

Table des matières

Remerciements

Introduction Générale

Le thème	2
Le contexte	2
La problématique	3
La contribution	5
Structure de la thèse	6

Chapitre 1. Aide à la Décision Médicale

1.1	Introduction	9
1.2	Décision et aide à la décision	10
1.2.1	<i>Le processus de décision</i>	12
1.2.2	<i>Les intervenants dans le processus d'aide à la décision</i>	13
1.2.3	<i>Les modèles de décision</i>	14
1.2.4	<i>La typologie des décisions</i>	14
1.3	Aide à la décision médicale	16
1.3.1	<i>Le raisonnement clinique</i>	17
1.3.2	<i>La décision en situation</i>	18
1.3.3	<i>La décision médicale</i>	18
1.4	Les systèmes d'aide à la décision médicale (SADM)	19
1.4.1	<i>Définitions</i>	20
1.4.2	<i>Les objectifs du SADM</i>	20
1.4.3	<i>Les principales fonctions du SADM</i>	21
1.4.4	<i>Structure d'un SADM</i>	22
1.4.5	<i>La typologie des SADM</i>	22
1.4.6	<i>Composantes du SADM</i>	25
1.4.7	<i>Les méthodologies d'aide à la décision médicale</i>	28
1.4.8	<i>Conclusion</i>	29
1.5	Aide à la décision médicale par RBC	29
1.5.1	<i>Le mode RBC</i>	29
1.5.2	<i>Les caractéristiques des systèmes RBC en médecine</i>	34
1.6	Synthèse de l'aide à la décision médicale par RBC	35
1.6.1	<i>Les limites de l'aide à la décision médicale par RBC</i>	37
1.6.2	<i>Conclusion</i>	37

1.7	Les intégrations du RBC	38
1.7.1	<i>Intégration RBC-Le raisonnement à base de règles (RBR)</i>	38
1.7.2	<i>Intégration RBC-Le raisonnement à base de modèles (MBR)</i>	39
1.7.3	<i>Intégration RBC-La recherche d'Information (IR)</i>	40
1.7.4	<i>Intégration RBC-La satisfaction de contraintes (CSP)</i>	40
1.7.5	<i>Intégrations RBC-L'analyse multicritères (AMC)</i>	40
1.7.6	<i>Intégration RBC-La fouille de données (FDD)</i>	41
1.8	Conclusion	43

Chapitre 2. Intégration RBC-FDD pour l'Aide à la Décision Médicale

2.1.	Introduction	45
2.2.	La fouille de données	45
2.2.1	<i>Les méthodes de la fouille de données</i>	45
2.3.	Extraction de Connaissances à partir de Données (ECD)	48
2.3.1	<i>Les tâches de l'ECD</i>	50
2.3.2	<i>Le processus ECD</i>	53
2.3.3	<i>Les étapes du processus ECD</i>	54
2.3.4	<i>La sélection des données</i>	54
2.3.5	<i>Le prétraitement des données</i>	54
2.3.6	<i>La transformation des données</i>	55
2.3.7	<i>La fouille de données</i>	55
2.3.8	<i>L'évaluation et la présentation</i>	55
2.4.	Apport de la fouille de données en aide à la décision médicale	56
2.5.	Synthèse de l'utilisation des méthodes de FDD en aide à la décision médicale	59
2.6.	Les limites de l'intégration	60
2.7.	Conclusion	61

Chapitre 3. Intégration RBC-AMC pour L'aide à La Décision Médicale

3.1	Introduction	64
3.2	L'analyse multicritères (AMC)	64
3.2.1	<i>Le paradigme multicritères</i>	65
3.2.2	<i>Les problématiques multicritères</i>	69
3.2.3	<i>L'analyse multicritères d'un problème décisionnel</i>	71
3.2.4	<i>La démarche multicritères d'un problème décisionnel</i>	71
3.2.5	<i>Les principales méthodes multicritères</i>	74
3.2.6	<i>Le choix de méthodes multicritères</i>	77
3.3	Synthèse de l'utilisation de l'AMC pour l'aide à la décision médicale	78
3.3.1	<i>Utilisation des méthodes du critère unique de synthèse</i>	78
3.3.2	<i>Utilisation des méthodes de surclassement</i>	78

3.4	Les limites de l'intégration RBC-AMC	79
3.5	Conclusion	79

Chapitre 4. Proposition de Systèmes d'Aide à la Décision Médical Intégrés

4.1	Introduction	81
4.2	Approche d'aide à la décision intégrée proposée	81
4.2.1	<i>Formalisation du problème médical</i>	81
4.2.2	<i>Le Modèle d'Aide à la Décision Médicale Intégré</i>	82
4.2.3	<i>Elaboration du modèle</i>	85
4.2.4	<i>Définition de la situation médicale</i>	85
4.2.5	<i>Le RBC</i>	85
4.2.6	<i>Le raisonneur</i>	87
4.3	Proposition d'un Système d'Aide à la Décision Intégré RBC-FDD	87
4.3.1	<i>Le problème médical considéré : Diagnostic des maladies</i>	87
4.3.2	<i>Intégration RBC-FDD</i>	89
4.3.3	<i>Le modèle d'aide a la décision proposé</i>	89
4.3.4	<i>Le système RBC^{FDD} proposé</i>	92
4.3.4.1	<i>La situation médicale</i>	93
4.3.4.2	<i>Le RBC</i>	93
4.3.4.3	<i>Le raisonneur FDD</i>	95
4.3.5	<i>Mise en œuvre</i>	97
4.3.5.1	<i>Expérimentation</i>	100
4.3.5.2	<i>Evaluation</i>	101
4.4	Proposition d'un Système d'Aide à la Décision Intégré RBC-AMC	102
4.4.1	<i>Le problème médical considéré : Proposition d'une thérapie</i>	102
4.4.2	<i>Le modèle d'aide à la décision proposé</i>	103
4.4.3	<i>Le système RBC^{AMC} proposé</i>	104
4.4.3.1	<i>Définition de la situation médicale</i>	105
4.4.3.2	<i>Le RBC</i>	106
4.4.3.3	<i>Le raisonneur AMC</i>	107
4.4.4	<i>Mise en œuvre</i>	112
4.4.4.1	<i>Expérimentation</i>	113
4.4.4.2	<i>Evaluation</i>	114
4.5	Conclusion	115

Conclusion Générale et Perspectives	116
--	------------

Références bibliographiques	121
------------------------------------	------------

Les figures	138
-------------	-----

Les tables	140
------------	-----

Introduction Générale

Introduction Générale

Le domaine de la décision est très riche tant en études théoriques qu'en systèmes informatiques dans divers domaines. Dans les années quatre-vingt sont apparus des outils d'aide à la décision, ayant permis de franchir un grand pas dans la prise de décision qui est passée du stade mono-décideur au multi-décideur, et passée aussi de l'optimisation d'une simple fonction économique à l'analyse multicritères. De nos jours, on a des systèmes développés à base de connaissances. On parla alors de systèmes d'aide à la décision médicale (SADM).

Les SADM sont « des applications informatiques dont le but est de fournir aux cliniciens en temps et lieux utiles les informations décrivant la situation clinique d'un patient ainsi que les connaissances appropriées à cette situation, correctement filtrées et présentées afin d'améliorer la qualité des soins et la santé des patients » [Renaud-Salis et al., 10 ; Serroussi et Bouaud, 14 ; Serroussi, et al., 13.a ; Serroussi, et al., 13.b].

En effet, l'aide à la décision a pour but d'aider le décideur dans des situations de recherche de la meilleure action qui a pour effet immédiat, l'amélioration d'une situation ou d'un résultat généralement positif.

À la faveur de l'existence d'un potentiel en informations médicales et d'un objectif permanent en matière d'amélioration de la qualité et de l'efficacité des soins prodigués aux malades, et profitant des outils informatiques mis à contribution (systèmes, logiciels, etc.), de nombreux travaux de recherche et de développement ont abouti à la création de systèmes experts précurseurs, et ultérieurement à des systèmes de résolution de problèmes médicaux. Cependant, les outils et méthodologies mises à contribution ont montré quelques limites, et une tendance à l'intégration, du savoir-faire du spécialiste dans les systèmes décisionnels, s'est fortement développée ces dernières années. Cette tendance a fait naître un fort développement de nouveaux systèmes médicaux basés sur de nouvelles approches de raisonnement.

Par rapport à notre thèse, on se focalise sur l'aide à la décision médicale guidée par de nouveaux modes de raisonnement en l'occurrence l'utilisation des connaissances.

Par souci de clarté, cette introduction se compose de plusieurs parties qui permettent d'exposer notre travail de recherche graduellement. Nous débutons par l'exposé du thème, son contexte et les relations qui le lient au domaine de l'extraction des connaissances à partir de données (ECD) et à la fouille de données (FDD). Divers aspects de la problématique générale de l'aide à la décision médicale seront exposés. Ensuite, nous abordons l'aspect systèmes d'aide à la décision médicale et continuons par un exposé de notre contribution au domaine et nous terminons par une conclusion de la présente thèse appuyée par une bibliographie.

1. Le thème

La thématique générale du domaine dans laquelle s'inscrit notre travail de recherche, est la conception de systèmes capables de produire de la connaissance pour guider l'aide à la décision médicale.

Le thème central autour duquel s'organisent nos activités de recherche, est constitué par la modélisation du processus d'aide à la décision médicale. Ce thème s'inscrit particulièrement au croisement de trois disciplines :

- la modélisation ;
- l'aide à la décision ;
- et le raisonnement médical pour la prise de décision.

Nous mettons l'accent sur l'orientation vers l'étude des possibilités d'apprentissage à partir d'un processus d'extraction de connaissance à partir de données (ECD) pour contribuer au processus général de raisonnement pour l'aide à la décision médicale. Notre travail, résulte d'une source assez variée de problématiques décisionnelles plus ou moins complexes dont nous n'abordons que quelques-unes, afin de faire aboutir notre tentative de solutions préconisées, et que sans cela, elles seraient trop vastes à étudier.

En effet , la modélisation et le traitement des données et des connaissances médicales, sont devenus des tâches très ardues du fait de la diversité des branches médicales allant de la simple consultation d'un patient jusqu'à l'interprétation automatique d'un scanner ou autres supports de données médicales. L'utilisation de ces stocks de données ne peut se faire par la simple lecture, si le travail de stockage n'est pas suivi par des procédures de traitement appropriées. Ceci fait, qu'un effort doit être soutenu du coté des procédures de traitement afin d'en profiter pleinement des connaissances qu'elles peuvent stocker et qui produiront une valeur ajoutée aux données de base.

Donc, l'aspect décision se trouva de fait projeté au-devant du stock de données manipulées à travers le SADM pour la recherche de solutions de diagnostic, thérapeutique ou de pronostic. Ce système, consiste à faire coopérer des procédures afin de produire des raisonnements complexes pour répondre à des questions posées en premier lieu par le décideur en l'occurrence le praticien de la santé. Ces systèmes ont été traités en abondance par les scientifiques, notamment des méthodologies, et des procédures appropriées notamment des méthodes de l'intelligence artificielle (IA), qui ont abouti à de nombreuses applications en médecine [Szolovits et al., 88 ; Marling et al., 05 ; Montani et al., 01 ; Pandey et Mishra, 10].

Le raisonnement à base de cas (RBC), qui est une méthodologie classique de recherche de cas similaires par analogie (ressemblance partielle), a été aussi mis à contribution dans la prise en charge des soins des patients. Son utilisation, a permis une avancée assez notable dans la résolution de problèmes liés au diagnostic, à la thérapie, ou au pronostic de maladies.

Dans la littérature, de nombreux travaux ont abordé l'aide à la décision médicale sous plusieurs angles :

- soit par utilisation des méthodes usuelles de l'IA ;
- soit par utilisation de manière classique du RBC.

Récemment, un autre axe majeur de la recherche s'est développé, c'est l'intégration de différents modes de raisonnement, appelé «raisonnement multimodal» [Deepti et al., 10 ; Pandey et Mishra, 10 ; Qi et al., 16 ; Verma et al., 14 ; Balakrishnan et al., 12 ; Bruland et al., 10]. Cette approche a été abordée timidement vers les années 1990 et depuis, c'est devenu un axe privilégié de recherche, vu les limites montrées par les approches classiques, particulièrement le RBC. En outre, vu que l'acte de décision a imposé la prise en compte des points de vue des différents acteurs de la décision, l'analyse multicritères combinée au RBC s'est vue aussi impliquée comme alternative pour la résolution de problématiques décisionnelles médicales [Angehrn et Dutta, 92 ; Bouhana et al., 11].

Dans ce contexte, notre thèse s'intéresse au raisonnement dans le cas d'un problème d'aide à la décision avec mise de l'accent sur l'intégration de la fouille de données, et ce pour guider un processus d'aide à la décision médicale.

2. La problématique

La nature des données médicales, la méthode de raisonnement classique RBC, la considération des points de vue des acteurs de la décision, la prise de décision et son exécution ont induit différentes problématiques avec des degrés d'importance aussi variés. Les considérer tous relève d'un grand défi qu'une thèse à elle seule ne peut relever. Seulement, nous tentons de résoudre quelques problématiques qui sont dans les limites de notre thèse.

Par rapport à la médecine, un décideur en l'occurrence le médecin lorsqu'il est en consultation d'un patient, il doit explorer tous les symptômes. À cela, s'ajoute le contexte du patient qui est lié à son état physique et mental. Ce contexte est aussi déterminant dans toute action visant l'amélioration de son état de santé. De ce fait, un patient se présentant pour une prise en charge médicale sera alors considéré comme *un cas médical* (patient) à résoudre *pris dans un contexte*. Ceci est pris comme un "tout", et le médecin est devant une *situation médicale*.

La situation médicale que nous préconisons est caractérisée par : une définition du cas (patient) plus ou moins complète, un contexte, un parcours de diagnostics/thérapies possibles ou proposés, et l'existence de préférences du médecin pour chaque diagnostic/thérapie [Lebraty, 06 ; Rasmussen, 86 ; Klein, 98 ; Lebraty et Guarnelli, 14 ; Sournia]. Le médecin définit alors la *situation médicale* comme étant : le patient, décrit par des *symptômes cliniques* et un *contexte* décrit par les *symptômes para-cliniques* (*signes spécifiques ou particuliers au patient*). Ces signes spécifiques peuvent être par exemple «une intolérance d'un régime salé», «une allergie à un produit donné», etc., ils peuvent indiquer par exemple qu'une thérapie souhaitée sera plus ou moins adaptée. Ainsi, le contexte est déterminant lors de la recherche d'une solution (aide à la décision) [Rasmussen, 86 ; Klein, 98].

Sur le volet RBC, il est important de souligner ses performances, néanmoins il apparaît qu'il a fait naître quelques problématiques liées à la base à la structure du cas, restée basique.

Ceci dit, la problématique que nous abordons est résumée aux quatre points suivants :

- a. Est-ce que les formalismes classiques, de représentation de données, suffisent-ils ou faut-il adopter une nouvelle structure. La structure classique du cas (problème, solution) n'est plus adéquate pour définir la situation médicale. De là, se pose la

question de la modélisation de cette situation médicale. Ceci, amène à réfléchir à un formalisme adéquat à travers des descripteurs pertinents, à savoir :

- les descripteurs cliniques ;
- les descripteurs para-cliniques.

Ce formalisme doit s'adapter aux différentes méthodes de raisonnement afin de faciliter sa manipulation, et doit être une valeur ajoutée pour contribuer à la résolution de situations médicales futures. La proposition d'un formalisme à travers un ensemble de descripteurs, est un travail de modélisation assez ardu, d'autant plus qu'il est bien reconnu aujourd'hui que les décisions de diagnostic/thérapie, liées à chaque patient doivent tenir compte des signes particuliers ou symptômes para-cliniques (effets secondaires de médicaments, personne âgée, allergie, etc.).

- b. Il est souvent fait état d'imprécision ou d'indétermination sur les données. De ce fait, il est parfois difficile de discriminer des situations, qui selon leurs contextes pourront être déclarées équivalentes ou incomparables. Donc, le décideur sera en fait devant des choix de solutions où il doit décider la prise en compte de l'une d'entre elles. Ceci fait que le décideur doit montrer des préférences ou priorités pour un diagnostic/thérapie. Ces préférences pourront orienter vers une solution compensatoire ou aider à élaborer une procédure de classement/groupement de solutions. Ainsi, la question comment tenir compte des préférences du décideur est posée, du moment que la structure classique du cas ne permet pas de définir explicitement cette notion.
- c. Le raisonnement classique du RBC ne permet pas de supporter facilement le raisonnement clinique, ni même par un raisonnement simple à base de règles, du fait qu'une règle est déjà écrite pour ne donner que sa conséquence. Alors, la question qui est posée, comment prendre en compte le raisonnement clinique dans la décision médicale. Ce mode de raisonnement doit être intégré dans un processus d'aide à la décision propre aux situations médicales. Ce qui fait que la décision médicale n'est plus appréhendée sous l'angle d'une comptabilisation de symptômes mais beaucoup plus de mode opératoire passant d'une étape à une autre sur la base d'un état initial et allant vers un état final.
- d. Comment définir une procédure de sélection de la "meilleure" solution ou solution optimale. D'un point de vue mathématique, la problématique de choix consiste à poser le problème avec objectif de recherche d'un sous-ensemble aussi restreint que possible des meilleures solutions. Mais par rapport à une situation médicale, comment définir cette procédure qui permet ce sous-ensemble aussi restreint que possible.

3. Contribution

Nous nous plaçons dans la position où le médecin (décideur) est face à une situation médicale et devra explorer les options possibles (Diagnostics/Thérapies) pour choisir la meilleure d'entre elles.

La situation médicale que nous préconisons, contient un contexte qui doit être pris en considération par le décideur (médecin). Celui-ci devra alors explorer les options possibles (diagnostic) pour proposer la meilleure thérapie.

À partir de telles situations, le médecin commence généralement par l'identification d'un état pathologique. Durant cette phase le médecin aura à interagir avec un système qui lui permettra en premier lieu d'utiliser un ensemble d'outils qui lui permettent de progresser de façon cohérente dans la définition de sa situation médicale avec la possibilité de noter ces préférences pour un diagnostic/thérapie. Ensuite, il déclenchera le processus de recherche de solution (diagnostic/thérapie) et se termine par la réalisation de la solution optimale. Il s'agit donc de développer des procédures adéquates pour une aide à la décision efficace.

Sur le plan évaluation de performance, il faut souligner la grande difficulté que présente ce point, pour un système d'aide à la décision médicale. La validation est une phase très importante et très complexe de l'élaboration de tels systèmes, vu que les méthodes d'évaluation sont multiples, parfois difficiles à utiliser.

Ainsi, la contribution principale de notre thèse dans le domaine d'aide à la décision médicale est la proposition d'un modèle décisionnel qui tient compte des différents aspects liés aux problématiques que nous venons de passer en revue. Cette approche assure une intégration du RBC avec d'autres méthodologies de raisonnement et permettant de mutualiser les performances pour résoudre une situation médicale par une aide à la décision médicale efficace.

Concrètement et par rapport aux problématiques que nous venons de passer en revue, notre contribution s'oriente vers quatre lignes directrices :

- a. définir et adopter une structure propre à la situation médicale qui palliera les limites d'une modélisation par le cas. Aussi, ce choix est motivé par :
 - la reproduction aussi proche que possible de l'état physique et mental du patient ;
 - la possibilité pour le décideur de considérer la pertinence des données pour leur prise en compte dans le processus d'aide à la décision.
- b. définir et adopter un mode de raisonnement intégré (RBC et autres) afin de supporter la recherche de solutions sur la base de la situation médicale en cours d'examen. Ce mode envisagé reposera sur la mutualisation des performances des modes de raisonnements impliqués.
- c. proposer un modèle d'aide à la décision médicale adoptant la situation médicale comme structure de base, et montrant toutes les phases du processus de l'élaboration jusqu'à la solution finale.
- d. Complémenter notre étude en explorant une autre approche par l'analyse multicritères (AMC) afin de mieux saisir l'aspect décisionnel.

Pour résumer, nous disons que les contributions mises en évidence dans cette thèse permettent de :

- définir et adopter une structure de données appropriées permettant de définir au mieux la situation médicale ;

- proposer un modèle décisionnel exploitant un raisonnement multimodal sur a base d'un raisonnement clinique ;
- proposer et tester un prototype exploitant cette approche en utilisant distinctement la FDD et l'AMC.

Pour les besoins de validation de notre approche nous avons adopté deux expérimentations distinctes :

- l'un utilisant la FDD ;
- l'autre utilisant l'AMC.

L'objectif principal est la prise en charge d'une situation médicale afin de la résoudre à travers une aide à la décision médicale guidée par un processus à base de connaissances.

4. Structuration de la thèse

Au chapitre I, nous introduisons l'aide à la décision et montrons le déroulement du processus de décision et ses principaux intervenants. Ensuite, nous passerons en revue l'aide à la décision médicale à travers ses différentes notions. Aussi, nous nous pencherons sur les SADM et les différentes méthodologies de raisonnement qu'ils utilisent, en particulier le RBC et ses contributions dans l'aide à la décision médicale qui ont été proposées par certains auteurs à travers différents travaux. Nous verrons aussi les différentes intégrations au RBC, qui ont été supportées par différents systèmes d'aide à la décision médicale. Le chapitre II est consacré à l'intégration de la FDD au RBC et le chapitre III à l'intégration de l'AMC au RBC. Au chapitre IV, nous exposerons notre contribution pour développer une aide à la décision médicale intégrée et sa mise en œuvre.

Chapitre I. Aide à la décision médicale.

La décision est le point de départ de l'étude des systèmes d'aide à la décision médicale. Elle définit les fonctionnalités et les étapes à mettre en œuvre pour atteindre les objectifs du processus de décision médicale. À ce titre, le premier chapitre constitue une présentation générale de la décision et de l'aide à la décision médicale. Nous présentons, d'abord les notions de base, le processus décisionnel et l'aide à la décision. Nous aborderons ensuite les systèmes d'aide à la décision médicale (SADM), leurs méthodologies, le principe du raisonnement à base de cas (RBC) qui a été largement utilisé en médecine, son utilisation et les limites qui ont favorisé les différentes intégrations avec d'autres méthodologies de raisonnement et enfin nous dresserons une synthèse.

Chapitre II. Intégration RBC-FDD pour l'aide à la décision médicale.

Ce chapitre sera consacré à l'utilisation de la FDD et à son intégration au RBC afin de guider le processus d'aide à la décision médicale. En fin de chapitre nous ferons une synthèse des travaux effectués dans cet axe.

Chapitre III. Intégration RBC-AMC pour l'aide à la décision médicale

Nous introduirons différents concepts de base de l'analyse multicritères auxquels nous ferons référence dans ce chapitre. Nous passerons en revue l'utilisation de cette approche en aide à la décision médicale et à son intégration au RBC et ferons une synthèse des différents travaux d'intégration.

Chapitre IV. Proposition de systèmes d'aide à la décision médicale intégrés.

Ce chapitre présente notre contribution dans l'aide à décision médicale. Après une étude de modélisation de la situation médicale, nous nous sommes basé sur les travaux de Lebraty, Guarnelli et Sournia [Lebraty, 06 ; Lebraty et Guarnelli, 14 ; Sournia], pour définir un modèle pour l'aide à la décision médicale. Ensuite, nous décrivons notre proposition d'un prototype d'aide à la décision médicale selon deux orientations :

- par intégration du RBC et de la FDD ;
- par intégration du RBC et de l'AMC.

Nous présentons les expérimentations effectuées visant à montrer les résultats que nous avons obtenus, et nous ferons une analyse pour montrer la pertinence de l'approche adoptée.

En conclusion de cette thèse, nous dressons les perspectives des travaux sur lesquels nous pensons nous investir dans le futur. En effet, la conception et le développement des systèmes d'aide à la décision médicale ont fortement évolué ces dernières années. Cependant, de nombreux progrès sont encore nécessaires dans le domaine des méthodologies de raisonnement ce qui influera beaucoup sur les traitements et par conséquent les résultats attendus des SADM.

Chapitre 1

Aide à la décision
médicale.

Chapitre 1

Aide à la décision médicale.

1.1	Introduction	9
1.2	Décision et aide à la décision	10
1.2.1	<i>Le processus de décision</i>	12
1.2.2	<i>Les intervenants dans le processus d'aide à la décision</i>	13
1.2.3	<i>Les modèles de décision</i>	14
1.2.4	<i>La typologie des décisions</i>	14
1.3	Aide à la décision médicale	16
1.3.1	<i>Le raisonnement clinique</i>	17
1.3.2	<i>La décision en situation</i>	18
1.3.3	<i>La décision médicale</i>	18
1.4	Les systèmes d'aide à la décision médicale (SADM)	19
1.4.1	<i>Définitions</i>	20
1.4.2	<i>Les objectifs du SADM</i>	20
1.4.3	<i>Les principales fonctions du SADM</i>	21
1.4.4	<i>Structure d'un SADM</i>	22
1.4.5	<i>La typologie des SADM</i>	22
1.4.6	<i>Composantes du SADM</i>	25
1.4.7	<i>Les méthodologies d'aide à la décision médicale</i>	28
1.4.8	<i>Conclusion</i>	29
1.5	Aide à la décision médicale par RBC	29
1.5.1	<i>Le mode RBC</i>	29
1.5.2	<i>Les caractéristiques des systèmes RBC en médecine</i>	34
1.6	Synthèse de l'aide à la décision médicale par RBC	35
1.6.1	<i>Les limites de l'aide à la décision médicale par RBC</i>	37
1.6.2	<i>Conclusion</i>	37
1.7	Les intégrations du RBC	38
1.7.1	<i>Intégration RBC-Le raisonnement à base de règles (RBR)</i>	38
1.7.2	<i>Intégration RBC-Le raisonnement à base de modèles (MBR)</i>	39
1.7.3	<i>Intégration RBC-La recherche d'Information (IR)</i>	40
1.7.4	<i>Intégration RBC-La satisfaction de contraintes (CSP)</i>	40
1.7.5	<i>Intégrations RBC-L'analyse multicritères (AMC)</i>	40
1.7.6	<i>Intégration RBC-La fouille de données (FDD)</i>	41
1.8	Conclusion	43

1.1 Introduction

A lors que les sciences de décision vont avoir pour objectif de trouver une décision optimale à partir d'une vision supposée objective de la réalité, l'aide à la décision va s'intéresser à la construction de décisions satisfaisantes en considérant toute la dimension subjective qui peut apparaître au cours d'un processus de décision. Tout d'abord, une distinction doit être faite entre l'homme d'étude et le décideur, c'est-à-dire entre l'expert des méthodes d'aide à la décision et la personne ou le groupe de personnes, chargé(e) de prendre les décisions.

L'aide à la décision va avoir pour objet de faire agir conjointement ces deux principaux acteurs afin de faire émerger les décisions. La prise de décision fait généralement intervenir des points de vue différents, voire contradictoire. La tâche de l'homme d'étude va alors être la modélisation des préférences du décideur en faisant émerger les différents points de vue, ou critères sur lesquelles on se base pour construire la décision relative à la situation décisionnelle ou problème décisionnel et qui vont être pris en compte dans le processus de décision. La combinaison de l'IA et des systèmes d'aide à la décision fournit une nouvelle assistance informatique aux décideurs en étendant leurs capacités de raisonnement dans des environnements complexes. Les systèmes d'aide à la décision dotés "d'intelligence", et l'expertise du domaine d'utilisation ont été profondément étudiés par de nombreux chercheurs.

Les techniques intelligentes sont utiles pour analyser les données et fournir des prévisions, quantifier l'incertitude, fournir facilement des informations et suggérer la marche à suivre. La figure 1.1 souligne comment la prise de décision est influencée par une gamme variée de techniques intelligentes

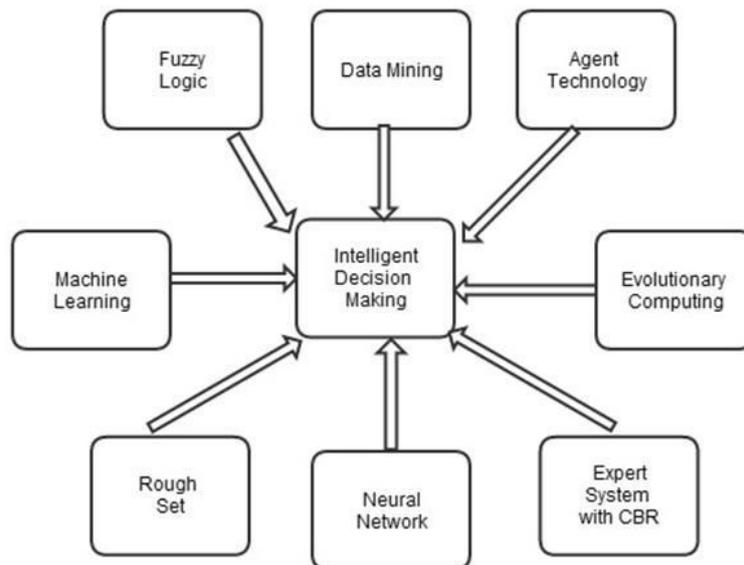


Figure 1.1. Les techniques intelligentes en aide à la décision [Das, 16].

1.2 Décision et aide à la décision

La prise de décision et son exécution sont, en effet, le but de tout problème décisionnel dans une organisation. Tout au long de la vie d'une organisation, des décisions sont prises continuellement en son sein par des décideurs. Les décisions sont souvent prises sur la base de faits se rapportant à un problème et des expériences vécues. Dans certaines situations plus ou moins délicates, le décideur se fait aider par des procédures automatiques afin de le guider dans ses choix de solutions et cette tâche de décision devient beaucoup plus ardue et coûteuse, il devient donc nécessaire d'utiliser des systèmes interactifs d'aide à la décision, notés : SIAD.

Pour les écoles rationalistes-analytiques, la décision est définie comme un choix entre plusieurs alternatives [Schneider, 96]. Pour d'autres, la décision concerne aussi le processus de sélection d'alternatives. Les approches cognitives traitent la décision comme le résultat d'un processus global de résolution de problèmes [Schneider, 96]. Ceci fait que le terme décision a plusieurs définitions. Il est utilisé pour désigner un acte, une action ou un processus de résolution de problème. D'autres auteurs proposent d'autres définitions, chacune reflète un point de vue différent.

Définition 1.

Roy et Bouyssou estiment que la décision est souvent présentée comme : « *le fait d'un individu (décideur) isolé qui exerce librement un choix entre plusieurs possibilités d'actions à un moment donné dans le temps* » [Roy et Bouyssou, 93].

Définitions 2.

Lévine et Pomerol définissent la décision comme suit : « *Une décision est une action qui est prise pour faire face à une difficulté ou répondre à une modification de l'environnement, c'est-à-dire pour résoudre un problème qui se pose à l'individu ou à l'organisation* » [Lévine et Pomerol, 89].

Définition 3.

Selon Mintzberg et al., « *une décision, qu'elle soit individuelle ou résultant d'un travail de groupe, peut être définie comme l'engagement dans une action, c.-à.d., une intention explicite d'agir* » [Mintzberg et al., 76].

Caractéristiques de la décision.

La décision est caractérisée par :

- *Son objet*

Il permet de distinguer les décisions stratégiques, tactiques et opérationnelles. La décision *stratégique* concerne les relations de l'entreprise avec le milieu, et porte essentiellement sur les choix de marchés et de produits afin d'obtenir une adaptation de l'entreprise à son milieu. La décision *tactique* est relative à la gestion des ressources qui sont l'acquisition, l'organisation et le développement. La décision *opérationnelle* qui porte sur l'exploitation

courante. Elle a pour objet de rendre le processus de transformation des ressources plus efficace.

- *Son échéance*

Celle-ci permet de distinguer :

- les décisions à court terme qui n'ont d'effet que sur une courte période ;
- les décisions à moyen terme qui engagent l'entreprise sur plusieurs exercices ;
- les décisions à long terme qui sont exceptionnelles.

- *Son degré de structure*

Le nombre et la complexité des paramètres intervenant dans un processus de décision peuvent être très variés. Lorsque les paramètres sont peu nombreux, aisément identifiables et quantifiables, il est possible de "formaliser" la décision, c'est-à-dire recourir à une procédure standard de résolution ou élaborer un modèle de prise de décision. Tout problème est alors soumis à une succession d'opérations exécutées dans un ordre précis et sous certaines contraintes, pour passer des informations de base aux choix définitifs.

Par ailleurs, l'homme est souvent confronté à des situations où il fait face à différentes hypothèses pour déterminer la meilleure des situations qui lui convient. Cette situation de choix entre plusieurs solutions, repose souvent sur l'optimisation d'un ou plusieurs critères objectifs. Ces situations de choix sont traitées en théorie de la décision. En effet, la théorie de la décision suppose explicitement les assertions suivantes :

- l'existence d'une meilleure décision que l'on peut atteindre moyennant, le temps et des ressources ;
- cette meilleure décision peut être optimale, si on arrive à optimiser un critère ;
- cette décision optimale est toujours joignable à travers un processus.

Cette théorie de la décision traite souvent ce genre de situation sous l'appellation *d'aide à la décision*.

Selon Roy, l'aide à la décision est considérée comme : « *une activité qui s'appuie sur des concepts rigoureux, des méthodologies, des modèles, et des techniques. Elle vise à éclairer les décisions devant être prises par un intervenant, sans pour autant dicter sa conduite* » [Roy, 85]. Elle accompagne le processus de décision en y apportant un éclairage sans se substituer à la prise de décision qui est du seul ressort de décideur. Schärliig voit que cette définition est simpliste, mais elle énonce la nécessité de prendre appui sur des modèles, et non pas de leur faire dire la solution, elle fait allusion à des éléments de réponse plutôt qu'à des réponses complètes et définitives, elle réfère à l'intervenant plutôt qu'au décideur, elle fait mention de la nécessité d'éclairer la décision plutôt que de déterminer quelle est la meilleure solution [Schärliig, 85].

De là, on peut dire que l'aide à la décision prend appui sur des modèles pour aider un acteur du processus de décision à obtenir des éléments de réponse aux questions qu'il se pose. Cette

aide à la décision peut aboutir à une prescription qui permet d'orienter vers une solution (décision). Elle est donc un processus qui utilise un ensemble d'informations disponibles, afin de formuler un problème et aboutir à une décision sur un objet précis. Seulement, Roy voit que la discipline de l'aide à la décision ne repose pas sur l'existence d'une vérité absolue. Et si cette vérité n'est pas supposée exister, l'objectif va être donc de guider et d'éclairer le décideur tout au long de son processus de décision [Roy, 92]. De ce fait, on ne cherchera plus à trouver "la meilleure décision" mais à accompagner le décideur en tentant de faire ressortir les aspects objectifs et ceux qui le sont moins, et apporter une justification aux décisions pour qu'il puisse mesurer sa situation et décider objectivement de ces choix en mettant en évidence des conclusions robustes par rapport à celles qui le sont moins.

Dans le même ordre d'idée, Roy proposa aussi la définition suivante : « *l'aide à la décision est l'activité de celui qui, prenant appui sur des modèles clairement explicités mais non nécessairement clairement formalisés, aide à obtenir des éléments de réponse aux questions que se pose un intervenant dans un processus de décision, éléments concourants à éclairer la décision et normalement à prescrire, ou simplement à favoriser, un comportement de nature à accroître la cohérence entre l'évolution d'un processus d'une part, les objectifs et le système de valeurs au service desquels cet intervenant se trouve placé d'autre part* » [Roy, 92].

1.2.1 Le processus de décision

Selon Chakhar et al., « *L'activité d'aide à la décision s'articule autour d'un processus de décision qui est un ensemble d'activités déclenché par un stimulus et aboutissant à un engagement spécifique à l'action* » [Chakhar et al., 05]. Le processus de décision peut être considéré comme un chemin qui part des données pour aller aux procédures de décision. La littérature concernant les concepts des différents processus de décision est vaste, cependant, le processus le plus diffusé est celui de Simon [Simon, 77]. Nous trouvons, également, d'autres processus tel que celui proposé par Mintzberg et al., [Mintzberg et al., 76], ou celui proposé par Tsoukias [Tsoukias, 03].

[a] Le modèle de Simon

Simon propose le processus de décision IDC (Intelligence, Modélisation, Choix). Ce processus se déroule selon 3 phases principales [Lévine et Pomerol, 89 ; Turban, 93 ; Power, 02]. Il est considéré comme étant le modèle le plus célèbre des processus décisionnels disponibles dans la littérature. Il est schématisé par la figure 1.2 [Alnafie, 16 ; Simon, 77] :

1. Information. Cela consiste à déterminer l'ensemble des données se rapportant à la situation décisionnelle.
2. Conception. A cette étape, les différentes alternatives qui forment l'ensemble des possibilités sont générées et les différentes solutions sont alors élaborées.
3. Choix. Il permet de restreindre l'ensemble des possibilités au sous ensemble de possibilités sélectionnées et qui sera en fait la solution.

On rajoute généralement une 4^{ème} étape pour le contrôle de la mise en œuvre de la décision et l'exercice éventuel d'actions correctives (feedback).

« Cette phase conduit à la recommandation d'une solution appropriée au modèle. Elle peut amener à la réactivation de l'une des trois phases précédentes ou, au contraire, à la validation de la solution. Après le choix, et dans la mesure où la décision s'intègre dans un processus dynamique ; la phase « review » nous semble extrêmement importante. De nouvelles informations pertinentes peuvent influencer tel ou tel choix, voir le modifier complètement. Une rétroaction (feedback) intelligente permet de corriger bien des erreurs et, sur le déroulement d'un processus décisionnel, elle conduit à des performances aussi bonnes que des stratégies compliquées sans rétroaction. Cette phase représente le retour du processus d'aide à décision à la réalité. La recommandation finale doit traduire le résultat fourni par le modèle d'évaluation dans le langage courant du client et du processus de décision dans lequel il est impliqué. » [Adla, 10].

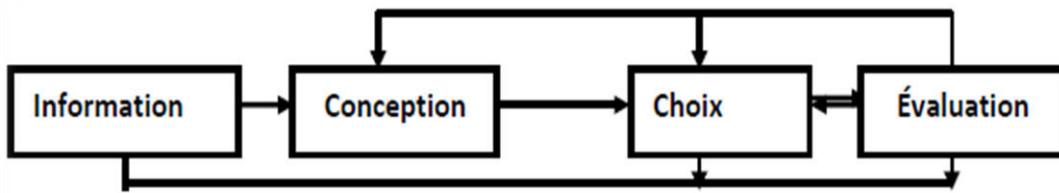


Figure 1.2. Le processus décisionnel selon Simon.

[b] *Le modèle de Mintzberg et al.,*

Ce processus décisionnel contient plusieurs activités regroupées en trois phases fondamentales [Chakhar et al., 05] :

1. Identification de la situation décisionnelle.
2. Développement des solutions possibles.
3. Sélection d'une solution à implémenter.

[c] *Le modèle de Tsoukias*

Tsoukias a introduit le concept de processus d'aide à la décision comme une extension au processus de décision. Pour l'auteur, le processus d'aide à la décision est subdivisé en trois phases [Chakhar et al., 05] :

1. Représentation du problème.
2. Formulation du problème.
3. Evaluation.

1.2.2 *Les intervenants dans le processus d'aide à la décision*

Le processus d'aide à la décision fait concourir plusieurs intervenants ou acteurs. Nous distinguons deux principaux : l'homme d'étude et le décideur. Néanmoins, d'autres acteurs peuvent être amenés à intervenir à des titres variés. L'étude des différents acteurs (typologie,

objectifs, interactions, etc.) constitue un aspect important à étudier pour analyser un processus de décision. Avant de les décrire nous donnerons la définition de Roy.

Selon Roy : « *Un individu ou un groupe d'individus est acteur d'un processus de décision si, par son système de valeurs, que ce soit au premier degré du fait des intentions de cet individu ou groupe d'individus ou au second degré par la manière dont il fait intervenir ceux d'autres individus, il influence directement ou indirectement la décision* » [Roy, 85].

Dans un processus de décision, il est possible de définir les principaux intervenants suivants [Roy et Bouyssou, 93] :

- Le décideur : la personne (ou les personnes) assistée(s) par l'aide à la décision et qui est aidée pour mieux exprimer ses préférences vis-à-vis d'une situation donnée.
- L'homme d'étude (l'analyste) : est un individu ou un groupe d'individus, qui a pour rôle d'établir un système de préférences, de définir le modèle d'aide à la décision, de l'exploiter afin d'obtenir des réponses et d'établir des recommandations pour conseiller le décideur sur les solutions envisageables.
- Le négociateur : mandaté par un décideur en vue de faire valoir la position de celui-ci dans une négociation et de rechercher une solution compromise.
- L'arbitre (juge) : intervient en se substituant aux acteurs dans la recherche d'une action compromise.

1.2.3 Les modèles de décision

D'un point de vue théorique, on peut distinguer trois grands types de modèles de prise de décision au sein d'une organisation [Simon, 83] :

- le modèle rationnel ou classique, où l'homme est considéré complètement informé pour prendre une décision optimale ;
- le modèle politique où les décisions sont plutôt l'objet de négociations entre les groupes ;
- le modèle psychologique où c'est l'optimalité qui est négociée pour parvenir selon Simon, avec le principe de rationalité limitée, à une décision plutôt satisfaisante.

1.2.4 La typologie des décisions

Les décisions prises au sein de l'organisation peuvent être hiérarchisées selon leur niveau leur échéance, leur objet ou leur nature. On trouve alors les classifications suivantes :

[a] Selon le niveau de la décision

Kast différencie trois niveaux de décision selon la hiérarchie dans la structure de décision : le niveau de planification (haute direction) ; le niveau de pilotage (direction) ; le niveau opérationnel (direction des opérations) [Kast, 02].

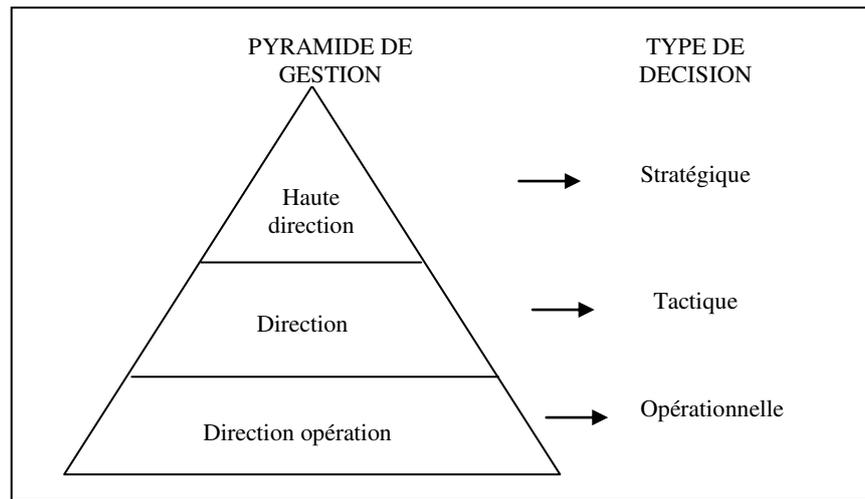


Figure 1.3. Les niveaux de décision [Kast, 02].

[b] *Selon l'échéance de la décision*

IGOR ANSOFF a proposé un classement des décisions en trois catégories [Ansoff, 1984] :

- La décision de court terme : c'est une décision qui engage l'avenir sur une courte période. De quelques jours à quelques mois (pas plus d'un an en général), par exemple, le choix d'un fournisseur occasionnel pour une faible quantité d'une pièce détachée.
- La décision de moyen terme : elle engage l'avenir sur une période d'une année et plus, par exemple, le remplacement d'une machine dans une usine.
- La décision de long terme : elle engage l'avenir de l'entreprise sur une longue période (5 ans, 10 ans, même plus). Elles sont souvent stratégiques, par exemple, l'implantation d'une usine dans une région.

[c] *Selon l'objet de la décision*

Ansoff définit la classification suivante [Ansoff, 84] :

- La décision stratégique : c'est une décision fondamentale, essentielle, qui engage l'avenir de l'entreprise à moyen et long terme. Elles concernent les relations de l'entreprise avec son environnement (par exemple, décision de produire un nouveau produit touristique pour une clientèle particulière). Elle doit être mûrement réfléchie, elle engage l'avenir. Ce type de décision est du ressort de la direction.
- La décision tactique : Elle se prend en situation moyenne dans la hiérarchie de l'entreprise. Les décisions de ce niveau sont des décisions de gestion qui assurent dans le moyen et le court terme, la réalisation des décisions stratégiques, par exemple, le choix d'un fournisseur après validation d'une commande de matière première.
- La décision opérationnelle : Elle se prend en bas de la pyramide hiérarchique de l'entreprise, et consiste à assurer le fonctionnement courant et constant de l'entreprise. C'est une décision de "routine" et généralement elle ne pose pas de

difficulté particulière. Par exemple, changement d'un fournisseur en cas d'indisponibilité d'un produit.

[d] *Selon la nature des variables de décision*

Simon a proposé une autre classification [Simon, 77] :

- Les décisions programmables : ce sont des décisions faciles à prendre qui portent sur des données quantitatives et peu nombreuses. Il est alors facile de formaliser la décision par l'élaboration d'une procédure d'exécution automatique.
- Les décisions non programmables : ce sont des décisions difficiles à prendre pour lesquelles les données sont qualitatives et nombreuses. Il est difficile de les inclure dans une procédure ou modèle mathématique. Elles répondent à un événement nouveau. Il est évident que ce genre de décision est plus coûteux en temps et financièrement.

Avec la croissance continue des connaissances médicales et l'arrivée de nouvelles maladies, le diagnostic est devenu complexe. Les méthodes classiques d'investigation médicales ont montrés des limites. Dès lors, l'IA a été mise à contribution dans le domaine médical vers les années 1970 [Shortliffe, 76], mais sans impact considérable, néanmoins beaucoup de travaux ont été menés et ont contribué à l'avancée de la recherche médicale. Vers les années 90, une étape a été franchie avec l'apparition des systèmes experts médicaux et de l'aide à la décision médicale [Szolovits et al., 88 ; Turban, 93], profitant de l'évolution de l'informatique avec l'arrivée des entrepôts de données médicales et des nouvelles technologies de l'information. L'aide à la décision médicale s'est alors imposée d'elle-même dans la pratique médicale courante, et puis devenue un axe majeure de l'informatique médicale.

Nous présentons dans la section suivante l'aide à la décision médicale, où nous notons une très grande avancée notamment par de nombreux travaux liés au diagnostic et au traitement des maladies comme le diabète, l'asthme et autres [Sivakumar, 07 ; Jha et al., 13 ; Marling et al., 08 ; Sefion et al., 03.a ; Sefion et al., 03.b].

1.3 Aide à la décision médicale

Les données et connaissances médicales sont devenues de plus en plus nombreuses et complexes. Le médecin ne peut donc mémoriser l'ensemble des connaissances médicales dont il a besoin dans sa pratique quotidienne. De ce fait, il doit être mieux fourni en moyens pour réaliser ces tâches liées à la prise en charge des patients. Parmi ces outils on trouve les systèmes d'aide à décision qui occupent une large espace de recherche en santé depuis de nombreuses années.

En fait, en médecine, la décision est considérée comme étant le centre de l'acte médical. Le processus de la décision médicale consiste entre autres à poser un diagnostic, une thérapie, un pronostic. Cet, acte médical est centré sur un raisonnement que le clinicien doit adopter afin d'aboutir à une action réfléchie. Ce raisonnement est destiné à soutenir cet acte. Ceci implique l'utilisation de divers données, informations, connaissances et méthodologies de raisonnement

clinique. Nous passerons en revue quelques notions de base sur ce raisonnement clinique qui nous permettent la compréhension du domaine que nous investiguons.

1.3.1 Le raisonnement clinique

Le raisonnement clinique est résumé selon Pelaccia et al., [Pelaccia et al., 11], et Kassirer [Kassirer, 10], comme suit : « *Le processus de raisonnement clinique est analytique (modèles hypothético-déductifs), non analytique (reconnaissance de la similitude avec un cas déjà vu) ou une combinaison des deux. Le modèle analytique est considéré comme suite d'étapes qui contiennent en premier lieu la génération des hypothèses de diagnostic, puis la recherche d'informations cliniques pour confirmer ou invalider ces hypothèses. Les informations cliniques recueillies peuvent en outre déduire de nouvelles hypothèses. Ce processus est effectué jusqu'à la confirmation ou l'élimination du diagnostic. Le modèle non analytique est également considéré comme la reconnaissance d'une situation clinique stockée dans la mémoire et qui correspond à l'expérience clinique. Cette expérience clinique contribue à générer des hypothèses, mais cette interaction n'est pas toujours positive, le rappel d'une situation clinique peut parfois perturber un objectif, mais peut également compléter l'analyse des signes observés* » [Pelaccia et al., 11 ; Kassirer, 10]. Pour notre approche, nous nous sommes basés sur un modèle non analytique pour étudier le raisonnement du médecin face à une situation pathologique. Le médecin utilise souvent sa compétence (le raisonnement) et les situations plus ou moins semblables déjà rencontrées (la mémoire). Ainsi, le raisonnement clinique du médecin implique les éléments mentionnés dans la figure 1.4, et qui sont déterminants dans la résolution de la présente situation clinique qui se pose devant lui.

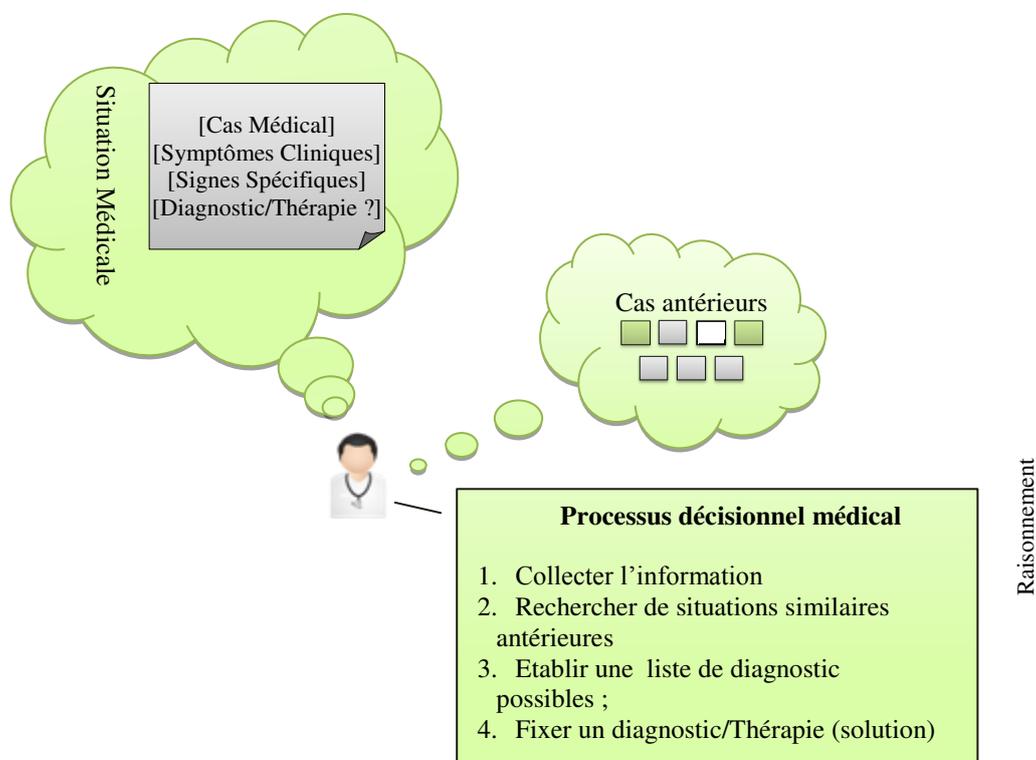


Figure 1.4. Le raisonnement clinique.

1.3.2 La décision en situation

Une notion importante en théorie de la décision est celle de la **décision en situation** [Lebraty et Guarnelli, 14]. L'intégration du concept de situation décisionnelle sera le fondement établissant le courant de la « décision en situation » [Rasmussen, 86 ; Klein, 98 ; Lebraty et Guarnelli, 14]. Dans cet ordre d'idée, « *l'analyse d'une décision doit intégrer le contexte dans lequel elle est prise. Le modèle décisionnel va se focaliser sur la reconnaissance, par le décideur, de la situation décisionnelle (Recognition-Primed Decision Model)* » [Lebraty, 06]. L'idée de cette approche est qu'elle n'étudie plus le processus cognitif indépendamment du contexte dans lequel il s'inscrit ou se déploie. Cette approche propose que la décision ne soit pas exécutée hors de son contexte, c'est-à-dire hors des règles dans laquelle la situation est décrite [Reason, 93]. Lebraty définit alors le contexte comme suit : « *L'ensemble des éléments, perçus par le décideur, qui exercent une contrainte sur la tâche gérée. Ainsi, le contexte est à la fois dépendant de la tâche et subjectif. Il peut être vu comme le savoir explicite et tacite permettant de mettre en œuvre les compétences du décideur dans une situation donnée* » [Lebraty, 06].

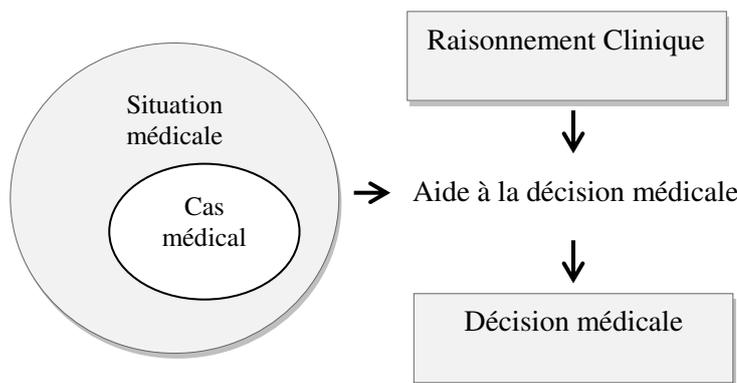


Figure 1.5. Éléments concourant à la décision médicale.

1.3.3 La décision médicale

À la prise de décision, le médecin doit agir sans connaître l'ensemble des données relatives à un patient et bien entendu toute la connaissance spécifique de la situation. Le médecin a souvent besoin d'aide afin d'établir une décision de qualité suite à un diagnostic médical. Donc, le diagnostic médical devient le préalable à toute décision. Ce processus de reconnaissance et de recherche de solution est long et délicat. Ce qui a conduit à la conception et au développement de systèmes ayant pour but d'appuyer la décision médicale. Ce qui est communément appelé : aide à la décision médicale.

Si nous voulons définir l'aide à la décision médicale, nous pouvons dire : « *c'est l'ensemble des techniques et outils de traitement de données permettant d'aider un processus d'établir une décision se rapportant à une situation médicale* ».

Ce processus a pour finalité d'aider un praticien de la santé à prendre des mesures adéquates lors de la prise en charge des patients. Cependant, cette aide est conditionnée par un diagnostic médical préalable à tout acte de décision pris par le médecin. Il s'agit donc

d'acquérir la connaissance à travers un contexte observable. Sournia définit le diagnostic médical comme suit : « *Démarche intellectuelle par laquelle une personne d'une profession médicale identifie la maladie d'une autre personne soumise à son examen, à partir des symptômes et des signes que cette dernière présente, et à l'aide d'éventuelles investigations complémentaires* » [Sournia].

En effet, un diagnostic médical représente une tâche difficile à réaliser parce qu'il repose sur la capacité de raisonnement du médecin, de son aptitude à discerner les symptômes. Ce diagnostic conditionne la décision médicale. Cette étape est d'autant plus difficile à cause des informations utilisées qui peuvent être entachées d'incertitude et d'autres formes d'imperfection. Cette incertitude peut provenir de diverses origines : erreur dans les données, ambiguïté dans la représentation des données, incertitude sur les relations entre les diverses données, etc. Ces difficultés ont conduit à la conception et au développement de systèmes d'aide au diagnostic ayant pour but d'assister les médecins dans l'élaboration de leurs diagnostics et la prescription de thérapies adéquates.

Un diagnostic médical représente donc, l'acte d'associer le nom d'une ou plusieurs maladies, à des signes observés (antécédents, symptômes) dans le cas d'un patient.

De ce fait, nous pouvons dire que le processus de diagnostic médical se déroule alors comme suit :

- (a) Le médecin constate les symptômes se manifestant chez un patient. A partir de ces symptômes, il formule des hypothèses de diagnostic initial.
- (b) Il procède à un examen initial du patient, qui lui permet d'augmenter le degré de confiance pour certaines hypothèses et le diminuer pour d'autres. En même temps, le médecin pose au patient des questions dont les réponses sont utiles à conforter ou rejeter une hypothèse initialement fixée.

Si le cas reste ambigu après ces étapes, le médecin cherche alors une autre source d'informations qui puisse apporter une information supplémentaire permettant d'éliminer l'ambiguïté. Celle-ci est généralement fournie à travers des examens complémentaires (analyses, radiographies, etc.). Souvent, ces analyses complémentaires viennent compléter les informations en sa possession. Si le médecin n'arrive toujours pas à établir un diagnostic complet et fiable, une dernière étape consiste à ce qu'il ait recours à l'étude d'une base de cas similaires traités par le passé afin d'établir une correspondance avec le cas actuel.

De ce fait, il a été possible de développer des systèmes centrés sur l'action médicale, permettant aux cliniciens de bénéficier des possibilités offertes par l'informatique et les méthodes de traitement de l'information, pour améliorer leurs connaissances, leurs décisions, et maîtriser leurs activités. Ces systèmes sont appelés : système d'aide à la décision médicale (SADM).

1.4 Les systèmes d'aide à la décision médicale (SADM)

Les systèmes experts étaient à l'origine des premiers SADM. Par la suite, les cliniciens ont montré un grand intérêt pour ces systèmes, dans leurs pratiques courantes en milieu médical (cabinet médical, laboratoire, hôpital et autres) [Osheroff, 09] :

- pour décider de certaines tâches de routine ;

- pour alerter les cliniciens de problèmes potentiels ;
- ou pour suggérer des examens aux cliniciens.

La littérature est très variée en définitions et dénominations pour les systèmes d'aide à la décision [Keen et al., 78 ; Alter 80, McKenney et Scott, 71]. Certains auteurs mettent l'accent sur le type de problèmes ou sur les fonctionnalités du système en question, d'autres mettent en valeur les composants ou les processus qu'ils utilisent. Pour une plus large lecture sur les SIAD, nous orientons le lecteur sur [Hamdadou et Bouamrane, 15].

Dans le milieu médical, les SIAD sont aussi appelé SADM [Serroussi et Bouaud, 14], d'autres les appellent SIADM [Ltifi et al., 10]. Dans notre travail nous utilisons l'appellation SADM, avec la précision que l'aspect interactivité est fortement présent dès lors que l'utilisateur est assuré de pouvoir effectuer les opérations suivantes qui lui permettent d'interagir avec le système soit pour introduire des données, lancer des processus ou faire des validations de résultats renvoyés par le système.

Jusqu'au début des années 1990, l'utilisation des SADM s'est limitée au milieu hospitalier. Au début des années 2000, leur usage est sorti du cadre de l'hôpital grâce au développement des Technologies de l'Information et de la Communication (TIC).

De nos jours, ils couvrent presque l'ensemble des activités médicales de diagnostic, de thérapie, de pronostic, etc.

1.4.1 Définitions

Définition 1.

Serroussi et Bouaud définissent les systèmes d'aide à la décision médicale (SADM) comme étant « *des applications informatiques dont le but est de fournir aux cliniciens, en temps et en lieu utiles, les informations décrivant la situation clinique d'un patient, ainsi que des connaissances appropriées à cette situation qui soient correctement filtrées et présentées afin d'améliorer la qualité des soins et la santé des patients* » [Serroussi et Bouaud, 14].

Berner précise aussi que pour « *améliorer la qualité des soins et la santé des patients les informations doivent également être : correctement filtrées et présentées ; fournies en temps et lieu utiles* » [Moreno, 15].

Définition 2.

Greenes définit l'aide à la décision médicale comme suit : « *l'utilisation des technologies de l'information et de la communication fournissant des connaissances pertinentes pour les soins de santé et le bien-être d'un patient* » [Mereno, 15 ; Greenes, 14].

Bates et al., précisent que les connaissances pertinentes sont des « *informations passives et actives, de référence, ainsi que des rappels, des alertes et des recommandations* » [Mereno, 15 ; Bates et al., 2003].

1.4.2 Les objectifs du SADM

Un SADM a les principaux objectifs suivants :

- proposer des données et des informations à la demande du clinicien lors des différentes activités de diagnostic et de recherche de thérapies;
- proposer un diagnostic, une thérapie, ou un pronostic ;
- alerter au bon moment pour éviter des événements indésirables.

De plus, et d'une manière générale, il peut intervenir sous diverses formes. On a alors :

- l'aide à la documentation des soins ;
- l'accès en ligne aux informations de référence ;
- la gestion de protocoles ou processus complexes.

1.4.3 Les principales fonctions du SADM

Les fonctions d'aide à la décision les plus courantes dans les systèmes d'aide à la décision médicale sont l'alerte et le rappel. Dans un environnement en temps réel, ces fonctions sont attachées aux dispositifs de surveillance pour fournir des alertes immédiates au fur et à mesure que la condition de déclenchement se produit. Par exemple, la surveillance d'oxygène et de pression artérielle dans un contexte difficile peuvent alerter les infirmières si l'état du patient dépasse un seuil fixé. Dans un contexte chronique, une simple analyse des résultats de laboratoire ou une alerte par email à l'intention du décideur sont des fonctions d'aide à la décision utiles. Certains systèmes d'aide à la décision médicale peuvent fournir des fonctions de reconnaissance et d'interprétation d'images. Ceux-ci sont extrêmement utiles dans des situations où des rapports de radiologie peuvent être interprétés et des alertes peuvent être générées pour attirer l'attention des médecins [What-when-how].

Table 1.1. Fonctions d'aide à la décision et exemples de problèmes cliniques (traduit [What-when-how]).

Fonction	Exemples de problèmes cliniques
Alerte	Basée sur les résultats de laboratoire avec différents niveaux personnalisables
Diagnostic	Identifier le diagnostic possible en fonction de l'historique, du physique, des résultats et des données saisies.
Rappel	Rappeler aux praticiens les ordres et leurs calendriers
Notification	Non-conformité, risques, événements anormaux et périodes de soins
Suggestion	Ajustements médicamenteux, les tendances et les dosages actuels de médicaments
Interprétation	Directives pour la situation actuelle - calendrier de tests-laboratoire, protocoles de soins
Prédiction	Prédire les résultats en fonction de certaines variables indépendantes
Assistance	Fournir un autre médicament suite à une interaction médicamenteuse ou à une allergie
Critique	L'utilisation d'une procédure médicale sur la base des directives médicales applicables en médecine et des antécédents médicaux du patient

1.4.4 Structure d'un SADM

Les systèmes d'aide à la décision médicale se composent de [Moreno, 15] :

- une base de connaissances construite à partir de recommandations et principes de la médecine ;
- des modèles de connaissances représentés selon des formalismes de représentation des connaissances et ;
- un moteur d'inférence utilisant des méthodes de raisonnement pour déduire l'aide à la décision de diagnostic, thérapeutique ou pronostic ;
- interface assurant la communication entre le clinicien et le système.

1.4.5 La typologie des SADM

Dans la large panoplie de systèmes ou prototypes développés, la littérature fournit plusieurs approches pour leurs classifications. Moreno cite par exemple deux types ceux basées sur la connaissance ou non [Moreno, 2015]. Nous n'en citerons que ces deux classifications à titre d'exemple.

[a] Selon le mode d'intervention

Ces systèmes sont fondés sur la manière dont ils interviennent dans le processus de décision en général, c.à.d., comment ils aident dans la prise de décision.

- *Les systèmes d'aide indirecte à la prise de décision ou systèmes d'assistance documentaire.*

L'accès à l'information pertinente fait partie du processus de la décision. Ainsi, l'accès aux résultats d'analyses biochimiques ou la consultation du dossier médical du patient constituent des aides indirectes à la décision. Cette aide intervient pour faciliter l'appréciation d'une situation par le médecin. Les systèmes de bases de données concernant les médicaments et leurs interactions sont des exemples de systèmes pouvant intervenir dans la décision médicale. Cependant, cette aide à la décision est restée au stade classique du stockage et retrait de l'information. Ces systèmes documentaires n'ont pas de méthode de raisonnement à proprement dit mais ils doivent gérer des bases de données.

- *Les systèmes de rappels automatiques ou systèmes d'alerte*

Ces systèmes permettent de rappeler au médecin des erreurs à ne pas commettre ou des éléments importants à prendre en compte lors de la prise de décision. L'assistance fournie n'est pas une aide au raisonnement ou à l'appréhension globale du cas du patient, mais plutôt un aide-mémoire fournissant une information utile et pertinente dans une situation clinique plus ou moins simple. Ainsi, par exemple, l'avertissement par une mise en garde lors de la description d'une posologie d'un médicament peut être aussi vu comme une aide précieuse dans la prise de décision.

Ces systèmes ne raisonnent pas véritablement mais l'aide devient plus personnalisée dans la mesure où le système tient compte des informations dont il dispose sur la situation envisagée. Les alertes peuvent être de nature différente comme par exemple, un protocole thérapeutique lorsqu'une pathologie est reconnue ou tout simplement fournir les valeurs normales des examens biologiques.

– *Les systèmes consultants*

Ils ont pour but de donner un avis de spécialiste devant une situation clinique précise, qu'elle soit de nature diagnostique ou thérapeutique. Les systèmes experts médicaux peuvent être classés dans cette catégorie. Ces systèmes raisonnent sur des situations médicales définies et fournissent à l'utilisateur des conclusions argumentées selon les méthodes de raisonnement employées. C'est dans cette catégorie que l'on note le plus de réalisations en matière de système d'aide à la décision.

[b] *Selon la représentation à base connaissances (symboliques)*

Ces systèmes mettent en avant la manière utilisée pour stocker leurs connaissances médicales, ce qui influe aussi sur leurs modes de raisonnement. La base de connaissances contient les règles et les relations entre les données écrites le plus souvent la forme de règles "Si Alors". Ces SADM sont composés de trois parties : la base de connaissances, un moteur d'inférence et une interface pour communiquer avec le clinicien utilisateur du système. Le moteur d'inférence utilise les règles et les données du patient pour déduire des "solutions" [Moreno, 15].

[c] *Selon la représentation à base de données (numériques)*

Ces systèmes utilisent l'IA à travers des algorithmes spécifiques (algorithmes génétiques, réseaux de neurones). Cependant, ils ne peuvent pas expliquer les raisons de leurs conclusions. Ces systèmes ne sont pas directement utilisés pour les diagnostics, pour des raisons de fiabilité et de responsabilité. Néanmoins, ils peuvent être utilisés en post-diagnostic [Moreno, 15]. On trouve alors :

– *Les systèmes d'aide au diagnostic*

Ces systèmes utilisent plusieurs manières plus ou moins complexes pour aider le médecin qui les sollicite [Moreno, 15] :

- suggérer un ensemble de diagnostics possibles ;
- prescrire des examens complémentaires tels que des imageries, des analyses biologiques ou autres ;
- montrer un schéma thérapeutique à suivre pour déterminer la maladie responsable, suite à un diagnostic préliminaire ;
- produire une information qui peut aider au suivi d'une pathologie ;
- produire un résumé médical permettant la prise en charge de malades chroniques.

– *Les systèmes d'aide à la décision thérapeutique*

On peut trouver :

- les systèmes d'aide à la prescription

Le médecin prescrit une liste de médicaments et le système effectue un certain nombre de vérifications (allergie, interaction médicamenteuse, etc.) à travers une base de médicaments et du dossier médical du patient, pour valider ou non la prescription. En cas de détection d'un problème (interaction, etc.), le système génère une alerte. Ces systèmes sont seulement destinés à sécuriser la prescription d'un médicament donné [Moreno, 15].

- Les systèmes d'aide à la stratégie thérapeutique

Ils aident le médecin à choisir le bon médicament à prescrire, en établissant un schéma thérapeutique qui nécessite un suivi dans le temps.

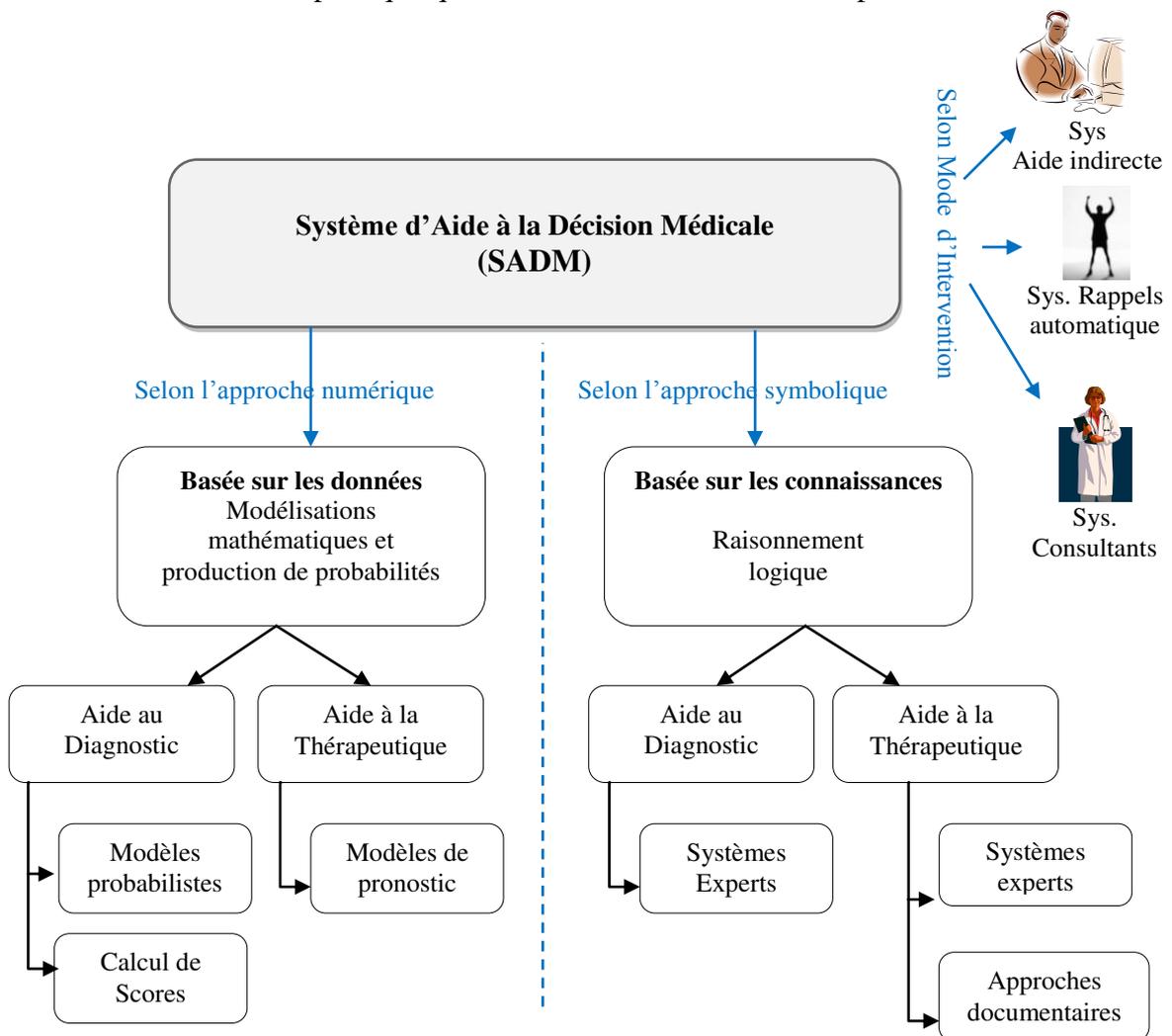


Figure 1.6. Typologie des SADM selon les approches utilisées .

1.4.6 Composantes du SADM

Les systèmes d'aide à la décision médicale ont fait l'objet de multiples réalisations. Depuis plus de vingt ans, les systèmes experts faisant appel aux méthodes de l'intelligence artificielle (IA), se multiplient en médecine comme dans d'autres domaines des sciences et des techniques. Avant eux, des approches plus traditionnelles, basées sur des méthodes statistiques ou probabilistes, avaient été largement développées. Les méthodes d'intelligence artificielle ont permis des avancées. Sur le plan technique, elles ont permis un saut qualitatif important dans la réalisation des systèmes complexes. Cette avancée ne concerne pas uniquement les applications médicales mais également les applications de gestion.

L'intelligence artificielle a donc permis de déboucher sur une véritable ingénierie de la connaissance. Elle se caractérise par une approche faisant une large part à la psychologie cognitive, à la représentation des connaissances symboliques et à la modélisation des processus de raisonnement. La figure ci-dessous montre la structure d'un système d'aide à la décision médicale qui comprend :

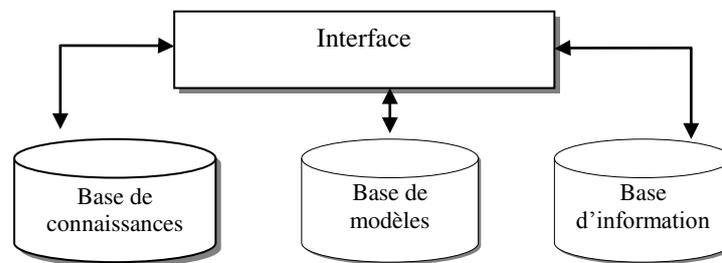


Figure 1.7. Structure d'un SADM [Holtzman 89].

– La base d'informations

Elle réunit les données sur le système, et assure les fonctions suivantes :

- mettre en corrélation les données de différentes sources ;
- rechercher des données suite à des requêtes ;
- faire des recherches complexes et des manipulations de données pour des requêtes.

– La base de modèles

Elle décrit le comportement du système. Elle se compose d'un ensemble de modèles et d'un système de gestion de ceux-ci. Les modèles peuvent être : des outils de recherche opérationnelle, des modèles statistiques ou autres.

– Une base de connaissances

Elle peut être un système à part entière et indépendant qui peut apporter une expertise en plus et spécifique au système en place. La base de connaissance regroupe pour sa part un ensemble de connaissances sur le domaine du problème.

- *Une interface Homme/Machine*

Elle assure la communication entre le système et l'utilisateur.

[a] *Les modèles de connaissances*

La base de connaissance regroupe l'ensemble des connaissances du domaine en question. Ces connaissances susceptibles d'être fournies à un système sont de divers types, on trouve, par exemple, l'interprétation des résultats d'examen biologiques, les connaissances anatomiques, physiopathologiques, épidémiologiques, taxonomiques (classification des maladies), les connaissances pharmacologiques et thérapeutiques. Szolovits et al., proposent de regrouper les modèles de connaissances utilisés en médecine en trois catégories [Szolovits et al., 88] :

- *Les modèles empiriques.*

Les connaissances empiriques concernent la connaissance des associations entre les maladies et les signes. Elles peuvent être fournies par un expert ou dérivées de l'analyse d'une base de données. Des connaissances de ce type sont très souvent utilisées dans les systèmes experts. Le meilleur exemple de ce type de connaissance et de sa mise en œuvre est certainement le système MYCIN [Shortliffe, 76] utilisant des règles de connaissance élaborées empiriquement par les experts. Exemple de règle de ce type :

Si couleur_urine="rouge" Alors patient="à haut risque"

- *Les modèles empiriques quantitatifs.*

Dans ce type de modèle, la connaissance est plus complexe. Elle combine une connaissance de type empirique, définie précédemment, et une connaissance modélisée de manière quantitative. Szolovits souligne l'intérêt de ce modèle et justifie l'usage d'un modèle empirique pour traduire une connaissance difficile à apprécier [Szolovits, 85]. Exemple de règle de ce type :

Si couleur urine = "rouge" et Dose=[20, 30] Alors cas="Salmoneia"

- *Les modèles physiologiques et physiopathologiques.*

Ce type de connaissances, permet de mettre en œuvre des raisonnements plus profonds décrivant, par des relations dont la sémantique est claire, les mécanismes qui sous-tendent des processus morbides. Les explications fournies à partir de connaissances causales sont plus faciles à comprendre. Les connaissances causales servent à identifier des états physiopathologiques. D'autres connaissances servent à classer le cas en fonction des états précédemment confirmés ou infirmés. Ce dernier type de connaissances, appliqué à des états identifiés et non à des données de départ, est utilisé à un niveau d'abstraction plus élevé que les règles de causalité. Il a pour but de produire des conclusions de diagnostic et de pronostic [Degoulet et Fieschi, 91].

[b] *Les formalismes de représentation des connaissances*

Les formalismes les plus utilisés dans les systèmes informatiques sont les règles de production et les objets structurés ou frames.

- *Les règles de production.*

Elles permettent de représenter simplement des connaissances qui s'expriment naturellement par des phrases conditionnelles, par exemple :

Si `taux_sucre_sang >= 126 g/l` Alors "le patient présente une glycémie"

La connaissance exprimée de cette façon déclarative, présente une simplicité d'expression et une facilité de compréhension liée à sa syntaxe.

De très nombreux systèmes, dont le plus célèbre est MYCIN, utilisent ce mode de représentation des connaissances et obtiennent des performances tout à fait intéressantes. Deux stratégies d'utilisation de ces règles peuvent être mises en œuvre. La stratégie guidée par les données, appelée "chaînage avant", utilisation de toutes les règles pour déduire tout ce qui est déductible. La stratégie guidée par le but, appelée en "chaînage arrière", à partir d'une proposition d'un but à atteindre, il ya construction de tous les chemins possibles.

- *Les objets structurés (frames).*

Les objets structurés sont tous les concepts qui se présentent à la pensée. Différentes connaissances et propriétés sont rattachées à chaque objet. On peut se représenter une structure d'objet comme un réseau de nœuds et de relations. Il est également possible de définir des liens dont la sémantique est précisée : relations entre classes et instances, ou entre ensembles et sous-ensembles.

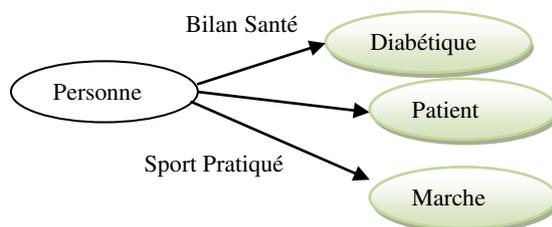


Figure 1.8. Exemple de représentation de la classe personne dans une application médicale.

A ce niveau, on peut opposer deux types de connaissances, la connaissance statique et la connaissance dynamique. La connaissance statique correspondant à la définition des concepts. La connaissance dynamique décrit la façon d'utiliser les concepts et leurs propriétés dans un raisonnement ou l'aspect empirique des règles est mis en évidence.

[c] *Les méthodes de raisonnement*

Les méthodes de résolution de problème et de raisonnement sont très variées et peuvent être mises en œuvre sur les différents modèles de connaissances. Ainsi, on peut avoir recours aux types de raisonnement suivants :

- Le raisonnement déductif, qui porte sur des données catégoriques ou des données incertaines et/ou imprécises et mettant en œuvre le principe de l'implication logique ou l'une de ses généralisations.

- Le raisonnement hypothético-déductif, ou une généralisation du raisonnement par l'absurde (par réfutation), permet de focaliser la recherche d'une solution à un problème déterminé.
- Le raisonnement qualitatif permet d'exprimer des connaissances de bon sens courant.
- Le raisonnement inductif et le raisonnement par analogie sont également utilisés.

Ces raisonnements peuvent être mis en œuvre aussi bien pour résoudre des problèmes où toutes les situations possibles sont énumérées a priori pour résoudre des problèmes où cette énumération n'est pas possible.

1.4.7 *Les méthodologies d'aide à la décision médicale*

1.4.7.1 *L'approche numérique*

Cette approche, basée principalement sur la mise en œuvre d'algorithmes numériques (comme l'analyse discriminante ou le théorème de Bayes) a été historiquement la première utilisée dans les systèmes d'aide à la décision [Bruland et al., 10].

1.4.7.2 *Les analyses discriminantes et les méthodes statistiques*

Ces méthodes sont appliquées sur un échantillon de cas (dont on connaît, par exemple, le diagnostic) pour déterminer la fonction discriminante. Schématiquement, si l'on suppose que l'on cherche à discriminer les sujets malades des non malades, on peut dire qu'il s'agit, dans un espace à p dimensions correspondant aux variables décrivant les patients, de trouver le plan qui sépare au mieux (au sens d'un certain critère) les points correspondant aux malades et les points correspondant aux sujets sains. Cette fonction obtenue est testée sur un autre échantillon de données afin d'évaluer sa validité. De nombreuses méthodes ont été proposées et appliquées dans différents domaines médicaux [Bellazi et Zupan, 08 ; Abbasi et Kashiyarndi, 06].

1.4.7.3 *Les systèmes bayésiens*

L'approche bayésienne a donné lieu à de nombreuses applications, parmi lesquelles celle de De-Dombal sur les douleurs abdominales aiguës est l'une des plus significatives. Ce système utilise un modèle bayésien pour calculer les probabilités des affections se manifestant par des douleurs abdominales aiguës. Chaque patient est défini par 35 à 40 variables et les 7 catégories diagnostiques connues du système sont l'appendicite, la diverticulose, la perforation d'ulcère duodéal, la pancréatite, l'occlusion du grêle et les douleurs abdominales non spécifiques [De Dombal et al., 1972].

1.4.7.4 *L'approche intelligence artificielle*

L'importance de la connaissance pour réaliser des tâches de manière intelligente fait l'objet de plusieurs réalisations en intelligence artificielle [Szolovits et al., 88 ; Koton, 88]. Le but est déduire de nouvelles conclusions ou solutions en utilisant de manière formelle les descriptions d'objets ou entités réelles représentées dans un formalisme adéquat, et qui se prêtent bien pour les traitements voulus. Deux modes de représentations sont alors utilisés : la représentation procédurale pour une connaissance si celle-ci traduit un algorithme. Sinon,

cette connaissance n'est pas formulable algorithmiquement, et là, c'est une représentation déclarative. Donc, avant tout traitement, il faudra réfléchir et choisir un mode de représentation convenable pour la représentation des états du système, et la représentation des connaissances utilisées pour produire les nouveaux états bien sûr par "dérivation" ou déduction [Szolovits et al., 88 ; Koton, 88].

1.4.8 Conclusion

En se basant sur le mécanisme de résolution de situations médicales, il est remarqué que le raisonnement du médecin repose en grande partie sur le fait que la situation courante a une probabilité qu'elle soit déjà traitée auparavant, et de ce fait le médecin va proposer une solution plus ou moins identique à celle déjà utilisée. Ce raisonnement est très proche d'une méthodologie de raisonnement très connue sous le nom du raisonnement à base de cas (RBC).

Ceci a motivé énormément des travaux sur ce mode de raisonnement dans le domaine médical et a conduit à la réalisation d'outils informatiques de résolution de problèmes décisionnels s'appuyant entièrement sur ce mode de raisonnement (RBC). Ces travaux ont des ramifications dans les domaines variés de l'intelligence artificielle : représentation des connaissances, classification, mesures de similarité, etc., ce qui en a fait un mode de raisonnement complexe mais largement utilisé en aide à la décision médicale.

1.5 Aide à la décision médicale par RBC

L'utilisation de l'approche RBC est très utilisée en médecine à cause du raisonnement utilisé et qui est proche de celui du clinicien face à une situation pathologique donnée. En effet, un clinicien utilise la même démarche dans la recherche d'une solution médicale en se basant sur sa mémoire pour essayer de se remémorer des cas précédents déjà vécus en consultation, et delà il peut facilement s'orienter vers une situation analogue et l'ajuster si possible à sa présente situation.

De plus, cette approche est tout à fait justifiée dans les domaines où le traitement ne repose pas sur une méthode algorithmique structurée ou non pour aboutir à une solution, mais repose seulement sur la connaissance stockée qui est la solution expérience passée.

1.5.1 Le mode RBC

Le raisonnement à base de cas (RBC) appartient aux méthodes de l'intelligence artificielle (IA). Ce mode consiste à résoudre un problème en s'appuyant sur une expérience passée. En outre, un des points clés pour un système RBC est la recherche de cas pertinents, d'où l'importance d'un procédé qui va étiqueter les cas de façon à ce qu'ils puissent être retrouvés au moment opportun. Cette opération d'étiquetage est une sorte d'indexation qui affecte aux cas des indices qui vont les représenter durant les différentes phases qui composent le système RBC.

Le RBC est donc un paradigme de résolution des problèmes. Au lieu de compter seulement sur la connaissance générale d'un domaine du problème, il s'appuie sur la remémoration de problèmes passés et résolus, appelés les *cas sources*, pour résoudre un problème courant, appelé *problème cible*, et une nouvelle expérience est maintenue chaque fois qu'un problème

a été résolu, la rendant immédiatement disponible pour de futurs problèmes [Aamodt et Plaza, 94]. Les bases du RBC ont été posées par les travaux de Minsky et Schank à la fin des années 70.

La théorie développée par Minsky présente la notion de "frame, script, ou schéma" qui correspond à une structure remémorée qui doit être adaptée pour correspondre à la réalité d'une nouvelle situation rencontrée. Nous avons ainsi une représentation particulière des connaissances servant de support de raisonnement sur le monde réel [Minsky, 81]. Schank s'inspire ensuite des travaux de Minsky et formule, pour la première fois, le paradigme RBC.

D'après Schank, le processus de compréhension correspond à un processus d'explication qui s'applique d'une manière itérative. Illustrons ceci par une situation typique se présentant à un médecin : "Un médecin ayant examiné un patient particulier dans son cabinet se rappelle un autre patient qu'il a déjà traité il y a peu de temps auparavant. Nous supposons que ce rappel a été déclenché par une ressemblance des symptômes cliniques, le médecin utilisera ces mêmes symptômes pour déterminer la maladie et le traitement pour le patient en cours d'examen".

[a] *La base de cas*

L'objectif du raisonnement à base de cas est de résoudre des problèmes courants à partir d'expériences passées. Le processus met en œuvre une base de cas composée d'expériences passées dans laquelle peuvent être recherchées des expériences similaires au problème à résoudre (problème courant).

Un cas est composé de deux parties : la *partie problème* et la *partie solution*. La partie problème est composée d'un ensemble d'indices qui déterminent dans quelle situation un cas est applicable et utile. Les problèmes résolus sont stockés dans la base de cas. Lorsqu'un nouveau problème se présente, ce problème est alors décrit par un cas dit cas cible où seule la partie problème est connue.

[b] *Le raisonnement*

Un système de raisonnement à base de cas se fonde sur la comparaison des cas nouveaux aux cas existants. Tout d'abord, un ensemble d'expériences sont stockées avec leurs solutions respectives. Ensuite, lorsqu'il y a une nouvelle expérience, il s'agit de la comparer à celles qui sont stockées. Suivant la proximité, la similarité de celle-ci avec telle ou telle expérience existante et stockée, une solution s'adaptera à ce nouveau cas.

Ce système s'enrichit au fur et à mesure des nouveaux cas rencontrés, c'est une sorte de mise à jour des données. De plus, la facilité de résolution d'un nouveau cas augmente en fonction du nombre d'expériences stockées dans la base de cas.

Le principe du raisonnement est le suivant : la résolution basée sur la réutilisation par analogie des expériences passées. Un cas courant est comparé aux cas précédemment enregistrés dans la base de cas. La comparaison est effectuée par le calcul d'une mesure de similarité. Les cas ayant obtenu les meilleurs scores par cette mesure sont déclarés plus proches voisins et sont sélectionnés pour construire une solution au problème courant. Cette solution est alors « révisée » par l'utilisateur, puis introduite dans la base de cas pour être réutilisable pour une prochaine. Ce cycle est répété à chaque nouvelle situation.

1.5.1.1 Le cycle du raisonnement.

Pantic ou Aamodt et Plaza définissent un cycle de raisonnement ayant 5 phases [Pantic, 05 ; Aamodt et Plaza, 94]. Ces principales phases sont décrites par la figure 1.9 : élaboration, remémoration, adaptation, révision et mémorisation. Ces étapes tournent autour d'une base de connaissances du domaine d'application. Chacune des étapes du cycle mobilise ces connaissances pour supporter la recherche de la solution du problème cible.

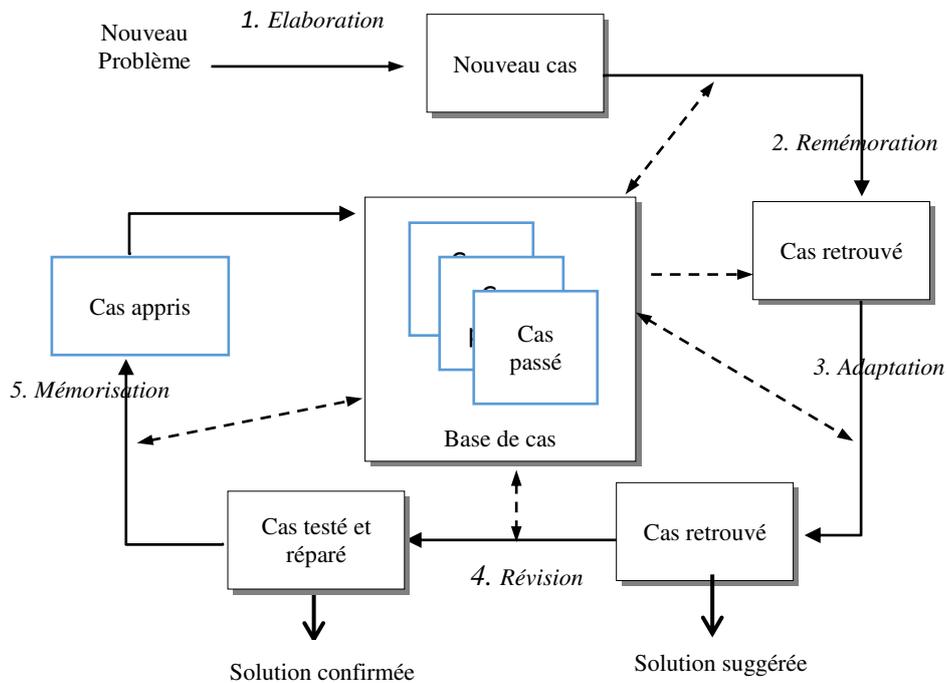


Figure 1.9. Principe de fonctionnement du RBC (cycle RBC), adapté de [Pantic, 05].

1. L'élaboration d'un nouveau problème (cas cible)

Elle représente l'acquisition et la modélisation des informations connues sur le nouveau problème (cas cible), pour lui donner une description initiale, de manière identique aux cas existants dans la base de cas.

2. La remémoration des cas (cas source)

C'est la recherche des cas similaires au cas cible, cela signifie la recherche des correspondances entre les descripteurs des cas de la base (cas source) et ceux du cas cible. Des mesures de similarités sont alors à définir sur les indices constituant la partie problème d'un cas. Les cas extraits de la base de cas sont appelés cas sources. Parmi les algorithmes de remémoration les plus utilisées, on trouve l'algorithme des *k-plus proches voisins* (*k-ppv* ou *k-nn* ; *nearest neighbors*).

▪ La méthode des *k-ppv*

L'algorithme des *k-plus proches voisins* est un algorithme dédié à la classification qui peut être étendu à des tâches d'estimation. Il consiste à déterminer pour chaque nouvel individu *Y* que l'on veut classer, la liste de ces plus proches voisins parmi les individus déjà classés.

L'individu Y est affecté à la classe qui contient le plus d'individus parmi ses plus proches voisins. Cette méthode nécessite de choisir une distance, la plus classique est la distance euclidienne, et le nombre k de voisins à prendre en compte.

Ce principe est appliqué dans le RBC comme suit : Un cas à résoudre est comparé à tous les cas (source) de la base de cas. On choisit pour le nouveau cas les k cas sources, plus proches dans la base de cas, au sens d'une distance choisie, par exemple, la distance euclidienne.

Pseudo Algorithme : K-ppv

- 1: Entrée : la valeur de k ; une mesure de similarité ; un échantillon de m exemples et leurs classes, une nouvelle instance Y
 - 2: Déterminer les k plus proches exemples de Y en calculant les distances
 - 3: Combiner les classes de ces k exemples en une classe c
 - 4: Sortie : la classe de Y est $c(Y)=c$
-

▪ ***La mesure de similarité***

La recherche des cas similaires au problème à résoudre est basée sur le concept de *similarité*. La mesure de similarité cherche des correspondances entre les descripteurs des cas sources et ceux du cas cible à l'aide d'un algorithme de recherche. L'objectif de cette mesure de similarité est de retrouver dans la base de cas, le cas similaire au problème actuel dans le sens qu'il soit facilement adaptable au cas cible. La mesure de similarité est une somme pondérée de calculs locaux de similarité pour chacun des descripteurs possibles des cas. Cette mesure de similarité est ajustable à deux niveaux :

- pour chaque descripteur, on doit définir la similarité entre les différentes valeurs possibles (par exemple, par des matrices quand les descripteurs prennent des valeurs discrètes ou par une fonction quand ils prennent des valeurs numériques) ;
- on doit définir les poids relatifs des différents descripteurs les uns par rapport aux autres. Cette mesure de similarité doit être définie de manière assez fine pour que les cas retrouvés aient des solutions assez proches d'une solution possible au cas cible.

Les mesures de similarité peuvent être :

- locales et établies au niveau des caractéristiques du cas, et généralement basées sur la notion de distance, et dépendent du type de descripteur (numérique, symbolique, taxonomique) ;
- ou globales et sont calculées au niveau des cas ou des objets en agrégeant les similarités locales,

On a alors les distances les plus courantes, Euclidienne, Manhattan, et Tchebychev. Mais, il est à noter qu'aucune mesure de similarité n'est parfaitement appropriée à tous les domaines [Gukhman, 65 ; Lesot et al., 08]

3. *L'adaptation des cas (cas source)*

C'est la réutilisation totalement ou partielle de la solution du cas trouvé le plus similaire, pour résoudre le nouveau problème. Cette étape a pour tâche de construire une solution Sol (cible) du problème cible en s'appuyant sur la solution Sol (source) du cas remémoré, appelé cas source et noté (source, Sol (source)). L'objectif de cette phase est de proposer une solution au problème courant (cas cible) en adaptant les solutions proposées par les cas sources. L'adaptation repose souvent sur l'utilisation des connaissances du domaine d'application. A l'issue de cette phase, une ou plusieurs solutions seront proposées pour le cas cible. En général, on retrouve deux approches d'adaptation de cas :

- L'approche transformationnelle : on obtient une nouvelle solution en modifiant des solutions antécédentes et en les réorientant afin de satisfaire le nouveau problème.
- L'approche dérivationnelle : en adaptant la méthode de génération de la solution. On garde, en fait, pour chaque cas passé, une trace des étapes qui ont permis de générer la solution. Pour un nouveau problème, une nouvelle solution est générée en appliquant l'une de ces d'étapes.

Une fois une adaptation trouvée, la solution est présentée à l'utilisateur. Peu de systèmes RBC font de l'adaptation complètement automatique. Pour la plupart des systèmes, une intervention humaine est nécessaire pour générer partiellement ou complètement une solution à partir d'exemples [Djebbar-Zaidi, 13].

4. *La révision de la solution proposée (solution cible)*

C'est l'évaluation de la solution proposée, ce qui sous-entend la possibilité d'une évaluation par le test dans un environnement réel ou simulé. Le retour d'information, suite au test, peut alors réorienter en cas d'échec de la solution proposée. Afin de procéder à la révision d'un échec, il est souvent utile d'expliquer cet échec en analysant les différences constatées entre les résultats des solutions obtenues et ceux qu'on aurait dû obtenir. Cette solution peut alors être testée. Si elle ne convient pas, il est possible de renseigner le système sur les causes de l'échec. Le système doit alors réviser ses connaissances sur le cas source ayant servi de base pour la résolution du problème et/ou sur les adaptations qui ont été effectuées. L'objectif de cette phase est de réviser les solutions proposées par la phase précédente en fonction de certaines règles et/ou heuristiques, qui dépendent du domaine de l'application. La phase de révision peut être faite par des experts dans le domaine de l'application ou d'une manière automatique. On peut alors considérer que la fonction d'apprentissage qui consiste à ajouter des nouveaux cas ou modifier des connaissances pour résoudre des situations d'échecs correspond à un apprentissage supervisé.

5. *La mémorisation d'un nouveau cas (cas cible).*

Une fois la révision faite et si le nouveau cas est d'un grand intérêt, il peut alors être enregistré pour enrichir la base de cas. On peut alors considérer que la fonction d'apprentissage qui consiste à ajouter de nouveaux cas ou modifier des connaissances pour résoudre des situations d'échecs correspond à un apprentissage supervisé.

Ces dernières phases du raisonnement (phases 4 et 5) sont généralement à la charge de l'expert du domaine responsable du système. C'est l'apprentissage d'un nouveau cas, qui pourra ainsi être utilisé pour la résolution de problèmes futurs. Cette phase va enrichir la base

de cas par les nouveaux problèmes résolus (cas cible auquel on a apporté une solution). En effet le cas résolu peut être rajouté à la base de cas pour être utilisé ultérieurement dans d'autres opérations de raisonnement. Cependant, avant d'ajouter ces cas, il faut juger la pertinence de cet ajout afin d'éviter, par exemple, l'ajout de cas redondants ce qui affectera les performances du système en termes de temps et de traitement sans pour autant améliorer la qualité des solutions apportées.

1.5.2 *Les caractéristiques des systèmes RBC en médecine*

Les caractéristiques des systèmes RBC en médecine sont nombreuses et peuvent être orientées selon deux fonctionnalités [Nilsson et Sollenborn, 04] :

[a] *Les caractéristiques orientées vers les objectifs*

– *Les systèmes de diagnostic.*

La majorité des systèmes de raisonnement à base de cas médicaux ont pour objectif le diagnostic. Ils tentent de fournir une aide aux praticiens dans la détermination d'un diagnostic, suivant différents degrés d'assistance.

– *Les systèmes de classification.*

Ces systèmes tentent d'identifier le groupe auquel appartient un cas. Le système de classification d'images en est un exemple.

– *Les systèmes de tutorat.*

Les systèmes RBC utilisent l'apprentissage par des exemples généralement réels de médecine et permettaient d'assigner à un système donné le *tutorat*. Ainsi un système de tutorat médical permet au clinicien d'accéder à des cas généralement réels ou parfois même fictifs et faire de l'apprentissage par les exemples.

– *Les systèmes de planification.*

Ils offrent une assistance dans la programmation de plan ou schéma thérapeutique comprenant plusieurs étapes.

[b] *Les caractéristiques orientées vers sa construction*

– Les systèmes hybrides

Ces systèmes hybrides tentent de mettre en œuvre une synergie entre raisonnement à base de cas et d'autres méthodologies de raisonnement.

– Les systèmes autonomes

Le degré d'autonomie est important pour un système de diagnostic. Il se fait donc en fonction du besoin de l'intervention du décideur dans le cycle de raisonnement et lors de l'évaluation des résultats.

1.6 Synthèse de l'aide à la décision médicale par RBC

Le RBC a été largement étudié dans la littérature sous divers aspects tant théoriques et qu'expérimentaux, notamment en explorant le domaine médical. Beaucoup de travaux concernant le RBC en aide à la décision médicale ont été menés dans les trois grands axes, à savoir le diagnostic, la thérapie et le pronostic (voir table 1.2), néanmoins nous citons quelques exemples illustratifs.

- *En diagnostic*

Althoff et al., ont décrit une approche pour développer les systèmes d'aide à la décision médicaux basés et l'ont adoptée pour réaliser un système d'aide à la décision en toxicologie au centre consultatif à Moscou pour diagnostiquer les cas d'empoisonnement par des psychotropes [Althoff et al., 98]. Jha et al., ont présenté une étude la détection et la prise en charge du diabète [Jha et al., 13]. Bareiss et al., ont développé un système pour le diagnostic cardiaque « PROTOS » [Bareiss et al., 88].

De Paz et al., ont également présenté un système d'aide à la décision basé sur le RBC pour le diagnostic de différents types de cancer [De Paz et al., 09]. Des systèmes sont créés pour le diagnostic de l'asthme, comme ADEMA [Sefion et al., 03.b], PROFORMA [Fox et al., 77], ou CARE-PARTNER [Bichindaritz et al, 03 ; Bichindaritz et al., 98].

- *En thérapeutique*

Le système CASIMIR a été développé pour le traitement du cancer du sein, [Bresson et Lieber, 00]. Marling et al., ont présenté une approche d'aide à la décision basée sur le RBC pour la gestion du diabète chez des patients atteints du diabète du type 1 [Marling et al., 08],

Par rapport à l'épidémie de l'asthme, un travail a été conduit pour comprendre cette pathologie, en essayant par exemple d'avoir un feedback à partir des données enregistrées régulièrement sur les consultations de médecine générale sur l'asthme [Kuilboer et al., 02].

Shanbezadeh et al., ont proposé un système d'aide à la décision pour le traitement de l'asthme [Shanbezadeh et al., 13]. D'autres travaux se sont vus orientés vers l'aide à la décision pour la prise en charge de cette pathologie [Alekosvska et Loskovska, 11]. Ceci, montre l'intérêt pour l'amélioration de la prise en charge des patients asthmatiques notamment en fournissant aux cliniciens des outils informatiques d'aide à la décision. Schwartz et al., ont aussi utilisé le RBC pour améliorer les soins en insulinothérapie [Schwartz et al., 08]. Song et al., ont proposé un système de radiothérapie qui utilise le RBC pour la planification de la dose pour le cancer de la prostate [Song et al., 07].

- *En pronostic*

Schmidt et Vorobieva ont présenté un système qui aide à expliquer les cas qui ne correspondent pas à une hypothèse médicale théorique utilisant le raisonnement basé sur des cas [Schmidt et Vorobieva, 05]. Saraiva et al., ont appliqué le RBR pour améliorer le processus de recherche du RBC. Ils ont utilisé les symptômes, les signes et les informations personnelles des patients en tant qu'entrées d'un modèle, ensuite ils appliquent le RBR pour définir les poids des attributs du cas qu'ils utilisent dans une fonction de similarité globale, et laissent le RBC converger vers la meilleure solution. La sortie du système présente la probabilité que le patient ait un type de cancer [Saraiva et al., 16]. Cette liste n'est pas exhaustive, mais elle montre la diversité de l'utilisation du RBC et souligne l'intérêt pour cette

approche à améliorer les soins des patients, en fournissant aux médecins des outils de traitement de données.

Table 1.2. Systèmes RBC et leurs domaines d'application (adaptée) [Begum et al., 11].

No	Author/system	Purpose-oriented properties	Application domain/context
01	McSherry/CaseBook	Diagnosis & classification	Contact lenses
02	De Paz/ExpressionCBR	Diagnosis & classification	Cancer diagnosis
03	Perner/Fungi-PAD	Classification, knowledge acquisition/management	Object recognition
04	Cordier/FrakaS	Diagnosis, knowledge acquisition/management	Oncology
05	Corchado/GerAmi	Planning, knowledge, acquisition/management	Alzheimer patients
06	Glez-Peña/geneCBR	Diagnosis & classification	Cancer classification
07	Perner/HEp2-PAD	Classification, knowledge acquisition/management	Image classifier
08	Schmidt/ISOR	Diagnosis & Planning	Endocrinology
09	Begum/IPOS	Diagnosis	Stress diagnosis
10	D'Aquin/KASIMIR	Diagnosis, classification, knowledge acquisition/management	Breast cancer
11	Bichindaritz/Mémoire	Diagnosis, planning, tutoring, knowledge acquisition/management	Biology & medicine
12	Montani/RHENE	Classification, planning, knowledge acquisition, management	hemodialysis
13	Kwiatkowska/Somnus	Diagnosis, planning, tutoring	Obstructive sleep apnea
14	Lorenzi/SISAIH	Diagnosis	Fraud detection in health care
15	Ochoa/SIDSTOU	Diagnosis, planning & tutoring	Tourette syndrome
16	Ahmed/Biofeedback	Planning	Stress management
17	Brien/ADHD	Classification, knowledge acquisition, management	Neuropsychiatry
18	Doyle/Bronchiolitis	Classification and tutoring	Bronchiolitis
19	O'sullivan/Dermatology	Diagnosis	Dermatology
20	Marling/Type-1 diabetes	Planning	Diabetes
21	Song/radiotherapy planning	Planning	Prostate cancer
22	Wu/Dietary counseling	Planning & Knowledge acquisition/management	Dietary counseling
23	Zhuang/Pathology	Classification, tutoring & knowledge acquisition/management	Pathology ordering
24	Ahn/Breast Cancer	Diagnosis	Breast cancer diagnosis
25	Huang/Chronic Diseases	Diagnosis, knowledge acquisition/management	Chronic diseases diagnosis
26	Chang/children Developmental	Diagnosis	Children with developmental delay
27	Houeland/Palliative care	Diagnosis & classification	Palliative care for long-term cancer
28	Nicolas/Melanoma	Diagnosis & classification	Melanoma
29	Topel/Metabolic disease	Diagnosis & Planning	Inborn Metabolic Disease
30	Arshadi/MOE4CBR	Classification	Biomedical Domain
31	Kurbalija/Multiple Sclerosis disease	Diagnosis	Multiple sclerosis disease
32	Obot/Hepatitis	Diagnosis	Hepatitis
33	CBSMS/Stress management	Diagnosis, classification & planning	Stress management
34	Yuan/HDCU	Classification, knowledge acquisition/management	Diabetes

1.6.1 *Les limites de l'aide à la décision médicale par RBC*

Bien que l'approche du RBC a été utilisée dans nombreuses situations médicales où elle était appropriée. Elle présente encore quelques limites liées principalement au domaine médical qui est un peu particulier à cause des types de données et des connaissances manipulées.

Gierl et al., citent quelques unes de ces limites en médecine [Zemirline, 08 ; Gierl et al., 98] :

- en raison du grand nombre d'attributs composant un cas médical, l'adaptation d'un cas est problématique. Néanmoins, les méthodes de généralisation et d'identification des attributs pertinents aident partiellement à remédier à cela ;
- en raison d'un grand nombre d'attributs décrivant la situation cible, des cas semblables peuvent être oubliés par le processus (cette situation peut se produire) et peuvent conduire à une décision moins robuste ;
- parfois, le RBC ne trouve pas nécessairement la solution concrète à un problème, alors il proposera seulement un ensemble de solutions possibles ;
- on constate en pratique un manque d'effort en acquisition de connaissances ;
- quand il s'agit de situations le développement des systèmes de RBC se heurte toujours à un problème d'ingénierie des connaissances, en particulier, les connaissances d'adaptation sont difficiles à modéliser, surtout médicales ;
- les systèmes de raisonnement à base de cas reposent sur des références (des cas exemples), ils ne peuvent fonctionner sans base de cas cliniques réels pour répondre à certains besoins des cliniciens ;
- un système RBC médical exige un grand niveau d'interaction avec l'utilisateur et surtout dans l'évaluation des résultats ;
- cette démarche utilise un modèle de domaine, d'où les difficultés de modélisation du savoir-faire des médecins (complexité des ontologies) ;
- comment formaliser et prendre en compte la démarche diagnostique ou thérapeutique des médecins ;
- les systèmes de raisonnement à base de cas présentent une acquisition automatique et incrémentale des connaissances à partir des cas ;
- vouloir adapter et utiliser le RBC dans le domaine médical est très fastidieux et complexe à cause des cas médicaux qui sont décrit par un nombre assez important de descripteurs ou attributs ;
- la capitalisation progressive des connaissances est faible dans le domaine médical. Les experts hésitent à partager leurs connaissances, acquises après des années de travail et ce pour diverses raisons tant professionnelles que personnelles [Dieng-Kuntz et al., 01].

1.6.2 *Conclusion*

Le raisonnement à base de cas (RBC) est une méthodologie puissante. C'est une méthodologie qui vise la réutilisation des expériences passées dans la résolution de nouveaux problèmes. Le RBC, comme méthodologie d'ingénierie des connaissances, peut être renforcée

dans les différentes étapes de son processus par la richesse d'autres méthodologies de raisonnement ou techniques c'est ce que nous appellerons : *intégration*.

Cette intégration a été largement déployée dans les systèmes de raisonnement multimodaux et elle s'est montrée bien adaptée notamment pour les travaux liés au domaine médical [Schmidt, et al., 01]. Cet intérêt pour cette approche multimodale impliquant le RBC remonte à plusieurs années et a récemment pris une ampleur notamment par les récents travaux des chercheurs [Bichindaritz et Marling, 10]. Il s'agit d'un axe majeur de la recherche sur le RBC dans différents domaines [Begum et al., 11 ; Bichindaritz et Marling, 10].

Des chercheurs ont proposé des solutions hybrides en combinant le RBC avec d'autres techniques comme le raisonnement à base de règles, et de nombreux travaux ont émergé de ces études [Verma et al., 14 ; Marling et al., 05]. Le premier système de raisonnement multimodal en médecine était CASE, il intègre le RBC avec un raisonnement à base de modèle (MBR) pour le diagnostic des insuffisances cardiaques [Marling et al., 05]. Le RBC a montré qu'il s'adapte bien aux intégrations au vue d'une multitude de travaux réalisés dans ce sens [Marling et al., 02 ; Marling et al., 05].

1.7 Les intégrations du RBC

L'intégration est une approche constructiviste visant à fournir des outils permettant de progresser dans la résolution d'un problème par le processus RBC. Cette intégration est une approche pour combler les lacunes inhérentes aux différentes phases de son cycle. Elle constitue une issue pour faire évoluer les différentes tâches du cycle RBC [Molines, 07]. La combinaison du RBC avec d'autres approches (raisonnement multimodal) représente une autre façon d'éviter des problèmes inhérents au RBC comme par exemple ceux liés à l'adaptation. Cette combinaison touche principalement la tâche de remémoration avec d'autres stratégies de raisonnement ou de recherche. La table 1.3 donne une liste non exhaustive des méthodes spécifiques utilisées, seulement nous avons voulu citer quelques exemples illustratifs par type d'intégration.

1.7.1 Intégration RBC-Le raisonnement à base de règles (RBR)

Comme il est bien connu, le raisonnement à base de règles (Rules Based Reasoning : RBR) consiste à utiliser les règles par un raisonnement en chainage avant ou chainage arrière, en exploitant les données pour arriver à une décision. Le RBR a été la première approche à être intégrée avec succès au RBC, c'est l'approche de raisonnement adopté par les systèmes experts classiques, tels que MYCIN. La combinaison RBC-RBR a reçu une attention particulière, car les règles suivent le formalisme de représentation de la connaissance explicite la plus adopté par les systèmes intelligents. En tant qu'outil efficace, le RBR s'est bien intégré au RBC et de nombreuses solutions ont été proposées.

Verma et al., ont proposé une solution hybride en utilisant un système basé sur la fouille de données (règles) et le RBC. Il consiste en une base de connaissances, un raisonnement à base de RBC et un sous-système de fouille de données pour proposer un modèle guidé par la connaissance pour produire actions recommandées aux utilisateurs. Cette combinaison vise à accroître la capacité à résoudre des problèmes et à améliorer la précision des actions suggérées [Verma et al., 14]. Cabrera et Edye, ont utilisé une intégration des règles et RBC

pour diagnostiquer les cas cliniques de méningite bactérienne aiguë. Ils proposent un système qui est initialement appliqué à la phase de pré-diagnostic avec des règles de diagnostic de base et si le stade de pré-diagnostic réussit, il y a une solution au problème qui est présenté à l'utilisateur, offrant par la même la possibilité de réviser nouveau cas, et si le cas n'est pas évident ou simple, le pré-diagnostic n'est pas applicable et le système passe à l'utilisation du RBC [Cabrera et Edye, 10].

CARE-PARTNER est un système interactif d'aide à la décision pour le suivi à long terme des patients transplantés de cellules osseuses. Il donne l'aide à la décision médicale aux centres de soin qui suivent les patients transplantés. Le système emploie un cadre multimodal de raisonnement qui combine le raisonnement à base de cas et le raisonnement à base de règles. Une des caractéristiques du système est qu'il emploie une base de connaissances riche en cas prototypes et des directives de pratique pour interpréter des cas médicaux.

1.7.2 Intégration RBC-Le raisonnement à base de modèles (MBR)

Le raisonnement à base de modèles (Models Based Reasoning : MBR) est une approche dans laquelle les connaissances générales sont représentées par la formalisation des relations mathématiques ou physiques présentes dans un problème d'un domaine donné. Le MBR a été combiné avec succès au RBC dans de nombreux domaines [Merida-Campos et Rollón Rico, 03]. La méthodologie MBR représente une alternative pour s'attaquer à certaines lacunes du RBC et de la complexité du domaine d'application. L'intégration RBC-MBR facilite généralement le processus d'adaptation, améliore les performances et l'efficacité des traitements [Koton, 88].

CASEY était le premier système RBC-MBR. Il utilise une base de cas patient et un modèle physiologique du cœur humain pour diagnostiquer les insuffisances cardiaques. Il a été interconnecté avec un programme d'insuffisance cardiaque à base du MBR déjà existant et utilise en plus un modèle physiologique du cœur pour faire correspondre les nouveaux cas aux anciens et déduire de nouveaux diagnostics à partir des anciens cas. Lorsque CASEY n'a pas trouvé une assez proche correspondance entre le cas courant et les anciens cas, il invoque le système MBR original pour résoudre le problème [Marling et al., 02].

PROTOS, est l'un des premiers systèmes RBC-MBR qui utilise un modèle multi-relationnel de connaissances pour diagnostiquer les maladies auditives et la remémoration des cas [Marling et al., 02].

Montani et al., ont essayé d'intégrer différentes méthodologies dans un système de raisonnement à base de modèles (MMR). Ce système a été utilisé dans l'appui de thérapie pour les patients diabétiques [Montani et al., 01]. Les auteurs affirment que la plupart des systèmes utilisant plus d'une méthode font cela seulement dans un mode exclusif, avec des méthodes fonctionnant simplement comme prolongements à une des autres méthodes. Montani affirme qu'un système de MMR a besoin d'une intégration beaucoup plus étroite des technologies pour obtenir l'ensemble des bénéfices d'une solution multimodale. Le système proposé essaie d'employer une intégration complète et d'utiliser le RBC, le raisonnement à base de règles et le raisonnement à base de modèles (RBM).

1.7.3 *Intégration RBC-La recherche d'Information (IR)*

Dans l'approche avec recherche de l'information (Information Retrieval : IR), on s'efforce d'obtenir une mémorisation parfaite, ou la récupération de toutes les informations pertinentes avec une parfaite précision. Beaucoup de procédures IR ont été appliquées dans différents systèmes.

Begum et al., ont suggéré une solution en utilisant un algorithme de recherche de correspondance flou, incorporé dans un système RBC pour sélectionner et récupérer un cas en plus de la vérification des contraintes avec notation. Cette méthode donne le cas le plus simple à adapter. Cette approche a été testée pour proposer un menu pour le traitement du stress [Begum et al., 09]. CAREPARTNER est un système qui intègre le RBC-RBR et l'IR pour aider les cliniciens au suivi à long terme des patients atteints de cancer qui ont subi des transplantations de moelle osseuse. Dans ce système, les cas contiennent des problèmes et des solutions spécifiques aux patients, les règles codent des directives pratiques standards et l'IR fournit aux cliniciens des documents pertinents la littérature médicale pour aider la pratique clinique [Marling et al., 05].

1.7.4 *Intégration RBC-La satisfaction de contraintes (CSP)*

La satisfaction de contraintes (Constraint Satisfaction Problem : CSP) est un paradigme de l'IA largement utilisé en aide à la décision. Dans cette approche, le problème est défini en tant que modèle CSP, puis le RBC est utilisé pour compenser les éléments manquants dans ce modèle. Les deux principales approches pour intégrer CSP et RBC ont été :

- l'utilisation du RBC pour initialiser le système CSP ;
- l'utilisation du CSP dans l'étape d'adaptation du RBC.

Dans la première approche, un cas similaire est récupéré par RBC sera utilisé pour positionner le processus CSP. Dans la deuxième approche, le CSP fournit à RBC une méthode spécifique pour accomplir l'adaptation [Marling et al., 02].

Sqalli et al., ont utilisé le CSP pour modéliser le problème et le RBC prend en charge le processus de mémorisation en fournissant la base de cas avec de nouveaux cas, ces cas ne sont pas médicaux seulement ils montrent quand même l'utilisation de cette approche d'intégration. Le RBC est également utilisé pour mettre à jour le modèle CSP et le rendre plus robuste pour résoudre plus de problèmes. Un premier système RBC-CSP était CADSYN, qui a utilisé des contraintes de conception pour l'adaptation de cas et la génération des dessins de structure de bâtiments [Sqalli et al., 99 ; Sqalli et Freuder, 98]. Un autre système, CHARADE a également utilisé cette combinaison pour gérer les urgences dans le combat des feux de forêts. Dans ce système, RBC est utilisé pour évaluer rapidement les situations d'urgence et CSP est utilisé pour déterminer comment exploiter au mieux les ressources disponibles pour gérer l'urgence [Marling et al., 05].

1.7.5 *Intégrations RBC-L'analyse multicritères (AMC)*

L'aspect multicritère a été un axe de développement dans l'aide à la décision médicale, en particulier par les travaux de Belacel. De nombreux travaux ont été menés dans cette direction notamment en médecine, parmi lesquels Belacel proposa une méthodologie pour l'utilisation

de l'aspect multicritère dans l'aide au diagnostic médical [Belacel, 03]. L'analyse multicritères a contribué donc à résoudre certaines limites du RBC en palliant certaines faiblesses auxquelles les méthodes classiques du RBC ne peuvent répondre.

Armaghan et Renaud ont utilisé l'intégration RBC-AMC pour étudier le diabète. Cette étude traite de l'opération "recherche" en utilisant le concept de décision multicritères dans la description du problème pour rechercher la solution dans un scénario basé sur des cas. Ils proposent d'utiliser l'acquisition de connaissances comme base pour rechercher des solutions d'aides à la décision multicritères non compensatoires [Armaghan et Renaud, 12]. Malekpoor et al., ont proposé une TOPSIS-RBC approche (Technique pour la Préférence d'Ordre par Similitude à la solution idéale). Au départ, RBC est utilisé pour extraire de base de données. Par la suite, les cas inférés sont évalués à l'aide de TOPSIS, une approche décisionnelle multicritères pour prescrire un plan de dose optimal. Cette méthode aidera les oncologues pour prescrire un plan de dose optimal pour le cancer de la prostate et éviter les effets secondaires du traitement [Malekpoor et al., 16]. Araujo de Castro et al., ont utilisé un modèle hybride basée sur l'analyse multicritères et le RBC pour le diagnostic de la maladie d'Alzheimer [Araujo de Castro et al., 09]. Li et Sun ont combiné l'analyse multicritères avec le RBC pour améliorer un processus de fouille de données pour la détection des maladies [Li et Sun, 09]. Erjaee et al., ont proposé une méthode spécifique basée sur le multicritère pour proposer une aide à la décision pour un traitement efficace de l'Helicobacter pylori Infection, chez les enfants [Erjaee et al., 12].

1.7.6 Intégration RBC-La fouille de données (FDD)

Les techniques de fouille de données ont également été utilisées de différentes façons pour faciliter le raisonnement à base de cas [Guo et al., 11 ; Bichindaritz, 15]. Dans la littérature, plusieurs contributions visant la combinaison des techniques de la fouille de données avec le processus du RBC ont été proposées.

Balakrishnan et al., ont proposé un système de prédiction de la rétinopathie à base de règles d'association déduite à travers l'algorithme Apriori et le raisonnement basé sur des cas. Les règles d'association sont utilisées pour analyser les profils dans l'ensemble de données et pour calculer la probabilité de rétinopathie tandis que le raisonnement basé sur des cas est utilisé pour la phase de recherche des cas similaires. Cette technique aborde le problème de la maintenance de cas-base en développant une nouvelle technique dite « association-based case reduction technique (ACRT) », pour réduire la taille de la base de cas afin d'améliorer l'efficacité tout en maintenant ou même en améliorant la précision du RBC [Balakrishnan et al., 12]. Sung et Seong, ont fait récemment une étude en se basant sur une méthode hybride, combinant des méthodes de fouille de données (règles d'association, arbres de classification) pour aider les médecins à faire une classification plus rapide et plus précise des maladies de douleurs thoraciques [Sung et Seong, 10].

Araujo de Castro et al., ont combiné le raisonnement à base de règles (RBR) et le RBC pour recommander des médicaments neuroleptiques pour les patients atteints d'Alzheimer [Araujo de Castro et al., 09].

CARE-PARTNER intègre le RBC-RBR et l'IR (Information Retrieval) pour aider les cliniciens à suivre à long terme les patients atteints de cancer qui ont subi des transplantations de moelle osseuse. Ici, les cas contiennent des problèmes et des solutions spécifiques aux

patients, les règles dictent les lignes directrices de pratique standard et l'IR fournit aux cliniciens les documents pertinents de la littérature médicale pour soutenir la pratique clinique sus nommée « evidence-based clinical practice » [Marling et al., 05].

Table 1.3. Systèmes développés avec le RBC et d'autres techniques [Begum et al., 11].

No	Author/System	Other Techniques Used In Conjunction With Cbr	Matching Techniques
01	McSherry/CaseBook	HDR (hypothetico-deductive reasoning)	Author's Defined Similarity Algorithm
02	De Paz/ExpressionCBR	NN & statistics	Nearest-neighbor and minkowski distance
03	Perner/Fungi-PAD	Image processing	Author's defined similarity measurement function
04	Cordier/FrakaS	None	Using adaptation knowledge
05	Corchado/GerAmi	Variational calculus	Hierarchical, multivariate conglomerates analysis and mahalanobis distance
06	Glez-Peña/geneCBR	RBR & Fuzzy logic	Author's defined fuzzy similarity metric
07	Perner/HEp2-PAD	Image processing & data mining	Euclidian distance, Nearest-neighbor
08	Schmidt/ISOR	Statistics	Keyword-based similarity
09	Begum/IPOS	Fuzzy logic	Fuzzy similarity, similarity matrix, euclidian distance, cosine similarity
10	D'Aquin/KASIMIR	Semantic Web, belief revision theory, Fuzzy logic & ergonomy	Matching of source (general) cases using adaptation knowledge
11	Bichindaritz/Mémoire	RBR, Data mining & statistics	Ontology assisted case matching including semantic information
12	Montani/RHENE	Temporal abstractions	Euclidian distance, nearest-neighbor
13	Kwiatkowska/Somnus	Fuzzy logic	Fuzzy logic, semiotic approach
14	Lorenzi/SISAIH	None	Nearest-neighbor
15	Ochoa/SIDSTOU	Data mining	Author's defined method
16	Ahmed/Biofeedback	Fuzzy logic	Fuzzy similarity matching, similarity matrix
17	Brien/ADHD	None	Modified nearest-neighbor matching
18	Doyle/Bronchiolitis	RBR	Nearest-neighbor
19	O'sullivan/Dermatology	KM & image processing	IR Metrics
20	Marling/Type-1 diabetes	RBR	Nearest-neighbor and similarity metric
21	Song/radiotherapy planning	Fuzzy logic, Dempster-Shafer Theory & simulated annealing	Fuzzy sets, distance function and author's defined similarity function
22	Wu/Dietary counseling	Data mining, rule based & ontology	Nearest-neighbor
23	Zhuang/Pathology	Data mining and clustering	Kohonen's self-organizing maps
24	Ahn/Breast Cancer	Genetic algorithms	Genetic algorithms, nearest-neighbor
25	Huang/Chronic Diseases	Data mining	Knowledge-Guide Method, weight Ratio Functionality
26	Chang/children Developmental	None	Nearest-neighbor
27	Houeland/Palliative care	Rule-based & probabilistic model-based method	Semantic matching
28	Nicolas/Melanoma	RBR	Normalized euclidian distance
29	Topel/Metabolic disease	None	Similarity tables, difference-based similarity functions
30	Arshadi/MOE4CBR	Spectral clustering & logistic regression	Modified nearest-neighbor
31	Kurbalija/Multiple Sclerosis disease	None	Case retrieval net
32	Obot/Hepatitis	Rule base & neural networks	Binary search algorithm
33	CBSMS/Stress management	RBR, textual information retrieval & fuzzy logic	Fuzzy similarity matching, modified distance function, similarity matrix
34	Yuan/HDCU	Support vector machine	Self-organizing map

1.8 Conclusion

Dans ce chapitre, nous avons passé en revue l'utilisation de RBC dans le domaine médical, à travers ses différentes utilisations tant en diagnostic qu'en thérapie.

Cette méthodologie s'est vu étendre notamment avec l'intégration d'autres méthodologies de raisonnement à son processus, particulièrement les méthodes de fouille de données et son application à travers différents domaines de la médecine. Nous constatons que les concepts de fouille de données recouvrent plusieurs dimensions et peuvent convenir à la mutualisation des différentes méthodes de traitement des données et des connaissances en vue de palier les limites d'autres procédures de calcul à différents problèmes à des échelles différentes. Cela peut aller de l'extraction de connaissances à la modélisation ou à la recherche de motifs intéressants. Dans notre étude, nous avons ciblé l'aide à la décision dans le domaine médical. Cependant, ce champ est assez général et vaste, il comporte plusieurs facettes, plusieurs problématiques, etc. De ce fait nous sommes orientés vers les systèmes d'aide à la décision médicale guidée par un processus de fouille de données. Ultérieurement, une autre application avec l'analyse multicritères sera étudiée afin de tester une autre approche d'intégration qui nous permettra d'approcher ce thème de la décision médicale intégrée selon deux approches d'intégration. Ceci afin d'étendre notre champ d'application et expérimenter notre approche d'intégration avec deux méthodologies de raisonnement.

Chapitre 2

Intégration RBC-Fouille de données pour l'aide à la décision médicale

Chapitre 2.

Intégration RBC-Fouille de données pour l'aide à la décision médicale

2.1.	Introduction	45
2.2.	La fouille de données	45
	2.2.1 <i>Les méthodes de la fouille de données</i>	45
2.3.	Extraction de Connaissances à partir de Données (ECD)	48
	2.3.1 <i>Les tâches de l'ECD</i>	50
	2.3.2 <i>Le processus ECD</i>	53
	2.3.3 <i>Les étapes du processus ECD</i>	54
	2.3.4 <i>La sélection des données</i>	54
	2.3.5 <i>Le prétraitement des données</i>	54
	2.3.6 <i>La transformation des données</i>	55
	2.3.7 <i>La fouille de données</i>	55
	2.3.8 <i>L'évaluation et la présentation</i>	55
2.4.	Apport de la fouille de données en aide à la décision médicale	56
2.5.	Synthèse de l'utilisation des méthodes de FDD en aide à la décision médicale	59
2.6.	Les limites de l'intégration	60
2.7.	Conclusion	61

2.1. Introduction

La médecine a eu recours à la méthodologie du RBC parce que ce mode de raisonnement est très proche du raisonnement humain. De ce fait, la médecine s'est approprié ce mode de traitement dans la recherche de solutions en aide à la décision médicale. Cependant, les limites montrées par ce mode de raisonnement se sont répercutées aussi sur les systèmes utilisant cette méthodologie en médecine. Les chercheurs se sont vu obligés de les résoudre en adoptant de nouvelles procédures et techniques palliatives notamment en intégrant les méthodes de fouille de données qui ont montré quelques avantages. Aussi, parce que la médecine utilise une immense quantité de données accumulées au cours des années dans de grandes bases de données qu'il fallait exploiter en aide à la décision.

2.2. La fouille de données

Historiquement, la fouille de données a pris naissance dans les secteurs qui manipulaient de grands volumes de données clients : banques, grande distribution, etc. Grâce à des méthodes d'analyse de données et de statistiques, on a utilisé la fouille de données depuis plus de 40 ans dans de nombreux secteurs d'activités. Mais si le concept est aujourd'hui au cœur du processus qui manipule de grands volumes de données brutes pour trouver des connaissances profitables, puisqu'il s'agit de l'extraction de connaissances pertinentes, qui mènent à la prise de décision.

Les méthodes statistiques analysaient surtout des données élémentaires "historisées" provenant de la gestion courante. La fouille de données se focalise sur les données qui circulent dans les systèmes d'informations des entités administratives, entreprises ou organisations et par la même le milieu hospitalier ou médical.

Définition 1.

« La fouille de données est l'ensemble des méthodes et techniques destinées à l'exploration et l'analyse de bases de données informatiques (souvent grandes), de façon automatique ou semi-automatique, en vue de détecter dans ces données des règles, des associations, des tendances inconnues ou cachées, des structures particulières restituant l'essentiel de l'information utile tout en réduisant la quantité de données » [Kantardzic, 11].

Définition 2.

Une autre définition communément admise est celle de Fayyad : *« La fouille de données est un processus non trivial qui consiste à identifier dans des données, des motifs nouveaux, valides, potentiellement utiles, et surtout compréhensibles et utilisables » [Fayyad et al., 96].*

2.2.1 Les méthodes de la fouille de données

Pour réaliser une tâche de fouille de données, on utilise souvent une ou plusieurs méthodes que l'on choisira en fonction de plusieurs considérations, telles que :

- la tâche à résoudre ;

- la nature et la disponibilité des données ;
- la finalité du modèle construit ; etc.

Ceci dit, celles qui sont présentées ci-dessous sont tout de même les plus importantes. Il faut noter qu'il n'y a pas de méthode meilleure, chacune présente des avantages et des inconvénients. On trouve alors [Dunham, 06 ; Han et al., 00] :

[a] *Les méthodes de visualisation*

Elles permettent l'analyse exploratoire avec comme objectif le dégagement de motifs, de structures, de synthèses, etc. Elles sont basées sur des graphiques qui facilitent l'interprétation des résultats. Les méthodes les plus utilisées sont : les graphiques de statistiques élémentaires (moyenne, écart type, variance), les histogrammes, les nuages de points, et les courbes.

[b] *Les arbres de décision*

Ce sont des structures qui représentent des ensembles de décisions. Ces décisions génèrent des règles pour la classification d'un ensemble de données, c'est donc une représentation graphique d'une procédure de classification, c'est-à-dire la prédiction de variables discrètes. Les nœuds internes de l'arbre sont des tests sur les attributs, et les feuilles sont les classes. Un arbre de décision peut donc être perçu comme étant un ensemble de règles qui mènent à une classe. Une correspondance est établie entre un objet décrit par un ensemble de caractéristiques (attributs) et un ensemble de classes disjointes. Chaque feuille de l'arbre dénote une classe et chaque nœud intérieur un test portant sur un ou plusieurs attributs, produisant un sous-arbre de décision pour chaque résultat possible du test. La construction de l'arbre se fait par un algorithme approprié. On trouve alors : ID3, CHAID, CART, QUEST, et C5.

[c] *Les réseaux de neurones*

Ce sont des modèles prédictifs qui utilisent des données existantes avec un résultat connu pour former un modèle pouvant être utilisé en prédiction avec des résultats inconnus. Un réseau neuronal est composé de groupes de nœuds (neurones) où chaque groupe de nœuds correspond à une couche. Il est formé par au moins trois couches : entrée, intermédiaire et sortie. Dans la couche entrée, chaque nœud correspond à une variable prédictive. Les valeurs internes des autres nœuds (des couches intermédiaires et de la couche sortie) sont calculées à travers une fonction de sommation. La couche sortie contient un ou plusieurs nœuds et les variables à prédire.

Le réseau peut avoir plusieurs couches intermédiaires (mais une seule entrée et une seule sortie), appelées aussi couches cachées. Chaque nœud de la couche j est relié à tous les nœuds de la couche $j+1$. A chaque arc est associé un poids (une valeur) W_{ij} , c'est le poids de l'arc entre le nœud i et le nœud j .

Les réseaux de neurones sont des outils très utilisés pour la classification, l'estimation, la prédiction et le groupement. Ils permettent de construire un modèle qui prédit la valeur d'une variable à partir d'autres variables connues appelées variables prédictives. Si la variable à prédire est discrète (qualitative) alors il s'agit d'une classification, si elle est continue (quantitative) il s'agit alors de régression. Les méthodes les plus utilisés sont : le Perceptron multicouches et les réseaux de Kohonen.

[d] *Les réseaux bayésiens*

Ce sont des modèles probabilistes graphiques. Il s'agit d'un graphe acyclique dirigé, où chaque nœud représente une variable continue ou discrète et les arcs représentent une dépendance probabiliste entre un nœud et ses parents. Si un arc relie un nœud Y à un nœud Z, alors Y est le parent de Z et Z est le descendant de Y. Chaque variable est indépendante des variables auxquelles elle n'est pas reliée. Les variables peuvent être continues ou discrètes. Chaque lien entre deux variables est pondéré par la valeur de la dépendance en probabilité. Ainsi, la valeur que porte l'arc reliant Y à Z est en fait $P(Z/Y)$. Les réseaux bayésiens nécessitent un nombre élevé de paramètres à estimer alors que, dans la plupart des cas, le nombre de données disponibles est faible.

[e] *Les machines à vecteur de support (SVM)*

Ils appartiennent à une classe d'algorithmes d'apprentissage initialement définis pour la discrimination, c'est-à-dire la prévision d'une variable qualitative initialement binaire. Elles ont été ensuite généralisées à la prévision d'une variable quantitative. Les SVM peuvent être utilisés pour résoudre des problèmes de discrimination (c'est-à-dire décider à quelle classe appartient un échantillon) ou de régression (c'est-à-dire prédire la valeur numérique d'une variable).

[f] *La méthode des k-plus proches voisins (k-ppv)*

C'est est une méthode dédiée à la classification qui peut être étendue à des tâches d'estimations. La méthode k-ppv est une méthode de raisonnement à partir de cas. Elle part de l'idée de prendre des décisions en recherchant un ou des cas similaires déjà résolus en mémoire. Elle décide de la classe à laquelle appartient un nouveau cas, en examinant les k cas qui lui sont similaires ou proches. Il n'y a pas d'étape d'apprentissage consistant en la construction d'un modèle à partir d'un échantillon d'apprentissage. C'est l'échantillon d'apprentissage qui conduit au modèle. On lui associe une fonction de distance et une fonction de choix de la classe en fonction des classes des voisins les plus proches.

[g] *La méthode k-moyenne (k-means)*

La méthode consiste à diviser les données en k groupes, k étant donné par l'utilisateur. Cette méthode commence par un groupement aléatoire des données (en k groupes), ensuite chaque objet est affecté au groupe le plus proche. Après l'exécution de la première itération, les moyennes des groupes sont calculées et le processus est répété jusqu'à stabilisation des groupes.

[h] *L'induction de règles*

C'est une technique qui permet d'identifier des profils, associations ou structures entre les items ou objets qui sont fréquents dans les bases de données. Autrement dit, il s'agit d'identifier les items qui apparaissent souvent ensemble lors d'un événement. Cette règle d'association est une règle de la forme : « Si X et Y Alors Z », règle dont la sémantique peut être énoncée : « Si X et Y apparaissent simultanément Alors Z apparaît ». Pour considérer et exprimer cette association sous forme d'une règle, il faut définir des quantités numériques qui vont servir à valider son intérêt, d'où : le support et la confiance. Le support est la fréquence d'apparition simultanée des éléments qui apparaissent dans la prémisse et la conclusion, soit :

support=fréquence (prémisse et conclusion), et la *confiance=fréquence (prémisse et conclusion) / fréquence (prémisse)*. Ainsi, les règles dont le support et la confiance sont assez élevés sont alors privilégiées. Les algorithmes les plus utilisés sont : Apriori, FP-Growth.

[i] *Les modèles de Markov cachés*

Les modèles de Markov cachés d'ordre 1 ou 2 (HMM1 et HMM2 pour Hidden Markov Models) sont utilisés pour la classification des différentes données temporelles ou spatiales. Contrairement aux algorithmes classiques qui fournissent une réponse exacte, les HMMs permettent un apprentissage automatique, ils interviennent par exemple dans de nombreux algorithmes d'analyse de séquences biologiques que ce soit pour, la détection de gènes et la détection de motifs exceptionnels.

[j] *La régression linéaire (méthode statistique)*

C'est une technique qui vise la prédiction de la valeur d'une variable continue. Son objectif est de définir le meilleur modèle qui associe une variable quantitative « Sortie » à plusieurs variables prédictives « Entrée ». Cette opération s'appelle ajustement du modèle aux données. Les modèles linéaires sont les plus fréquemment utilisés. C'est ce qu'on appelle la régression linéaire. La relation qui relie une variable à prédire Y à p autres variables prédictives (X_i), est une équation de régression souvent sous cette forme :

$$Y = a_0 + a_1 X_1 + a_2 X_2 + \dots + a_p X_p$$

Les méthodes les plus utilisées sont : la régression simple et la régression multiple.

[k] *La régression logistique*

C'est une technique statistique qui permet de calculer la probabilité de survenue d'un événement à partir d'un ensemble de variables prédictives. Mais la régression logistique ne converge pas toujours vers une solution optimale.

[l] *Les algorithmes génétiques.*

Ce sont des techniques d'optimisation de type méta-heuristique fondées sur les principes de l'évolution biologique. Ils ne constituent pas une méthode de fouille de données à part entière et ne ciblent directement aucune tâche. Ils viennent aider le processus de fouille de données. Ce sont des heuristiques qui guident la recherche de bons modèles dans un espace de recherche très vaste. Les algorithmes génétiques se basent sur les principes de sélection, enjambement, et mutation qui sont des notions issues de la génétique. L'inconvénient majeur des algorithmes génétiques est le temps d'exécution qui est assez long. Il n'y a pas de garantie quant à l'obtention de la solution optimale au problème posé en un temps fini, et sa complexité augmente en fonction du nombre de règles utilisées.

2.3. Extraction de Connaissances à partir de Données (ECD)

Avec la numérisation des données des différents services hospitaliers, les praticiens ressentent le besoin de croiser les données issues de ces différents services afin de déduire de nouvelles informations qui leur permettent de diagnostiquer des cas qui présentent certaines complexités [Kargupta et al., 97], ce croisement de données sera appelé : Extraction de Connaissances à

partir de données (ECD), ou Knowledge Discovery from Data (KDD). Ce concept d'ECD remonte à 1989, mais les premières conférences sur le sujet datent de 1995. Ce concept est apparu avec l'explosion des quantités d'informations stockées suite au progrès important des méthodes de traitement et des supports de stockage.

Donc, l'ECD vise à découvrir, dans les grandes quantités de données, des connaissances précieuses qui peuvent aider à comprendre les données ou à prédire un comportement futur de ces données. La Fouille de données qui est l'opération clé du processus ECD utilise depuis son apparition plusieurs outils de statistiques et d'intelligence artificielle pour atteindre ses objectifs. Elle se situe à l'intersection de nombreuses disciplines (figure 2.1), comme l'apprentissage automatique, les technologies de bases de données, les statistiques, la représentation des connaissances, l'intelligence artificielle, les systèmes experts, etc., [Kodratoff, 96].

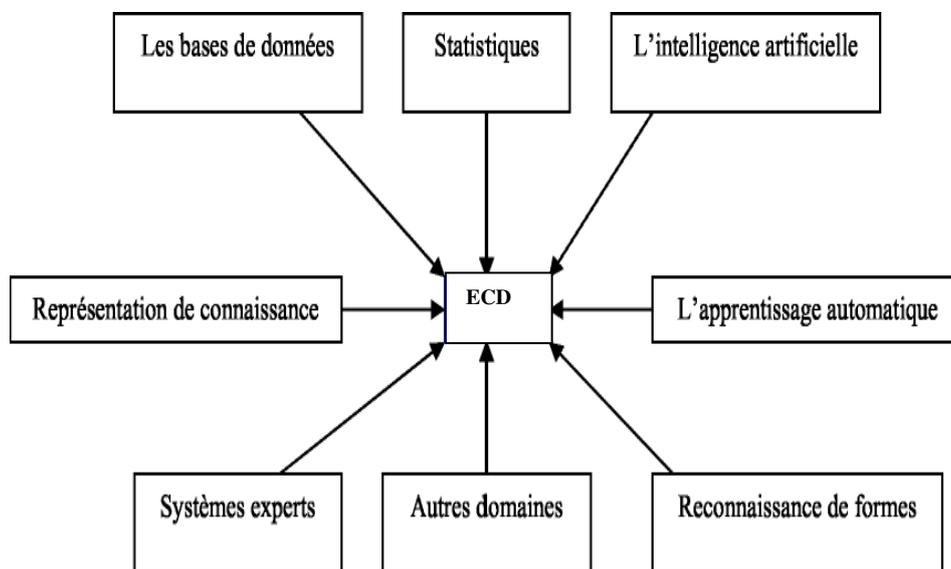


Figure 2.1. Disciplines co-fondatrices de la FD [Kodratoff, 98].

Dans son prolongement vers d'autres disciplines, nous trouvons l'ECD médicales, qui est basée sur la notion de croisement de données médicales afin d'aider le praticien de la santé à comprendre et à prendre en charge un patient, comme par exemple, trouver et expliquer les causes à effet d'un médicament afin de promouvoir une thérapie bien précise ou de valider l'efficacité d'un médicament.

Définition 1.

Fayyad et al., définissent ce concept comme « un processus non trivial qui construit un modèle valide, nouveau, potentiellement utile et au final compréhensible, à partir de données. » [Fayyad et al, 96].

Définition 2.

Zighed et al., présentent l'ECD comme : « un processus itératif et interactif d'analyse d'un grand ensemble de données brutes afin d'extraire des connaissances exploitables par un utilisateur analyste qui y joue un rôle central » [Zighed et al., 01].

Dans un processus ECD, il existe différents intervenants dont l'utilisateur, l'expert en fouille de données, l'analyste de données et l'analyste du domaine d'application [Buchner et al., 97 ; Simoudis, 96] :

- L'utilisateur est la personne à qui est destiné le système ECD. Le plus souvent l'utilisateur apparaît comme étant un expert du domaine et qui possède certaines notions en fouille de données lui permettant d'interagir dans les différentes étapes du processus.
- L'expert en fouille de données est la personne qui met en place le processus ECD. Souvent, il est assisté par l'utilisateur et l'analyste de données.
- L'analyste de données s'occupe de sélectionner et de transformer les données pour les préparer au processus.
- L'analyste du domaine est un expert qui peut analyser les résultats afin de les valider.

Dans certains processus d'ECD, il n'est fait référence qu'à l'utilisateur ce qui sous-entend que l'utilisateur remplit plusieurs rôles [Fayyad et al., 96].

2.3.1 Les tâches de l'ECD

La tâche représente le but, ou l'objectif, d'un processus d'ECD. Fayyad et al., distinguent dans la pratique deux grandes familles de tâches réalisées en ECD : la description et la prédiction [Fayyad et al., 96.b].

- la description se concentre sur la recherche de caractéristiques générales relatives aux données fouillées, ce sont des motifs (modèles, schémas ou règles) décrivant ces données, ils doivent être compréhensibles et interprétables par l'utilisateur ;
- la prédiction consiste à utiliser des attributs dans la base de données pour prédire des valeurs futures ou inconnues d'autres variables considérées.

Ces tâches de description et de prédiction sont réalisées à l'aide d'une variété de méthodes de fouille de données. Nous présentons ici un bref aperçu sur quelques unes de ces méthodes.

Formellement, la fouille de données est considérée comme l'étape centrale du processus d'ECD. Pour cela, on lui associe de nombreuses méthodes vu la diversité des objectifs qui lui sont assignés. Parmi celles-ci nous citons [Dunham, 06 ; Han et al., 00] :

[a] La description

Parfois, les analystes essaient simplement de trouver une manière de décrire des tendances cachées dans les données. Les descriptions de ces tendances ou modèles servent à expliquer ou vérifier un fait. Cela permet souvent une exploitation supplémentaire en vue de fournir des explications. Par exemple : "ceux qui ont le plus de diplômes sont les plus susceptibles d'avoir un poste de responsabilité". La technique la plus appropriée à cette tâche est : les règles d'associations

[b] *La classification*

La classification consiste à examiner des caractéristiques d'un élément nouvellement présenté afin de l'affecter à une classe prédéfinie. Elle est utilisée pour prédire les valeurs discrètes ou nominales (homme / femme, rouge / vert / bleu, ...etc. La classe est un attribut particulier à valeurs discrètes. Un exemple de tâche de classification est par exemple l'attribution d'un type de diabète I ou II à un patient. Une des techniques les plus appropriées à la classification est : les arbres de décision.

[c] *Le groupement*

Le groupement par similitude consiste à réunir les objets qui vont naturellement ensemble. Un groupe maximise la similarité de ces objets et minimise la similarité des objets ne lui appartenant pas. Pour ce faire, le processus de fouille de données utilise des fonctions de distance. Ces fonctions évaluent les distances existantes entre les entités à grouper. De nombreuses fonctions de distance sont disponibles, mais les plus fréquemment utilisées sont celles qui calculent les distances euclidiennes. Cependant, il faut noter qu'il est difficile de calculer ces distances pour des valeurs symboliques telles que des chaînes de caractères : un codage ou une transformation est alors nécessaire. En effet, il n'y a pas de variable cible pour le groupement. La tâche de groupement ne cherche pas à classer, estimer, ou prédire la valeur d'une variable cible, mais plutôt à segmenter l'ensemble des objets en sous-groupes relativement homogènes à l'aide de mesures de distance. Le groupement est une tâche d'apprentissage "non supervisée" car on ne dispose d'aucune autre information préalable que la description des objets à grouper. Lorsque les groupes ont été construits, d'autres techniques ou une expertise doivent dégager leur signification et leur éventuel intérêt. Les méthodes de groupement se divisent en deux types : le groupement basé sur les partitions et le groupement hiérarchique. La technique la plus connue du premier type est la méthode des k-moyenne (k-means).

[d] *L'estimation*

Elle consiste à estimer la valeur d'un champ à partir des caractéristiques d'un objet. Contrairement à la classification, le résultat d'une estimation permet d'obtenir une variable continue. L'estimation est similaire à la classification, sauf que la variable cible est numérique plutôt que catégorique. Les modèles sont construits en utilisant des données, qui fournissent la valeur de la variable cible, ainsi que les "prédicteurs". Par exemple : "l'estimation de la pression artérielle d'un patient, basée sur son âge, son indice de masse corporelle". L'estimation peut être aussi utilisée dans un but de classification. Par exemple, on peut estimer le revenu d'une personne selon divers critères (type de véhicule, profession, type d'habitation, etc.), ensuite définir des tranches de revenus pour classer les individus. La technique la plus appropriée à l'estimation est : les réseaux de neurones.

[e] *La prédiction*

Cela consiste à estimer une valeur future en fonction de valeurs anciennes sauvegardées (historique). La prédiction est semblable à la classification et l'estimation, sauf que pour la prévision, les résultats se situent dans l'avenir. La seule méthode pour mesurer la qualité de la prédiction est d'attendre les résultats. Un exemple de cette tâche appliquée au marketing, est la prédiction du prix d'un article après deux mois. Les méthodes de classification et

d'estimation peuvent être utilisées en prédiction. Les techniques les plus appropriées à la prédiction sont : les arbres de décision et les réseaux de neurones.

[f] *La recherche d'associations*

C'est la tâche la plus intéressante en FDD. Elle est plus connue sous le nom de "l'analyse du panier de la ménagère". Elle consiste à déterminer les attributs qui sont liés. L'exemple type est la détermination des articles (le pain, le lait, les biscuits, etc.) qui se retrouvent ensemble sur un même ticket de supermarché. Cette tâche peut être effectuée pour identifier des opportunités de vente et concevoir des groupements attractifs de produit.

[g] *L'analyse d'exception et de déviation*

Dans cette tâche, on tente de dégager et d'étudier des exceptions ou des surprises contenues dans les données, comme par exemple les objets ne pouvant être classés par une classification. Ces objets peuvent révéler des explications utiles dans certaines situations.

[h] *La visualisation*

Elle aide l'utilisateur à acquérir et accroître ses connaissances et à guider son raisonnement grâce à ses capacités d'analyse. Habituellement, c'est par visualisation que se fait le post-traitement des modèles de connaissance. Dans le premier type de visualisation, l'utilisateur ne connaît pas forcément ce qu'il cherche dans les données, il essaye de chercher des modèles, des motifs, ou plus généralement des hypothèses qu'il veut démontrer. Au deuxième type de visualisation, l'utilisateur a une hypothèse qu'il veut tester et confirmer. Ce type de visualisation est dérivé directement des statistiques, et ne convient pas vraiment au principe même de la fouille de données (bien qu'il en fasse partie). En fouille de données, il existe trois types de visualisation :

- *Sélection de sous-espaces.* Il s'agit d'une sélection de règles ou de groupes qui contiennent un item ou un ensemble d'items choisi par l'utilisateur. Dans le cas des règles d'association, l'utilisateur peut sélectionner les règles qui contiennent un ou plusieurs items dans la partie gauche ou dans la partie droite ou bien dans les deux parties à la fois. Ce type de visualisation est convenable en cas de grande quantité d'informations traitées.
- *Ordonnement.* C'est un tri selon un ou plusieurs critères particuliers, tels que la taille des données, leur importance, etc. Pour trier les groupes par exemple, on peut traiter le critère de l'importance des données de la manière suivante. Plus le générateur est de petite taille, plus il est plus important. Puis, plus le nombre de lignes du jeu de données qui comportent les similarités est grand, plus il est important.
- *Affichage par les outils graphiques.* C'est une visualisation des résultats sous formes de graphiques, de matrices, d'histogrammes, etc. Ceci permettra de mieux interpréter les résultats.

Tous ces types de visualisation peuvent être combinés ensemble pour construire des outils de visualisation des différents motifs établis par un processus d'ECD.

Table 2.1. Les tâches de l'ECD.

Tâches de l'ECD	Caractéristiques / Objectifs	Méthodes de fouille de données utilisées
Description	<ul style="list-style-type: none"> • Il s'agit de décrire les données pour essayer de découvrir et de comprendre le processus qui est à leur origine. 	<ul style="list-style-type: none"> • Stat. élémentaire ; • Histogramme ; • moyenne, écart-type ; • ACP....
Estimation	<ul style="list-style-type: none"> • Consiste à estimer la valeur d'une variable à valeurs continues à partir des valeurs d'autres attributs. 	<ul style="list-style-type: none"> • Régression ; • Réseaux de neurones ; • k-ppv.
Prédiction	<ul style="list-style-type: none"> • Consiste à prédire la valeur future d'un attribut en fonction d'autres attributs ; • Se base sur le présent pour trouver des résultats dans le futur ; • Assimilable à l'estimation mais les objets sont classés en fonction d'un comportement futur prédit. 	<ul style="list-style-type: none"> • Arbre de décision ; • Réseaux de neurones ; • Réseaux bayesiens.
Classification	<ul style="list-style-type: none"> • Consiste à examiner les caractéristiques d'un objet et lui attribuer une classe ; • Les classes sont connues à l'avance avec des profils particuliers. 	<ul style="list-style-type: none"> • k-ppv ; • Arbre de décision ; • Réseaux de neurones ; • Algo. Génétique ; • HMM.
Groupement	<ul style="list-style-type: none"> • Il s'agit de grouper des objets en se basant sur leurs similarités ; • Les objets sont les plus similaires dans un groupe et moins similaires entre deux groupes ; • La similarité peut être calculée pour différents types de données. Elle dépend des données utilisées et du type de similarité recherchée. 	<ul style="list-style-type: none"> • k moyennes ; • Réseaux de neurones.
Recherche d'associations	<ul style="list-style-type: none"> • Déterminer les attributs qui sont corrélés, i.e. découvrir des relations plus fines entre les données. 	<ul style="list-style-type: none"> • A-priori ; • AIS ; • FP-Growth.

2.3.2 Le processus ECD

Le nombre d'étapes défini dans un processus d'ECD est variant selon les auteurs, néanmoins nous présentons un processus plus ou moins standard et le plus utilisé avec ces différentes étapes [Han et al., 00].

. Le modèle du processus de fouille de données que nous présentons se décompose en plusieurs étapes : la sélection des données, le prétraitement, la transformation, la fouille de données et enfin l'évaluation et la présentation des résultats.

Le modèle du processus d'ECD que nous avons retenu se décompose en plusieurs phases :

1. La sélection ou la création d'un ensemble de données à étudier ;
2. Le prétraitement qui permet d'éliminer le bruit et traiter les données manquantes ;
3. La transformation ou la définition des structures optimales de représentation des données ;
4. La fouille de données à l'aide de paramètres appropriés ;
5. l'interprétation et l'évaluation durant laquelle les éléments extraits sont analysés pour aboutir à des connaissances stockées dans une base de connaissances.

Il faut noter qu'il y a la possibilité d'un retour à une étape antérieure afin de réviser le processus.

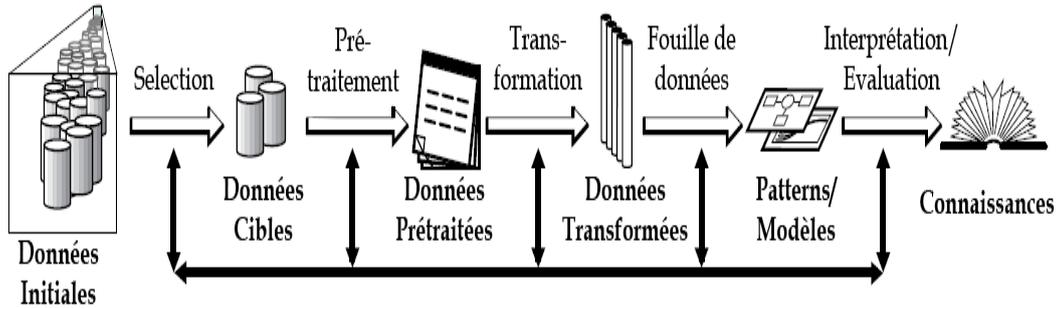


Figure 2.2. Schéma global de l'ECD d'après Fayyad et al., [Fayyad et al., 96].

A partir de ce schéma, nous pouvons dire que la fouille de données n'est qu'une étape de traitement dans le processus ECD. Elle est le "pivot" du processus. En fait, c'est une étape de traitement des données qui va donner les motifs ou modèles qui seront exploités ultérieurement par l'utilisateur. Elle consiste à appliquer des algorithmes d'apprentissage sur les données afin d'extraire des connaissances valides.

2.3.3 Les étapes du processus ECD

2.3.3.1. La sélection des données

Cette étape ne se limite pas à la seule sélection des données qui vont être exploitées. Elle comprend également l'analyse du problème à résoudre [Dunham, 06 ; Han et al., 00 ; Fayyad et al., 96], ce qui permet d'en déduire le ou les types de données qui sont exploités, ainsi que les méthodes qui pourraient être utilisées pour accomplir cette tâche.

2.3.3.2. Le prétraitement des données

Cette seconde étape permet d'affiner les données. Si la base de données est bien construite, le prétraitement de données peut permettre d'améliorer les résultats lors de la fouille de données. Les données à analyser par les méthodes de fouille de données sont parfois incomplètes, inconsistantes, erronées, incompatibles entre elles ou inadaptées [Han et al., 00]. Ces données sont courantes et se retrouvent régulièrement dans les bases de données. Donc il faut les préparer du point de vue forme, type et contenu. Plusieurs procédures sont alors nécessaires. On trouve alors les procédures de nettoyage, de transformation et les procédures de réduction.

- La procédure de nettoyage.

Le nettoyage des données consiste à "retravailler" ces données bruitées, en supprimant certaines, et en modifiant d'autres de manière à tirer le meilleur profit. Le but de ces deux opérations est de générer de "nouvelles données retravaillées" pour faciliter leur exploitation future. Pour le traitement des données manquantes, plusieurs méthodes permettent d'accomplir cette opération, et le choix entre elles dépend des données et de l'objectif de l'étude.

- *La réduction des données*

Cette procédure permet une réduction de la représentation des données tout en sauvegardant leur intégrité. Les méthodes de réduction les plus connues sont :

- *agrégation des données cibles* : agrège les données pour construire un cube de données, ceci permet de visualiser les données de façon multidimensionnelle ;
- *réduction dimensionnelle* : détecte les attributs qui ne sont pas intéressants, faiblement intéressants ou récurrents afin de les supprimer ;
- *compression des données* : codage qui permet la réduction de la taille des données ;
- *discrétisation et génération de concept hiérarchique* : remplace les valeurs bruitées des attributs par des niveaux conceptuels plus élevés.

2.3.3.3. *La transformation des données*

Cette procédure transforme les données sous une forme appropriée aux méthodes de fouille de données. Les méthodes que nous retrouvons dans cette procédure sont les suivantes :

- *l'agrégation* : permet de regrouper des données saisies à des périodes différentes. Ceci permet d'avoir une vue d'ensemble sur toute la durée de l'acquisition.
- *la généralisation* : remplace les données primaires par des concepts supérieurs en utilisant des hiérarchies de concepts. Par exemple : l'attribut type "nom_rue" peut être généralisé à un concept supérieur tel que "adresse".
- *la normalisation* : permet de regrouper les valeurs d'attributs dans un intervalle bien délimité afin de faciliter l'interprétation des données.
- *l'ajout d'attributs* : consiste à ajouter des attributs afin de les rendre les données compatibles avec certaines méthodes de fouille de données.

2.3.3.4. *La fouille de données*

C'est l'application d'une méthode adéquate sur les données prêtes à l'emploi. Du fait d'une grande diversité des données exploitées, il en résulte un nombre important de méthodes de fouille de données. Ces dernières sont issues de divers domaines comme par exemple : les statistiques, l'analyse de données, l'apprentissage automatique, etc. En outre, certaines de ces méthodes peuvent être combinées afin de réduire les inconvénients de l'une ou l'autre. Le choix des méthodes de fouille de données est fonction d'une part, des besoins exprimés par l'utilisateur et d'autre part, des données exploitées.

2.3.3.5. *L'évaluation et la présentation*

C'est l'évaluation des motifs (modèles) extraits, qui est faite à ce stade. Celle-ci permet de mesurer l'intérêt de ces motifs. Ensuite, une présentation des résultats à l'utilisateur grâce à différentes techniques de visualisation. Ce n'est qu'à partir de la présentation que l'on peut employer le terme de connaissance à condition que ces motifs soient validés par l'expert du domaine. On distingue alors deux modes de validation : statistique et par expertise. Pour certains problèmes, on peut associer les deux modes de validation [Young, 94] :

– *la validation statistique*

Cela consiste à utiliser des méthodes de base de statistique descriptive. L'objectif est d'obtenir des informations qui permettront de juger le résultat obtenu, ou d'estimer la qualité par le biais des données d'apprentissage. Cette validation peut être obtenue par :

- le calcul des moyennes et variances des attributs ;
- le calcul de la corrélation entre certains champs ;
- ou la détermination de la classe majoritaire dans le cas de la classification.

– *la validation par expertise*

Elle fait appel à un expert dans le domaine duquel proviennent les données. Celui-ci jugera la pertinence des résultats produits. Dans le domaine médical, par exemple, le motif extrait doit être facile à comprendre, pour cela une première validation doit être effectuée par un expert médical qui jugera de la compréhensibilité du motif avant de le présenter au médecin pour exploitation.

La validation complète d'un système à base d'ECD consiste alors à s'assurer de l'adéquation entre la connaissance modélisée dans la base de connaissances et la connaissance de l'expert. Deux techniques de validation peuvent être envisagées :

– *La validation par l'examen des résultats obtenus à partir de la base de connaissances.*

Elle permet de mesurer l'adéquation entre la connaissance réelle et sa modélisation en confrontant les résultats du système et ceux de l'expert sur un ensemble de problèmes constituant un jeu de tests. On suppose dans ce type de validation que l'expert possède une description du monde réel, sous forme d'un ensemble de problèmes caractéristiques accompagnés de leurs solutions.

Pour ce type de validation, de nombreuses méthodes sont utilisées comme le hold-out, sous-échantillonnage aléatoire (random sub-sampling), validation croisée (cross-validation) et bootstrap [Beleites et al., 16]. Cependant, des mesures de performance peuvent être utilisées pour analyser les modèles prédictifs. Ils sont basés sur quatre valeurs de la matrice de confusion telles que résumées dans la figure 2.3 : vrai positif (TP), faux positif (FP), vrai négatif (TN) et faux négatif (FN). De plus, à un autre niveau de vérification, l'évaluation de la performance de l'ensemble du test est estimée par le taux d'erreur (error rate), et il sera estimé à l'aide de mesures d'erreurs communes. Ainsi, il est utilisé la sensibilité (sensitivity), la spécificité (specificity), la précision (accuracy) telles qu'elles sont définies ci-dessous:

Table 2.2. Les mesures pour l'évaluation de modèles [Beleites et al., 16].

Mesure		Formule de calcul
Sensibilité (<i>Sensitivity</i>)	pourcentage de cas positif qui sont correctement identifiés comme positifs.	$TP/(TP + FN)$
Spécificité (<i>Specificity</i>)	pourcentage de cas correctement identifiés comme négatifs.	$TN/(TN + FP)$
Précision (<i>Accuracy</i>)	pourcentage de cas correctement testés (Réellement Positifs ou réellement Négatifs).	$(TP + TN) / (TP + FP + TN + FN)$
Valeur prédictive positive (<i>Positive Predictive Value</i> : PPV)	pourcentage de cas testés positifs et réellement positifs.	$\frac{TP}{TP + FP}$
Valeur prédictive négative (<i>Negative Predictive Value</i> : NPV)	pourcentage de cas testés négatifs et réellement négatifs.	$\frac{TN}{TN + FN}$

		Actual		
		Positive	Negative	
Test	Positive	True Positive (TP)	False Positive (FP)	PPV
	Negative	False Negative (FN)	True Negative (TN)	NPV
		Sensitivity	Specificity	

Figure 2.3. Matrice de confusion 2 x 2 dimensions.

- La validation par l'étude de la cohérence de la base de connaissances.

La détection d'incohérence d'une base de connaissances suppose que soit donné un modèle conceptuel du monde réel pour l'étude de la cohérence. Ce modèle peut être considéré plus explicite que la connaissance issue de la base de connaissances. C'est à l'aide de ce modèle que nous pouvons juger de la cohérence de la base de connaissances sans faire appel à un expert.

Table 2.3. Les méthodes de fouille de données.

Tâches de l'ECD	Types				Méthodes de FDD									
	Descriptive	Prédictive	Supervisé	Non Supervisé	H M M	Régression linéaire	S V M	k-moyenne	k-ppv	Réseau de neurones	Arbre de décision	Réseau bayésien	Algo. génétiques	Règles d'association
Description	•		•		•	•								
Estimation		•	•			•				•				
Prédiction		•	•		•	•				•				
Classification		•	•		•	•	•		•	•	•	•	•	•
Groupement	•			•				•	•	•				
Recherche Associations	•	•		•								•	•	•

2.4. Apport de la fouille de données en aide à la décision médicale

La médecine a eu recours à la fouille de données en raison de l'immense quantité de données accumulées au cours des années dans de grandes bases de données. Ainsi, l'apport de la fouille de données en médecine a été très grand, notamment par la mise à disposition d'outils spécifiques d'analyse de données médicales pour l'aide à la décision [Barigou et al., 12].

Aussi, il est très intéressant de pouvoir trouver des relations intéressantes entre les entités, comme par exemple le lien qui peut exister entre la prise d'un médicament et un effet secondaire, afin de déduire des effets indésirables d'une médication.

La fouille de données a donc été très utile dans de nombreux travaux et systèmes d'aide à la décision médicale. Nous en citerons quelques exemples.

[a] En diagnostic

La recherche de patients pouvant être soumis à des schémas thérapeutiques bien déterminés est parmi les premières applications réalisées. Une catégorisation des patients est faite pour cibler les patients par groupes. La recherche des facteurs de risque pour certaines pathologies comme le diabète, est effectuée à partir de données des études épidémiologiques, avec choix d'un traitement approprié et individuel [Zorman et al., 02], Mokeddem et al., ont présenté une nouvelle approche pour le diagnostic des maladies coronariennes, cette nouvelle méthode est fondée sur les algorithmes génétiques et la classification naïve bayésienne [Mokeddem et al., 2014].

[b] *En thérapeutique*

L'association de médicaments avec leurs effets secondaires sur une catégorie de patients est une tâche qui a été abordée très tôt en fouille de données médicales [Prather, et al., 97]. La recherche d'association entre thérapeutique et une symptomatologie donnée a été aussi abordée, suivi d'une étude sur les effets indésirables [Prather, et al., 97]. La prédiction des maladies cardio-vasculaires chez un patient donné [Podgorelec, 05]. Certains praticiens ont besoin de faire un suivi d'une catégorie de patients afin de promouvoir les traitements les plus adaptés à ces derniers [Ramirez et al., 00, Richards et al., 01].

[c] *En pronostic*

On recherche à prédire le temps de rétablissement après une opération, en fonction des données du patient (âge, poids, etc.) [Schmidt et Gierl, 02]. Huang et al., utilisent un raisonnement à base de cas couplé aux arbres de décision et une recherche d'association à base de cas pour le pronostic des maladies chroniques [Huang et al., 07].

2.5. Synthèse de l'utilisation des méthodes de FDD en aide à la décision médicale

Les méthodes de fouille de données citées précédemment ont été largement utilisées dans de nombreuses études en médecine, en particulier, les arbres de décision et les réseaux de neurones.

Les arbres de décision

Allayous et al., ont utilisé les arbres de décision pour déterminer les variables impliquées dans la sévérité et la récurrence de la crise de séquestration splénique aiguë (augmentation brutale de la taille de la rate, et chute du taux de l'hémoglobine). Cette méthode fournit un outil de diagnostic qui améliore le traitement médical et la qualité des soins pour les patients atteints de cette pathologie [Allayous et al., 08]. Park et al., ont utilisé l'algorithme C4.5 pour construire un arbre de décision pour découvrir les principales causes du diabète de type II. Cet algorithme génère un ensemble de règles pour le diagnostic et la prédiction du diabète [Park et al., 06].

Les règles d'association

Balakrishnan et al., proposent un système de prédiction de la rétinopathie basé sur les règles d'association en utilisant l'algorithme Apriori et le raisonnement à base de cas. Les règles d'association sont utilisées pour analyser les motifs dans les données et pour calculer la probabilité de rétinopathie alors que le raisonnement à base de cas est utilisé pour récupérer des cas similaires. Cette technique aborde le problème de la maintenance de cas-base en développant une nouvelle technique, ACRT (Association-based Case Reduction Technique), pour réduire la taille de la base de cas afin d'améliorer l'efficacité tout en maintenant ou même en améliorant la précision du RBC [Balakrishnan et al., 12].

Sung et Seong ont récemment réalisé une étude fondée sur la construction d'une méthode hybride, combinant des méthodes de fouille de données (règles d'association, arbres de

décision) pour aider les cliniciens à classifier des maladies de la douleur thoracique de manière précises [Sung et Seong, 10].

Les réseaux de neurones

Malyshevskaya a étudié le problème du diagnostic du cancer à l'aide des réseaux de neurones. L'objectif de cette étude était de classifier les différents types de cellules utilisés pour déterminer le risque de cancer [Malyshevskaya, 09]. Sivakumar a présenté une méthode basée sur les réseaux neuronaux pour classifier les patients atteints de rétinopathie diabétique (complications courantes du diabète). Cet algorithme génère un ensemble de règles pour le diagnostic et la prédiction du diabète [Sivakumar, 07].

La régression logistique

Kiezun et al., ont utilisé la méthode de régression logistique pour aider les cliniciens dans le diagnostic de l'infection du myocarde (une infection peut provoquer une inflammation du myocarde qui est le muscle cardiaque) chez des patients présentant une douleur thoracique [Kiezun et al., 09].

L'implication des méthodes de fouilles de données a été très remarquée de par les nombreuses études et réalisations dans le domaine médical. Seulement, les données médicales sont assez variées et complexes et souvent elles sont représentées dans des structures plates dans les bases de données habituelles sans les liens nécessaires qui peuvent donner un peu plus de sens dans leur existence dans un endroit commun, en l'occurrence la base de données.

De plus, les situations médicales (voir introduction générale. Sect. 3) auxquelles nous faisons référence dans notre travail sont assez variées, ce qui conditionne l'application de telle ou telle méthode qui souvent mets l'utilisateur devant un autre problème celui de la bonne préparation des données qui à lui seul est un autre travail ardu. De ce fait, nous le voyons à travers les différents travaux qu'il y a eu des tentatives assez variées d'utilisation de méthodes, ce qui montre que rien n'est définitivement établi par rapport à l'intégration des méthodes de fouille de données au RBC et que celles-ci montrent toujours des limites. Nous passerons quelques-unes en revue dans la section suivante afin d'avoir une idée sur les difficultés d'intégration.

2.6. Les limites de l'intégration

L'utilisation combinée de RBC et des méthodes de fouille de données a augmenté d'une manière considérable dans le domaine médical. Cependant cette intégration a montré des limites qui poussent les chercheurs du domaine à trouver des solutions. Parmi ces limites, nous notons [Pandey et Mishra, 10 ; Pan et al., 07 ; Lamiche, 13] :

- Le premier problème concerne la spécificité des données médicales : il s'agit de données particulières, complexes, hétérogènes, hiérarchiques souvent imprécises, subjectives, entachées de valeurs manquantes, ou incomplètes.

- Toutes ces lacunes au niveau des données font que le processus de fouille de données devient assez laborieux en traitement surtout en phase de sélection et prétraitement des données de base.
- L'accumulation de très grande quantité de données (étude épidémiologiques et pharmaceutiques, génomique, imagerie, etc.). Cet accroissement des connaissances médicales a entraîné une augmentation du nombre de paramètres nécessaires à la description d'une situation décisionnelle prise en charge par le médecin.
- Il devient de plus en plus difficile, pour un médecin, de maîtriser la définition du cas, voir simplement d'y accéder en temps utile par un simple processus de fouille de données.
- D'une façon générale, la gestion d'objets complexes est devenue nécessaire pour sa prise en charge par un processus RBC avant de passer le relais vers un processus de fouille de données.
- Comment opérer une fouille d'objets complexes (Cas), si ceux-ci présentent des images, des textes, etc., mais aussi des concepts reliés entre eux.
- La considération des points de vue des acteurs de la décision est un aspect qui n'est pas pris en charge par la fouille de données et qui aussi important que les données fouillées elles même.

2.7. Conclusion

L'intégration des méthodes de fouille de données à la méthodologie RBC, appliquée au domaine médical a été une expérience assez enrichissante, mais aussi avec des difficultés inhérentes au domaine d'application mais aussi aux limites de certaines méthodes cités ci-dessus.

- Au niveau de l'intégration, celle est conditionnées surtout par la qualité des données, qui elle même conditionne la qualité des motifs extrait qui vont et viennent entre le RBC et l'ECD.
- Au niveau de la fouille des données proprement dite, la qualité des données et la méthode de prise en compte des données manquantes ont une influence sur les résultats obtenues par ces techniques.

Le second problème concerne la validation du résultat de tout le processus d'intégration par les experts. Chose qui n'est pas facilement aisée du moment que nous avons deux validations simultanées :

- en premier, la validation du résultat de la fouille de données ;
- en second, la validation de l'aide à la décision proposée par le processus en entier.

Les deux validations nécessitent un niveau d'expertise assez élevé de la part du médecin. Chose que l'on ne trouvera pas facilement chez le corps médical.

De ce fait, une certaine prudence doit être de mise surtout lorsqu'il s'agit de systèmes qui peuvent être utilisés dans la pratique courante des médecins et qui ne sont pas forcément familiarisé avec les procédures de fouille de données, qui nécessitent souvent des

connaissances assez particulières et non générales sur les types de données et leurs manipulations.

Sur un autre volet, il y a un aspect très important en aide à la décision qui n'est pas pris en charge par la fouille de données. C'est l'aspect choix du décideur.

En effet, le décideur a des préférences, des évaluations et des choix. Ces aspects ne sont pas pris en compte par les méthodes de fouille de données, ces dernières s'appliquent "froidement" sur des stocks de données. De ce fait, quelque soit la méthode, appliquée en support au RBC, elle ne pourra donner que des résultats plus ou moins "distancés" par rapport à la réalité ou aux attentes du décideur.

Au final, on se retrouve avec des combinaisons de méthodes mais avec une possibilité d'écart par rapport à la réalité parce que les choix du décideur ne sont pas suffisamment pris en considération par le processus d'aide à la décision.

A cet effet, une autre approche plus pragmatique et proche de la réalité du décideur semble être plus envisageable, c'est le raisonnement à base de cas avec une collaboration de l'AMC.

Cette approche a été abordée, par intégration au RBC, du fait que celle-ci est formalisée à travers des méthodes bien posées théoriquement et ne présente pas d'énormes difficultés quant à son utilisation technique. Nous l'aborderons dans le prochain chapitre.

Chapitre 3

Intégration RBC-AMC
pour l'aide à la décision
médicale

Chapitre 3.

Intégration RBC-AMC pour l'aide à la décision médicale

3.1	Introduction	64
3.2	L'analyse multicritères (AMC)	64
3.2.1	<i>Le paradigme multicritères</i>	65
3.2.2	<i>Les problématiques multicritères</i>	69
3.2.3	<i>L'analyse multicritères d'un problème décisionnel</i>	71
3.2.4	<i>La démarche multicritères d'un problème décisionnel</i>	71
3.2.5	<i>Les principales méthodes multicritères</i>	74
3.2.6	<i>Le choix de méthodes multicritères</i>	77
3.3	Synthèse de l'utilisation de l'AMC pour l'aide à la décision médicale	78
3.3.1	<i>Utilisation des méthodes du critère unique de synthèse</i>	78
3.3.2	<i>Utilisation des méthodes de surclassement</i>	78
3.4	Les limites de l'intégration RBC-AMC	79
3.5	Conclusion	79

3.1 Introduction

Une très grande partie des problèmes de décision se caractérise par la diversité des points de vue qui sont souvent contradictoires et qui mesurent des choses de nature différente. Aussi, sont nombreuses les situations concrètes où les conséquences sont suffisamment complexes pour qu'une seule fonction objectif (un seul critère) ne puisse appréhender toute l'information nécessaire à la comparaison globale des actions (solutions). La résolution de ce type de problèmes doit tenir compte simultanément de tous les points de vue jugés pertinents par le décideur et de plusieurs critères : c'est l'analyse multicritères (AMC).

C'est vers la fin des années soixante qu'il y a eu la naissance de ce paradigme qui regroupe une grande partie des notions de l'optimisation monocritère. Les chercheurs de ce nouveau mode de raisonnement développent des moyens de travail nouveaux pour résoudre des problématiques décisionnelles d'un nouveau genre différentes du paradigme monocritère. Toutefois, de nouvelles questions sont posées, par exemple, comment pondérer les critères et comment les agréger. En effet, le paradigme multicritères se caractérise par un schéma de pensée qui tient compte de plusieurs critères qui sont à considérer dans le processus décisionnel. Ce paradigme trouve sa justification dans le constat qu'il est très difficile, sinon impossible, d'optimiser tous les points de vue devant lesquels le décideur prend sa décision.

De là, on a essayé à développer des modèles plus ou moins formalisés dans la perspective d'améliorer, de faciliter et d'accompagner le décideur dans le déroulement du processus décisionnel.

L'ensemble des méthodes et modèles développés en analyse multicritères ont un but commun qui vise à aider le décideur à prendre une décision qui le satisfait et en l'intégrant dans une démarche décisionnelle en lui offrant la possibilité de progresser vers une solution.

3.2 L'analyse multicritères (AMC)

Un problème de décision monocritère est un problème du type :

optimiser { $g(x) : x \in A$ } où A est l'ensemble des actions admissibles et g est la fonction critère à optimiser.

Ce modèle traduit généralement un problème bien structuré et bien défini mathématiquement, qui s'impose à la fois au décideur et à l'homme d'étude, exemple :

optimiser { $Prix(x) : x \in A$ }

Lorsque les actions potentielles d'un problème de décision ne sont pas évaluées par un critère unique, mais par un ensemble de critères qu'on désigne par $g_1, g_2 \dots, g_m$, et que le décideur souhaite optimiser simultanément, le problème posé sera alors de la forme :

optimiser { $g_1(x), g_2(x), \dots, g_m(x) : x \in A$ }

La principale difficulté d'un problème multicritères est qu'il s'agit d'un problème sans solution objective. Il n'existe pas, en général, une action meilleure que toutes les autres, et simultanément sur tous les critères. Le concept de solution optimale, un postulat de base de

l'approche monocritère n'a donc pas de sens dans un contexte multicritères. Il s'agit donc d'aider le décideur à progresser vers une action de compromis [Roy, 91].

Objectif de l'analyse multicritères.

Face à la complexité de diverses situations où l'être humain est appelé à décider, il fait intervenir un ou plusieurs critères dans un environnement parfois aléatoire. Le décideur ne peut évaluer correctement un problème pour le résoudre d'autant plus si plusieurs paramètres entrent en jeu pour évaluer la situation problématique et décider d'une action (solution). Ceci fait que le décideur choisit l'analyse multicritères parce que le décideur n'a pas la faculté de faire une évaluation en prenant en compte plusieurs paramètres parfois conflictuels. Ceci, « dépasse » un peu de ces capacités mentales. Cependant, il s'oriente vers l'AMC afin de l'aider à faire cette « combinaison » d'option par des méthodes prouvées et pouvant assurer le calcul combinatoire ou parfois récursif afin de lui faire sortir une option la meilleure. Donc, l'objectif est celui de trouver la solution optimale si l'on peut dire par la prise en compte de paramètres suivants : le problème, les solutions envisagées et les critères d'évaluation sur lesquels il se base pour chercher une solution

3.2.1 Le paradigme multicritères

De nombreuses approches traitent des problèmes de décision. Les approches classiques (utilité, théorie des jeux, analyse coût-bénéfices, programmation mathématique, etc.) ont aidé à traiter des problèmes de décision en optimisant un objectif (coût, distance, etc.). L'aide multicritère à la décision est apparue pour traiter plusieurs classes de problèmes de décision (choix, tri, rangement...) s'exprimant à l'aide de critères, et des préférences du décideur. Cette approche procède à la fois à l'aide d'une démarche couplée à des méthodes pratiques pour solutionner des problèmes de décision.

3.2.1.1 Action

C'est une représentation de l'élément de solution qui contribue à la décision, par exemple, l'achat d'un produit, le choix d'un site pour implanter une usine, etc., constituent des actions. Quand les actions sont exclusives, on utilise le terme "variante", "alternative en anglais". Les actions potentielles ont pour objet de délimiter le champ des solutions possibles. On peut faire apparaître deux types d'actions potentielles : les actions *réelles* et les actions *fictives*. Les premières correspondent à une réalité susceptible d'être appréhendée par le décideur, par exemple, une Peugeot 206 est une action réelle. On peut néanmoins vouloir considérer des actions qui ne correspondent à aucune réalité existante mais qui permettent quand même d'éclairer les décisions, ce sont les actions fictives. Elles vont servir de base pour effectuer des comparaisons. Si on reste dans le domaine automobile, on peut dire qu'une "voiture familiale" ou une "voiture sport" sont des actions fictives.

Les notions de Préférence, Indifférence, Incomparabilité et Surclassement.

L'activité d'aide à la décision passe par la comparaison des actions entre elles, en utilisant les critères. Vincke ne considère que deux relations : la préférence et l'indifférence, notées respectivement I et P ou (\sim et $>$) [Vincke, 89].

– Préférence

Cette relation permet de traduire une situation dans laquelle il existe des raisons claires et suffisantes pour mettre en évidence une préférence entre deux actions *a* et *b*. On notera :

$a P b$: une situation dans laquelle *a* est préférée à *b*.

Cela signifie que *a* est meilleure que *b*. De part la sémantique associée à cette relation, il est naturel de considérer cette relation comme étant irréflexive et asymétrique.

– Indifférence

Cette relation traduit une situation dans laquelle il n'existe pas de raisons suffisamment fortes pour confirmer une préférence dans un sens ou dans l'autre. On notera :

$a I b$: une situation d'indifférence entre *a* et *b*.

Cela signifie que *a* et *b* sont tellement proches et il est difficile de dire que l'une est meilleure que l'autre. Cette relation est généralement considérée comme étant réflexive et symétrique. Ces deux relations apparaissent ainsi comme complémentaires. Lorsque deux actions *a* et *b* sont indifférentes, il n'est pas possible d'affirmer une préférence dans un sens ou dans l'autre. De même, lorsqu'il existe une préférence entre *a* et *b*, les deux actions ne peuvent être indifférentes.

$a P b$ et $b P a$

– Incomparabilité

Il peut exister des situations où le décideur ne peut s'exprimer en faveur d'une action ou d'une autre sans pour autant être indifférent : on parlera de situations d'incomparabilité.

$a R b$: *a* est incomparable avec *b*

Cela signifie que *a* et *b* sont tellement différentes l'une de l'autre qu'il est difficile de pouvoir les comparer. Cette relation est naturellement considérée comme symétrique et irréflexive. Comme pour les autres relations utilisées en modélisation des préférences, nous considérerons que cette relation n'est pas nécessairement transitive.

– Surclassement

Le concept de surclassement est dû à Roy [Roy, 85]. Une relation de surclassement est une relation *S* définie dans *A* telle que $a S b$ s'il y a suffisamment d'arguments pour admettre que *a* est au moins aussi bonne que *b*, sans qu'il y ait de raison importante de refuser cette affirmation.

$a S b$: "*a* est au moins aussi bonne que *b*" $a S b \Leftrightarrow C(a, b) \text{ et } D(a, b)$

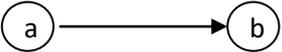
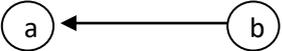
On dit qu'une action *a* surclasse une action *b* si : *a* est au moins aussi bonne que *b* relativement à une majorité de critères (condition de concordance : $C(a, b)$), sans être trop nettement plus mauvaise relativement aux autres critères (condition de non-

discordance : $D(a, b)$, c'est-à-dire, il n'y a pas de critère qui émet son veto pour $a S b$.

Table 3.1. Les situations possibles de comparaison de 2 actions.

Situation	Définition	Relation binaire (propriétés)	Notation
Préférence Stricte	Existence de raisons claires et positives qui justifient une référence significative en faveur de l'une (identifiée) des deux actions.	P : relation asymétrique (irréflexive).	$a P b$
Préférence faible	Existence de raisons claires et positives qui infirment une préférence stricte en faveur de l'une (identifiée) des deux actions mais ces raisons sont insuffisantes pour en déduire soit une préférence stricte en faveur de l'autre soit une indifférence entre ces deux actions (ces raisons ne permettent donc pas d'isoler l'une des deux situations précédentes comme étant la seule appropriée).	Q ("quasi") : relation asymétrique (irréflexive).	$a Q b$
Indifférence	Existence de raisons claires et positives qui justifient une équivalence entre les deux actions.	I : relation symétrique et réflexive.	$a I b$
Incomparabilité	Absence de raisons claires et positives justifiant l'une des trois situations précédentes.	R (refus de se prononcer): relation symétrique irréflexive.	$a R b$
Surclassment	Existence d'arguments pour admettre qu'une action est meilleure que l'autre.		$a S b$

Table 3.2. Les situations possibles de comparaison de 2 actions.

Situations	Relations	Représentations
$a S b$ <i>et</i> non $b S a$	$a P b$	
non $b S a$ <i>et</i> $b S a$	$b P a$	
$a S b$ <i>et</i> $b S a$	$a I b$	
non $a S b$ <i>et</i> non $b S a$	$a R b$	

3.2.1.2 Critère

Schärliig propose la définition la suivante : « un critère est une référence par rapport à laquelle on mesure la conséquence d'une action, en d'autres termes un critère exprime plus ou moins les préférences du décideur relativement à un attribut donné » [Schärliig, 85].

Selon Vincke, « *Un critère est une fonction g , définie sur l'ensemble \mathcal{A} des actions, qui prend ses valeurs dans un ensemble totalement ordonné, et qui représente les préférences du décideur selon un point de vue* » [Vincke, 89]. Lorsque le problème repose sur la considération de plusieurs critères, nous les notons g_1, \dots, g_n . L'évaluation d'une action a suivant le critère j est notée $g_j(a)$.

Un critère peut donc être défini comme le moyen de modéliser un point de vue. Cependant, plusieurs aspects d'une action peuvent concourir à un même point de vue. Par exemple, si l'on s'intéresse au point de vue confort d'une automobile plusieurs aspects doivent être pris en compte comme la suspension, la tenue de route, le niveau sonore, etc.

Un critère est alors une fonction g définie sur \mathcal{A} et prenant ses valeurs dans un ensemble totalement ordonné \mathcal{R} ou $g : \mathcal{A} \rightarrow \mathcal{R}$, et qui représente les préférences du décideur selon un point de vue. On distingue alors plusieurs types de critères : le vrai critère, le pseudo critère, le quasi critère.

Chaque action a de \mathcal{A} sera donc représentée par un vecteur $(g_1(a), \dots, g_n(a))$ que l'on appelle vecteur de performances. L'ensemble \mathcal{A} des actions sera représenté par une matrice appelée matrice de performance. Il existe alors différents types de critères. Le vrai critère, le pseudo critère, le quasi-critère et le pré-critère.

3.2.1.3 Poids

Il est assez courant en analyse multicritères que le décideur pense qu'un critère est plus important qu'un autre pour des raisons diverses parmi lesquelles ses préférences personnelles. Nous appelons *poids* (w), cette mesure de l'importance relative entre les critères telle qu'elle est vue par le décideur. Néanmoins cette mesure n'est pas toujours déterminée facilement par le décideur. Ils doivent être le reflet des points de vue des différents acteurs concernés par la prise de décision. Ces derniers « *s'appuient sur un système de perception, de valeurs et d'opinions différentes pour expliciter leurs préférences. Plusieurs méthodes de pondération de critères existent et elles peuvent beaucoup influencer le résultat final de l'analyse* » [Pomerol et Barba-Romero, 93]. Eastman et al., notent une difficulté couramment rencontrée dans la résolution des problématiques multicritères est l'affectation des poids aux différents critères conduisant à la prise de décision surtout lorsque le nombre de critères est assez élevé [Eastman et al., 93].

3.2.1.4 Matrice de performance

Face à un problème de décision, le décideur est amené à juger et à évaluer les actions potentielles, en utilisant les critères. Un critère est donc un facteur de jugement utilisé pour mesurer et évaluer une action. Il diffère de la notion de variable dans la mesure où un critère est relié aux préférences du décideur alors qu'une variable ne l'est pas nécessairement.

Dans le cas fréquent où l'analyse des conséquences des actions potentielles conduit à construire plusieurs critères, c'est l'analyse multicritères qui permet de donner des réponses au problème posé. Pour chaque action considérée et pour chaque critère, un seuil de préférence (p), d'indifférence (q) et un seuil de veto (v) sont estimés. Chaque critère se voit attribuer un poids (w) traduisant sa contribution dans la décision finale. Le résultat de l'analyse des

conséquences est présenté dans une matrice de performance, également appelée matrice d'évaluation ou matrice de jugements.

Table 3.3. Matrice de performance.

	g_1		g_j
a_1	$g_1(a_1)$					$g_j(a_1)$
...						
a_n	$g_1(a_n)$					$g_j(a_n)$
Poids	w_1					w_j

Table 3.4 Exemple de matrice de performance (pour le choix d'implantation d'usine).

Actions	Eloignement centre ville	Surface	Avis Architecte (note).
Site 1	80	133	6
Site 2	100	210	2
Poids	2	2	1

3.2.2 Les problématiques multicritères

La problématique peut être perçue comme étant une orientation de l'investigation qu'on adopte pour un problème de décision donné. Elle exprime les termes dans lesquels le décideur ou l'homme d'étude pose le problème et traduit le type de la solution qu'il souhaite obtenir. Roy distingue trois problématiques de base dont les caractéristiques sont résumées ci-dessous. Ainsi, tout problème décisionnel multicritères doit se ramener nécessairement à l'une d'entre elles. Nous avons alors : les problématiques de choix, de tri, de rangement et de description, respectivement notées, P_α , P_β , P_γ et P_δ [Roy, 85].

[a] Problématique de choix (P_α)

Elle consiste à sélectionner un sous ensemble aussi restreint que possible de l'ensemble des actions \mathcal{A} , contenant les meilleures actions. L'idéal est d'obtenir une seule et meilleure action. Mais à cause de la nature conflictuelle des critères, il est préférable de fournir au décideur quelques actions qui représentent différentes variantes de la "meilleure action". Formellement, le résultat de ce type de situation décisionnelle, est un sous ensemble $\mathcal{A}' \subset \mathcal{A}$. Exemple : choix du meilleur site pour une installation industrielle.

[b] *Problématique de tri ($P\beta$)*

Elle consiste à affecter chaque action à un ensemble de catégories prédéfinies. Cette formulation est adéquate lorsque le problème de décision consiste à examiner chaque action indépendamment des autres (en tenant compte que des caractéristiques intrinsèques de chaque action) dans le but de proposer une recommandation parmi un ensemble de recommandations spécifiées auparavant. Chaque recommandation peut être associée à une catégorie. Le problème de décision est alors vu comme un tri des actions potentielles en différentes catégories prédéfinies. La procédure de tri doit être définie de telle sorte que chaque action est affectée à une et une seule catégorie. Comme par exemple l'affectation de différents sites (chaque site est une action) à un type d'installation industrielle particulier parmi différentes installations.

[c] *Problématique de rangement ($P\gamma$)*

Elle consiste à ranger les différentes actions en allant de la meilleure action à la moins bonne avec éventuellement des ex aequo. Cette problématique est intéressante lorsque les actions sont à différencier selon leur intérêt relatif. Il est à noter qu'en pratique, le rangement peut être nécessaire seulement pour les actions les plus intéressantes. Exemple, la maintenance de plusieurs sites industriels en commençant par le site le plus urgent.

[d] *Problématique de description ($P\delta$)*

Elle consiste simplement à décrire les actions et leurs conséquences et non pas à les comparer comme c'est le cas avec les trois autres problématiques précédentes. Ici, il n'existe pas une solution. Exemple : identifier et décrire, dans le seul but d'apprendre et se préparer à d'éventuels incidents.

Table 3.5. Les problématiques décisionnelles.

Problématique	Objectif	Prescription	Opération
P_α (Alpha)	Sélectionner un sous ensemble aussi restreint que possible de A contenant les meilleures actions.	Sous ensemble $A' \subset A$	Choix
P_β (Beta)	Affecter chaque action de A à une et une seule des catégories définies selon de normes préétablies.	Partition de A	Tri
P_γ (Gamma)	Ordonner les actions de A de la moins bonne à la mauvaise.	Ordre partiel sur A	Rangement
P_δ (Omega)	Décrire les actions et leurs conséquences	Ordre partiel sur A	Description

Notons que plusieurs auteurs ne retiennent pas la problématique de description, étant donné qu'il n'existe pas de méthodes particulières pour cette problématique [Schärliig, 85].

3.2.3 L'analyse multicritères d'un problème décisionnel

3.2.3.1 La formulation multicritères d'un problème décisionnel de choix

Un problème décisionnel est la définition d'une situation à analyser qui demande une aide à la décision. Cette situation problématique est décrite à l'aide d'attributs qui sont les l'objet de la décision et les différents critères utilisés pour une évaluation de la future aide à la décision (solution). Pour cela, le décideur utilisera : différentes actions (solutions possibles), des critères, et des pondérations pour ces derniers. De ce fait, un problème multicritères sera défini comme suit :

Problème Décisionnel Multicritères (A_p , C_k , W_i)

Où : $A = \{a_1, a_2, \dots, a_n\}$, l'ensemble des actions potentielles (ou alternatives).

$C = \{c_1, c_2, \dots, c_m\}$, les critères d'évaluation ($c_j(a_i)$ =évaluation de a_i selon le critère c_j).

$W = \{w_1, w_2, \dots, w_k\}$, les poids des critères.

Ce problème décisionnel multicritères sera représenté par une matrice de jugement qui permettra de représenter l'évaluation quantitative ou qualitative des différents critères qui entreront dans la recherche de la solution "aide à la décision".

3.2.4 La démarche multicritères d'un problème décisionnel

Différentes démarches existent pour conduire une situation de décision multicritères. Chaque démarche met l'accent sur certains aspects aux dépendants d'autres et, par conséquent, chacune aura ses avantages et ses inconvénients [Laaribi, 00 ; Chakhar, 06]. Deux importantes écoles portent leur regard sur l'aide multicritères à la décision :

- l'école américaine basée sur l'approche *top-down* (du "haut vers le bas") à partir des objectifs qu'il définit [Keeney et Raiffa, 93];
- l'école française basée sur l'approche *"bottom-up"* (du « bas vers le haut ») à partir des conséquences qu'ils préconisent [Roy, 85 ; Roy et Vanderpooten, 96].

La démarche de Keeney et Raiffa consiste à construire une structure hiérarchique ayant à son premier niveau l'objectif global qui est "éclaté" en sous-objectifs qui, à leur tour, sont "éclatés" en sous-sous-objectifs jusqu'à ce qu'il soit atteint un niveau mesurable qualifié d'attributs [Keeney et Raiffa, 93]. Quant à l'approche de Roy "bottom-up", il identifie toutes les conséquences pouvant résulter de la mise en œuvre des actions, que l'on structure en dimensions puis en axes de signification autour desquels sont construits les critères [Roy, 85].

Chakhar précise que : « Les méthodes d'aide à la décision développées selon la première approche sont très différentes de celles développées selon la deuxième approche, mais elles ne s'opposent pas car elles s'appliquent à des problèmes différents. Elles sont donc plutôt

complémentaires » [Chakhar, 06]. D'après Vansnick, « les deux écoles se basent sur le même modèle de décision pour l'application de leurs méthodes » [Vansnick, 90].

Les problèmes de décision multicritères opèrent habituellement en 3 phases [Ham, 08 ; Maystre et al., 94]. Ben Mena note que les deux premières sont communes pour toutes les méthodes multicritères à l'inverse des deux dernières qui dépendent de la méthode choisie. Il propose la démarche suivante [Ben Mena, 00] :

1. Création d'une liste d'actions potentielles

Au cours de cette étape, on établit une liste des actions potentielles qui vont rentrer en concurrence. Cette liste n'est pas exhaustive et définitive. Elle peut évoluer tout au long de l'étude (suppression ou ajout d'actions).

2. Modéliser les préférences du décideur (Création d'une matrice de performance)

– Création d'une liste de critères à prendre en considération

Il s'agit d'élaborer la liste des critères à prendre en considération. Un critère peut être plus important qu'un autre. Cette importance relative est exprimée par un poids.

– Evaluer les actions selon les critères et création de la matrice de performance

N'importe quelle méthode multicritères agit sur la matrice des performances. Il s'agit de juger chaque action par rapport à chacun des critères, c'est la matrice de performances. Il est de dimension $(n \times m)$ qui, pour chaque action A_1 de l'ensemble des actions A , on fait correspondre un vecteur $G = (G_1(A_1), G_2(A_1), \dots, G_n(A_1))$. Ces valeurs numériques sont souvent appelées performances. Deux ensembles sont considérés, le premier A est celui des actions, le second C est celui des critères :

$$A = \{A_1, A_2, A_3, \dots, A_n\}; \quad C = \{C_1, C_2, C_3, \dots, C_m\},$$

avec n le nombre des actions et m le nombre des critères. S_{ij} désigne la performance ou l'évaluation de l'action A_i par rapport au critère C_j .

3. Agrégation des préférences

Pour définir une solution (action) qui fait émerger une préférence commune (qui jouit globalement des meilleures évaluations), les jugements doivent être agrégés, c'est ce qu'on appelle l'agrégation des préférences qui consiste en « *une opération permettant d'obtenir des informations sur la préférence globale entre les actions potentielles, à partir d'informations sur les préférences par critères.* » [Maystre et al., 94].

Les méthodes multicritères diffèrent selon leurs façons de traiter cette dernière étape. Les trois premières étapes sont pratiquement communes à toutes les méthodes. La différence entre elles, se trouve fondamentalement dans la façon de réaliser les deux dernières étapes relatives à l'évaluation et à l'agrégation, c'est-à-dire dans la façon d'évaluer chacune des solutions en fonction des critères retenus.

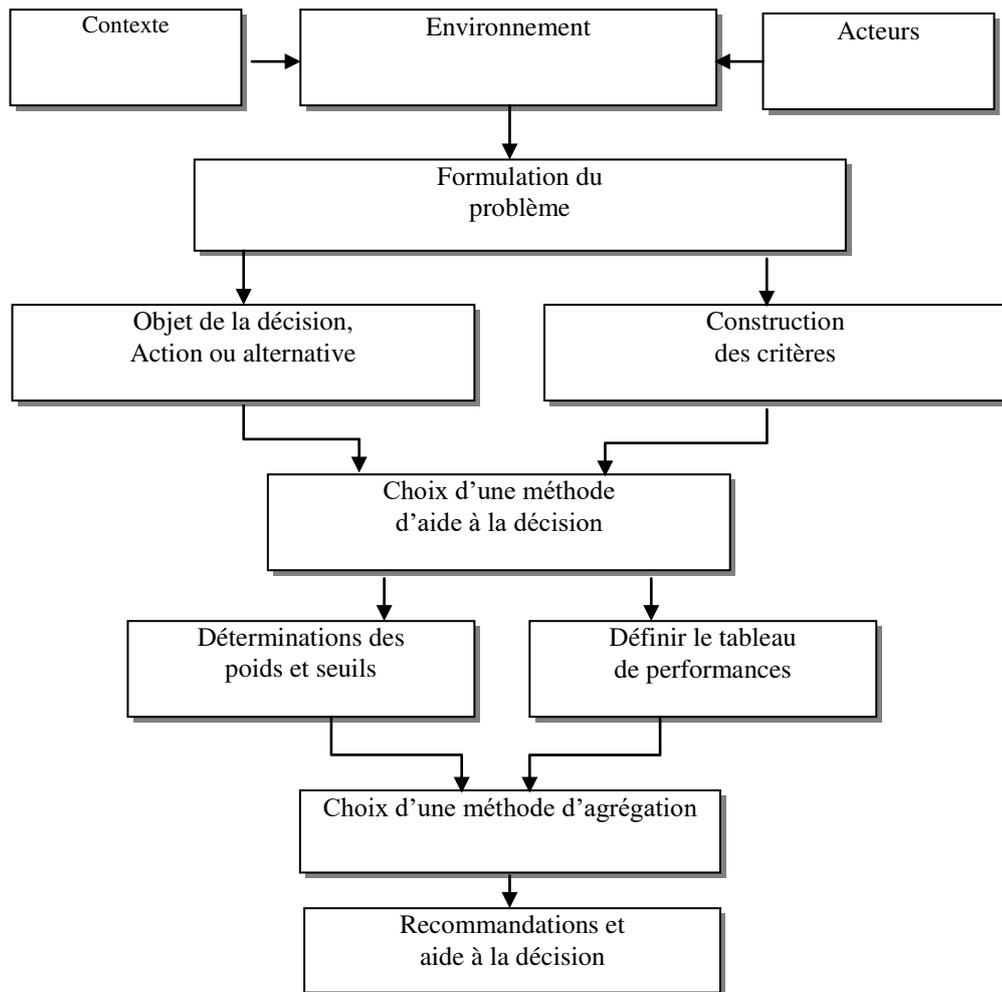


Figure 3.1. Les étapes d'une méthode multicritères [Nafi et Wery, 10].

3.2.4.1 L'agrégation

Il s'agit d'établir un modèle des préférences globales, c'est-à-dire une représentation formalisant de telles préférences relativement à un ensemble \mathcal{A} d'actions potentielles que l'homme d'étude juge approprié au problème d'aide à la décision [Ben Mena, 00].

De ce fait, un problème multicritères va manipuler un ensemble noté $\mathcal{A}\{a_1, a_2, \dots, a_n\}$ où a_i est l'action i . Les actions sont évaluées et comparées sur la base d'un ensemble des critères d'évaluation g_1, g_2, \dots, g_m . L'évaluation d'une action a selon un critère g_j est notée $g_j(a)$.

Ainsi, pour définir une solution (action) qui fait émerger une préférence commune (qui jouit globalement des meilleures évaluations), les jugements ou évaluations partielles (i.e. par rapport à chaque critère) doivent être agrégés en une évaluation globale en utilisant un mécanisme d'agrégation approprié. Cette agrégation consiste à utiliser un mécanisme mathématique d'agrégation pour combiner les différentes évaluations partielles (i.e. par rapport à chaque critère) en une évaluation globale (i.e. qui tient compte de tous les critères). Les procédures d'agrégation sont, en quelque sorte, la caractéristique principale des méthodes multicritères. Les méthodes multicritères diffèrent selon leurs façons de traiter cette opération (agrégation).

3.2.4.2 Les méthodes d'agrégation selon l'approche du critère unique de synthèse

Roy ainsi que Maystre et al., l'appellent "*approche du critère unique de synthèse évacuant toute incomparabilité*" [Roy, 85 ; Maystre et al., 94]. Vincke nomme cette approche "*théorie de l'utilité multicritères*" [Vincke, 89] d'autres les désignent sous l'appellation *méthodes d'agrégation complète*. Cette approche est la plus classique. Scharlig note que ces méthodes peuvent s'avérer les seules utilisables en pratique [Scharlig, 85]. On cherche à agréger les n critères afin de les réduire en un critère unique et obtenir une fonction à critère unique qui synthétise l'ensemble des critères, ceci implique que les critères soient tous mesurables. On suppose aussi que les jugements sont transitifs, exemple : $a S b, b S c$ alors $a S c$.

On trouve alors les principales méthodes : la somme pondérée, la moyenne pondérée, combinaison linéaire pondérée (WLC : Weighted Linear Combination), AHP (Analytical Hierarchy Process) de Saaty, moyenne par pondération contrôlée (OWA : Ordred Weighted Averaging), goal programming, utilités additives (UTA), analytic hierarchy process (AHP), et multi attribute utility theory (MAUT).

3.2.4.3 Les méthodes interactives selon l'approche du jugement local interactif

Elles sont appelées *méthodes d'agrégation locale interactive avec itérations essai-erreur*. Ces méthodes reposent sur on interaction régissant l'enchaînement d'étapes de dialogues et de traitement permettant de cheminer a travers des essais et erreurs et sur la base de jugement local, vers des éléments de solutions. On compare alors les actions deux à deux et on vérifie si, selon certaines conditions préétablies l'une des deux actions surclasse l'autre ou pas, et à partir de toutes ces comparaisons, on tente ensuite de réaliser une synthèse et élaborer une proposition de quelques alternatives, puis reprise de l'analyse en boucle. Parmi ces méthodes on a : la programmation linéaire multiple et UTA interactive.

3.2.4.4 Les méthodes de surclassement selon l'approche du surclassement de synthèse

Les méthodes appartenant à cette approche, d'inspiration française, sont appelées également les *méthodes d'agrégation partielle*. Selon Roy, ces méthodes acceptent l'incomparabilité entre les différentes actions lorsque l'affirmation d'une préférence ou d'une indifférence s'avère être insuffisamment justifiée (notion de préférence). Ce type de méthodes s'applique aux cas où l'ensemble des actions est fini. On cherche à comparer des actions par couple, et à établir des relations de surclassement entre ces éléments sur la base d'un indice de surclassement. Electre, Prométhée, Oreste, Macbeth sont les méthodes les plus connues dans cette catégorie.

3.2.5 Les principales méthodes multicritères

Les méthodes d'analyse multicritère sont souvent classées sur la base de l'ensemble des actions A en deux catégories : méthodes discrètes et méthodes continues. Roy a regroupé ces dernières dans trois catégories principales représentant chacune d'elles une approche différente [Roy, 85]. Ces catégories sont présentées ci-dessous (figure 3.2). De nombreuses méthodes ont vu le jour depuis la typologie présentée par Roy mais celle-ci demeure

pertinente et ces méthodes peuvent être associées aux catégories proposées [Figueira et al., 05 ; Figueira et al., 16].

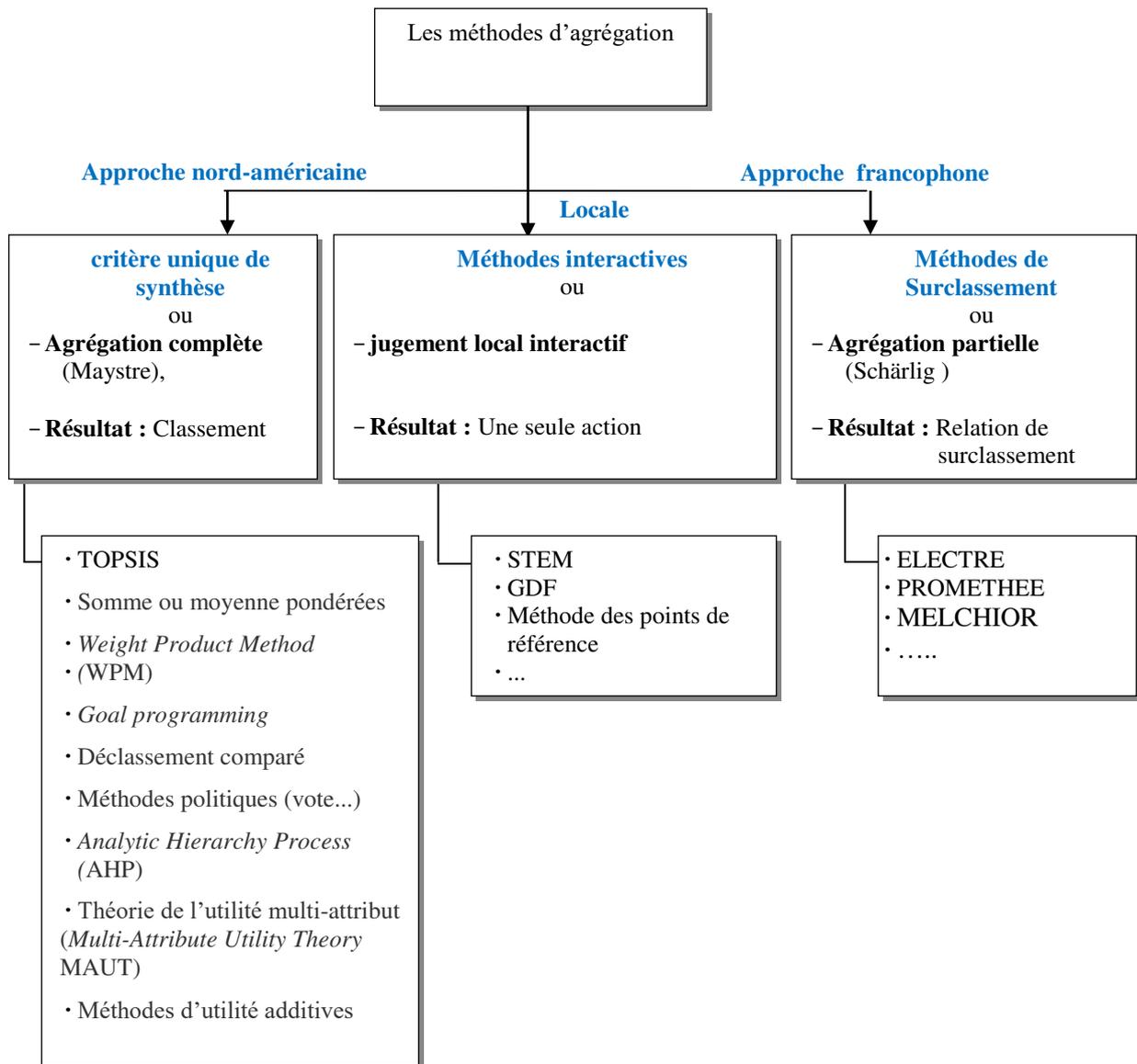


Figure 3.2. Les méthodes d'agrégation.

3.2.5.1 Les méthodes PROMETHEE

Prométhée a été proposée pour la première fois en 1982 par Brans [Brans, 82]. Le principe de la méthode Prométhée consiste à établir un processus de comparaison numérique de chaque action par rapport à toutes les autres actions. Le résultat de cette comparaison permet le classement ordonné des actions [Brans et al., 84]. Dans cette méthode, deux traitements mathématiques particuliers sont proposés : le premier permet de ranger les actions en un pré-ordre partiel et qui mène à l'incomparabilité (méthode Prométhée I), le second permet de ranger les actions potentielles selon un pré-ordre total (Prométhée II).

Table 3.6. Les méthodes PROMETHEE.

Méthode	Principe
PROMETHEE I	<p>La méthode peut être appliquée à n'importe quel ensemble de variantes dénombrable A et commence par calculer les différences entre chaque paire d'alternatives par rapport à chaque critère. Une fonction de préférence est ensuite appliquée à l'ensemble de ces différences afin de les convertir en degrés de préférence qui prennent des valeurs entre 0 et 1.</p> <p>Une valeur de 0 signifie que le décideur ne considère pas la différence entre les deux alternatives pour être significative et une valeur de 1 signifie que la différence est assez forte pour le décideur afin de préférer la meilleure alternative. Enfin la méthode agrège ces degrés de préférence pour chaque solution alternative, afin de définir soit un score par critère (appelé flux net uni critère) soit un score global (appelé le flux net). PROMETHEE I donne un classement partiel des actions.</p>
PROMETHEE II	<p>On utilisera Prométhée II si on souhaite disposer d'un rangement complet de toutes les actions. Ce rangement est obtenu en rangeant les actions dans l'ordre décroissant des Φ (actions). PROMETHEE II : donne un classement total des Actions</p>
PROMETHEE III	<p>PROMETHEE III est une extension de PROMETHEE II dans laquelle la notion d'indifférence est amplifiée. En effet, le préordre complet PROMETHEE II laisse relativement peu de place aux indifférences, étant donné qu'elles résultent d'égalités entre les flux nets des actions. Le plus souvent, PROMETHEE fournit un ordre complet sur l'ensemble des actions, sans aucune indifférence.</p>
PROMETHEE IV	<p>Dans certains cas, le problème posé n'est pas de sélectionner une action particulière ou de ranger l'ensemble des actions de la meilleure à la moins bonne, mais au contraire de sélectionner un sous-ensemble d'actions. La problématique est d'un type plus complexe, noté P_{α, A_n}. Elle consiste à choisir A actions parmi n, le nombre A étant fixé à l'avance, ou à déterminer selon les cas. Dans PROMETHEE VI, il est aussi proposé au décideur de fixer des intervalles dans lesquels les poids peuvent varier.</p>

3.2.5.2 Les méthodes ELECTRE.

Electre est une famille de méthodes conçues par Roy. Elles se basent sur les mêmes concepts fondamentaux de l'analyse multicritères mais diffèrent dans leurs fonctionnements ainsi que dans le type de la problématique traitée. On trouve alors les notions de sur-classement ainsi que les notions de concordance et de discordance. Dans ce qui suit, nous survolons les différentes méthodes de la famille *Electre* à savoir : *Electre I*, *Electre II*, *Electre III*, *Electre IV*, *Electre Is*, *Electre Tri* [Roy et Boyssou, 93].

Table 3.7. Les méthodes ELECTRE.

Méthode	Principe
Electre I	Cette méthode permet de résoudre les problèmes multicritères de choix. Elle permet d'identifier le sous-ensemble d'actions offrant le meilleur compromis possible, on vise également à retenir la meilleure décision ou solution. Souvent utilisée dans le choix de solution ou de décision concurrentes, afin d'identifier le sous ensemble de solution le plus performant sur la base des critères considérés.
Electre II	Cette méthode relève de la problématique décisionnelle γ (procédure de classement). Il s'agit donc d'essayer de classer toutes les actions de la "meilleure" jusqu'à la "moins bonne". L'approche utilisée reste toujours la même, elle est fondée sur la concordance et la discordance.
Electre III	La méthode Electre III relève de la problématique décisionnelle γ (procédure de classement) : son but est de classer les actions potentielles, depuis les "meilleures" jusqu'aux "moins bonnes". Pour se faire, Electre III traite une matrice d'évaluation contenant des actions et des pseudos critères. Les traitements de surclassement munis sur cette matrice permettront d'établir un pré-ordre final partiel.
Electre IV	Cette méthode relève aussi de la problématique décisionnelle γ (procédure de classement) témoigne d'une sophistication de plus en plus poussée. Par rapport à cette méthode, il n'y a plus de poids attribué à chaque critère, et l'abandon de l'hypothèse de sur-classement, qui rend inutiles les notions de concordance et de discordance.
Electre TRI	Cette méthode relève de la problématique β (procédure d'affectation), pose le problème en termes d'attribution de chaque action à une catégorie prédéfinie. Ainsi, des actions de référence sont utilisées pour segmenter l'espace des critères en catégories. Cette une méthode intéressante dans la mesure où elle permet une comparaison différente des actions potentielles, non plus entre elles, mais par rapport à une action de référence.
Electre IS	Elle relève aussi de la problématique décisionnelle α , c'est une adaptation d'Electre I à la logique floue, permettant d'utiliser des pseudo-critères. Pour choisir la "meilleure" action potentielle, une partition des actions potentielles A en deux sous-ensembles doit être réalisée, comme dans Electre I, c'est dans le noyau (sous-ensemble des actions non-surclassés) que se trouve la "meilleure" action. La construction de ces partitions nécessite l'utilisation de la relation de sur-classement

3.2.6 Le choix de méthodes multicritères

Cette étape dépend de la nature du problème posé. Plusieurs méthodes ont été développées, la table 3.8 identifie certaines méthodes en fonction de la problématique étudiée.

Table 3.8. Choix de la méthode multicritères [Nafi et Wery, 10 ; Rogers et al., 00].

Critères	Problématique / Méthode multicritères		
	$P\alpha$	$P\gamma$	$P\gamma$
Vrai critère	Electre I		Electre II
Pseudo critère	Electre IS	Electre Tri	Electre III, IV

3.3 Synthèse de l'utilisation de l'AMC pour l'aide à la décision médicale

L'analyse multicritères a contribué à résoudre certaines limites du RBC. Ainsi, de nombreuses études ont été menées dans ce sens. On a essayé d'améliorer l'approche traditionnelle du RBC pour améliorer les systèmes d'aide à la décision par diverses méthodes. Nous citons quelques travaux.

3.3.1 Utilisation des méthodes du critère unique de synthèse

Malekpoor et al., ont proposé une approche TOPSIS-RBC. Initialement, le RBC est utilisé pour extraire les cas pertinents de la base de données. Par la suite, les cas inférés sont évalués en utilisant TOPSIS (technique de préférence de commande par similarité avec la solution idéale : une technique de prise de décision multicritères) pour prescrire un plan de dose optimal. Cette méthode aidera les oncologues à faire un meilleur compromis entre les mesures de similarité, le taux de réussite et les effets secondaires du traitement [Malekpoor et al., 16]. Bouhana et al., ont utilisé une combinaison du RBC et AHP pour la recherche d'itinéraire [Bouhana et al., 11].

3.3.2 Utilisation des méthodes de surclassement

Armaghan et Renaud ont utilisé l'intégration RBC-AMC pour étudier le diabète. Cette étude traite de l'opération "Remémoration" en utilisant le concept de décisions multicritères dans la description du problème pour rechercher la solution dans un scénario basé sur des cas. Ils proposent d'utiliser l'acquisition des connaissances comme base pour la recherche de solutions à partir d'aides à la décision multicritères non compensatoires [Armaghan et Renaud, 12]. Li et Sun ont combiné le RBC et AMC pour améliorer un processus de fouille de données pour la détection de la maladie [Li et Sun, 09]. Erjaee et al., ont proposé une méthode spécifique basée sur des critères multiples pour proposer une décision pour un traitement efficace contre l'infection *Helicobacter pylori* chez les enfants [Erjaee et al., 12].

3.4 Les limites de l'intégration RBC-AMC

L'intégration RBC-AMC a été largement étudiée en médecine. Cependant cette intégration a montré quelques lacunes. Parmi lesquelles, nous notons [Moreno, 15] :

- L'évolution constante des connaissances en santé.
- les difficultés d'interopérabilité au sein des systèmes d'information médicaux.
- Concentration de l'approche à résoudre à l'aspect recherche du RBC. En effet, les études se sont attelées à toucher l'aspect recherche de cas en se concentrant sur la description du cas pour essayer de résoudre une lacune RBC.
- Parfois, une translation amoindrie du cas médical en problème AMC ou d'autres aspects liés à l'AMC sont négligés tels que la pondération des critères qui est parfois liée aux attributs même du cas.
- La jonction RBC-AMC n'est bien prise en charge convenablement faisant parfois deux processus successifs et non intégrés.

3.5 Conclusion

Décider d'un traitement médical, adopter un régime lors d'une infection ou diagnostiquer une maladie, ont été des sujets, parmi tant d'autres, largement traités en aide à la décision par RBC. Avec l'introduction de l'approche multicritères ceci a facilité plus ou moins la résolution de certaines problématiques liées à la remémoration de cas similaires. Néanmoins, l'utilisation de cette approche par combinaison du RBC et de l'AMC reste tributaire d'une bonne translation du cas médical en problème multicritères sans cette opération de translation "définition", l'aspect multicritères ne donnera pas les résultats escomptés. La plupart des hybridations se sont intéressées à résoudre les lacunes du RBC sans pour autant se concentrer sur la qualité des résultats obtenus. Il semble que d'autres facteurs entrent en jeu, tels que la pondération qui est aussi une autre problématique au niveau AMC. Ces facteurs peuvent en effet affecter la décision finale du système à l'étude.

Afin d'évaluer les résultats dans cette approche RBC-AMC, le modèle de prise de décision doit être adapté non seulement pour résoudre les lacunes du RBC mais tenir aussi compte des contraintes de l'AMC. La jonction RBC-AMC doit être bien prise en charge pour tenir compte simultanément des exigences respectives au RBC et à l'AMC.

Chapitre 4

Proposition de systèmes
d'aide à la décision intégrés

Chapitre 4.

Proposition de systèmes d'aide à la décision intégrés

4.1	Introduction	81
4.2	Approche d'aide à la décision intégrée proposée	81
4.2.1	<i>Formalisation du problème médical</i>	81
4.2.2	<i>Le Modèle d'Aide à la Décision Médicale Intégré</i>	82
4.2.3	<i>Elaboration du modèle</i>	85
4.2.4	<i>Définition de la situation médicale</i>	85
4.2.5	<i>Le RBC</i>	85
4.2.6	<i>Le raisonneur</i>	87
4.3	Proposition d'un Système d'Aide à la Décision Intégré RBC-FDD	87
4.3.1	<i>Le problème médical considéré : Diagnostic des maladies</i>	87
4.3.2	<i>Intégration RBC-FDD</i>	89
4.3.3	<i>Le modèle d'aide à la décision proposé</i>	89
4.3.4	<i>Le système RBC^{FDD} proposé</i>	92
4.3.4.1	<i>La situation médicale</i>	93
4.3.4.2	<i>Le RBC</i>	93
4.3.4.3	<i>Le raisonneur FDD</i>	95
4.3.5	<i>Mise en œuvre</i>	97
4.3.5.1	<i>Expérimentation</i>	100
4.3.5.2	<i>Evaluation</i>	101
4.4	Proposition d'un Système d'Aide à la Décision Intégré RBC-AMC	102
4.4.1	<i>Le problème médical considéré : Proposition d'une thérapie</i>	102
4.4.2	<i>Le modèle d'aide à la décision proposé</i>	103
4.4.3	<i>Le système RBC^{AMC} proposé</i>	104
4.4.3.1	<i>Définition de la situation médicale</i>	105
4.4.3.2	<i>Le RBC</i>	106
4.4.3.3	<i>Le raisonneur AMC</i>	107
4.4.4	<i>Mise en œuvre</i>	112
4.4.4.1	<i>Expérimentation</i>	113
4.4.4.2	<i>Evaluation</i>	114
4.5	Conclusion	115

4.1 Introduction

Le présent travail de recherche a pour objectif d'adopter une méthodologie pour l'intégration du *RBC* et une autre méthode de raisonnement que nous appellerons dans notre étude : le *Raisonneur*. L'approche d'intégration est utilisée pour soutenir le processus de raisonnement du *RBC* particulièrement dans la phase recherche de situations similaires (remémoration).

Nous explorons une nouvelle approche qui utilise l'intégration *RBC-Raisonneur* pour proposer une stratégie de remémoration qui permet, au final, de choisir la meilleure solution à partir d'un ensemble de situations déjà résolues. Cette approche contribuera à remédier certains inconvénients liés au fait de trouver plusieurs cas similaires, et par conséquent plusieurs solutions plus ou moins acceptables. Nous utilisons un modèle décisionnel basé sur l'idée d'intégration de deux modes de raisonnements qui peuvent être complémentaires. Ainsi, et avant de détailler notre approche nous ferons remarquer que nous utiliserons les notions citées au chapitre 1 (1.3.1) ayant trait à la décision médicale.

4.2 Approche d'aide à la décision intégrée proposée

4.2.1 Formalisation du problème médical

Dans une étude précédente [Mansoul et Atmani, 16], nous avons défini la situation médicale où habituellement le médecin est en consultation d'un patient et devra explorer le diagnostic possible pour prescrire la meilleure thérapie.

Guidés par cette définition et par les notions que nous avons mises en avant, à savoir : le *raisonnement clinique* et la *décision en situation* (1.3.1). La situation médicale que nous préconisons est décrite par le décideur (médecin) devant le diagnostic d'une situation et il devra explorer les solutions possibles (diagnostic) pour choisir la meilleure thérapie. La situation médicale est donc caractérisée par : une définition du problème plus ou moins complète, une étude exhaustive des diagnostics/thérapies possibles et l'existence de préférences individuelles pour chaque diagnostic/thérapie. En plus, de l'existence de signes spécifiques pour chaque patient, par exemple, «patient âgé», «allergie à la pénicilline», etc.

Ces signes spécifiques vont orienter ou indiquer une thérapie souhaitée (par exemple, un patient âgé peut être moins capable de suivre un régime salé). En outre, il est bien reconnu aujourd'hui que les décisions de diagnostic liées à chaque patient doivent tenir compte du contexte relatif défini à travers les signes spécifiques. Ainsi, le médecin définit une situation pathologique avec u symptômes cliniques et v signes spécifiques. Du point de vue modélisation, la situation médicale devient alors un problème médical défini par des descripteurs et une solution «diagnostic/thérapie» considérée. Pour formaliser cette situation médicale, nous utilisons la structure suivante :

Situation Médicale = {Symptômes cliniques, Signes Spécifiques,
Diagnostic/Thérapie_Proposée}.

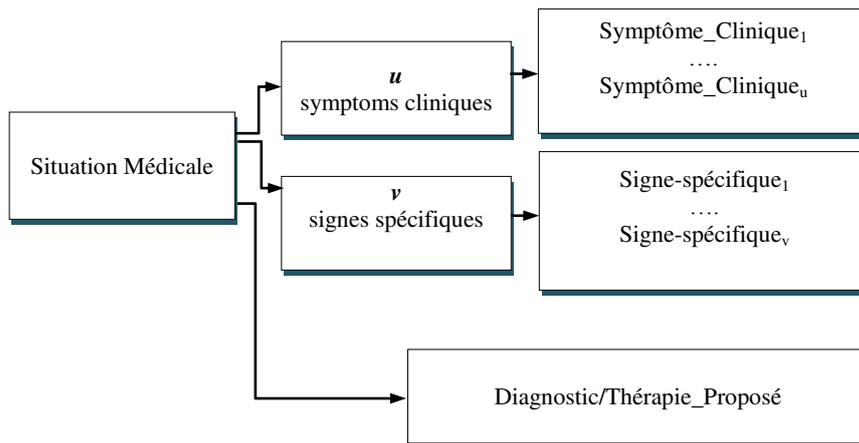


Figure 4.1. La situation médicale.

4.2.2 Le Modèle d'Aide à la Décision Médicale Intégré

L'intégration se base sur l'idée de mutualisation des procédures du *RBC* et du *Raisonneur* pour soutenir le RBC dans sa tâche de recherche de situations similaires ou proches de la situation décisionnelle courante. Sur le plan opérationnel de cette intégration *RBC-Raisonneur*, le raisonneur lui est assigné la tâche principale de « réduction de l'ensemble des cas éligibles à la bonne solution pour faire aboutir la solution ». Autrement, chaque raisonneur intègre ses propres fonctionnalités, ses propres données par rapport à la situation médicale, et sa propre interface pour communiquer avec l'utilisateur « médecin ». L'intégration entre eux, est effectuée par un bloc faisant office de pont. Celui-ci permet l'échange des données, sous une forme qui convient aux deux parties *RBC et Raisonneur*, qui vont coopérer lors de la recherche de la meilleure solution (*Diagnostic/Thérapie*), (figure 4.2).

Notre démarche repose sur l'application d'une stratégie d'intégration pour réaliser une "meilleure" aide à la décision. Pour cela, nous avons suivi la démarche ci-dessous :

- définir la situation médicale ;
- initier le RBC ;
- faire appel au Raisonneur pour guider la recherche de solution et échanger les données avec le celui-ci ;
- proposer une solution ;
- si la solution est acceptée, il ya mémorisation de la nouvelle situation médicale avec sa solution pour une utilisation ultérieure, sinon le décideur pourra revenir à une étape antérieure pour réviser sa situation médicale ou revoir une étape particulière de la présente démarche.

L'aspect interactivité est assuré à travers le dialogue entre l'utilisateur et le système dans réalisation des tâches suivantes :

- la définition de la situation médicale ;
- la modification des données entrées suite à des révisions de la situation médicale ;

- l'élaboration du cas médical pour l'aspect RBC ;
- l'interaction avec le raisonneur à travers la définition du problème médical
- le paramétrage et le choix des procédures adéquates aux traitements envisagés
- l'évaluation et la validation des résultats des traitements avant leur adoption finale.

Cette démarche de résolution suivra le Modèle d'Aide à la Décision Médicale Intégré ci-dessous :

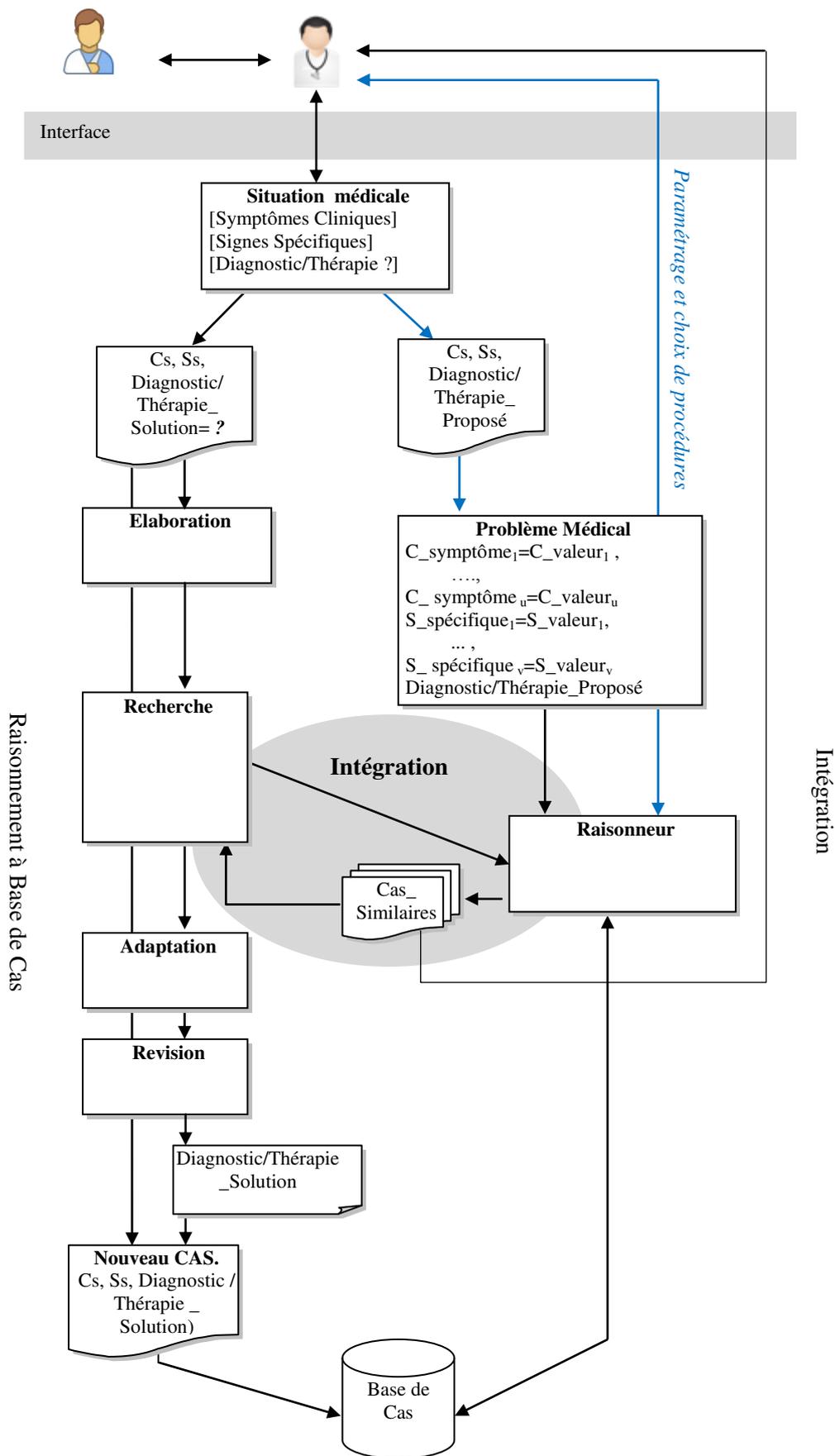


Figure 4.2. Modèle d'Aide à la Décision Médicale Intégré.

4.2.2.1 Elaboration du modèle

Pour établir le modèle permettant de fixer une stratégie de résolution d'une situation médicale, nous devons définir quelques éléments de base qui servent de support à cette situation afin de mieux la décrire dès le départ et continuer le cheminement jusqu'à atteindre le but, celui de l'établissement d'une aide à la décision médicale.

4.2.2.2 Définition de la situation médicale

C'est la compréhension et la perception de la situation. Le médecin analyse le contexte de la situation et définit les objectifs de la décision. Ainsi, le médecin définit sa situation médicale avec les symptômes cliniques (*Cs*) et les signes spécifiques (*Ss*). Ensuite, il peut proposer un diagnostic/thérapie possible (*Diag /Therapy_Pr*) qui sera pris en compte comme avis médical sur la situation en question, et qui est basé sur l'évaluation personnelle du médecin (décideur), sans pour autant que cet avis médical ne soit une solution à ce stade.

En fait, le résultat de cette procédure est exploité par les deux axes du model décisionnel : le *RBC* et le *Raisonneur*. Cette procédure comprend donc, la définition du nouveau cas pour la partie RBC et la définition du problème médical pour le raisonneur.

Table 4.1. Structure de la situation médicale.

Medical_Situation
<hr/>
Clinical_symptom
[C_Symptom ₁ =C_value ₁]
...
[C_Symptom _u =C_value _u]
End_clinical_symptom
Specific_Signs
[S_Sign ₁ =S_value ₁]
...
[S_Sign _v =S_value _v]
End_Specific_Signs
Diagnostic/Therapy_Pr= " "
End_Medical_case

4.2.2.3 Le RBC

[a] Construction du cas

C'est l'étape de construction du nouveau cas avec une solution (Diagnostic/Thérapie_Solution) inexistante encore à cette étape. La situation médicale mènera à la définition formelle de la structure suivante, pour sa prise en compte par le raisonneur :

Table 4.2. Structure du Cas Médical.

Medical_case
Clinical_symptom
[C_Symptom ₁ =C_value ₁]
...
[C_Symptom _u =C_value _u]
End_clinical_symptom
Specific_Signs
[S_Sign ₁ =S_value ₁]
...
[S_Sign _v =S_value _v]
End_Specific_Signs
Diagnostic/ Therapy _Solution=" "
End _Medical_case

[b] Remémoration

Ce processus a une tâche principale : la recherche de similitude. C'est la recherche des n cas les plus proches au cas proposé, en utilisant une mesure de similarité. Le processus sélectionnera les cas les plus proches ou similaires (Cas_Similaires) à partir de la base de cas. Ces cas plus proches seront initialement considérés les plus pertinents pour entamer le processus de recherche de la solution à la situation médicale. La méthode k - nn est utilisée pour la simplicité de sa mise en œuvre.

[c] Adaptation

Après la récupération, le système évalue le degré de similitude des cas sélectionnés avec le cas courant. Le degré de similitude détermine, si une adaptation est nécessaire ou si la solution est utilisée telle qu'elle est. Enfin, il y a proposition d'une solution sur la base des cas récupérés.

[d] Révision

L'utilisateur commence par réviser la solution récupérée pour refléter les différences entre le nouveau cas et le(s) cas récupéré(s). En fin de compte, la (les) solution(s) du (des) cas est adaptée pour devenir une solution du problème actuel à valider par le médecin.

[e] Mémorisation

Lorsqu'une solution (Diagnostic/Thérapie_Solution) pour le nouveau problème est trouvée, une nouvelle expérience est faite. Le médecin est invité à conserver le nouveau cas, s'il le souhaite, avec sa solution. Ceci augmentera alors l'expérience de résolution d'autres situations par le processus.

4.2.3 Le raisonneur

[a] Etape information

C'est la définition du problème médical pour la partie Raisonneur. Plusieurs informations relatives aux symptômes peuvent intervenir dans l'évaluation par le médecin. Les données et évaluations récoltées sur la situation médicale, seront intégrées dans le système pour créer le problème médical. Ensuite, il y a évaluation par le décideur des différents paramètres du problème, intervenants dans les traitements (Etape. b), en fonction de la procédure adoptée par le raisonneur.

Table 4.3. Le problème médical (Structure).

Problème Médical
C_symptôme ₁ =C_valeur ₁ ,
.....
C_symptôme _u =C_valeur _u
S_spécifique ₁ =S_valeur ₁ ,
....
S_spécifique _v =S_valeur _v
Diagnostic/Thérapie_Proposé

[b] Etape traitement

Le raisonneur entamera ses opérations pour faire aboutir la solution. Principalement, il y aura les opérations suivantes :

1. lancement de la procédure appropriée (méthode multicritères) ;
2. proposition d'une solution (s) ;
3. révision (si possible) de la solution.

4.3 Proposition d'un Système d'Aide à la Décision Intégré

RBC-FDD

Dans cette section, nous expérimentons une nouvelle méthode en utilisant la collaboration entre le RBC et la FDD, pour proposer une stratégie qui aide la tâche de recherche de la meilleure solution à travers une méthode de fouille de données : le groupement (Clustering).

4.3.1 Le problème médical considéré : Diagnostic des maladies de la colonne vertébrale

Nous nous proposons d'utiliser un ensemble de données médicales se rapportant aux maladies de la colonne vertébrale de patients orthopédiques, nous projetons d'utiliser la base de données "Presumptive Diagnosis of Diseases of Orthopaedic Patients" que nous nommerons :

UCI_PDDOP¹. En fait, c'est un ensemble d'enregistrements de patients, chez qui on a diagnostiqué des maladies orthopédiques. Chaque patient est en effet enregistré avec ensemble de données sur six caractéristiques biomécaniques utilisées pour classer les patients orthopédiques en 3 classes ("Normal", "Hernia" ou "Spondylolisthese"). Chaque patient est alors décrit par six attributs biomécaniques dérivés de la forme et de l'orientation du bassin et de la colonne lombaire, dans l'ordre suivant : incidence pelvienne, inclinaison pelvienne, angle de lordose lombaire, pente sacrale, rayon pelvien et degré de spondylolésistance. Tous ces attributs sont les caractéristiques biomécaniques des maladies orthopédiques. La notation suivante est utilisée pour les étiquettes de classe : H pour "Hernia", S pour "Spondylolisthesis" et N pour "Normal".

```

48.10923638,14.93072472,35.56468278,33.17851166,124.0564518,7.947904861,Hernia
74.37767772,32.05310438,78.77201304,42.32457334,143.5606905,56.12590603,Spondylolisthesis
89.68056731,32.70443487,83.13073216,56.97613244,129.9554764,92.02727682,Spondylolisthesis
52.86221391,9.410371613,46.98805181,43.4518423,123.0912395,1.856659161,Normal
.....

```

Figure 4.3. Echantillon² de "Presumptive Diagnosis of Diseases of Orthopaedic Patients".

Aux fins de notre expérimentation, nous avons transformé les attributs de «Presumptive Diagnosis of Diseases of Orthopaedic Patients Data Set» en descripteurs de la base de cas, où chaque cas (patient) est décrit par les descripteurs X_1 , X_2 , ..., X_6 , et nous associons le descripteur cible Y au diagnostic. La table 4.4 présente ces descripteurs.

Table 4.4. Les descripteurs de la base de données [UCI_PDDOP].

Descripteur	Descripteur biomécanique	Type
X_1	pelvic incidence	numérique
X_2	pelvic tilt	numérique
X_3	lumbar lordosis angle	numérique
X_4	sacral slope	numérique
X_5	pelvic radius	numérique
X_6	grade of spondylolisthesis	numérique
Y	Diagnosis	H (Hernia), S (Spondylolisthesis), N (Normal).

¹ UCI_PDDOP : <http://archive.ics.uci.edu/ml/datasets/Vertebral+Column#>

² Chaque donnée élémentaire est séparée par une virgule.

4.3.2 Intégration RBC-FDD

La contribution de la fouille de données.

Dans ce travail, nous expérimentons une nouvelle méthode en utilisant la collaboration entre RBC et l'ECD (fouille de données) pour proposer une stratégie pour la recherche de solution qui permet de choisir la meilleure solution à partir d'un ensemble réduit de solutions trouvées par fouille de données, en l'occurrence la méthode de groupement "Clustering".

4.3.3 Le modèle d'aide à la décision proposé

Donc, pour rechercher une solution (diagnostic), nous devons suivre un modèle décisionnel pratique qui peut être facilement assimilé à un modèle décisionnel médical comme suit :

- recueillir des informations sur la situation médicale ;
- prendre en compte d'un diagnostic possible, s'il est proposé par le décideur (médecin) ;
- définir les descripteurs les plus pertinents qui seront utilisés par la FDD ;
- évaluer la solution (Diagnostic_Pr) proposé ;
- réviser (si possible) ou accepter la solution choisie.

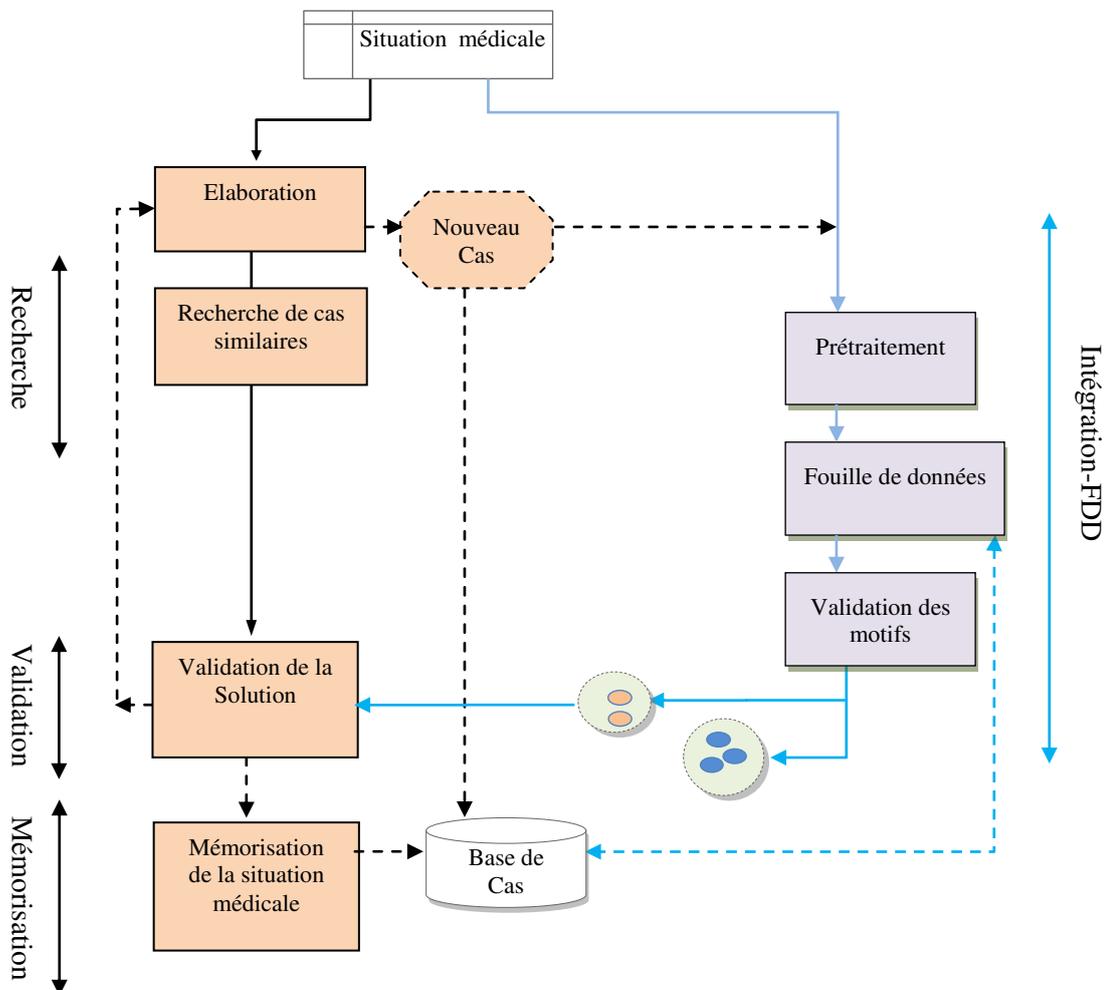


Figure 4.4. Le modèle d'aide à la décision médicale RBC-FDD proposé.

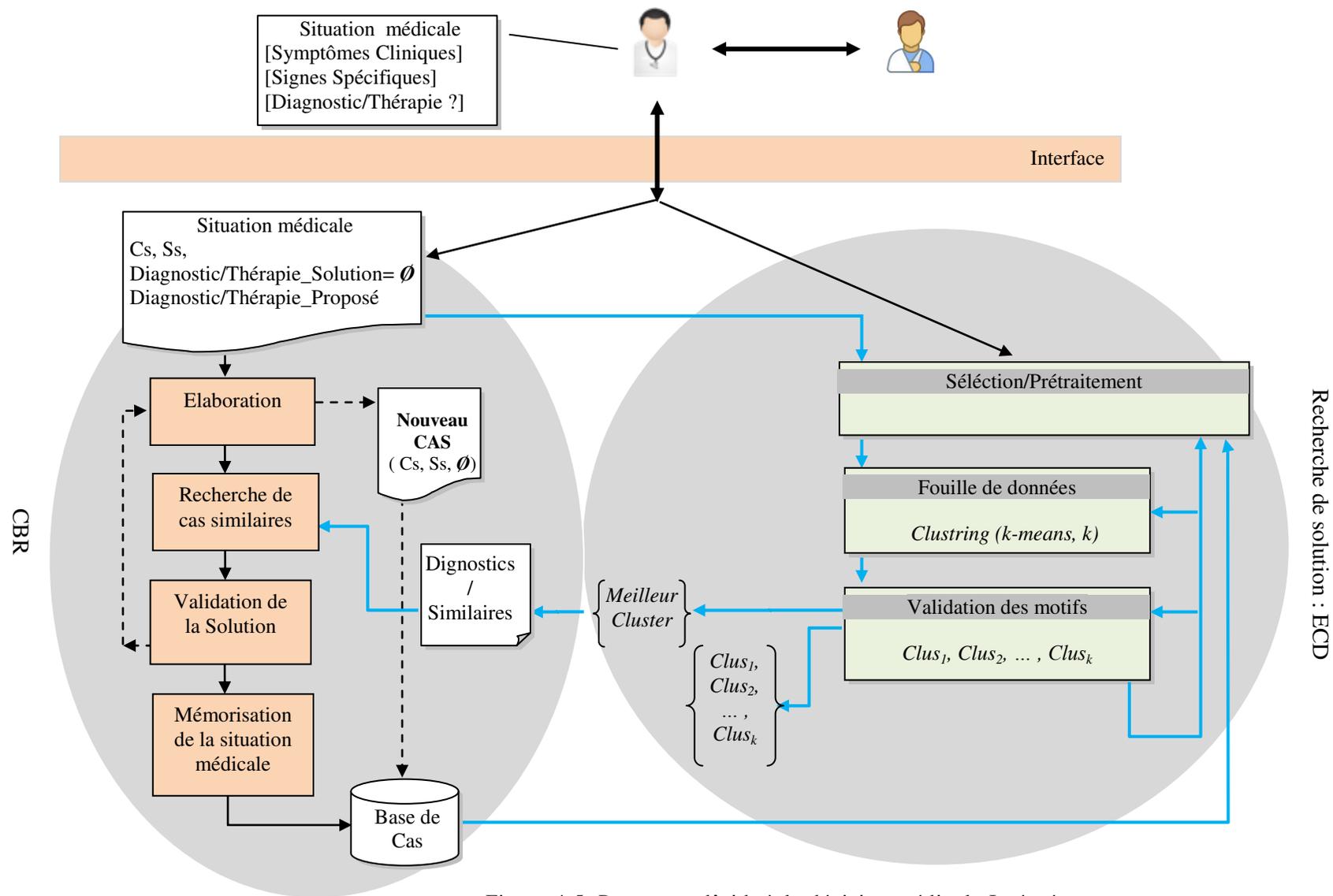


Figure 4.5. Processus d'aide à la décision médicale Intégré.

4.3.4 Le système RBC^{FDD} proposé

Nous proposons un système interactif d'aide à la décision médicale, défini comme un processus complet qui comprend un ensemble d'éléments et de routines pertinentes afin d'assurer les fonctions principales du système et aider à prendre les décisions appropriées. Une description du système est illustrée par la figure 4.6 qui montre schématiquement l'intégration de tous les traitements, de l'acquisition d'informations sur la situation de décision jusqu'à l'aide à la décision.

Ce système sera supporté par un modèle décisionnel comme indiqué dans la figure suivante :

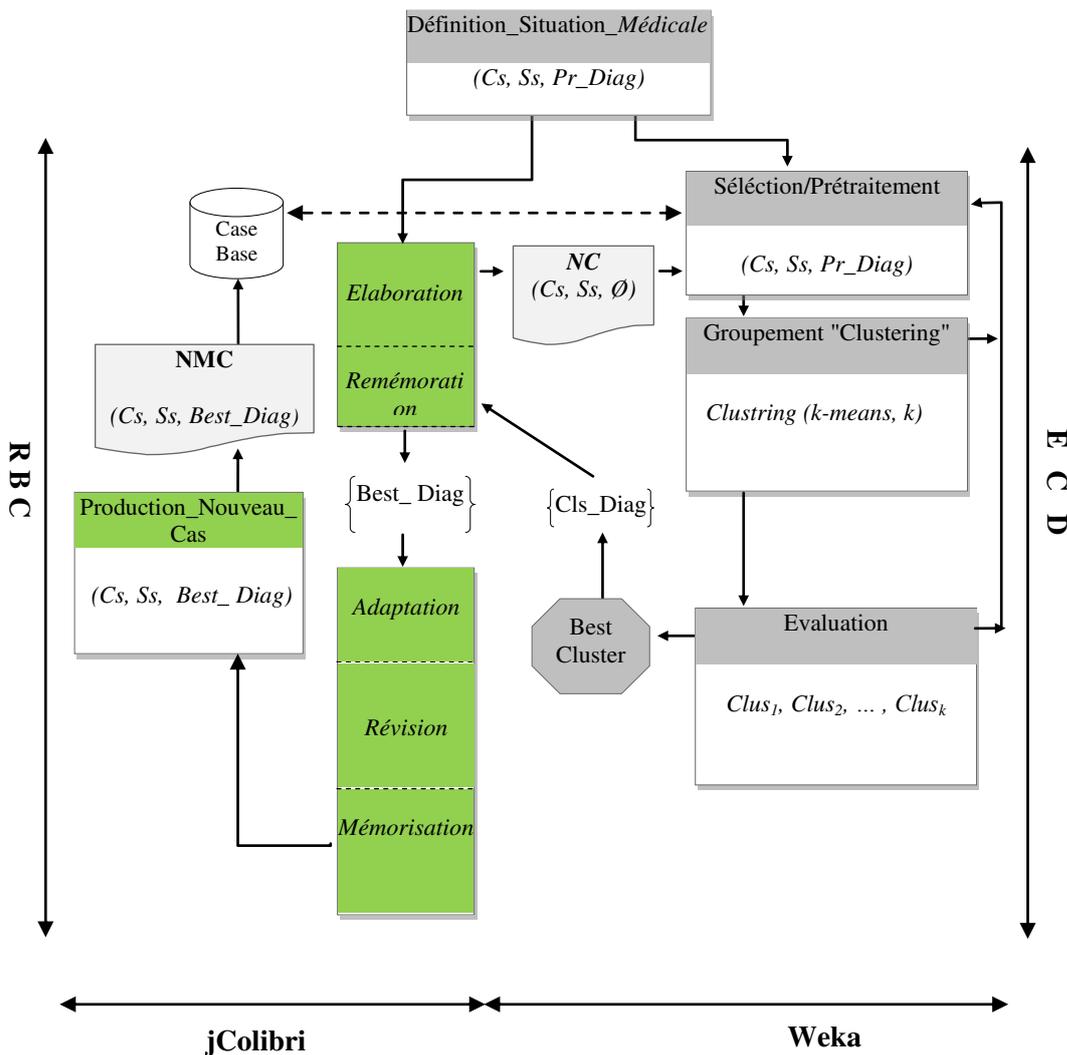


Figure 4.6. Le système RBC^{FDD} proposé.

- Cs, Ss, Pr_Diag : Clinical symptoms, Specific Signs, Proposed Diagnosis
- $Cls_Diags, Best_Diag$: Closest Diagnosis, Best Diagnosis
- $NC(Cs, Ss, \emptyset)$: New_Case (Clinical symptoms, Specific Signs, Diagnosis= $\{\emptyset\}$)
- $NMC(Cs, Ss, Best_Diag)$: New Medical Case (Clinical symptoms, Specific Signs, Best_Diagnosis)

4.3.4.1 La situation médicale

Le médecin distingue le contexte de la situation, définit les objectifs de la décision et propose s'il le souhaite un diagnostic possible qui sera pris en compte pour la présente situation médicale. En fait, cette procédure est commune aux deux axes du processus décisionnel : le RBC et la FDD.

4.3.4.2 Le RBC

(i) Elaboration

C'est la définition du nouveau cas. Le médecin définit sa situation médicale avec les symptômes cliniques (Cs) et les signes spécifiques (Ss). Ces derniers vont contribuer à construire le nouveau cas $NC (Cs, Ss, \emptyset)$ avec une solution (diagnostic) inexistante à ce stade. Compte tenu de notre contexte médical (choix de la méthode de contraception), un cas médical sera créé, et ayant structure suivante :

Table 4.5. Le cas médical orthopédique constitué.

Medical_Case
<hr/>
Clinical_Symptoms
Pelvic_incidence=value ₁
Pelvic_tilt=value ₂
Lumbar_lordosis_angle=value ₃
Sacral_slope=value ₄
Pelvic_radius=value ₅
Grade_of_spondylolisthesis=value ₆
End Clinical_Symptoms
Specific_Signs
End_Specific_Signs
Diagnosis= “ ”
END_Medical_Case

(ii) Recherche

Ce processus a une tâche principale : la recherche de similitude. C'est la recherche des n cas les plus proches au cas proposé en utilisant une mesure de similarité. La similarité locale ne tient compte que des symptômes qui ont été vérifiés, grâce à la consultation faite par le médecin. Il les considère comme plus importants ou assez pertinents pour la définition de son cas. Nous utilisons la méthode k -nn. Le processus sélectionnera alors, les cas les plus proches ou similaires (Best_Cluster), à partir de la base de cas.

Ce meilleur groupe (cluster) sera considéré comme étant *l'espace réduit* qui permettra de déduire par la suite, le meilleur diagnostic (Best_Diag). Ce cluster est finalement envoyé au RBC pour filtrage et proposition de la meilleure (Best_Diag) qui sera considérée comme solution. De ce fait, le médecin aura les charges suivantes:

- attribution d'une valeur à k pour la méthode k -nn ;

- lancement du processus RBC qui sera pris en charge par le pseudo algorithme qui suit.

Pseudo Algorithme. RBC

```

1: Input :Cls_Diag ← ∅
2: NC (Cs, Ss, ∅)
3: Initialize k
4: Retrieve(NC, Closest_Cases) using k-nn
5: If Closest_Cases≠∅ then
    For each Current_Case in Closest_Cases
        For i=1 to n
            { Cls_Diag ← Cls_Diag ∪ Current_Case (Casei, Diagi)}
        Endfor
    Endfor
Else
    Cls_Diag ← ∅
EndIf
6: RECEIVE(Cls_Diag)
7: Reuse(Best_Diag)
8: Revise(Best_Diag)
9: Retain(Best_Diag, result)
10: If result = "yes" Then NMC(Cs, Ss, Best_Diag)= PNMC (Cs, Ss, Best_Diag)
    Else
        NCM(∅, ∅, ∅)
    Endif
11: Output : NMC(Cs, Ss, Diag)

```

Cls_Diags:Closest Diagnostics PNMC:Production_Nouveau_Cas_Medical
NMC: New_Medical_Case NC: New_Case

(iii) Adaptation

Après la récupération, le système évalue le degré de similitude des cas sélectionnés avec le cas courant. Le degré de similitude détermine, si une adaptation est nécessaire ou si la solution récupérée peut être réutilisée directement. Ce processus d'adaptation est fait pour mieux adapter le cas cible. Enfin, il propose une solution à partir des cas récupérés.

(iv) Révision

Il commence par adapter la (s) solution (s) récupérée (s) pour refléter les différences entre le nouveau cas et le (s) cas (s) récupéré (s). En fin de compte, la (les) solution (s) du (des) cas est adaptée pour devenir une solution du problème actuel à valider par le médecin.

(v) Mémorisation

Lorsqu'une nouvelle solution (réussie) pour le nouveau problème est trouvée, une nouvelle expérience est faite, qui peut être stockée dans la base de cas pour augmenter sa compétence pour résoudre d'autres situations ou cas et le médecin est invité s'il souhaite conserver le nouveau cas avec sa solution.

4.3.4.3 Le raisonneur FDD

Dans cette section, nous expérimentons une nouvelle méthode en utilisant la collaboration entre le RBC et la FDD pour proposer une stratégie de recherche qui permet de choisir la meilleure solution à partir d'un espace de solutions trouvées par une procédure de fouille de données : le *groupement* «Clustering».

Le processus de groupement «Clustering».

Nous avons choisi une approche rationnelle pour la tâche de remémoration : au lieu d'une recherche massive de cas qui est la recette classique du raisonnement, nous analysons les cas de manière rationnelle et nous concentrons la recherche sur des périmètres particuliers avec des cas spécifiques qui font l'objet de suspicion ou ce que nous appellerons « cas potentiels ». Notre objectif est de trouver les cas les plus proches sur tous les cas traités précédemment afin d'éviter une complication à la phase d'adaptation et de la rendre pénible. En effet, nous n'allons pas rassembler tous les cas les plus proches, mais plutôt se concentrer sur un petit périmètre de cas similaires. Donc, nous devons procéder autrement que par une méthode classique. Ainsi, notre méthode consiste à :

- réduire le périmètre de recherche ;
- remémoration à partir de ce périmètre réduit.

De là, l'opération de regroupement est guidée principalement par les symptômes cliniques (*Cs*) et les signes spécifiques (*Ss*), pour réduire le sous-ensemble de cas éligibles à la recherche de solution par similitude. Ainsi, la réduction des solutions d'espace de recherche pour l'étape de *remémoration* devient une opération simplifiée pour le processus RBC. Ce choix peut clairement rendre la *remémoration* optimale en termes de calculs en ce qui concerne uniquement une solution intéressante où seulement les cas pris dans des circonstances comparables sont récupérés. De là, l'ensemble des cas les plus proches (*Closest_Cases*) est reçu du RBC pour l'associer au diagnostic proposé (*Pr_Diag*) que l'utilisateur a déjà définis dans la situation médicale. Ce processus sera traité par les étapes suivantes :

(i) Sélection

- Les symptômes cliniques seront pris automatiquement dans l'opération de sélection des attributs participant à l'opération parce que nous jugeons qu'à la base ce sont eux même qui définissent le problème médical.
- Le médecin choisira en plus le ou les attributs qui pourront se joindre à la liste des attributs du groupement. Ainsi, le médecin aura à sélectionner à partir de la liste des signes spécifiques, ceux qu'il jugera nécessaires à faire participer au groupement.

En effet, les signes spécifiques aident à mieux restreindre l'espace des cas éligibles, afin de ne conserver que ceux qui ne répondent qu'à certaines contraintes définies par le médecin. Ensuite, les étapes suivantes (ii) et (iii) seront lancées successivement.

(ii) Prétraitement

À cette étape, il y a préparation des données pour l'opération de groupement par un nettoyage ou autres traitements appropriés.

(iii) Le Groupement (Clustering)

Nous utilisons une procédure de sélection que nous avons appelée "Clustering", ayant pour objectif la génération du meilleur groupe (Best_cluster). Cette procédure sera guidée par le principe que nous résumons comme suit :

Principe : groupement < Clustering >

- 1: insérer le cas courant dans la base de cas;
 - 2: définir les attributs parmi C_s et S_s , les descripteurs qui participent au groupement
 - 3: initialiser³ $k = 2$;
 - 4: lancer le groupement de la base de cas avec k . Ceci permettra de partitionner la base en deux groupes : les cas "candidats" (C_Cases), et les cas "non_candidats" (C_Cases)⁴.
-

Le médecin doit vérifier les signes spécifiques qu'il veut impliquer dans le regroupement, c'est-à-dire, qu'il considère significatif par rapport à sa situation. Ensuite, une automatique et momentanément (jusqu'à la fin du processus de regroupement) insertion du cas actuel dans la base du cas est effectuée, ceci afin de le considérer dans la procédure de groupement. Viendra en dernier lieu, le groupement qui est guidé par les signes spécifiques pour sélectionner un sous-ensemble de cas, mais nécessairement précédé d'un prétraitement qui consiste à vérifier les données ou autre traitement spécifique.

En initialisant k à 2, l'opération de groupement divise la base de cas en deux sous ensembles: "cas candidats" (C_Cases) et "cas non candidats" (NC_Cases). Le groupe éligible à RBC est maintenant prêt.

L'ensemble du processus de FDD sera traité par le pseudo-algorithme suivant :

Pseudo Algorithme. Clustering

- 1: Input: New_Case(C_s, S_s, \emptyset), Case_Base
- 2: Output: Best_Cluster
- 3: Begin.
- 4: Initialize k with 2 for k -means method
- 5: Insert_momentarily (New_Case, Case_Base)
- 6: Features_Selection(C_s, S_s)

³ Pour avoir seulement deux Clusters.

⁴ Nous pouvons également répéter le test par rapport à $k > 2$ pour réduire de plus en plus l'espace de recherche des cas. Le cas actuel sera automatiquement dans le Cluster (C_Cases). Pour avoir seulement deux Clusters.

```

7: k-means_Clustering(k, Case_Base, C_s, S_s, C_Cases, NC_Cases)
8: Accept_or_Refuse(C_Cases)
9: If Accept C_Cases Then go to 10
    Else return to step 6 or go to 11
    Endif
10: Best_Cluster=C_Cases
11: End

```

(iv) Evaluation

Une fois l'opération de fouille de données terminée, le système présentera le résultat, et le médecin aura à le valider. Si, ce dernier accepte le groupement proposé, le système transmet le groupe « cluster » accepté au processus CBR afin de continuer son traitement, sinon la solution est rejetée et le médecin peut alors refaire une nouvelle expérimentation après avoir modifié, si possible, ses paramètres pour la nouvelle expérimentation. Le système doit alors prendre en compte cette révision du problème médical avant de relancer une nouvelle fois le processus de fouille de données.

4.3.5 Mise en œuvre

L'approche proposée a été appliquée à un ensemble de données médicales de l'Ensemble de données sur les colonnes vertébrales de patients orthopédiques, nous avons réalisé une plateforme développée en programmant les différents modules dont le module d'interconnexion (JColibri-Weka) en Java (JBuilder). Nous utilisons la plate-forme JCOLIBRI pour construire la base de cas et toutes les opérations relatives au CBR et WEKA pour le Clustering. L'environnement JBuilder a été utilisé pour programmer des procédures basiques d'interconnexion entre les deux environnements. Cette interconnexion permet le passage des données entre les deux environnements et les expérimentations.

- *L'environnement JBuilder.* JBuilder est un IDE (Integrated Development Environment), c'est-à-dire, un logiciel qui simplifie la programmation en proposant un certain nombre de raccourcis et d'outils d'aide à la programmation.
- *La plateforme JColibri [Recio-Garcia et al., 2014].* C'est une plateforme open source d'applications. Elle comprend une hiérarchie de classes Java et des fichiers XML. La plateforme est organisée autour des éléments principaux suivants : les tâches et méthodes, la base de cas, la gestion du cas et les méthodes de résolution de problèmes. Dans notre étude, la plateforme JColibri a été utilisée afin de procéder au raisonnement à base de cas, c'est-à-dire la recherche de cas similaires. JColibri traite le processus de résolution de problèmes comme le montre la figure suivante :

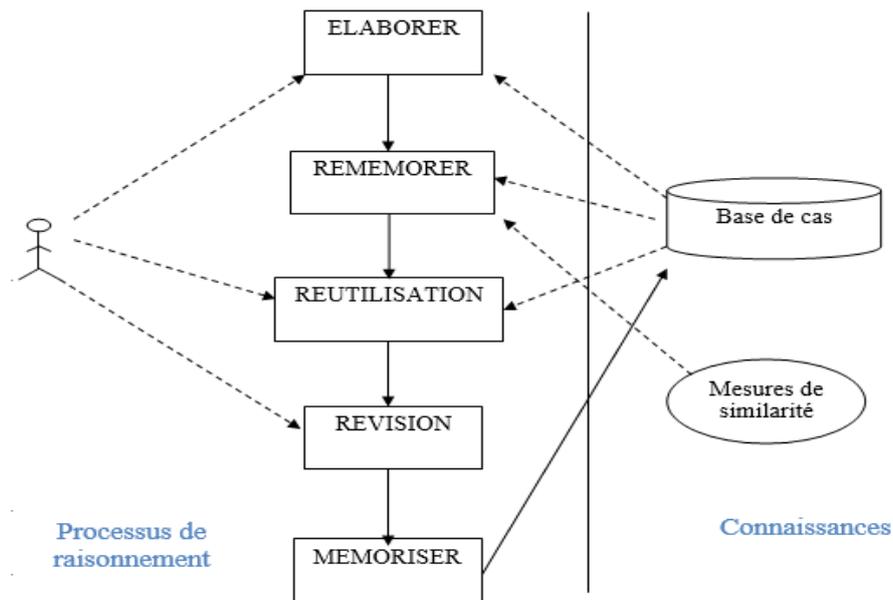


Figure 4.7. Processus RBC adapté de Assali et al. [Assali et al., 09].

- La plateforme WEKA [Hall et al., 09].

Weka (Waikato Environment for Knowledge Analysis) est un ensemble d'outils permettant de manipuler et d'analyser des fichiers de données, implémentant la plupart des algorithmes d'intelligence artificielle, dont les arbres de décision et les réseaux de neurones, le groupement, la classification et les règles d'association.

- Les principaux acteurs de la plateforme.

Pour simplifier les tâches, nous avons considéré un seul principal acteur :

- *Le médecin (décideur)*. Il est l'utilisateur principal de la plate forme. Il l'utilise pour faire de l'aide à la décision médicale.

Afin de montrer les principales tâches de l'aide à la décision nous présentons quelques diagrammes UML sur lesquels nous nous sommes basés pour développer notre plateforme.

- Les diagrammes UML

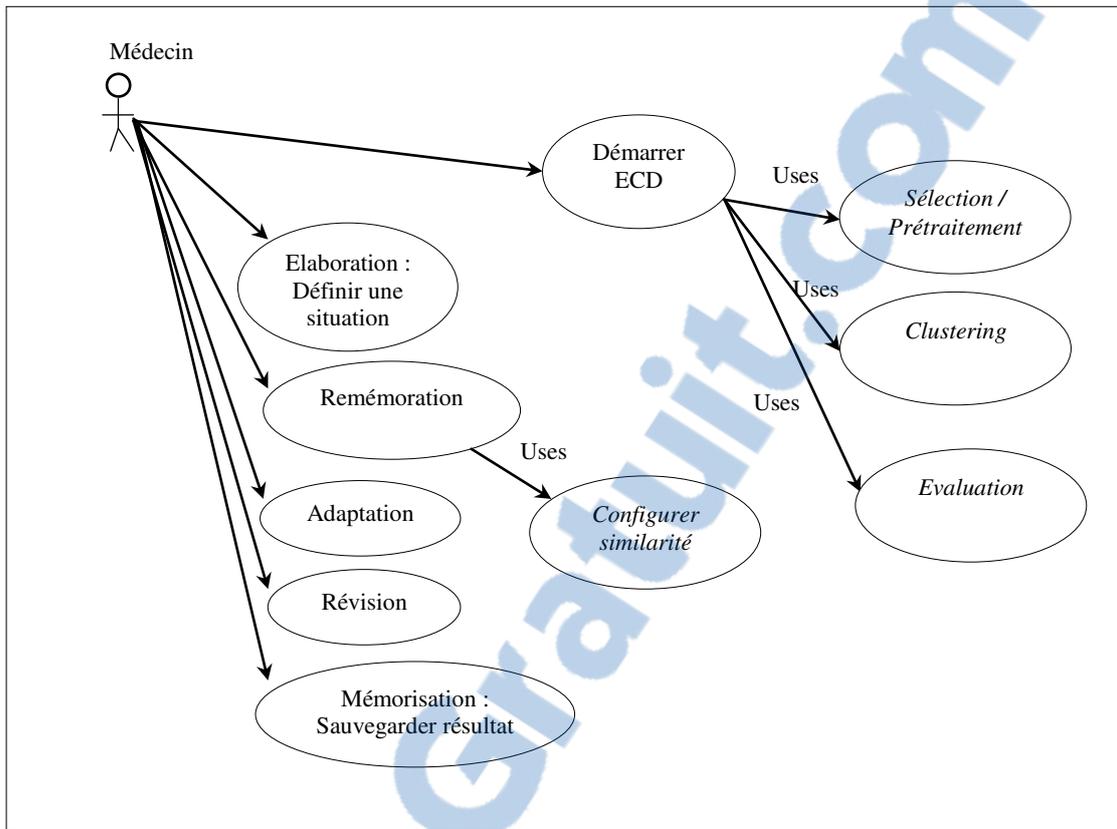


Figure 4.8. Diagramme de cas d'utilisation (FDD).

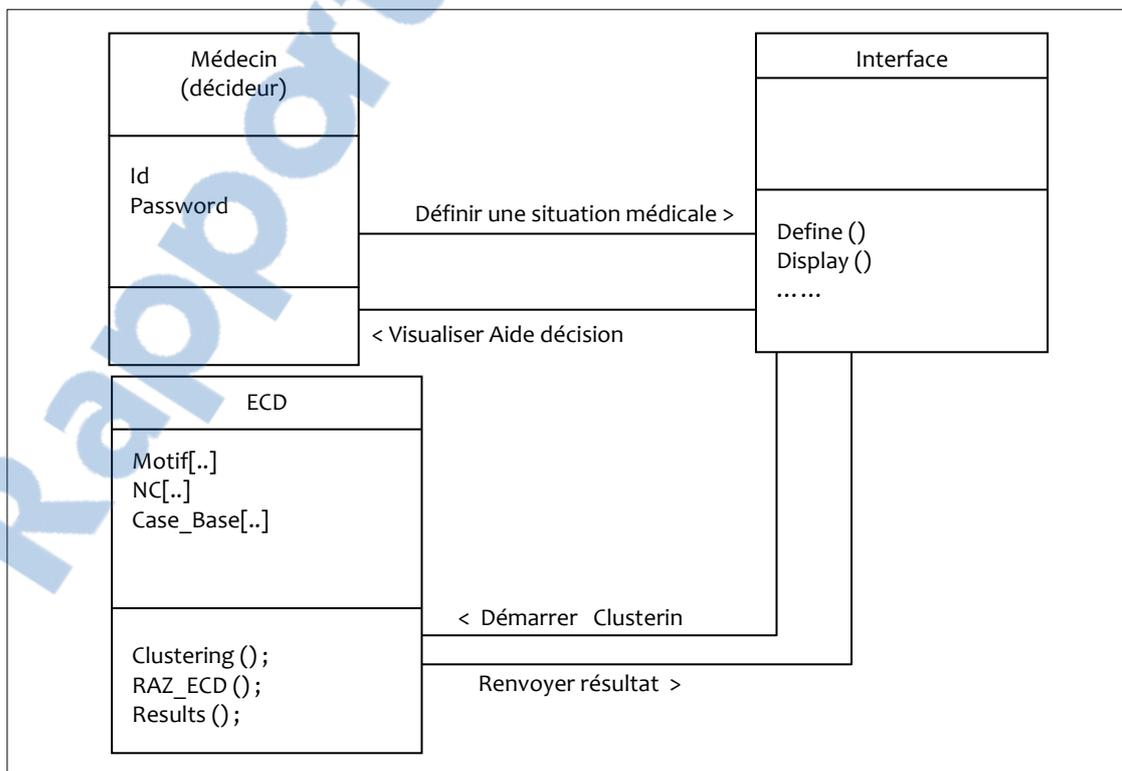


Figure 4.9. Le diagramme de classes (FDD).

4.3.5.1 Expérimentation

Pour les besoins de notre expérimentation, nous avons transformé l'échantillon « Presumptive Diagnosis of Diseases of Orthopaedic patients Data Set » en une base de cas nommée Ω_N . Il contient n cas ω_i , $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ où chaque cas est composé de l'ensemble X_1, X_2, \dots, X_6 , appelés attributs descriptifs. Ensuite, nous associons un attribut cible Y correspondant au diagnostic qui prend valeur dans l'ensemble $Y^5 = \{ "H", "S", "N" \}$.

Table 4.6. La base de cas Ω_N .

ω	$X_1(\omega)$	$X_2(\omega)$	$X_3(\omega)$	$X_4(\omega)$	$X_5(\omega)$	$X_6(\omega)$	$Y(\omega)$
ω_1	63.0278175	22.55258597	39.60911701	40.47523153	98.67291675	-0.254399986	Hernia
...							
ω_i	44.529051	9.433234213	51.99999999	35.09581679	134.7117723	29.10657504	Spondylolisthesis
...							
ω_n							

Après la construction de la base de cas Ω_N , on subdivise cette base de cas en une base d'apprentissage Ω_L (80% de Ω_N) et une base test Ω_T (20% de Ω_N). La table 4.7 montre les bases partielles.

Table 4.7. Bases de cas partielles Ω_L, Ω_T

Base de Cas Ω_N	Base d'apprentissage Ω_L 80%	Base Test Ω_T 20%
310	248	62

Pour évaluer l'efficacité et vérifier l'évolutivité de l'approche proposée, les expériences ont été réalisées comme suit :

Nous avons considéré 20 cas tirés aléatoirement de la base test Ω_T sans aucune hypothèse de diagnostic. Une comparaison de chaque cas de Ω_T se fait avec la base d'apprentissage Ω_L comme suit :

⁵ "H"="Hernia", "S"=" Spondylolisthesis", "N"="Normal".

Pseudo-Algorithmme : Compute Matching

```

1:  Input: Testing Base  $\Omega_T$ 
2:  Output: TC, TD
3:  Begin
4:  For Each case of  $\Omega_T$ 
    Executer Conditionnal_Structure (5)
    If Concordance Then TC=TC+1
    Else TD=TD+1
    Endif
5:  End_For
6:  End

```

Avec la structure conditionnelle (5), nous calculons le taux de concordance correcte. Ce taux représente le nombre de cas correctement identifiés dans la base d'apprentissage Ω_A et diagnostiqué de manière identique dans la base test Ω_T . Les résultats du test sont présentés dans la table 4.8.

$$\begin{array}{l}
 \forall \omega_i \in \Omega_P \text{ or } \omega_i \in \Omega_T \\
 \text{et} \\
 \forall \omega_j \in \Omega_A
 \end{array}
 \left\{ \begin{array}{l}
 \text{Si } Y(X(\omega_i)) = Y(X(\omega_j)) \text{ Alors "Concordance"}^6 \\
 \text{Sinon "Discordance"}
 \end{array} \right. \quad [5]$$

4.3.5.2 Evaluation

Nous calculons le taux d'erreur de chaque type de diagnostic. D'après les résultats présentés dans la table 4.8, nous notons que le taux d'erreur est relativement faible (inférieur à la moyenne), ce qui indique que notre système tend à donner des réponses proches de la réalité de la base de données initiale.

Table 4.8. Comparaison des résultats sur 3 ensembles de données de test.

Nombre de cas testé à partir de Ω_T	Type de diagnostic ⁷ des cas testés (à l'origine)	Cas mal classés	Taux erreur (%)
20	"N"	4	20
20	"H"	7	35
20	"S"	8	40

Selon ces résultats, nous notons que le taux de bonne concordance sur la valeur de diagnostic est relativement élevé par rapport à la moyenne, ce qui indique que le système fournit des

⁶ Concordance et Discordance sur la valeur de l'attribut Diagnostic (Y)

⁷ "H"="Hernia", "S"="Spondylolisthesis", "N"="Normal".

résultats proches de la réalité telle que déclarée dans la base test Ω_T , en particulier pour les cas avec un diagnostic "N". Nous notons également que le taux de bonne reconnaissance est supérieur à la moyenne, ce qui indique que le modèle décisionnel adopté a tendance à reconnaître et à faire une bonne reconnaissance de diagnostic.

4.4 Proposition d'un Système d'Aide à la Décision Intégré RBC-AMC

Nous mettons à contribution l'analyse multicritères pour appuyer le RBC dans la recherche de solution. A cet effet, l'AMC permettra de guider le processus vers un espace de recherche de solution assez réduit qui évitera au processus de passer par un ensemble large de cas similaires qui en fin de compte ne serviront pas tous à avoir une solution optimale. La démarche que nous préconisons adopte un champ d'application médical à savoir : l'usage des méthodes contraceptives chez les femmes.

4.4.1 Le problème médical considéré : Proposition d'une thérapie (une méthode de contraception)

Nous projetons d'utiliser des données sur le choix de méthodes contraceptives. Ces données proviennent d'une base de données appelée : « The 1987 National Indonesia Contraceptive Prevalence Survey ». Cette base de données est le résultat d'une enquête réalisée en Indonésie sur la prévalence des contraceptifs chez les femmes en 1987 [UCI_CMCDS⁸]. L'échantillon est composé de femmes mariées qui ne sont pas enceintes ou ne savent pas si elles l'étaient au moment de l'enquête. Ces données indiquent les méthodes contraceptives utilisées par les femmes : "No-Use Method"="1", "Long-Term Method"="2" ou "Short-Term Method"="3". La figure 4.10 donne un aperçu d'un échantillon de données.

24,2,3,3,1,1,2,3,0,1
45,1,3,10,1,1,3,4,0,1
43,2,3,7,1,1,3,4,0,1
42,3,2,9,1,1,3,3,0,1
.....

Figure 4.10. Echantillon⁹ de "Contraceptive Method Choice Data Set" [UCI_CMCDS].

Chaque femme mariée est décrite par un ensemble de dix descripteurs, dont le dernier contient la méthode de contraception utilisée. Les descripteurs utilisés sont les suivants :

⁸ UCI_CMCDS : <https://archive.ics.uci.edu/ml/datasets/Contraceptive+Method+Choice>

⁹ Chaque donnée élémentaire est séparée par une virgule

Table 4.9. Les descripteurs de la base de données [UCI_CMCDs].

Descripteur	Valeur
Wife's age	
Wife's education	1=low, 2, 3, 4=high
Husband's education	1=low, 2, 3, 4=high
Number of children ever born	
Wife's religion	0=Non-Islam, 1=Islam
Wife's now working?	0=Yes, 1=No
Husband's occupation	1, 2, 3, 4
Standard-of-living index	1=low, 2, 3, 4=high
Media exposure	0=Good, 1=Not good
Contraceptive method used	1="No-Use Method" 2="Long-Term Method " 3="Short-Term Method "

4.4.2 Le modèle d'aide à la décision proposé

Nous nous sommes basés sur les bases du raisonnement clinique non analytique (Chapitre 1 ; 1.3.1) pour proposer un modèle d'aide à la décision. Le modèle adopté (figure 4.11) montre la réalisation du processus jusqu'à aboutissement de la meilleure *Thérapie_Solution*. Le modèle que nous suivons découle de la situation décrite ci-dessous, décrite par le décideur (médecin) : définition du problème plus ou moins complète, large revue des solutions possibles (Thérapies) et l'existence de préférences individuelles pour chaque solution.

Ainsi, notre approche repose sur une situation typique qui suit le modèle non analytique du raisonnement clinique fondé théoriquement sur ces hypothèses :

- le médecin a toutes les informations nécessaires sur la situation médicale ;
- le médecin connaît les critères pertinents et toutes les thérapies avec les conséquences de chacune, ces critères et thérapies sont évalués en fonction de leur importance ;
- les critères ainsi listés sont stables et ayant des évaluations qui ne changent pas dans le temps ;
- le médecin choisit la thérapie qui garantit les résultats les plus acceptables.

Donc, pour chercher une solution (Diagnostic/Thérapie_Solution), nous avons fait une intégration de l'analyse multicritères (Raisonneur AMC), et nous avons défini un modèle décisionnel assimilable à un modèle d'aide multicritères à la décision médical, comme suit:

- recueillir des informations sur la situation médicale ;

- prendre en compte une liste des thérapies possibles proposée ;
- Pondérer les critères par un système de valeurs adéquat ;
- évaluer chaque thérapie selon les critères ;
- rechercher la meilleure Thérapie_Solution ;
- réviser (si possible) ou accepter la Thérapie_Solution.

Ce modèle décisionnel supportera les traitements indiqués par la figure suivante :

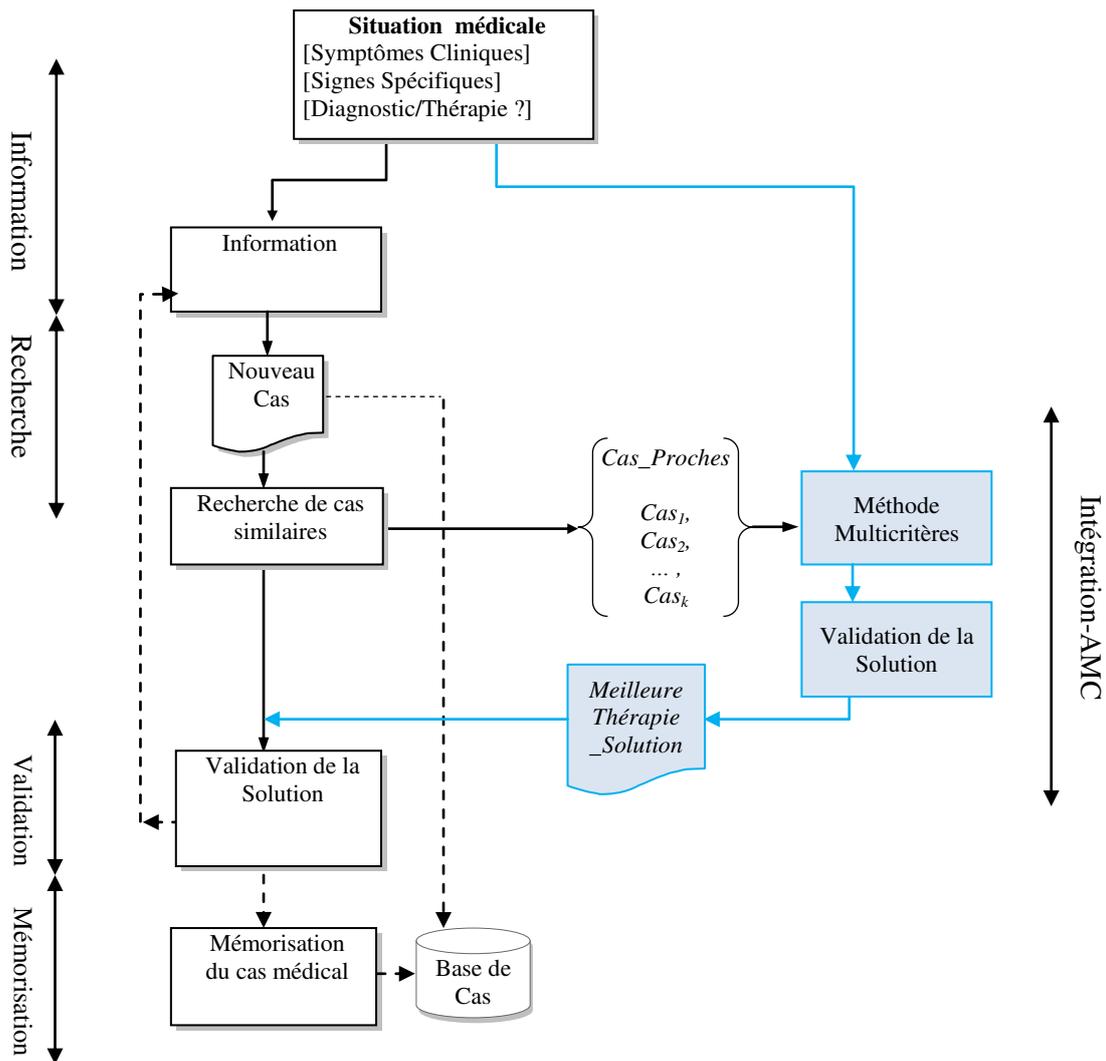


Figure 4.11. Le modèle d'aide à la décision médicale RBC-AMC proposé.

4.4.3 Le système RBC^{AMC} proposé

Nous proposons un système interactif d'aide à la décision médicale, défini comme un processus complet qui comprend un ensemble d'éléments et de routines adéquates afin d'assurer les fonctions principales du système et aider à prendre les décisions appropriées. Une description du modèle est illustrée à la figure 4.12, qui montre schématiquement l'intégration de tous les traitements, de l'acquisition d'informations sur la situation médicale

jusqu'à l'aide à la décision finale.

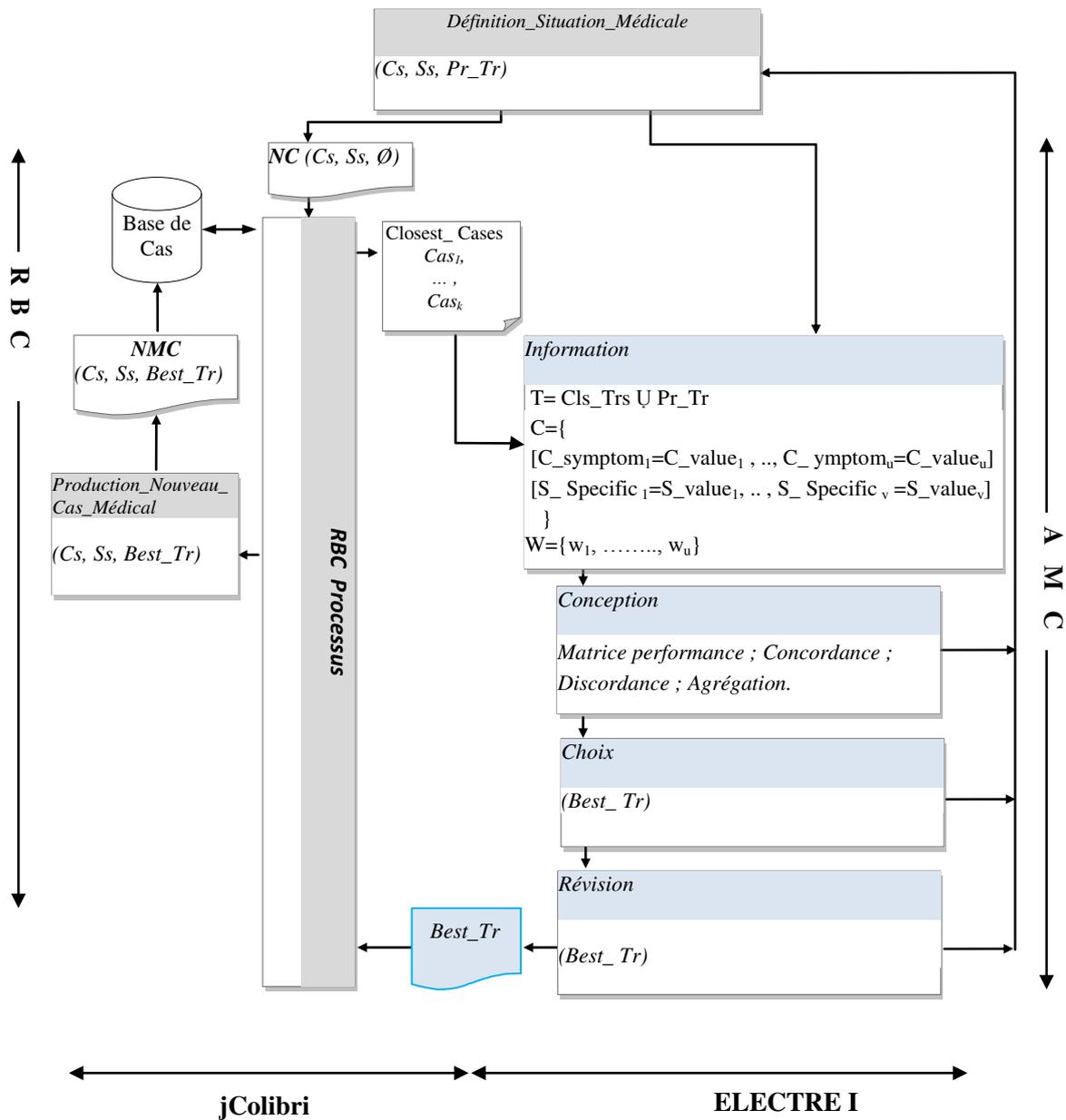


Figure 4.12. Le système RBC^{AMC} Proposé.

- Cs, Ss, Pr_Tr : Clinical_Symptoms, Signs_Specific, Proposed_Therapy
- $W, C, Cls_Trs, Best_Tr$: Weightings, Criteria, Closest Therapies, Best_Therapy
- $NC(Cs, Ss, \emptyset)$: Nouveau_Cas(Clinical_Symptoms, Signs_Specific, Therapy= $\{\emptyset\}$)
- $NMC(Cs, Ss, Best_Tr)$: New_Medical_Case(Clinical_Symptoms, Signs_Specific, Best_Therapy)

4.4.3.1 Définition de la situation médicale

Cette étape est initiée par le médecin. C'est la compréhension et la perception de la situation (diagnostic). En effet, les données recueillies par le médecin sont réparties entre ces 2 processus, le RBC et AMC. Du côté RBC, elle prépare la phase élaboration, et de l'autre (AMC), elle permettra d'entamer l'étape information.

4.4.3.2 Le RBC

(a) Elaboration

Cette opération est étroitement liée au domaine d'utilisation considéré. Compte tenu de notre contexte médical (choix de la méthode de contraception), un cas médical sera créé. Il aura la structure suivante :

Table 4.10. Structure du cas médical orthopédique.

Medical_Case
<hr/>
Clinical_symptom
End_Clinical_symptom
Specific_Signs
[Wife's age=value ₁] [Wife's education=value ₂] [Husband's education=value ₃]
[Number of children ever born=value ₄] [Wife's religion=value ₅] [Wife's now working?=value ₆]
[Husband's occupation=value ₇] [Standard-of-living index=value ₈] [Media exposure =value ₉]
End_Specific_Signs
Therapy=" "
End_Therapy
End_Medical_Case

(b) Remémoration

Ce processus a une tâche principale : *la recherche de cas similaires*. Nous utilisons la méthode *k-nn* pour la simplicité de sa mise en œuvre. Le processus sélectionnera les cas les plus similaires, à partir de la base de cas et déduira les thérapies qui ont été considérées pour ces derniers. Ces cas sont initialement envoyés au processus AMC pour les traiter et proposer la solution (Best_tr). À cette étape, le médecin :

- attribue la valeur *k* pour la méthode *k-nn* ;
- lance la procédure de traitement en l'occurrence le pseudo algorithme Recherche_Cas_Similaires, ci-dessous.

Pseudo Algorithme : Recherche_Cas_Similaires

```
1: Input : Cls_Tr ← ∅
2: Elaboration.NC (Cs, Ss, ∅)
3: Initialization k
4: Rememoration (NC, Closest_Cases) using k-nn
5: If Closest_Cases ≠ ∅ then
  For each Current_Case in Closest_Cases
  For i=1 to n
    {Cls_Tr ← Cls_Tr ∪ Current_Case (Case, therapyi)}
  Endfor
Endfor
Else
```

```

    Cls_Tr← ∅
  Endif
6: Call AMC_Process (Cls_Tr)
7: Adaptation (Best_Tr)
8: Revision (Best_Tr)
9: Memorization(Best_Tr, result)
10: If result = "yes" then
    NMC(Cs, Ss, Best_Tr)= PNM (Cs, Ss, Best_Tr)
    Else
    NMC(∅, ∅, ∅)
  Endif
11: Output : NMC(Cs, Ss, Best_Tr)

```

Cls_Tr : Closest Therapies
 PNM : Production_New_Medical_Case
 NMC : New_Medical_Case
 NC : New_Cas

(c) Adaptation.

A ce stade du processus et selon le principe adopté, cette opération est simplifiée par l'adoption pure et simple de la meilleure solution (*Best_Tr*) trouvée, ce qui évitera au système de lancer une opération de vérification de degré de similitude et autres traitements. De ce fait, cette opération se trouve être juste un passage simple à l'opération suivante.

(d) Révision

Aucune révision n'est nécessaire et la solution proposée devient une solution au problème actuel, à valider par le médecin. Cette solution peut alors être testée et évaluée. Si elle ne convient pas, le médecin peut alors l'invalidier et revoir son problème médical de nouveau. Il doit alors réviser le problème médical (cas source) ayant servi de base pour la résolution du problème et relancer une nouvelle expérimentation. Si la solution est d'un intérêt, il peut alors la valider définitivement avec le cas, et l'enregistrer pour enrichir la base de cas.

(e) Mémorisation

La solution pour le nouveau problème est trouvée et validée : Une nouvelle expérience est faite, elle est stockée dans la base de cas pour enrichir la base de cas et augmenter la capacité à résoudre d'autres situations ultérieurement.

4.4.3.3 Le raisonneur AMC

Le raisonneur AMC opère selon le principe de l'analyse multicritères, en procédant par les étapes clés du processus, à savoir : Information, conception, choix et révision.

(a) Information

Le médecin vérifiera et validera les symptômes cliniques et les signes spécifiques (Cs, Ss). Ces données contribuent à définir le problème multicritères médical (PMM) comme suit :

Définition du problème multicritères médical.

C'est l'examen des thérapies possibles pour résoudre le problème, les différents critères qui sont tous déterminants et les pondérations. Par conséquent, notre problème multicritères sera défini comme suit :

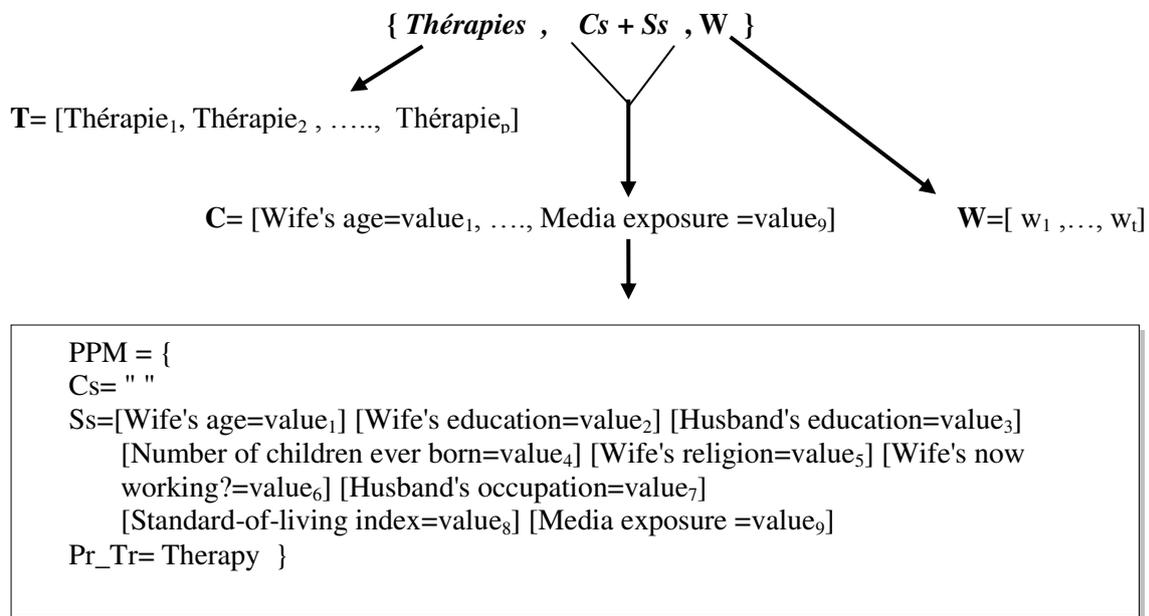


Figure 4.13. Structure du Problème Multicritères Médical.

En outre, cette procédure permet au médecin de poursuivre la définition de son problème multicritères médical par :

- la proposition d'une thérapie possible et appropriée (Pr_Tr) ;
- l'identification des critères d'évaluation des thérapies (dans notre cas : les descripteurs) ;
- l'affectation de pondérations aux critères (dans notre cas, ils sont tous à 1, donc équivalents) ;
- l'évaluation de chaque thérapie selon les critères définis et la proposition d'une solution.

(b) Conception

Au début de cette étape, un ensemble de cas) les plus proches ($Closest_Cases$, est reçu du processus RBC pour les joindre à la thérapie proposée (Pr_Tr) qui pourrait être proposée par le décideur (médecin) lors de la définition de la situation médicale décrite ci-dessus (figure 4.1).

Une fois le problème multicritères médical défini, le médecin décidera de la méthode multicritères convenable à utiliser. Pour notre étude, Electre I est appliquée. Elle permet de résoudre une problématique décisionnelle de type α , en identifiant le sous-ensemble (aussi petit que possible) de solutions offrant la meilleure thérapie possible ($Best_Tr$). Cette étape est

initiiée par le pseudo algorithme suivant :

Pseudo Algorithme : AMC

```
1: Input : Cls_TrS
2: If Pr_TrS ≠ ∅ Then Cls_TrS ← Cls_TrS ∪ Pr_TrS EndIf
3: Define_Criteria (Criteria)
4: Define_Weights (Weights)
5: Electre_I (Cls_TrS, Criteria, Weights)
6: Return to CBR_Process
```

La méthode Electre I. Cette méthode proposée par Roy [Roy, 85], permet de résoudre les problèmes multicritères de choix. Elle permet d'identifier le sous-ensemble d'actions offrant le meilleur compromis possible. Souvent utilisée dans l'identification d'un sous-ensemble de solutions le plus restreint possible, sur la base des critères considérés. Dans le cas de cette méthode, on définit de vrai-critères.

- *Principe de la méthode :* On considère un ensemble A de m actions, qui représentent l'objet de la décision, dont le but est d'identifier un sous-ensemble d'actions offrant un meilleur compromis parmi l'ensemble de départ. On définit pour chaque critère une fonction d'évaluation j, g (où $j=1$ à n , n est le nombre de critères), pour chaque critère, on évalue un poids k_j qui augmente avec l'importance du critère. Du fait des techniques de calcul utilisées dans cette méthode, il faut transformer toutes les performances des actions en notes. Celles-ci varieront sur des échelles dont la longueur évoluera de la même façon que les poids accordés aux critères. C'est la *matrice de performance*.
- *Concordance/Discordance.* La matrice de performance établie, le calcul de l'indice de concordance est fait en considérant chaque critère comme un critère vrai. L'indice de discordance sera établi en mesurant, pour chaque critère dans chaque couple d'actions, l'éventuelle différence discordante entre les deux actions, en ne retenant que la plus grande pour ce couple, et en la divisant par la plus grande longueur d'échelle. Cela garantit un indice de discordance compris entre 0 et 1.

En plaçant chaque action à la fois en ligne et en colonne, on établit les matrices de concordance et de discordance, dont la diagonale ne présente aucune valeur. Il faut alors définir un seuil de concordance et un seuil de discordance. Ces seuils permettront de réaliser les tests de concordance et de discordance. Le premier indique une valeur minimale à dépasser, le second une valeur maximale à ne pas outrepasser.

- *Seuil de concordance.* Le seuil de concordance pour deux actions a et b est noté par $C(a,b)$, compris entre 1 et 0, il mesure la pertinence de l'assertion « a surclasse b », comme suit :

$$C(a,b) = \frac{\sum_{\forall j: g_j(a) \geq g_j(b)} k_j}{K} \text{ avec } K = \sum_{j=1}^n k_j \quad [1]$$

- *Seuil de discordance.* Le seuil de discordance $D(a,b)$ est défini par :

$$D(a,b) = 0 \quad \text{si } \forall j, g_j(a) \geq g_j(b) \quad [2]$$

Sinon

$$D(a,b) = \frac{1}{\delta} \max_j [g_j(b) - g_j(a)] \quad [3]$$

Avec δ est la différence maximale entre le même critère pour deux actions données.

- *Le surclassement.* La relation de sur-classement pour Electre I est construite par la comparaison des seuils de concordance et de discordance à des seuils limites de concordance C et de discordance D , selon la relation suivante :

$$a S b \Leftrightarrow C(a,b) \geq \hat{c} \text{ et } D(a,b) \leq \hat{d}.$$

- *Le graphe de sur-classement.* Il visualise la relation de sur-classement pour l'ensemble des couples des actions. La théorie des graphes est ici utilisée pour représenter les relations de sur-classement.

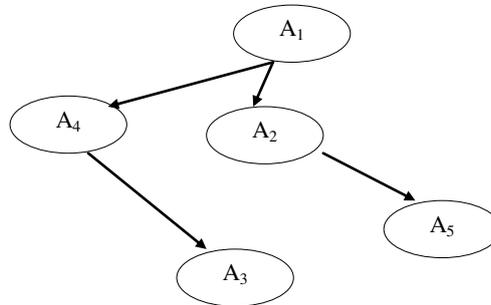


Figure 4.14. Exemple d'un graphe de sur-classement.

- Organigramme de la méthode Electre I (le surclassement).

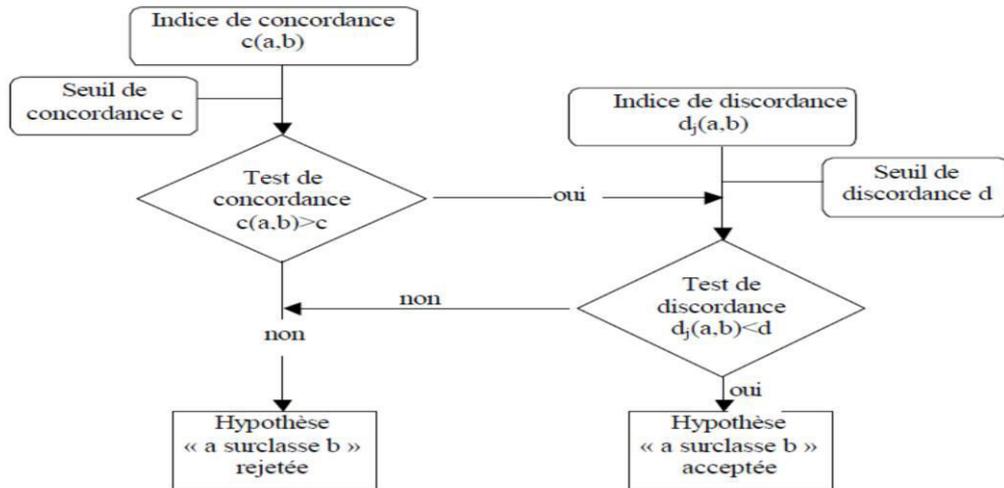


Figure 4.15. Organigramme du surclassent [Maystre et al., 94].

Pseudo Algorithme : AMC

```

1: Input : Performance_Matrix
2: Calculate Concordance_Matrix
3: Calculate Discordance_Matrix
4: Set concordance_threshold_C, discordance_threshold_D
5: Result=Calculate Outranking_actions
6: IF Resultat_Satisfaction Then Solution=Result
   Else Revenir à 4
   Endif
7: Output : Solution
  
```

(c) Choix

Le système sélectionne les thérapies possibles et les propose au médecin. Ensuite, celui-ci effectue une évaluation selon son degré de satisfaction et il décidera de prendre en compte ce que lui a été proposé et de valider la solution qui sera considérée pour le nouveau cas.

(d) Révision

Pour l'opération de révision, le décideur (médecin) n'a pas besoin de données factuelles mais il a simplement besoin d'une évaluation médicale de la solution proposée. Donc, le médecin peut réexaminer la situation s'il estime que la thérapie proposée ne le satisfait pas, en revenant à une étape précédente (information, conception) du processus AMC, pour d'éventuelles reformulations qu'il juge nécessaires pour redéfinir son problème médical.

4.4.4 Mise en œuvre

L'approche proposée a été appliquée aux données sur la prévalence de l'utilisation des contraceptifs chez les femmes indonésiennes en 1987. Nous avons réalisé une plateforme développée dans le même environnement de développement de l'expérimentation précédente. Sauf que nous utilisons la plate-forme JCOLIBRI pour construire la base de cas et toutes les opérations relatives au CBR et AMC (un ensemble de routines) pour la partie Analyse multicritères.

- *Frame_Work AMC.*

AMC est un ensemble de procédures qui assurent la réception des cas envoyés par RBC, ensuite le paramétrage des procédures propres au problème multicritères. Ces procédures sont initiées par le médecin. On alors : *Receive_Cases* ; *Information_Step* ; *Electre_I* et *Find_Best_Solution*.

- *Les principaux acteurs de la plateforme.*

Pour simplifier les tâches, nous avons considéré 1 principal acteur, comme dans la précédente expérimentation.

- *Le médecin (décideur).* Il est l'utilisateur principal de la plate forme. Il l'utilise pour l'aider à faire de l'aide à la décision médicale.

Afin de montrer les principales taches de l'aide à la décision nous présentons quelques diagrammes UML sur lesquels nous nous sommes basés pour développer notre plateforme.

- *Les diagrammes UML*

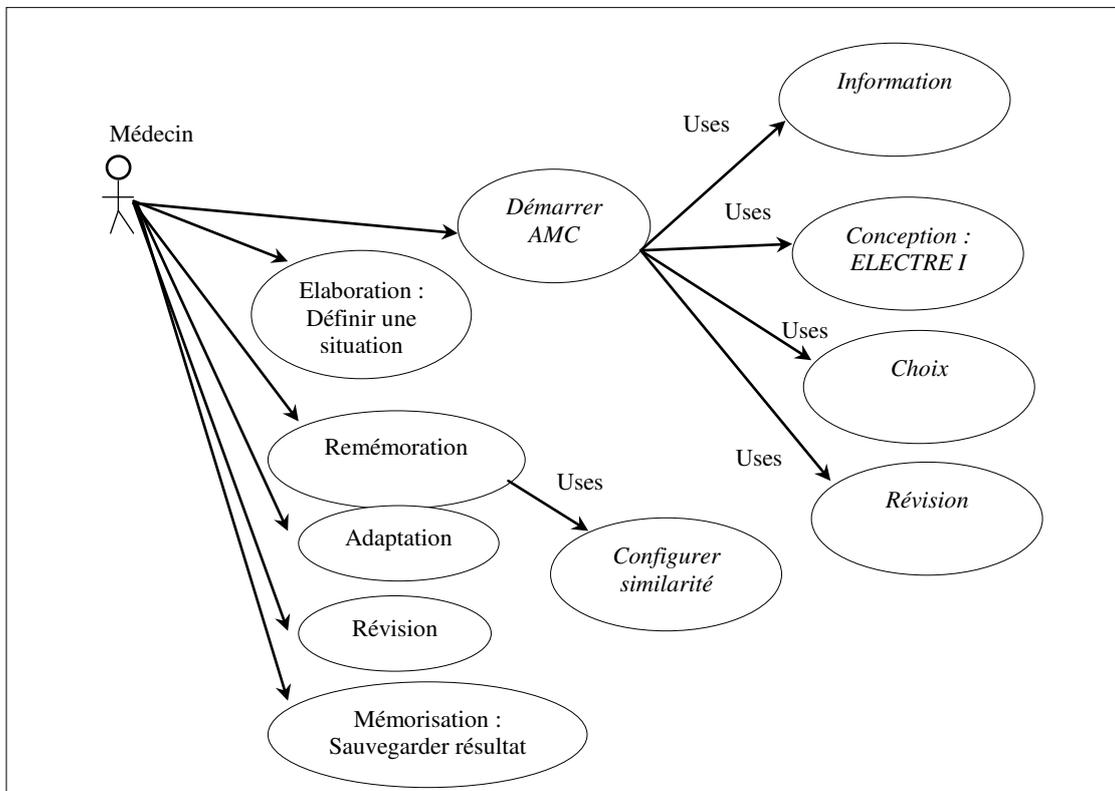


Figure 4.16. Diagramme des cas d'utilisation (AMC).

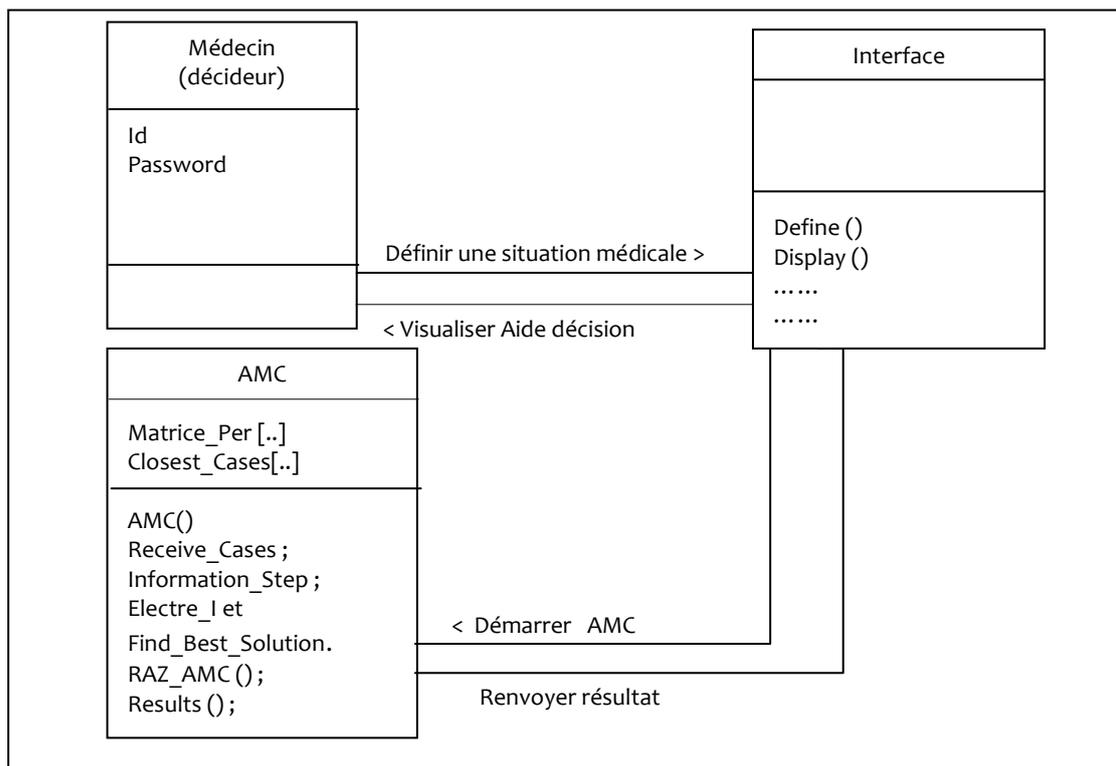


Figure 4.17. Le diagramme de classes (AMC).

4.4.4.1 Expérimentation

Pour les besoins de notre expérimentation, nous avons transformé l'échantillon «Contraceptive Method Choice Data Set sample» en une base de cas nommée Ω_N . Elle contient n cas ω_i , $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ où chaque cas est décrit par l'ensemble X_1, X_2, \dots, X_9 , appelés attributs descriptifs. Ensuite, nous avons associé un attribut cible Y correspondant à la méthode de contraception utilisée, qui prend valeurs dans l'ensemble $Y = \{\text{"NUM"}, \text{"LTM"}, \text{"STM"}\}$.¹⁰

Après la construction de la base de cas Ω_N , nous l'avons partitionné selon la méthode de contraception utilisée, en une base d'apprentissage Ω_A (60% de Ω_N), une base test Ω_T (40% de Ω_N) et des bases partielles, $\Omega_0, \Omega_1, \Omega_2$, (table 4.11).

Table 4.11. Les bases partielles $\Omega_A, \Omega_T, \Omega_0, \Omega_1, \Omega_2$

Ω_A Base Apprent. 60%	Ω_T Base Test 40%	Ω_0 "NUM"	Ω_1 "LTM"	Ω_2 "STM"
884	589	1214	106	153

¹⁰ NUM="No-Use Method", LTM= "Long Term Method" et STM= "Short Term Method".

Pour évaluer l'efficacité de l'approche proposée, des expériences ont été réalisées selon deux techniques ou nous avons utilisé la structure conditionnelle (4) pour vérifier la concordance (diagnostic similaire) des résultats entre chaque cas présenté au système provenant de base test du médecin Ω_P ou de la base test interne au système Ω_T . Chaque cas présenté au système est comparé aux cas de la base d'apprentissage Ω_A pour vérifier la concordance. Enfin, nous calculons les taux (%) de concordance et de discordance. Ces taux représentent le nombre de cas correctement identifiés aux bonnes méthodes de contraception utilisées ("NUM", "LTM" ou "STM") ou également identifiés à tort dans la base de cas Ω_A .

$$\forall \omega_i \in \Omega_P \text{ ou } \omega_i \in \Omega_T \left\{ \begin{array}{l} \text{Si } Y(X(\omega_i)) = Y(X(\omega_j)) \text{ Alors "Concordance"} \\ \text{Sinon "Discordance"} \end{array} \right. \quad [4]$$

et
 $\forall \omega_j \in \Omega_A$

1st technique : Les cas sont introduits par le médecin (Ω_P) et vérifiés sur la base d'apprentissage Ω_A

Nous avons introduit des valeurs pour définir 12 cas qui sont supposés être "NUM", 12 cas avec l'hypothèse "LTM" et 12 cas avec l'hypothèse "STM". Chaque cas présenté par le médecin est comparé aux cas de la base d'apprentissage Ω_A , pour avoir un résultat (en l'occurrence la thérapie proposée).

2nd technique : les cas sont sélectionnés depuis Ω_T et vérifiés sur Ω_A

Nous avons considéré 12 cas pris aléatoirement à partir de la base test Ω_T sans aucune hypothèse sur le diagnostic. Une comparaison est faite entre chaque cas venant de Ω_T avec les cas de la base d'apprentissage Ω_A .

4.4.4.2 Evaluation

Les résultats de la table 4.12 montrent que le taux de concordance, relatif à la méthode de contraception, est supérieur à la moyenne, ce qui démontre que notre approche reconnaît bien la méthode de contraception utilisée. On peut remarquer ce bon résultat à travers, par ex., la thérapie "NUM", lorsque $k \geq 7$ et aussi avec "STM" le résultat est de 75% de bonne concordance. Nous notons également que le taux de concordance est relativement élevé (> 58%) pour notre approche.

Aussi, par rapport à la moyenne, ces résultats montrent que l'approche adoptée identifie avec un pourcentage élevé, un cas qui a "NUM", "LTM" ou "STM", comme déclaré dans la base test Ω_T ou Ω_P .

De même, pour la discordance, qui indique dans quelle mesure l'approche identifie à tort la méthode contraceptive utilisée ?. Dans la table 4.12, nous notons un résultat inférieur ou égal à 40%.

Autre constatation : selon les deux techniques, les résultats indiquent avec quelle précision notre approche n'identifie pas correctement la méthode de contraception utilisée ou à quel taux le test de reconnaissance de la méthode de contraception est faux. Notre approche donne un résultat inférieur à 40%, sur les valeurs de test globales, ce qui représente un pourcentage

relativement intéressant.

Table 4.12. Résultats de l'expérimentation.

Taille de la base Test (Cas)	Méthode contraceptive utilisée ¹¹ .	Taille de Ω_A (Cas)	Concordance %				Discordance ¹² %
			k=4	k=7	k=12	k=15	
1 ^{re} technique							
12	"NUM"	884	66	75	58	58	≈ 36
	"LTM"		50	66	66	66	≈ 38
	"STM"		58	58	75	75	≈ 34
2 ^{me} technique							
12	"NUM"	884	58	66	75	66	≈ 34
	"LTM"		66	58	66	66	≈ 36
	"STM"		58	58	66	58	≈ 40

4.5 Conclusion

Nous avons essayé de présenter une vision globale du déploiement du RBC avec d'autres méthodologies. Les modèles hybrides sont effectivement une tendance vers la résolution de certaines lacunes du RBC. Pour cela, nous avons expérimenté deux approches l'une en utilisant une méthode de fouille de données qui a montré quelques résultats encourageants. Une autre approche a été testée en utilisant l'aspect AMC, qui elle aussi a montré des résultats acceptables. Cependant, lors de notre étude nous avons testé l'aspect théorique de l'intégration comme principe de base avec un teste dans un contexte médical. Néanmoins, l'une ou l'autre intégration, doivent être prise en charge par le développement de tout un environnement cohérent et complet qui prendra en considération tout le processus d'aide à la décision avec l'intégration du décideur qui restera l'acteur clé et qui pourra réellement valider l'approche et ne restera pas au stade des résultats numériques.

¹¹ selon la base de test.

¹² sur les valeurs de test globales de k.

Conclusion Générale & Perspectives

Conclusion Générale & Perspectives

Le domaine médical est très riche en données. Leur manipulation par de simples procédures devient très fastidieuse. De ce fait, beaucoup de travaux de recherche utilisant ces stocks de données médicales ont produit des théories et des prototypes de systèmes d'aide à la décision médicale. Ce qui a encore fait évoluer l'aide à la décision dans ce domaine, c'est l'évolution des techniques informatiques et les systèmes de surveillance et de supervision, qui ont offert de nouvelles vues aux praticiens de la santé. Toutefois, les problèmes liés au diagnostic des maladies doivent être encore plus approfondis par des techniques de modélisation et de traitement de données et des connaissances dans un but de reconnaissance des phénomènes pathologiques.

Très tôt, les SADM ont été utilisés dans le domaine médical. De par leur capacité de stockage, de gestion, d'analyse, de modélisation et d'affichage de données, ils se présentaient alors comme les outils les plus adéquats pour appréhender les problèmes de décision médicale. Néanmoins, les techniques informatiques liées au domaine médical souffrent encore de plusieurs lacunes dans le domaine de l'aide à la décision, ce qui a fait que beaucoup de SADM sont restés au stade de prototype. La fouille de données a été aussi très utilisée pour comprendre les phénomènes pathologiques et proposer des solutions en aide à la décision. Récemment l'utilisation de l'analyse multicritères a contribué à faire des avancées dans ce domaine.

Récemment et dans une seconde phase de recherche dans ce domaine, l'intégration de méthodologies de raisonnement a fait son entrée, néanmoins cette intégration a montré aussi certaines limites car les problèmes de décision liés au domaine médical sont complexes et de nature multicritères. Le traitement de ces problèmes requiert d'abord une bonne définition du problème médical. De plus, l'analyse et la proposition de solution pour la décision dans ces situations rendent le cas médical en lui-même une problématique.

(I) LES LIMITES

La mise en place d'outils de fouille de données repose fondamentalement sur le volume des données mais aussi sur la qualité des attributs qui sont impliqués dans les opérations effectuées. Aussi, vu le caractère interdisciplinaire de notre travail de recherche (traitement des connaissances ; aide à la décision), il est important d'avoir des données qui vont bien ces deux sens. Cependant, tout au long de notre travail, nous avons été confrontés à de nombreux freins. Il s'agit entre autres, de la disponibilité des données expérimentales.

Sur le volet aide à la décision, l'intégration a eu un apport confirmé et considérable à l'aide à la décision au vu du nombre important de travaux d'intégration RBC, AMC, FDD. Cependant ces travaux présentent plusieurs limites, se rapportant à l'aide à la décision par RBC en général et aux autres modes de raisonnement aussi. Ces limites qui les empêchent ou limitent leurs diffusions au-delà du contexte académique :

De manière générale, nous notons :

- L'absence de méthodologies concernant les modes d'intégration qui n'est pas encore suffisamment confirmé dans le sens où il n'existe pas de critères de choix de la méthode à intégrer dans un domaine ou un problème donné.
- Nécessité pour le décideur, en l'occurrence le médecin, d'avoir des connaissances approfondies dans le domaine du traitement des connaissances pour ne pas dire en FDD.
- Le raisonnement classique du RBC ne permet pas de supporter facilement le raisonnement clinique, ni même par un raisonnement simple à base de règles, du fait qu'une règle est déjà écrite pour ne donner que sa conséquence.

De manière spécifique, nous notons :

- L'évaluation des solutions déduites n'est pas faite de manière uniforme entre les différentes méthodologies. On trouve par exemple que l'une (FDD) évalue des motifs extraits et l'autre (AMC) évalue des actions (solutions).
- L'évaluation de la situation médicale est faite différemment en FDD et en AMC. La FDD à travers les symptômes et l'AMC à travers les critères.
- De là, se pose la question relative aux formalismes classiques de représentation de données est ce qu'ils sont adéquats. Ce qui pose aussi la problématique de la modélisation des situations médicales en général.
- Ces préférences pourront guider vers une solution optimale ou aider à élaborer une procédure de classement/groupement de solutions. Ainsi, la question comment tenir compte des préférences du décideur est posée, du moment que la structure classique du cas ne permet pas de définir explicitement cette notion.

Sur le volet FDD, le point suivant mérite une attention :

- La considération des points de vue des acteurs de la décision est un aspect qui n'est pas pris en charge par la fouille de données.

Sur le volet AMC, ce point aussi doit attirer l'attention :

- Par rapport à une situation médicale, comment définir une procédure qui permet ce sous-ensemble aussi restreint que possible, c.à.d., la "meilleure" solution ou solution optimale.

(II) CONTRIBUTIONS

Nos travaux de recherche se placent dans le cadre de la théorie de la décision, de l'extraction des connaissances avec comme cadre pratique la maladie et la production de diagnostic. Dans notre travail de thèse, notre effort a porté sur la proposition de solutions conceptuelles, méthodologiques et informatiques afin de pallier certaines limites, et par voie

de conséquence, promouvoir les SADM par une nouvelle approche d'intégration pour l'aide à la décision médicale. Notre choix portait sur la mise en place d'un environnement décisionnel sur la base d'un processus d'extraction des connaissances, d'analyse et de déduction pour les praticiens de la santé afin de répondre à des questions découlant de situations médicales de consultation de malades.

Nous avons proposé et testé deux stratégies d'intégration RBC-FDD et RBC-AMC. Plus précisément, la stratégie est centrée sur la recherche de solutions dans un espace de recherche assez réduit ce qui nous emmènera à restreindre le nombre de situations médicales similaires à la situation en cours de traitement par le praticien de la santé. Ainsi, l'objectif de départ était de fournir aux praticiens qui interviennent dans le domaine des consultations médicales :

- Des vues fonctionnelles : description des caractéristiques des patients consultés.
- Des vues opérationnelles : traitement liée à une pathologie en vue d'une aide à la décision médicale.

Une solution possible consiste en l'intégration du RBC à la FDD ou à l'AMC, et laisser l'analyse et l'appréciation au médecin (décideur)

Le modèle décisionnel que nous avons conçu utilise la théorie de la décision à base d'un raisonnement médical. Ce raisonnement considère le patient comme étant le sujet du problème médical analysé, et pour le résoudre nous avons suivi l'approche rationnelle comme le point de départ de notre réflexion, avec prise en compte du raisonnement clinique considéré dans le milieu médical. Les dimensions de la situation médicale ont été identifiées comme étant les symptômes cliniques et les signes spécifiques du patient. Ce modèle décisionnel a été évalué et validé par la mise en place d'outils intégrés de traitements adéquats correspondants aux attentes des praticiens de la santé qui sont plus intéressés par les résultats (aide à la décision).

L'étude bibliographique nous a permis de constater que les méthodes de FDD et de l'AMC, sont diverses, et ont reçu beaucoup d'attention de par les travaux en aide à la décision. Cependant, sont elles toutes bien adaptées au domaine médical ?, le constat le plus naturel est que ces méthodes sont sujettes à des "limitations informatiques" de par l'existence du phénomène combinatoire et autres limitations propres aux méthodes. Pour faciliter l'intégration des approches de raisonnement, nous avons opté pour une stratégie opérant en deux étapes :

1. Réduction de l'espace de recherche.
2. Générer la solution au problème dans cet espace réduit.

Cette stratégie permet de réduire considérablement les traitements et éviter d'avoir recours à un nombre assez conséquent de solutions qui nuisent à la qualité du résultat obtenu (aide à la décision).

Cependant nous devons réaliser les opérations annexes à cette stratégie, à savoir :

- La définition des structures de données adéquates.
- La mise en place d'une base de cas.

- limiter le champ d'application au seul domaine médical, notre terrain d'investigation fixé dès le début de notre étude, avec une situation médicale assez réduite en descripteurs.

Les différentes expérimentations ont été réalisées en utilisant des algorithmes (k-means, k-ppv) qui nous ont permis d'aboutir à des résultats. Cependant, nous avons pu identifier les limites des méthodes de groupement à savoir le choix des attributs qui rentrent dans l'opération, ceci impactera les résultats attendus.

Dans ce sens et en se basant sur l'approche d'intégration que nous avons adoptée. Nous avons pu tester les modèles décisionnels intégrés conçus à cet effet, et obtenu les résultats suivants :

- Par rapport à l'approche RBC-AMC (voir 4.342), les résultats obtenus démontrent bien que le modèle décisionnel suivi identifie avec un bon pourcentage une bonne thérapie (méthode de contraception utilisée) telle que déclarée dans les bases de test. Ce qui démontre que notre approche reconnaît bien la thérapie.
- De même que pour l'approche RBC-FDD. Les résultats présentés (voir 4452) ont montré que le taux de bonne reconnaissance du diagnostic est relativement au-dessus de la moyenne, ce qui indique que notre approche tend à faire un bon diagnostic pour les cas soumis à partir des bases de test.

Ceci montre que les modèles décisionnels adoptés, ont tendance à donner des résultats encourageants. Ce qui peut nous mener à envisager les perspectives suivantes

(III) PERSPECTIVES

Au terme de nos travaux de recherche et par rapport au domaine que nous avons investis, et des résultats obtenus, les perspectives que nous dessinons sont liées d'une part aux limites auxquelles nous avons été confrontés et aux résultats encourageants obtenus.

Cependant, les améliorations qui pourraient être apportées à notre approche vont être guidées par les lignes directrices ci-dessous :

De manière générale.

- Utiliser une cohorte qui peut se trouver au niveau des infrastructures sanitaires nationales.
- Développer environnements personnalisés, ainsi, des interfaces adaptées (structure et contenu) à chaque besoin (par utilisateur/et par spécialité) pourraient être dégagées, bien que ce point n'as pas relation directe avec l'aide à la décision, mais ce point peut être envisagée à long terme.

La FDD peut être valorisée à travers les points suivants :

- Adapter la méthode de FDD pour réduire les motifs en se concentrant sur les attributs les plus pertinents (une approche récursive pourrait permettre d'éliminer les attributs "non potentiels" au cours de l'exécution de la méthode de groupement) ;
- Accessoirement une intégration d'un outil de visualisation graphique des résultats, ne pourra qu'être une amélioration dans leurs présentations.

- Tester les potentialités de la FDD en l'étendant à d'autres méthodes qui peuvent convenir à l'aide à la décision médicale, telle que la recherche de règles d'association.
- Faire une étude assez approfondie sur les méthodes de sélection des attributs pertinents à faire participer à la FDD, ce qui permettra de renforcer la pertinence des résultats attendus.
- Néanmoins, un nouveau problème apparaît : comment choisir la méthode à appliquer dans un problème médical donné alors que le praticien de la santé n'est pas tout à fait bien formé aux techniques du traitement de l'information.

Pour l'AMC, elle peut être valorisée à travers :

- La notion de pondération des critères qui pourra être étudiée en profondeur pour pouvoir fournir un problème multicritères médical qui reflète au mieux la situation médicale. Et nous supposons que l'aspect évaluation des critères sera même influencé par la spécialité dans le domaine médical en question.

Perspectives à court terme

La finalisation du prototype. Le prototype développé ne comporte pas toutes les solutions proposées dans le cadre de cette thèse. Notre première préoccupation dans le futur concerne l'incorporation de ces solutions pour la génération d'autres types d'actions.

Adoption d'une méthode d'évaluation des actions (solutions). Cette méthode pourra proposer par exemple une évaluation plus ou moins standard qui pourra être adoptée quelle que soit la méthode de raisonnement qui participe à l'intégration. Le développement d'une telle méthode pour assister le décideur lors du choix des paramètres à prendre en considération lors de l'évaluation de sa situation médicale. Cette méthode d'évaluation ou de pondération des différents paramètres est très importante pour la réalisation d'une bonne aide à la décision.

Perspectives à long terme

Revoir la modélisation de la situation médicale par une approche formelle. Cette structure devra être validée pour adoption en aide à la décision médicale, au lieu d'utiliser une structure classique propre à un raisonnement classique.

Faire une étude et adopter une méthode ou un principe pour le choix des attributs rentrant dans l'opération de FDD médicales.

Faire une étude sur la pondération des critères dans le domaine médical. La construction des critères et leur pondération varient bien d'un praticien à un autre. Les critères sont construits sur la base d'un ensemble de conséquences qui résultent des actions envisagées. Dans ce cas, nous envisageons d'explorer l'aspect pondération afin de construire un bon ensemble de critères bien pondérés qui peut être très utile lors de la modélisation d'un problème multicritères médical.

Références bibliographiques

[Aamodt et Plaza, 94]

Aamodt, A., and Plaza, E. *Case-based reasoning : Foundational issues, methodological variations, and system approaches*. AI communications, Vol. 7, no 1, pp. 39-59.

[Abbasi et Kashiyarndi, 06]

Abbasi, M. M., and Kashiyarndi, S. *Clinical Decision Support Systems: A discussion on different methodologies used in Health Care*. Marlaedalen University Sweden.

[Alnafie, I6]

Alnafie E. Vers une nouvelle approche pour l'élicitation des préférences dans la méthodologie multicritères d'aide à la décision. Thèse de doctorat. Université d'Oran I Ahmed Ben Bella.

[Adla, I0]

Adla, A. *Aide à la facilitation pour une prise de Décision Collective: Proposition d'un Modèle et d'un Outil*. Thèse de doctorat. Université de Toulouse, Université Toulouse III-Paul Sabatier.

[Agrawal et al, 93]

Agrawal, R., Imieliński, T., & Swami, A. *Mining association rules between sets of items in large databases*. ACM SIGMOD Record, Vol. 22, no 2, p. 207-216.

[Agrawal et Srikant, 95]

Agrawal, R., and Srikant, R. *Mining sequential patterns*. In Data Engineering. Proceedings of the Eleventh International Conference on (pp. 3-14). IEEE.

[Allayous et al., 08]

Allayous, C., Diara, J.P., Emilion, R., & Marianne-Pepin, T. *Decision trees for the severity and recurrence of acute splenic sequestration in sickle cell disease*.

[Alter, 80]

Alter, S. *Decision support systems: current practice and continuing challenges* (No. 04; HD30. 23, A5.).

[Althoff et al., 98]

Althoff, K.D., Bergmann, R., Wess, S., Manago, M., Aurio, E., Larichev, O.I., Bolotov, A., Zhuravlev, Y.I., & Gurov, S.I. *Case-Based Reasoning for Medical Decision Support Tasks: The INRECA Approach*. Artificial Intelligence in Medicine Journal, Vol. 12, No. 1, 25-41

[Aleksavska et Loskovska, 11]

Aleksavska S.L., and Loskovska, S. *Architectural and data model of clinical decision support system for managing asthma in school-aged children*. Minnesota State University Mankato, Minnesota USA.

[Angehrn et Dutta, 92]

Angehrn, A. A., Dutta, S. *Integrating case-based reasoning in multi-criteria decision support systems*. INSEAD.

[Ansoff, 84]

Ansoff, H.I, "Stratégie du développement de l'entreprise", 1er Ed, Ed d'organisation, Paris, 1984.

[Armaghan et Renaud, 12]

Armaghan, N., and Renaud, J. *An application of multi-criteria decision aids models for Case-Based Reasoning. Information Sciences.* Vol 210. 55-66.

[Araujo de Castro et al., 09]

Araujo de Castro, A. K., Pinheiro, P. R., Dantas Pinheiro, M. C. *Towards the Neuropsychological Diagnosis of Alzheimer's Disease: A Hybrid Model in Decision Making.* WSKS, CCIS 49, pp. 522–531.

[Assali, et al., 09]

Assali, A. A., Lenne, D., Debray, B., & Bouchet, S. COBRA: Une plate-forme de RàPC basée sur des ontologies. In IC 2009 (pp. 277-288).

[Austin et al., 96]

Austin, T., Iliffe, S., Leaning, M., & Modell, M. *A prototype computer decision support system for the management of asthma.* Journal of Medical Systems, Vol. 20, no. 1, pp. 45-55.

[Costa, 96]

Costa, C. A. B. E. *Les problématiques de l'aide à la décision : Vers l'enrichissement de la trilogie choix-tri-rangement.* Recherche opérationnelle/Operations Research, 30(2) :191–216.

[Barigou, et al., 12]

Barigou, F., Atmani, B., & Beldjilali, B. *Using a Cellular Automaton to Extract Medical Information from Clinical Reports.* Journal of Information Processing Systems, 8(1).

[Balakrishnan et al., 12]

Balakrishnan, V., Shakouri, M.R., & Hoodeh, H. *Integrating association rules and case-based reasoning to predict retinopathy.* Maejo International Journal of Science and Technology, Vol. 6, No 3, pp. 334-343.

[Bates et al., 03]

Bates, D.W., Kuperman, G. J., Wang, S., Gandhi, T., Kittler, A., Volk, L., Spurr, C., Khorasani, R., Tanasijevic, M., & Middleton, B. *Ten Commandments for Effective Clinical Decision Support : Making the Practice of Evidence-based Medicine a Reality.* Journal of the American Medical Informatics Association : JAMIA, 10(6) :523–530.
URL <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC264429/>

[Bareiss et al., 88]

Bareiss, E., Porter, E., Bruce, W., & Wie, C.C. *Protos: An exemplar-based learning apprentice.* International Journal of Man-Machine Studies, Vol. 29, no 5, pp. 549-561.

[Bellazzi et Zupan, 01]

Bellazzi, R. and Zupan, B. Intelligent data analysis—special issue. Methods Archive, 40(5), 362-364.

[Begum et al., 09]

Begum, S., Ahmed, M.U., Funk, P., Xiong, N., & Von Schéele, B. *A case-based decision support system for individual stress diagnosis using fuzzy similarity matching.* Computational Intelligence. Vol. 25. No 3, 180-195.

[Begum et al., 11]

Begum, S., Ahmed, M., Funk, P., Xiong, N., & Folke, M. *Case-Based Reasoning Systems in the Health Sciences: A Survey of Recent Trends and Developments.* IEEE Transactions on systems, man, and cybernetics part c: applications and reviews, Vol. 41 No 4, pp. 421-434

[Belacel, 99]

Belacel, N. *Méthodes de classification multicritère méthodologie et applications à l'aide au diagnostic médical*. Thèse de doctorat. Université Libre de Bruxelles. Belgique.

[Belacel, 03]

Belacel, N. *Multicriteria Decision Analysis Methodology for Medical Diagnosis*, Aid. 2003.

[Ben mena, 00]

Benmena, S. *Introduction aux méthodes multicritères d'aide à la décision*. Biotechnologie, Agronomie, Société et Environnement, Vol. 4, no 2, pp. 83-93

[Berner, 09]

Berner, E.S. *Clinical decision support systems: State of the Art*. AHRQ Publication No. 09-0069-EF. Rockville, Maryland: Agency for Healthcare Research and Quality.

[Beleites et al., I6]

Beleites, C., Salzer, R., & Sergo, V. "Validation of soft classification models using partial class memberships: An extended concept of sensitivity & co. applied to grading of astrocytoma tissues," *Chemometrics and Intelligent Laboratory Systems*, vol. 122, pp. 12-22.

[Bello-Tomás et al, 04]

Bello-Tomás, J.J., González-Calero, P.A., & Díaz-Agudo, B. *Jcolibri: An object-oriented framework for building cbr systems*. In *European Conference on Case-Based Reasoning* (pp. 32-46). Springer, Berlin, Heidelberg.

[Bellazi et Zupan, 08]

Bellazi, R., and Zupan, B. *Predictive data mining in clinical medicine : current issues and guidelines*. *International journal of medical informatics*. Vol. 77. 8197

[Bichindaritz et Marling, I0]

Bichindaritz, I., Marling, C. *Case-based reasoning in the health sciences: Foundations and research directions*. *Computational Intelligence in Healthcare 4*. Springer Berlin Heidelberg, 127-157

[Bichindaritz et Montani, 09]

Bichindaritz, I., and Montani, S. *Introduction to the special issue on case-based reasoning in the health sciences*, *Computational intelligence*, Vol. 25, issue 3, pp.161-194

[Bichindaritz et al, 03]

Bichindaritz, I., Moinpour, C., Kansu, E., Donaldson, G., Bush, N., & Sullivan, K. M. *Case based reasoning for medical decision-support in a safety critical environment*. In *AIME*, pages 314–323.

[Bichindaritz, et al., 98]

Bichindaritz, I., Kansu, E., & Sullivan, K. M. *Case-based reasoning in care-partner : Gathering evidence for evidence-based medical practice*. *j-LECT-NOTES-COMPSCI*, I488 :334–345.

[Bichindaritz, I5]

Bichindaritz, I. *Data Mining Methods for Case-Based Reasoning in Health Sciences*. In : *ICCBR (Workshops)*, pp. 184-198, 2015.

[Bouhana et al., I1]

Bouhana A., Abed M., & Chabchoub, H. *An integrated Case-Based Reasoning and AHP method for personalized itinerary search*. *Logistics*, 4th International Conference on. IEEE, 460-467

[Bouyssou, 93]

Bouyssou, D. *Décision Multicritère ou Aide Multicritère?* Newsletter Of The European Working Groupe, Series2, P2.

[Brans, 82]

Brans J.P. *Elaboration d'instruments d'aide à la décision: méthode PROMETHEE*, Colloque d'aide à la décision, Université Laval, Québec.

[Brans et al., 84]

Brans. J.P., Mareschal, B., & Vincke, Ph. *PROMETHEE : A new family of outranking methods in multicriteria analysis*. In J.P. Brans, editor, *Operational Research'84*, pages 408–421, North Holland, Elsevier Science Publishers, B.V.

[Bruland et al., 10]

Bruland, T., Aamodt, A., & Langseth, H. *Architectures Integrating Case-Based Reasoning and Bayesian Networks for Clinical Decision Support*. IFIP Advances in Information and Communication Technology, Vol. 340, 82-91, DOI: 10.1007/978-3-642-16327

[Bresson et Lieber, 00]

Bresson, B., and Lieber, J. *Raisonnement à partir de cas pour l'aide au traitement du cancer du sein*. In journées ingénierie des connaissances (pp. 189-196)..

[Buchner et al., 97]

Buchner A.G., Anand S.S., & Hughes J.G. *Data mining in manufacturing environments : Goals, techniques and applications*. Studies in Informatics and Control, 6(4) :319–328.

[Cabrera et Edye, 10]

Cabrera. MM., and Edye, E.O. *Integration of rule based expert systems and case based reasoning in an acute bacterial meningitis clinical decision support system*, arXiv preprint arXiv:1003.1493. Available: <https://arxiv.org/ftp/arxiv/papers/1003/1003.1493.pdf>

[Chakhar et al., 05]

Chakhar. S., Mousseau, V., Puscaddu, C., & Roy, B. *Decision map for spatial decision making*. In *The 9th International Computers in Urban Planning and Urban Management*. Conference (CUPUM'05), London, UK, 29 July-1 August.

[Chakhar, 06]

Chakhar, S. *Cartographie décisionnelle multicritères : formalisation et implémentation informatique*, thèse de Doctorat, Université Paris Dauphine.

[Chiang et al., 15]

Chiang,I.J., Shieh, M.J., Hsu, J.Y., & Wong, J.M. *Building a Medical Decision Support System for Colon Polyp Screening by Using Fuzzy Classification Trees*. Applied Intelligence Vol 22, Number 1, 61-75,

[Cleret et al., 01]

Cleret, M., Le Beux, P., & Le Duff, F. *Les systèmes d'aide à la décision médicale*. Les Cahiers du numérique. Vol 2. 125-154

[Coiera, 13]

Coiera, E. *Guide to Health Informatics*,3rd Edition. Hodder Arnold, 3e edition

[Das, 16]

Das, T.K. *Intelligent Techniques in Decision Making: A Survey*. Indian Journal of Science and Technology, Vol 9, N 12, 2016.

[Deepti et al., 10]

Deepti, J., Rani, A., & Rani, R. *A Framework for Medical Diagnosis using Hybrid Reasoning*. Proceedings of the International Multi Conference of Engineers and Computer Scientists, Vol. 1

[Degoulet et Fieschi, 91]

Degoulet P., et Fieschi M. *Traitement de l'information médicale, Méthodes et applications hospitalières*. Collection : Manuels Informatiques Masson – Entreprise, 320 pages.

[De Dombal et al., 72]

De Dombal, F. T., Leaper, D. J., Staniland, J. R., McCann, A. P., & Horrocks, J. C. *Computer-aided diagnosis of acute abdominal pain*. Br Med J, 2(5804), 9-13.

[De Paz et al., 09]

De Paz. F. J., Rodriguez, S., Bajo, J., & Corchado, M.J. *Case-based reasoning as a decision support system for cancer diagnosis : A case study*. International Journal of Hybrid Intelligent Systems, Vol. 6.

[Dieng-Kuntz et al., 01]

Dieng-Kuntz, R., Corby, O., Gandon, F., Giboin, A., Golebiowska, J., Matta, N., & Ribière, M. *Méthodes et outils pour la gestion des connaissances : une approche pluridisciplinaire du Knowledge management*. Dunod, 2ème édition, Paris.

[Dilou et al., 10]

Dilou, A., Rekik, M., & Pasquier, M. N. *Visualisation en Fouille de Données*.

[Djebbar-Zaidi, 13]

Djebbar-Zaidi, A. *Optimisation de la recherche d'un cas Bayésien*. Thèse de doctorat. Université Badji Mokhtar de Annaba.

[Dunham, 06]

Dunham, M. H. *Data mining: Introductory and advanced topics*. Pearson Education India.

[Eastman et al., 93]

Eastman, J.R., P.A.K. Kyem, J. Toledano, and W. Jin. GIS and decision making. Technical report, The United Nations Institute for Training and Research (UNITAR), Geneva.

[Erjaee et al., 12]

Erjaee, A., Bagherpour, M., Razeghi, S., Dehghani, S. M., Imanieh, M. H., & Haghghat. M. *A multi-criteria decision making model for treatment of Helicobacter pylori infection in children*. Hong Kong J Paediatr, Vol. 17, no. 4, pp. 237-42

[Eva, 05]

Eva, K.W. *What every teacher needs to know about clinical reasoning*. Med Educ, Vol. 39, pp.98-106.

[Fayyad et al, 96]

Fayyad, M., Piatetsky-Shapiro, G., & Smyth, P. *From data mining to knowledge discovery : An overview*. Advances in Knowledge Discovery and Data Mining, pages 1–34.

[Figueira et al., 05]

Figueira, J., Greco, S., & Ehrgott, M. *Multiple Criteria Decision Analysis : State Of The Art Surveys*, Springer's International series. 2005.

[Figueira et al., 16]

Figueira, J. R., Mousseau, V., & Roy, B. *ELECTRE methods*. In Multiple Criteria Decision Analysis (pp. 155-185). Springer, New York, NY.

[Fox et al., 77]

Fox, J., Johns, N., Rahmanzadeh, A., & Thomson, R. *PROFORMA : a general technology for clinical decision support systems*. Computer methods and programs in biomedicine, Vol. 54, no. 1, pp.59-67.

[Garg et al., 05]

Garg, A. X., Adhikari, N. K. J., McDonald, H., Rosas-Arellano, M. P., Devereaux P. J., Beyene, J. Sam, J., & Haynes, R. B. *Effects of computerized clinical decision support systems on practitioner performance and patient outcomes : a systematic review*. JAMA, 293(10) :1223–1238.

[Gierl et al., 98]

Gierl, L., Bull, M., & Schmidt, R. *Cbr in medicine, case-based reasoning technology*. In Case-Based Reasoning Technology (pp. 273-297). Springer, Berlin, Heidelberg.

[Greenes, 14]

Greenes, R. A. *Clinical Decision Support : The Road to Broad Adoption*. Academic Press, Amsterdam : Boston, 2e édition.

[Guo et al., 11]

Guo, Y., Hu, J., & Peng, Y. *Research on CBR system based on data mining*. Applied Soft Computing, Vol. 11, No 8, 5006–5014

[Han et al., 00]

Han, J., Kamber, M., & Pei, J. *Data mining: concepts and techniques (the Morgan Kaufmann Series in data management systems)*. Morgan Kaufmann.

[Hall et al., 09]

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I.H. *The WEKA data mining software: an update*. ACM SIGKDD explorations newsletter, Vol. 11 No 1, 10-18

[Hamdadou et al., 15]

Hamdadou, J., Bouamrane, K., & Naoui, O. E. K. *Méthodologie multicritères d'aide à la décision*. Document pédagogique : cours. Université Oran I.

[Hamdadou, 08]

Hamdadou, J. *Un modèle pour l'apprise de décision en aménagement du territoire : une approche multicritères et une approche de négociation*. Thèse de doctorat en informatique université oran I.

[Hardin et Chieng., 07]

Hardin, J.M., and Chieng, D.C. *Data Mining and Clinical Decision Support Systems Clinical Decision Support Systems*. Health Informatics, Section 1, 44-63, DOI: 10.1007/978-0-387-38319-4_3

[Ha et Joo, 10]

Ha, S. H., and Joo, S. H. *A Hybrid Data Mining Method for the Medical Classification of Chest Pain*. International Journal of Computer and Information Engineering, Vol. 4, No 1, 33-38

[Holtzman, 89]

Holtzman S. *Intelligent Decision Systems*. Addison Wesley, 1989.

[Huang et Tzeng, 11]

Huang, J. J., Tzeng, G. H. Multiple attribute decision making: methods and applications. CRC press.

[Huang et al., 07]

Huang, M-J., Chen, M-Y. & Lee, S-C. Integrating data mining with case-based reasoning for chronic diseases prognosis and diagnosis, *Expert Systems with Applications*, Vol. 32, pp.856–867.

[Hwang et Yoon, 81]

Hwang, C. L., and Yoon, K. *Methods for multiple attribute decision making*. (pp. 58-191). Springer, Berlin, Heidelberg.

[Jha et al., 13]

Jha, M. K., Pakhira, D., & Chakraborty, B. *Diabetes detection and care applying CBR Techniques*. International Journal of Soft Computing and Engineering, Vol. 2, Issue-6.

[Kargupta et al., 97]

Kargupta, H., Stafford, B., & Hamzaoglu, I. *Web based parallel/distributed medical data mining using software agents*. Los Alamos National Lab., NM (United States).

[Kaplan, 01]

Kaplan, B. *Evaluating informatics applications - clinical decision support systems literature review*. International Journal of Medical Informatics, Volume 64, Issue 1, Pages 15-37

[Kast, 2002]

Kast, R. *La théorie de la décision*. La Découverte, « Repères », 2002, 128 pages. ISBN : 9782707137692. URL : <https://www.cairn.info/la-theorie-de-la-decision--9782707137692.htm>

[Kassirer, 10]

Kassirer, J. P. *Teaching Clinical Reasoning: Case-based and Coached*. Academic Medicine, Vol. 85, No. 7.

[Kawamoto, 05]

Kawamoto, K., Houlihan, C. A., Balas, E. A., Lobach, D. F. *Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success*. Bmj, 330(7494), 765.

[Kantardzic, 11]

Kantardzic, M. *Data mining: concepts, models, methods and algorithms*. John Wiley & Sons.

[Keen and Scott-Morton, 78]

Keen P. et Scott-Morton M., *Decision Support Systems: an organizational perspective*, Addison-Wesley Publishing, 1978.

[Keeney et Raiffa, 93]

Keeney, R. L., and Raiffa, H. *Decisions with multiple objectives: preferences and value trade-offs*. Cambridge university press.

[Kodratoff, 96]

Kodratoff, Y. *L'extraction de connaissances à partir des données: Un nouveau sujet pour la recherche scientifique*. In : INFORSID. Congrès. pp. 3-25.

[Kodratoff, 98]

Kodratoff, Y. *Techniques et outils de l'extraction de connaissances à partir des données*, Signaux, vol. 92, pp 38–43.

[Kong et al., 08]

Kong, G., Xu, D.L., Yang, J.B. *Clinical decision support systems: a review on knowledge representation and inference under uncertainties*. International Journal of Computational Intelligence Systems, Vol.I, No. 2, 159-167

[Kumar et al., 09]

Kumar, K.A., Singh, Y., & Sanyal, S. *Hybrid approach using case-based reasoning and rule-based reasoning for domain independent clinical decision support in ICU*. Expert Systems with Applications, Vol. 36, no. 1, pp. 65-71

[Kiezun et al., 09]

Kiezun, A., Lee, I.T.A., & Shomron, N. *Evaluation of optimization techniques for variable selection in logistic regression applied to diagnosis of myocardial infarction*. Bioinformatics. vol. 3, no 7, p. 311.

[Kuo et al., 01]

Kuo, W.J., Chang, R.F., Chen, D.R., & Lee, C.C. *Data mining with decision trees for diagnosis of breast tumor in medical ultrasonic images*, Breast cancer research and treatment 66 : 51-57

[Keeney, 92]

Keeney, R.L. *Value-focused thinking : A path to creative decision*. Harvard University Press, Cambridge, USA.

[Klein, 98]

Klein, G. *Sources of Power: How people make decisions*. Cambridge, Mass: MIT Press.

[Kuilboer et al., 02]

Kuilboer, M.M., Van Wijk, M.A.M. Mossveld, M., Van der Does, E., Ponsioen, B.P., de Jongste, J.C., Overbeek, S. E., & Van der lei, J. *Feasibility Of Asthma Critic, a decision-support system for asthma and COPD Which generates patient-specific feedback on routinely recorded data in general practice*. Family practice, Vol. 19, no. 5, pp.442-447.

[Kim et Han, 01]

Kim, K.S., and Han, I. *The cluster-indexing method for case-based reasoning using selforganizing maps and learning vector quantization for bond rating cases*. In: Expert systems with Application 21, 147-156.

[Korhonen et Laakso, 86]

Korhonen, P. J., and Laakso, J. *A visual interactive method for solving the multiple criteria problem*. European Journal of Operational Research, 1986, vol. 24, no 2, p. 277-287.

[Koton, 88]

Koton, P. *Reasoning about evidence in causal explanations*, in: Proceedings of the Seventh National Conference on Artificial Intelligence, AAI Press. Menlo Park, CA, pp. 256–263.

[Ltifi et al., 10]

Ltifi, H., Ben Ayed, M., Kolski, C., & Alimi, A. M. *Démarche centrée utilisateur pour la conception de SIAD basés sur un processus d'ECD, application dans le domaine de la santé*. Journal d'Interaction Personne-Système, Vol. I, Num. 1, Art. 1

[Laaribi, 00]

Laaribi, A. *SIG et Analyse Multicritère*. Hermes Science Publications, 1er Ed, Paris.

[Lebraty, 06]

Lebraty, J. F., *Les systèmes décisionnels*. Akoka, A, Comyn-Wattiau, I. Encyclopédie de l'informatique et des systèmes d'information, Vuibert, pp.1338-1349, 2006.

[Lebraty et Guarnelli, 14]

Lebraty, J. F., et Guarnelli, J. *Décider en situation: un état de l'art*. In Actes du 19ème du Colloques de l' Association Information et Management (AIM). Le Management à l'épreuve des Systèmes d'Information: enjeux, défis et perspectives,.

[Li et Sun, 09]

Li. H., and Sun, J., *Hybridizing principles of the Electre method with case-based reasoning for data mining: Electre-CBR-I and Electre-CBR-II*. European Journal of Operational Research, Vol. 197, no. 1, pp. 214-224

[Lévine et Pomerol, 89]

Lévine P., et Pomerol, J. *Systèmes interactifs d'aide à la décision et systèmes experts*, Editions Hermès.

[Lobach et al., 07]

Lobach, D.F., Kawamoto, K., Anstrom, K.J., Russell, M.L., Woods, P., & Smith, D. *Development, deployment and usability of a point-of-care decision support system for chronic disease management using the recently approved HL7 decision support service standard*. Stud Health Technol In-form. Vol. 129. 861-5.

[Liu et Ke, 06]

Liu, D.R., Ke, C.K. *Knowledge support for problem-solving in a production process: A hybrid of knowledge discovery and case-based reasoning*. In: Expert Systems with Applications.

[Gierl et al., 98]

Gierl, L., Bull, M., & Schmidt, R. *Cbr in medicine*. In Case-Based Reasoning Technology, pages 273–298. Springer, Berlin, Heidelberg

[McKenney et Scott, 71]

McKenney, J. L., & Scott, M. M. *Management decision systems: computer-based support for decision making*. Harvard Business School Press.

[Mansoul et Atmani, 16]

Mansoul, A., and Atmani, B. *Clustering to Enhance Case-Based Reasoning*, in Modelling and Implementation of Complex Systems. Springer International Publishing, 2016, pp. 137-151.

[Mansoul et Atmani, 17]

Abdelhak, M., & Baghdad, A. *Combining Multi-Criteria Analysis with CBR for Medical Decision Support*. Journal of Information Processing Systems, 13(6).

[Maystre et al., 94]

Maystre, L.Y., J. Pictet, & Simos, J. *Méthodes multicritères Electre*. Presses Polytechniques et Universitaires Romandes, Lausanne, Suisse.

[Marling et al., 05]

Marling. C., Rissland, E., Aamodt, A. *Integrations with case-based reasoning*. The Knowledge Engineering Review, Vol. 20 No 3, 241-245

[Marling et al., 02]

Marling, C., Sqalli, M., Rissland, E., Muñoz-Avila, H., & Aha, D. *Case-Based Reasoning Integrations*. AI magazine, Vol. 23, No 1, p. 69.

[Marling et al., 08]

Marling, C., Shubrook, J., & Schwartz, F. *Case-Based Decision Support for Patients with Type 1 Diabetes on Insulin Pump Therapy*. 9th European Conference. ECCBR 2008. Springer-Verlag, Berlin. 325-339.

[Musen et al., 14]

Musen, M. A., Middleton, B., & Greenes, R. A. *Clinical decision-support systems*. In Biomedical informatics (pp. 643-674). Springer, London.

[Malyshevska, 09]

Malyshevska, K. *The usage of neural networks for the medical diagnosis*. International Book Series. Information Science and Computing, 77-80

[Malekpoor et al., 16]

Malekpoor, H., Mishra, N., Sumalya, S., & Kumari, S. *An efficient approach to radiotherapy dose planning problem: a TOPSIS case-based reasoning approach*. International Journal of Systems Science: Operations & Logistics, p. 1-9

[Merida-Campos et Rollón Rico, 03]

Merida-Campos, C., et Rollón Rico, E. *CBR and MBR techniques: review for an application in the emergencies domain*. Available : <https://upcommons.upc.edu/bitstream/handle/2117/97020/R03-36.pdf>

[Mokeddem et al., 14]

Mokeddem, S., Atmani, B., & Mokaddem, M. A new approach for coronary artery diseases diagnosis based on genetic algorithm. International Journal of Decision Support System Technology (IJDSST), 6(4), 1-15.

[Montani, 08]

Montani, S. *Exploring new roles for case-based reasoning in heterogeneous AI systems for medical decision support*. Applied Intelligence, Vol. 28 No 3, 275-285

[Montani et al., 01]

Montani, S., Magni, P.A., Roudsari, V., Carson, E. R., & Bellazzi, R. *Integrating different methodologies for insulin therapy support in type 1 diabetic patients*. In AIME'01 : Proceedings of the 8th Conference on AI in Medicine in Europe, pages 121–130, London, UK, Springer-Verlag.

[Moreno, 15]

Moreno, M. *Développement des systèmes d'aide à la décision dans les cabinets de médecine générale en France*. Thèse pour obtenir le grade de Docteur en Médecine. <https://petale.univ-lorraine.fr/notice/view/univ-lorraine-ori-34885>

[Mintzberg et al., 76]

Mintzberg, H., Raisinghani, D., & Theoret, A. *The structure of unstructured decision process*. Administrative Science Quarterly, 246–275.

[Mintzberg, 82]

Mintzberg, H. *Structure et dynamique des organisations*, Ed d'organisation, Paris.

[Minsky, 81]

Minsky, M. *A Framework for Representing Knowledge*. In: Mind Design. Haugeland (ed). Boston: MIT Press.

[Musen et al., 06]

Musen, M.A, Shahar, Y., & Shortliff, E.H. *Clinical Decision-Support Systems*. Biomedical Informatics Health Informatics, UNIT II, 698-736, DOI: 10.1007/0-387-36278-9_20.

[Molines, 07]

Molines, N. *SIG et analyse multicritère : des outils au service de l'amélioration du processus décisionnel des projets autoroutiers*. CRENAM U, St-Etienne/ CRG U, Laval (Québec).

[Melvin, 12]

Melvin, A. *Decision Making using the analytic hierarchy process (AHP) and SAS/IML*. SESUG, Paper SD-04.

[Nafi et Werey, 10]

Nafi, A., et Werey, C. *Aide à la décision multicritères : introduction aux méthodes d'analyse multicritère de type ELECTRE*. Module d'ingénierie financière, ENGEES, 2009, vol. 2010.

[Nilsson et Sollenborn, 04]

Nilsson, M., and Sollenborn, M. *Advancements and trends in medical casebased reasoning : An overview of systems and system development*. In FLAIRS Conference.

[Noori, 15]

Noori, B. *Developing a CBR system for marketing mix planning and weighting method selection using fuzzy AHP*. Applied Artificial Intelligence, Vol. 29, no 1, p. 1-32

[Osheroff, 09]

Osheroff, J. A., *Improving Medication Use and Outcomes With Clinical Decision Support : Step-By-Step Guide*. HIMSS.

[Ozernoy, 92]

Ozernoy, V.M. *Choosing the best multiple criteria decision-making method*. INFOR, 32(2):159–171

[Pantic, 05]

Pantic, M. *Introduction to Machine Learning & Case-Based Reasoning*. London: Imperial College.

[Pan et al., 07]

Pan, R., Yang, Q., & Pan, S.J. *Mining competent case bases for case-based reasoning*. In: Artificial Intelligence 171, 1039–1068.

[Pandey et Mishra, 10]

Pandey, B., & Mishra, R. B. *Data mining and CBR integrated methods in medicine: a review*. International Journal of Medical Engineering and Informatics, 2(2), 205-218.

[Pandey et Mishra, 09]

Pandey, B., & Mishra, R. B. *Knowledge and intelligent computing system in medicine*. Computers in biology and medicine, 39(3), 215-230.

[Park et al., 06]

Park, Y. J., Kim, B. C., & Chun, S. H. *New knowledge extraction technique using probability for case-based reasoning: application to medical diagnosis*. Expert Systems, 23(1), 2-20.

[Pelaccia et al., 11]

Pelaccia. T., Tardif, J., Tribby, E., Ammirati, C., Bertrand, C., & Charlin, B. *Comment les médecins raisonnent-ils pour poser des diagnostics et prendre des décisions thérapeutiques? Les enjeux en médecine d'urgence*. Annales françaises de médecine d'urgence (1), 77-84

[Podgorelec, 05]

Podgorelec, V., Hericko, M., & Rozman., I *Improving mining of medical data by outliers prediction*. In CBMS, pages 91-96.

[Pomerol et Barba-Romero, 93]

Pomerol. J.Ch., et S. Barba-Romero. *Choix multicritère dans l'entreprise*. Hermès, Paris.

[Power, 02]

Power, D. J. *Decision support systems: concepts and resources for managers*. Studies in Informatics and Control, 2002, vol. 11, no 4, p. 349-350.

[Prabusankarlal et al., 15]

Prabusankarlal. K.M., Thirumoorthy, P., & Manavalan, R. *Assessment of combined textural and morphological features for diagnosis of breast masses in ultrasound*. Human-centric Computing and Information Sciences, Vol. 5, no 1, p. 1-17.

[Prather, et al., 97]

Prather, J.C., Lobach, D. F., Goodwin, L.K, Hales, J. W., Hage, M.L., & Hammond. W. E. *Medical data mining : Knowledge discovery in a clinical data warehouse*. In n 1997 Annual Conference of the American Medical Informatics Association, Philadelphia.

[Qi et al., 16]

Qi. J., Hu, J., and Peng, Y. *Hybrid weighted mean for CBR adaptation in mechanical design by exploring effective, correlative and adaptative values*. Computers in Industry, Vol. 75, p. 58-66.

[Ramirez et al., 00]

Ramirez J. C., , Cook, D. J., Peterson, L. L., Peterson, D. M. *An event set approach to sequence discovery in medical data*. Intell. Data Anal., 4(6) :513-530.

[Rasmussen, 86]

Rasmussen, J. Information processing and human-machine interaction. An approach to cognitive engineering.

[Rialle, 94]

Rialle. V. *Décision et Cognition en Biomédecine: modèles et Intégration*. Thèse de doctorat. Université Joseph-Fourier-Grenoble I.

[Richards et al., 01]

Richards, G., Rayward-Smith, V. J., Sönksen, P. H., Carey, S., & Weng, C. *Data mining for indicators of early mortality in a database of clinical records*. Artificial intelligence in medicine, 22(3), 215-231.

[Reason, 93]

Reason, J. *L'erreur Humaine*, Presses Universitaires de France 2ème Ed.

[Renaud et al., 10]

Renaud. J.L., Lagouarde, P., & Darmoni, S. *Etude des systèmes d'aide à la décision médicale*. France.

[Recio-Garcia, et al., I4]

Recio-Garcia, J. A., Diaz-Agudo, B., Belen, & Gonzalez-Calero, P. A. *jcolibri2 Tutorial jcolibri2: A framework for building Case-based reasoning system*. Science of Computer Programming, 79, 126-145.

[Renaud-Salis et al., I0]

Renaud-Salis, J., Lagouarde, P., & Darmoni, S.J. *Étude des systèmes d'aide à la décision médicale*, Haute Autorité de Santé (HAS).

http://www.has-sante.fr/portail/jcms/c_1021245/systemes-informatiques-d-aide-a-la-decision-medicale

[Rogers et al., 00]

Rogers, M., Bruen, M. & Maystre, L. Y. *ELECTRE and decision support, method and applications in engineering and infrastructure investment*. Kluwer Academic Publisher, ISBN 0- 7923-8647-7, USA.

[Roy, 85]

Roy, B. *Méthodologie Multicritères d'Aide à la Décision*. Collection Gestion, Série Production et techniques quantitatives appliquées à la gestion, Economica, Paris.

[Roy, 91]

Roy, B. *The outranking approach and the foundations of ELECTRE methods*. Theory Decision, pp. 49-73.

[Roy, 92]

Roy, B. *Science de la décision ou science de l'aide à la décision ?*. Revue Internationale de Systémique, 6(5) :497-529

[Roy et Bouyssou, 93]

Roy, B., et Bouyssou, D. *Aide multicritère à la décision, méthodes et cas*. Paris : Economica, 1993.

[Roy et Vanderpooten, 96]

Roy, B., and Vanderpooten, D. *The European school of MCDM : Emergence, basic features and current works*. Journal of Multi-Criteria Decision Analysis, 5(1) :22-38,

[Royes, 04]

Royes, G. F. *A Hybrid Fuzzy-Multicriteria-CBR Methodology for Strategic Planning Support*. Processing NAFIPS'04, Annual Meeting of the Fuzzy Information, Vol. I, 208-213

[Saraiva et al., I6]

Saraiva, R., Perkusich, M., Silva, L., Siebra, C., & Perkusich, A. *Early diagnosis of gastrointestinal cancer by using case-based and rule-based reasoning*, Expert Systems with Applications, Vol. 61, pp. 192-202.

[Schärli, 85]

Schärli, A. *Décider sur plusieurs critères, panorama de l'aide à la décision multicritère*. Collection Diriger l'entreprise, Presses Polytechniques et Universitaires Romandes, Lausanne, Suisse

[Schneider, 96]

Schneider, D. K. *Modélisation de la démarche du décideur politique dans la perspective de l'intelligence artificielle*. Ecole polytechnique fédérale de Lausanne, vol. 440.

[Schwartz et al., 08]

Schwartz, F.L., Shubrook, J. H., & Marling, R. *Use of case-based reasoning to enhance intensive management of patients on insulin pump therapy.* Journal of diabetes science and technology, Vol. 2, no. 4, pp. 603-611.

[Schmidt et al., 01]

Schmidt, R., Montani, S., Bellazzi, R., Portinale, L., & Gierl, L. *Cased-based reasoning for medical knowledge-based systems.* International Journal of Medical Informatics, Vol. 64 No 2, pp 355-367

[Schmidt et Vorobieva, 05]

Schmidt, R., and Vorobieva, O. *Adaptation and medical case-based reasoning focusing on endocrine therapy support,”* Artificial Intelligence in Medicine, 10th Conference on Artificial Intelligence in Medicine, pp. 300–309.

[Schmidt et Gierl, 02]

Schmidt, R and Gier, L. Prognostic model for early warning of threatening influenza waves. In In German Workshop on Experience Management, pages 34–46. GWEM'02, 2002.

[Sefion et al., 03.a]

Sefion, I., Ennaji, A., Gailhardou, M., & Canu, S. *Aide à la décision médicale Contribution pour la prise en charge de l'asthme.* Ingénierie des Systèmes d'Information Vol. 8 No. 1, 11-32.

[Sefion et al., 03.b]

Sefion, I., Ennaji, A., & Gailhardou, M. *ADEMA: A System to Help Physicians in the Asthma Health Care.* In FLAIRS Conference, pp. 82-86.

[Serroussi et Bouaud, I4]

Séroussi, B., et Bouaud, J. *Systèmes informatiques d'aide à la décision en médecine: panorama des approches utilisant les données et les connaissances.* Pratique Neurologique-FMC, 2014, vol. 5, no 4, p. 303-316.

[Serroussi, et al., I3.a]

Séroussi, B., Le Beux P., & Venot A. *L'aide au diagnostic médical. In: Informatique médicale, e-Santé.* Springer, Paris

[Serroussi, et al., I3.b]

Séroussi, B., Bouaud, J., Duclos, C., Dufour, J.C., & Venot, A. *L'aide à la décision thérapeutique.* In : Informatique médicale, e-Santé. Springer, Paris. p. 175-198.

[Shortliffe, 76]

Shortliffe, E.H. *Computer-Based Medical Consultation : MYCIN.* New York. American Elsevier.

[Sittig, 08]

Sittig, D.F., Wright, A., Osheroff, J.A., Middleton, B., Teich, J.M., Ash, J.S., Campbell, E., Bates, D.W. *Grand challenges in clinical decision support.* Journal of Biomedical Informatics Vol 41, Issue 2, Pages 387-392.

[Simon, 77]

Simon, H. A. *The new science of management decision.* Prentice Hall, New Jersey, systems development in an emerging economy. Decision Support Systems

[Simon, 83]

SIMON, H. A. *Administration et processus de décision*, traduction française de Administrative Behavior, 1947. Economica, Paris.

[Simoudis, 96]

Simoudis, E., Reality check for data mining. Expert: Intelligent systems and their applications, vol. 11, no 5, p. 26-33.

[Sivakumar, 07]

Sivakumar, R. *Neural Network Based Diabetic Retinopathy Classification Using Phase spectral Periodicity components*. ICGST-BIME Journal, Vol. 7, No 1, 23-28.

[Sharareh et al., 10]

Sharareh, R. Kalhori, N., Nasehi, M., & Zeng, X.J. *A Logistic Regression Model to Predict High Risk Patients to Fail in Tuberculosis Treatment Course Completion*. International Journal of Applied Mathematics. Vol. 40.

[Shanbezadeh et al., 13]

Shanbezadeh, M., Soltani, T., & Ahmadi, M. *Developing a Clinical Decision Support Model to Evaluate the Quality of Asthma Control Level*. Middle-East Journal of Scientific Research 14, N° 3, 387-393.

[Szolovits et al., 88]

Szolovits, P., Patil, R.S., Schwartz, W.B. *Artificial intelligence in medical diagnosis*. Annals of internal medicine, 108(1), 80-87.

[Sung et Seong, 10]

Sung, H.H., and Seong, H.J. *A Hybrid Data Mining Method for the Medical Classification of Chest Pain*. International Journal of Computer and Information Engineering. Vol.4

[Sqalli et Freuder, 98]

Sqalli, M.H., and Freuder, E.C. *Integration of CSP and CBR to compensate for incompleteness and incorrectness of models*. in : The AAI-98 Spring Symposium on Multimodal Reasoning.

[Sqalli et al., 99]

Sqalli, M.H., Purvis, L., & Freuder, E.C. *Survey of applications integrating constraint satisfaction and case-based reasoning*. in : PACLP99: The First International Conference and Exhibition on the Practical Application of Constraint Technologies and Logic Programming, pp. 19-21.

[Song et al., 07]

Song, X., Petrovic, S., & Sundar, S. *A case-based reasoning approach to dose planning in Radiotherapy*, 7th International Conference on Case-based Reasoning ICCBR, pp.348-357.

[Thokala et Praveen, 12]

Thokala, P., and Praveen, A. *Multiple criteria decision analysis for health technology assessment*. Value in Health, Vol. 15, no. 8, pp. 1172-1181

[Tsoukias, 03]

Tsoukiàs, A. *From decision theory to decision aiding methodology*. DIMACS Technical report 2003-21, Rutgers University.

[Turban, 93]

Turban, E. Decision Support and Expert Systems. Macmillan, New York.

[Vansnick, 88]

Vansnick, J.C. *Principes et applications des méthodes multicritères*. Technical report, Université de Mons-Hainaut.

[Vansnick, 90]

Vansnick, J.C. *Measurement theory and decision aid*. In C. Bana e Costa, editor, *Readings in Multiple Criteria Decision Aid*, pages 81–100. Springer-Verlag, Berlin.

[Verma et al., 14]

Verma, L., Srinivasan, S., & Sapra, V. *Integration of rule based and case-based reasoning system to support decision making*. *Issues and Challenges in Intelligent Computing Technics (ICICT)*, International Conference on. IEEE, pp. 106-108.

[Velasquez et Hester, 13]

Velasquez, M., and Hester, P. T. *An Analysis of Multi-Criteria Decision Making Methods*. *International Journal of Operations Research*, Vol. 10, no. 2, pp. 56-66

[Vincke, 94]

Vincke, P. *Comment choisir une méthode d'aide à la décision ?*. *Cahiers du CERO*, 339–343.

[Vincke, 89]

Vincke, P. *L'aide multicritère à la décision*. Éditions de l'Université de Bruxelles, Bruxelles

[Wang et al., 16]

Wang, H., Sun, B., & Shen, X. *Hybrid similarity measure for retrieval in case-based reasoning systems and its applications for computer numerical control turret design*. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, pp. 0954405416654432

[Watson, 99]

Watson, I. *Case-based reasoning is a methodology not a technology*. In: *AI-CBR*, University of Salford, United Kingdom.

[Weber, 15]

Weber, S. *Clinical Decision Support Systems and How Critical Care Clinicians Use Them*, *Journal of Healthcare Information Management*, Vol. 21, No. 2

[Xu Li, 96]

Xu Li, D. *An integrated rule-and case-based approach to AIDS initial assessment*. *International journal of bio-medical computing*, Vol. 40, No 3, 197-207

[Young, 94]

Young, G.A. *Bootstrap : More than a stab in the dark*. *Statistical Science*, 9(3) :382–415.

[Zighed et al, 01]

Zighed, D.A., Kodratoff, Y., & Napoli, A., *Extraction de connaissance à partir d'une base de données*. *Bulletin LFIA*, Vol. 1.

[Zhuang et al., 09]

Zhuang, Z.Y., Churilov, L., Burstein, F., & Sikaris, K. *Combining data mining and case-based reasoning for intelligent decision support for pathology ordering by general practitioners*. *European Journal of Operational Research*, Vol. 195, No 3, pp-662-675.

[Zemirline, 08]

Zemirline, A. *Définition et fusion de systèmes diagnostic à l'aide d'un processus de fouille de données: Application aux systèmes diagnostics* (Doctoral dissertation, Télécom Bretagne), 2008.

[Zorman, et al., 02]

Zorman, M., Masuda, G., Kokol, P., Yamamoto, R., & Stiglic, B. *Mining diabetes database with decision trees and association rules*. In CBMS, pages 134–, 2002.

Webographie

[OpenClinical, 15]

OpenClinical, Knowledge management.
<http://www.openclinical.org>
<http://www.openclinical.org/aisinpracticeDSS.html>

[Sournia]

http://ancien.serpsy.org/formation_debat/diagnostic/diagnostic_medical.html

[UCI_CMCDS]

Contraceptive Method Choice Data Set Online. Available :
<https://archive.ics.uci.edu/ml/datasets/Contraceptive+Method+Choice>

[UCI_PDDOP]

Vertebral Column Data Set Online. Available :
<http://archive.ics.uci.edu/ml/datasets/Vertebral+Column#>

[What-when-how]

<http://what-when-how.com/medical-informatics/medical-decision-support-systems-and-knowledge-sharing-standards/>

[Concepts_Stat]

<http://www.statsoft.fr/concepts-statistiques/classifications/classifications.php#mesure>

Les figures

Figure 1.1. Les techniques intelligentes en aide à la décision [Das, 16].	9
Figure 1.2. Le processus décisionnel selon Simon.	13
Figure 1.3. Les niveaux de décision [Kast, 02].	15
Figure 1.4. Le raisonnement clinique.	17
Figure 1.5. Eléments concourant à la décision médicale.	18
Figure 1.6. Typologie des SADM selon les approches utilisées .	24
Figure 1.7. Structure d'un SADM [Holtzman 89].	25
Figure 1.8. Exemple de représentation de la classe personne dans une application médicale.	27
Figure 1.9. Principe de fonctionnement du RBC (cycle RBC), adapté de [Pantic, 05].	31
Figure 2.1. Disciplines co-fondatrices de la FD [Kodratoff, 98].	49
Figure 2.2. Schéma global de l'ECD d'après Fayyad et al., [Fayyad et al., 96].	54
Figure 2.3. Matrice de confusion 2 x 2 dimensions.	57
Figure 3.1. Les étapes d'une méthode multicritères [Nafi et Wery, 10].	73
Figure 3.2. Les méthodes d'agrégation.	75
Figure 4.1. La situation médicale.	82
Figure 4.2. Modèle d'Aide à la Décision Médicale Intégré.	87
Figure 4.3. Echantillon de "Presumptive Diagnosis of Diseases of Orthopaedic Patients".	88
Figure 4.4. Le modèle d'aide à la décision médicale RBC-FDD proposé.	90
Figure 4.5. Processus d'aide à la décision médicale Intégré.	91
Figure 4.6. Le système RBC^{FDD} proposé.	92
Figure 4.7. Processus RBC adapté de Assali et al. [Assali et al., 09].	98
Figure 4.8. Diagramme de cas d'utilisation (FDD).	99
Figure 4.9. Le diagramme de classes (FDD).	99
Figure 4.10. Echantillon de "Contraceptive Method Choice Data Set" [UCI_CMCDs].	102
Figure 4.11. Le modèle d'aide à la décision médicale RBC-AMC proposé.	104
Figure 4.12. Le système RBC^{AMC} Proposé.	105
Figure 4.13. Structure du Problème Multicritères Médical.	108

Figure 4.14. Exemple d'un graphe de sur-classement.	110
Figure 4.15. Organigramme du surclassent [Maystre et al., 94].	111
Figure 4.16. Diagramme des cas d'utilisation (AMC).	112
Figure 4.17. Le diagramme de classes (AMC).	113

Les tables

Table 1.1. Fonctions d'aide à la décision et exemples de problèmes cliniques	21
Table 1.2. Systèmes RBC et leurs domaines d'application (adaptée) [Begum et al., 11].	36
Table 1.3. Systèmes développés avec le RBC et d'autres techniques [Begum et al., 11].	42
Table 2.1. Les tâches de l'ECD.	53
Table 2.2. Les mesures pour l'évaluation de modèles [Beleites et al., 16].	57
Table 2.3. Les méthodes de fouille de données.	58
Table 3.1. Les situations possibles de comparaison de 2 actions .	67
Table 3.2. Les situations possibles de comparaison de 2 actions.	67
Table 3.3. Matrice de performance.	69
Table 3.4 Exemple de matrice de performance (pour le choix d'implantation d'usine).	69
Table 3.5. Les problématiques décisionnelles.	70
Table 3.6. Les méthodes PROMETHEE.	76
Table 3.7. Les méthodes ELECTRE.	77
Table 3.8. Choix de la méthode multicritères [Nafi et Wery, 10 ; Rogers et al., 00].	78
Table 4.1. Structure de la situation médicale.	85
Table 4.2. Structure du Cas Médical.	86
Table 4.3. Le problème médical (Structure).	87
Table 4.4. Les descripteurs de la base de données [UCI_PDDOP].	88
Table 4.5. Le cas médical orthopédique constitué.	93
Table 4.6. La base de cas Ω_N .	100
Table 4.7. Bases de cas partielles Ω_L, Ω_T	100
Table 4.8. Comparaison des résultats sur 3 ensembles de données de test.	101
Table 4.9. Les descripteurs de la base de données [UCI_CMCDS].	103
Table 4.10. Structure du cas médical orthopédique.	106
Table 4.11. Les bases partielles $\Omega_A, \Omega_T, \Omega_0, \Omega_1, \Omega_2$	113
Table 4.12. Résultats de l'expérimentation.	115

ملخص

لقد حقق دعم القرار تقدماً ملحوظاً في المجال الطبي من خلال أنظمة دعم القرار. والتي هي "تطبيقات الكمبيوتر التي تزود الممارسين الصحيين ببيانات ومعلومات مفيدة تصف الحالة السريرية للمساعدة في تحسين جودة رعاية المرضى."

نظراً لاعتماد أساليب التفكير المستندة إلى الحالات (CBR) على نطاق واسع في هذه الأنظمة ، فإننا نشهد مؤخراً تكاملها مع طرق التفكير الأخرى ، وهي تحديد البيانات (DM) ، وتحليل متعدد المتغيرات (MCA) ، وقواعد القرار ، وما إلى ذلك. في هذا السياق ، فإن الهدف من هاته الأطروحة هو تصميم وتجريب وتقييم النهج الجديدة من خلال دمج CBR مع DM و CBR مع MCA.

وأظهرت التجارب التي أجريت أن هناك صلة بين النهج المتبع وأكدت من أن نموذج صنع القرار المقترح يميل إلى توفير نتائج مرضية.

الكلمات المفتاحية:

نظام دعم القرار الطبي ، تحليل متعدد المتغيرات ، استخراج البيانات ، التفكير القائم على الحالة .

Abstract

Decision support has made remarkable progress in the medical field through decision support systems called MDSS. These are "computer applications that provide health practitioners with useful data and information describing a clinical situation to help improve the quality of patient care."

As case-based reasoning (CBR) is widely adopted in these systems, we are now witnessing its integration with other modes of reasoning, namely data mining (DM), multicriteria analysis (MCA) decision, etc. In this context the goal of our thesis is the design, experimentation and evaluation of new approaches by integrating CBR with DM and CBR with MCA.

The experiments carried out showed the relevance of the adopted approach and confirm that the proposed decision-making model tends to provide satisfactory results.

Keywords: Medical decision support system, Multicriteria analysis, Data mining, Case-based reasoning.

Résumé

L'aide à la décision a connu une avancée très remarquable dans le domaine médical à travers les systèmes d'aide à la décision nommés SADM. Ce sont des "applications informatiques fournissant aux praticiens de la santé des données et des informations utiles, décrivant une situation clinique dans le but d'aider à améliorer la qualité des soins des patients".

Le raisonnement à base de cas (RBC) étant largement adopté dans ces systèmes, nous assistons récemment à son intégration avec d'autres modes de raisonnement à savoir la fouille de données (FDD), l'analyse multicritères (AMC), les règles de décision, etc. Dans ce contexte l'objectif de notre thèse est la conception, l'expérimentation et l'évaluation de nouvelles approches par intégration du RBC avec la FDD et du RBC avec l'AMC.

Les expérimentations effectuées ont montré la pertinence de l'approche adoptée et confirment que le modèle décisionnel proposé a tendance à fournir des résultats satisfaisants.

Mots clés : Système d'Aide à la décision médicale, Analyse multicritères, Fouille de données, Raisonnement à base de cas.

Résumé

L'aide à la décision a connu une avancée très remarquable dans le domaine médical à travers les systèmes d'aide à la décision nommés SADM. Ce sont des "applications informatiques fournissant aux praticiens de la santé des données et des informations utiles, décrivant une situation clinique dans le but d'aider à améliorer la qualité des soins des patients". Le raisonnement à base de cas (RBC) étant largement adopté dans ces systèmes, nous assistons récemment à son intégration avec d'autres modes de raisonnement à savoir la fouille de données (FDD), l'analyse multicritères (AMC), les règles de décision, etc. Dans ce contexte l'objectif de notre thèse est la conception, l'expérimentation et l'évaluation de nouvelles approches par intégration du RBC avec la FDD et du RBC avec l'AMC. Les expérimentations effectuées ont montré la pertinence de l'approche adoptée et confirment que le modèle décisionnel proposé a tendance à fournir des résultats satisfaisants.

Mots clés :

Système d'Aide à la décision médicale; Analyse multicritères; Fouille de données; Raisonnement à base de cas; RBC; Règles; AMC; Groupement; FDD; CBR .