utilisation des réseaux bayésiens dans le cadre de l'extraction de règles d'association, Université de Lyon.
- [17] Ghada Trabelsi, Philippe Leray, Mounir Ben Ayed, Adel M. Alimi . Evaluation des algorithmes d'apprentissage de structure pour les réseaux bayésiens. Université de Nantes, France, 2014.
- [18] Fradj Ben Lamine, Karim Kalti, Mohamed Ali Mahjoub. Etude de Modèles à base de réseaux Bayésiens pour l'aide au diagnostic de tumeurs cérébrales. conférence, Paris (France), Juin 2012.
- [19] Olivier François, Philippe Leray. Apprentissage de structure des réseaux bayésiens et données incomplètes. Université de Rouen, France, 2005.
- [20] Olivier François . De l'identification de structure de réseaux bayésiens à la reconnaissance de formes à partir d'informations complètes ou incomplètes, Université de Rouen, 2006.
- [21] Olivier Lebeltel, Pierre Bessière, Julien Diard , Emmanuel Mazer. Programmation bayésienne des robots, Laboratoire GRAVIR / IMAG, France.
- [22] Laffosse Jean-Michel. Tumeurs des Primitives et Secondaires. Manuel de programme, Clinique Médipôle Garonne, 45 Rue de Gironis, 31000 Toulouse, France, 2016.
- [23] Steven A. Goldman, Nimish A. Mohile. Lymphomes primitifs du cerveau, Université of Rochester Medical Center, juin 2018.
- [24] Yannick Bridé . Méthodes de construction de réseaux bayésiens, mémoire master, université de Strasbourg UFR Mathématiques et Informatique, Août 2016.

- [25] Will Monroe, Chris Plech. Parameter Learning, Université de stanford, août 2017.
- [26] Zaabot Zohra . Traitement d'Images et Reconnaissance de Formes, mémoire de magister, Université Mouloud Mammeri, Tizi-Ouzou, 2012.