



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

N° d'ordre : SIOD 15 / M2 /2019

Mémoire

Présenté pour obtenir le diplôme de master académique en

Informatique

Parcours : SIOD (Informatique Décisionnelle et Optimisation)

Prédiction des déclencheurs du cancer du sein chez la femme

Par :

DJEDDOU DJALAL EDDINE

Soutenu le 07 juillet 2019, devant le jury composé de :

| | | |
|---------------------------|-------------|------------|
| Mme Slatnia Siham | Grade M C B | Président |
| Mr Zerarka Mohamed Faouzi | Grade M C A | Rapporteur |
| Mme Youkana Imene | Grade M C B | Examineur |

Résumé

Le cancer du sein est le cancer le plus fréquent chez la femme, et son incidence est en progression permanente.

L'efficacité de la lutte contre la maladie passe par le dépistage précoce, qui est un objectif majeur pour améliorer le pronostic de cette cause de mortalité.

Afin de rendre le dépistage plus performant, nous proposons un modèle prédictif des facteurs de risque conduisant à la maladie, par la méthode de classification associative des motifs fréquents.

Mots Clés: glande mammaire, tumeur, néoplasie, masse, nodule, néoformation, carcinome mammaire, in situ, invasif, malin, bénin cancer du sein, dépistage, prolifération, facteurs de risque, lésion, data mining, la classification associative, motifs fréquents.

Abstract

Breast cancer is the most common cancer in women, and its incidence is growing steadily.

The effectiveness of the fight against the disease requires early detection, which is a major objective to improve the prognosis of this cause of death.

In order to make screening more efficient, we propose a predictive model of the risk factors leading to the disease, using the associative classification method of frequent patterns.

Keywords: mammary gland, tumor, néoplasma, mass, nodule, neoformation, mammary carcinoma, in situ, invasif, malignant, benign, Breast cancer, screening, proliferation, risk factors, lesion, data mining, the associative classification, the frequent patterns.

Remerciements

Je tiens tout d'abord à remercier dieu qui m'a donnée courage et patience a réalisé ce modeste travail.

Je souhaite remercier du fond du cœur mon encadreur Dr Zerarka Mohamed Faouzi, qui m'a admis l'idée de ce projet et le rendre réalisable et a permis la poursuite de mon parcours vers la thèse.

Je tiens à remercier ma femme Dr Salem, oncologue qui a été à mon côté dès la première idée jusqu'à achèvement.

Je remercie également tous ceux qui m'ont soutenu et aidé à la réalisation de ce modeste travail.

Tables des matières

| | |
|---|----|
| 1. Introduction générale : | 1 |
| Chapitre I : La Classification | |
| 1. Introduction : | 3 |
| 2. Le Data mining : | 4 |
| 2.1. Définition1 : | 4 |
| 2.2. Définition2 : | 4 |
| 3. Le processus Extraction de Connaissances par les Données. | 6 |
| 3.1. Définition : | 6 |
| 3.2. Etapes du processus ECD : | 6 |
| 3.2.1. Acquisition des données : | 7 |
| 3.2.2. Prétraitement : | 7 |
| 3.2.3. Construction du model : | 7 |
| 3.2.4. Evaluation des résultats : | 7 |
| 4. Méthodes de classifications associatives : | 8 |
| 4.1. Classification et classement : | 8 |
| 4.2. Principe de la classification : | 8 |
| 4.3. Les types de la classification : | 8 |
| 4.3.1. La classification supervisée : | 8 |
| 4.3.2. La classification non supervisée (clustering) : | 8 |
| 5. Définitions : | 9 |
| 5.1. Base de données formelle | 9 |
| 5.2. Motif | 10 |
| 5.3. Itemset | 10 |
| 5.4. Itemset | 10 |
| 5.5. Support d'un Itemset | 10 |
| 5.6. Support d'une règle d'association | 11 |
| 5.7. Confiance | 11 |
| 5.8. Itemset Fréquent | 11 |
| 5.9. Transaction | 11 |
| 6. Les motifs fréquents : | 11 |
| 7. Les règles d'association : | 11 |
| 7.1. Définition 1 : | 11 |
| 7.2. Définition 2 : | 12 |
| 8. Algorithme Apriori : | 12 |
| 9. Conclusion : | 13 |
| Chapitre II : Cas d'étude | |
| 1. Introduction : | 14 |
| 2. Anatomie descriptive et fonctionnelle | 14 |
| 2.1. Situation du sein : | 14 |
| 2.2. La forme : | 16 |
| 2.3. Le poids : | 16 |
| 2.4. La consistance : | 16 |
| 3. Structure du sein : | 16 |
| 3.1. Le tissu adipeux et conjonctif : | 17 |
| 3.2. La peau et la plaque aréolo-mamelonnaire : | 17 |
| 3.3. Le mamelon : | 17 |

| | | |
|----------|--|----|
| 3.4. | L'aréole : | 17 |
| 4. | La glande mammaire : | 18 |
| 4.1. | Vascularisation artérielle et veineuse du sein : | 20 |
| 4.2. | Réseaux lymphatiques : | 21 |
| 4.2.1. | Le réseau lymphatique cutané : | 21 |
| 4.2.2. | Le réseau lymphatique glandulaire : | 22 |
| 4.2.3. | Les ganglions axillaires : | 22 |
| 5. | Histoire et définition du cancer : | 22 |
| 5.1. | Qu'est-ce qu'un cancer ? : | 22 |
| 5.2. | Le cancer du sein : | 24 |
| 5.3. | Les différentes formes de cancer du sein : | 25 |
| 5.3.1. | In situ ou invasif : | 25 |
| 5.3.2. | Tumeurs bénignes et tumeurs malignes : | 25 |
| 5.4. | Corrélation entre l'anatomie et les images mammographiques : | 25 |
| 5.5. | Les anomalies de la plaque areolo-mamelonnaire : | 26 |
| 5.5.1. | Changements cutanés : | 26 |
| 5.5.2. | Masse : | 27 |
| 5.5.3. | Écoulement mamelonnaire : | 27 |
| 5.6. | Nodule mammaire : | 27 |
| 6. | Les déclencheurs de cancer du sein : | 27 |
| 6.1. | Facteurs de risque : | 29 |
| 6.1.1. | L'âge : | 29 |
| 6.1.2. | Antécédents personnels : | 30 |
| 6.1.2.1. | Antécédents de cancer du sein : | 30 |
| 6.1.2.2. | Affections bénignes du sein : | 30 |
| 6.1.2.3. | L'hyperplasie canalaire simple : | 30 |
| 6.1.2.4. | L'hyperplasie canalaire atypique : | 30 |
| 6.1.2.5. | Cicatrice radiaire : | 31 |
| 6.1.2.6. | L'adénose sclérosante : | 31 |
| 6.1.2.7. | Les atypies épithéliales planes : | 31 |
| 6.1.2.8. | Les lésions papillaires : | 31 |
| 6.1.2.9. | Exposition à des radiations médicales : | 32 |
| 6.1.3. | Antécédent familiaux : | 32 |
| 6.1.4. | Prédisposition génétique : | 33 |
| 6.1.5. | Les facteurs de risque liés aux modes de vie : | 33 |
| 6.1.5.1. | La consommation de tabac : | 33 |
| 6.1.5.2. | La consommation d'alcool : | 34 |
| 6.1.5.3. | Le surpoids/ L'obésité : | 34 |
| 7. | Conclusion : | 34 |

Chapitre III : La conception du système

| | | |
|----------|---|----|
| 1. | Introduction : | 35 |
| 2. | La conception du système : | 35 |
| 2.1. | La conception générale du système : | 35 |
| 2.2. | Conception Détaillée : | 36 |
| 2.2.1. | Sous système d'acquisition des données cibles : | 36 |
| 2.2.1.1. | Acquisition des informations acteurs : | 37 |
| 2.2.1.2. | Acquisition des informations des patients : | 38 |
| 2.2.1.3. | Acquisition des informations des habitudes de vie du consultant : | 40 |

| | | |
|--|---|----|
| 2.2.1.4. | Acquisition des informations des antécédents du consultant : | 41 |
| 2.2.1.5. | Acquisition des informations de l'examen clinique du consultant : | 42 |
| 2.2.1.6. | Acquisition des informations de pré-orientations du consultant : | 43 |
| 2.2.1.7. | Acquisition des informations des examens des creux axillaires : | 45 |
| 2.2.1.8. | Acquisition des informations des examens sein droit : | 46 |
| 2.2.1.9. | Acquisition des informations des orientations : | 47 |
| 2.2.2. | Le sous-système du Prétraitement | 49 |
| 2.2.2.1. | Les entrées du Sous système | 49 |
| 2.2.2.2. | La sortie du Sous système | 49 |
| 2.2.3. | La classification..... | 50 |
| 2.2.4. | Le sous-système de la visualisation..... | 53 |
| 3. | Conclusion..... | 53 |
| Chapitre IV : L'implémentation et la validation | | |
| 1. | Introduction : | 54 |
| 2. | L'environnement matériel/logiciel : | 54 |
| 2.1. | Matériel | 54 |
| 2.2. | Logiciel..... | 54 |
| 2.2.1. | Système d'exploitation utilisé | 54 |
| 2.2.2. | Langage de Programmation utilisé..... | 54 |
| 2.2.3. | L'environnement utilisé..... | 55 |
| 2.2.4. | Les SGBDS utilisé..... | 56 |
| 4. | L'implémentation : | 56 |
| 4.1. | Sous-système d'acquisition des données cibles..... | 56 |
| 4.2. | Sous-système de prétraitement | 64 |
| 4.2.1. | La sortie du sous-système..... | 65 |
| 4.3. | Sous-système de la classification | 66 |
| 4.4. | Les tests et la visualisation des résultats..... | 67 |
| 4.4.1. | La présentation du sous-système d'acquisition des données cibles..... | 68 |
| 4.4.2. | La présentation du sous-système Prétraitement..... | 69 |
| 4.4.3. | Le sous-système de la classification..... | 71 |
| 4.4.4. | Le sous-système de la visualisation du modèle | 72 |
| 5. | Conclusion..... | 73 |
| Conclusion générale | | 74 |
| Référence | | 76 |

Listes des figures

- Figure01** : processus de prise de décision
- Figure02** : architecture d'un système type de data mining
- Figure03** : processus ECD
- Figure04** : situation du sein
- Figure05** : angle pariéto mamelonnaire
- Figure06** : schéma clinique du sein
- Figure07** : structure de sein
- Figure08** : la plaque aréolomamelonnaire
- Figure09** : la glande mammaire
- Figure10** : l'alvéole
- Figure11** : réseau lymphatique
- Figure12** : réseau lymphatique cutané
- Figure13** : formation d'une tumeur
- Figure14** : anomalie de la plaque aréolomamelonnaire
- Figure15** : prévalence de l'atteinte familiale chez les femmes ayant développé un CS
- Figure16** : répartition des incidences du CS par tranche d'âge chez les femmes Alger 2016
- Figure17** : l'architecture générale du système
- Figure18** : unité d'acquisition des informations des acteurs
- Figure19** : unité fonctionnelle identification du consultant
- Figure20** : diagramme de structure d'acquisition des informations des patients
- Figure21** : diagramme de structure du processus de consultation
- Figure22** : unité fonctionnelle habitudes de vie
- Figure23** : diagramme de structure habitudes de vie
- Figure24** : unité fonctionnelle antécédents personnel et familiaux du patient
- Figure25** : diagramme de structure antécédent patient
- Figure26** : unité fonctionnelle examen clinique
- Figure27** : diagramme de structure examen clinique
- Figure28** : unité fonctionnelle pré-orientations
- Figure29** : diagramme de structure pré-orientations
- Figure30** : unité fonctionnelle examens des creux axillaires
- Figure31** : diagramme de structure examen des creux axillaire
- Figure32** : unité fonctionnelle examen clinique sein droit / gauche
- Figure33** : diagramme de structure examen clinique sein
- Figure34** : unité fonctionnelle des orientations
- Figure35** : diagramme de structure post-orientations
- Figure36** : architecture du sous-système de prétraitement
- Figure37** : architecture du sous-système de la classification
- Figure38** : le processus de l'extraction des motifs fréquent
- Figure39** : le processus de l'extraction des règles solides
- Figure40** : l'architecture du sous-système de visualisation
- Figure41** : exemple code java

Figure42 : fenêtre NETBEANS
Figure43 : connexion PostgreSQL
Figure44 : la classe acteur
Figure45 : code source de la classe acteur
Figure46 : la classe login
Figure47 : unité identification patient classe Action_acteur avant acquisition de données
Figure48 : code source Action_acteur
Figure49 : unité habitudes de vie avant acquisition des données
Figure50 : unité habitudes de vie avant acquisition des données
Figure51 : unité information générales, habitude de vie, antécédents après acquisition de données
Figure52 : unité d'examen clinique avant acquisition des données
Figure53 : unité d'examen clinique après acquisition des données
Figure54 : unité pré-orientations avant acquisition des données
Figure55 : unité pré-orientations après acquisition des données
Figure56 : unité examen des creux axillaires avant acquisition des données
Figure57 : unité examen des creux axillaires après acquisition des données
Figure58 : unité examen clinique sein droit / gauche
Figure59 : unité post-orientations
Figure60 : descripteur avant l'exécution
Figure61 : descripteur après l'exécution
Figure62 : code source méthode descripteur
Figure63 : l'acquisition des valeurs support et confiance
Figure64 : l'extraction des motifs fréquents et les règles solides
Figure65 : l'interface principale du système proposé
Figure66 : sous système d'acquisition des données cible
Figure67 : identification des médecins
Figure68 : sous système de prétraitement
Figure69 : le sous-système de la visualisation
Figure70 : histogramme des règles
Figure71 : courbe des règles

Introduction générale

Estimé à plus d'un million de cas par an, le cancer du sein est le premier cancer de la femme dans le monde.

Il fait partie des cancers les plus graves avec une mortalité non négligeable ; sur 100 femmes qui meurent d'un cancer, 25 sont atteintes de cancer du sein. [17]

En Algérie Près de 40 % des cancers féminins sont représentés par le cancer du sein. Son incidence continue son ascension et se rapproche inexorablement de l'incidence mondiale (80 nouveaux cas pour 100 000 femmes). [13]

La lutte contre la maladie se répartie globalement suivant trois grands axes, l'amélioration des traitements utilisés une fois la maladie détectée, la détection de la maladie au plutôt, et enfin la mise en œuvre de mesures permettant d'éviter d'être atteinte par la maladie (le dépistage).ces moyens manquent dans les pays en développement où la majorité des cas sont diagnostiqués à des stades avancés.

Les études scientifiques observationnelles ont permis d'identifier certaines caractéristiques propres à l'individu, qui peuvent être en rapport avec le risque de développer un cancer du sein, on appelle ces caractéristiques des facteurs de risque.

La meilleure connaissance de ces facteurs de risques avec leurs impacts sur les individus permet de prédire l'apparition de la maladie et d'améliorer les moyens préventifs, en utilisant des modèles où score calculant le risque individuel.

Il existe déjà des index prédictifs tel que les modèles (**Gail, Claus, IBIS, BRCAPRO, BOADICEA**). Une analyse de ces modèles, montre des insuffisances qui rendent ces modèles non adaptés à toutes les populations, de plus, des facteurs connus comme à risque ne sont inclus dans aucun de ces modèles, à savoir : l'activité physique, l'utilisation d'oestro progestatifs avant la ménopause, l'utilisation d'un traitement hormonal substitutif (THM) après, une partie de la génétique familiale ou somatique (dont le père des femmes suivies), la consommation d'alcool ou l'ostéoporose...etc. [18]

Notre système propose un modèle prédictif des facteurs de risque de développement du cancer du sein chez la femme.

Afin de réaliser ce projet, nous nous sommes focalisé sur quatre (04) chapitres structurés comme suit :

- ✚ Au premier chapitre nous avons fait une étude sur la prédiction des modèles par les motifs fréquents, où on a commencé par la définition du data mining, le processus ECD et ses différentes phases, ensuite les méthodes de

classifications associatives pour finir par l'algorithme Apriori utilisé dans notre projet.

- ✚ Le deuxième chapitre a concerné notre cas d'étude, on l'a entamé par une étude anatomique et descriptive de la glande mammaire, ensuite une définition de la maladie du cancer, le cancer du sein, pour le finir par les déclencheurs de la maladie.
- ✚ Le troisième chapitre était dédié à l'étude conceptuelle globale et détaillée de notre modèle.
- ✚ Le quatrième chapitre a fait l'objet de la description de l'implémentation du système conçu, des tests et de la validation du modèle obtenu.

Nous terminerons ce mémoire par une conclusion générale de notre projet ainsi que les différentes perspectives qui nous semblent intéressantes pour continuer ce travail.

Chapitre 1

La prédiction des modèles par les
motifs fréquents

1. Introduction :

Avec l'explosion des quantités de données et connaissances multiples stockées sur différents supports, il est devenu nécessaire d'introduire les techniques de fouille de données pour la résolution, la compréhension, la modélisation et l'anticipation des problèmes.

Dans ce chapitre, on propose de donner quelques définitions et principes de data mining, les étapes ECD, la classification, la segmentation et particulièrement la prédiction de modèle par l'approche associative ou par les motifs fréquents et l'extraction des règles d'association.

2. Le Data mining :

2.1. Définition1 :

C'est l'ensemble des algorithmes, méthodes et technologies inspirés de plusieurs autres disciplines, propres ou non au data mining pouvant servir à remplacer ou à aider l'expert humain ou le décideur dans un domaine spécifique dans le cadre de prise de décision, et ce en fouillant dans des bases de données décisionnelles des corrélations, des associations, des comportements homogènes, des formules de lien entre indicateurs, des spécifications par rapport à une thématique bien déterminée, etc. Cet ensemble de techniques peut être une étape fondamentale dans tout processus ECD (KDD) ou bien l'intervention de l'intelligence artificielle, la reconnaissance de forme, et la statistique décisionnelle dans tout processus de Business Intelligence afin de transformer tout système d'aide à la décision en un système décisionnel au vrai sens du mot. [19]

2.2. Définition2 :

C'est un ensemble de techniques et processus permettant de fouiller dans des bases de données décisionnelles des corrélations, des associations, des comportements homogènes, des connaissances nouvelles, des spécification par rapport à une thématique bien déterminée, ces techniques sont une étape fondamentale dans le processus ECD (KDD) ,le data mining utilise des outils tel que les régression, règles d'associations, le Clustering, les arbres de décision et les réseaux de neurones issus de différentes spécialités tel que les statistiques, l'intelligence artificielle, afin de transformer tout système d'aide à la décision en un système décisionnel tout en réduisant la quantité de données. [20]

Data Mining et aide à la décision

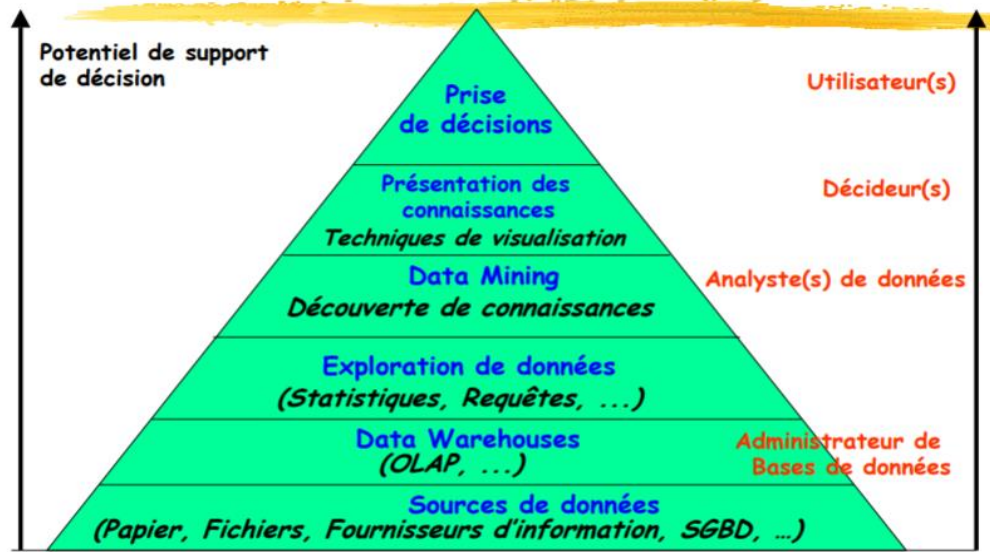


Figure1: processus de prise de décision

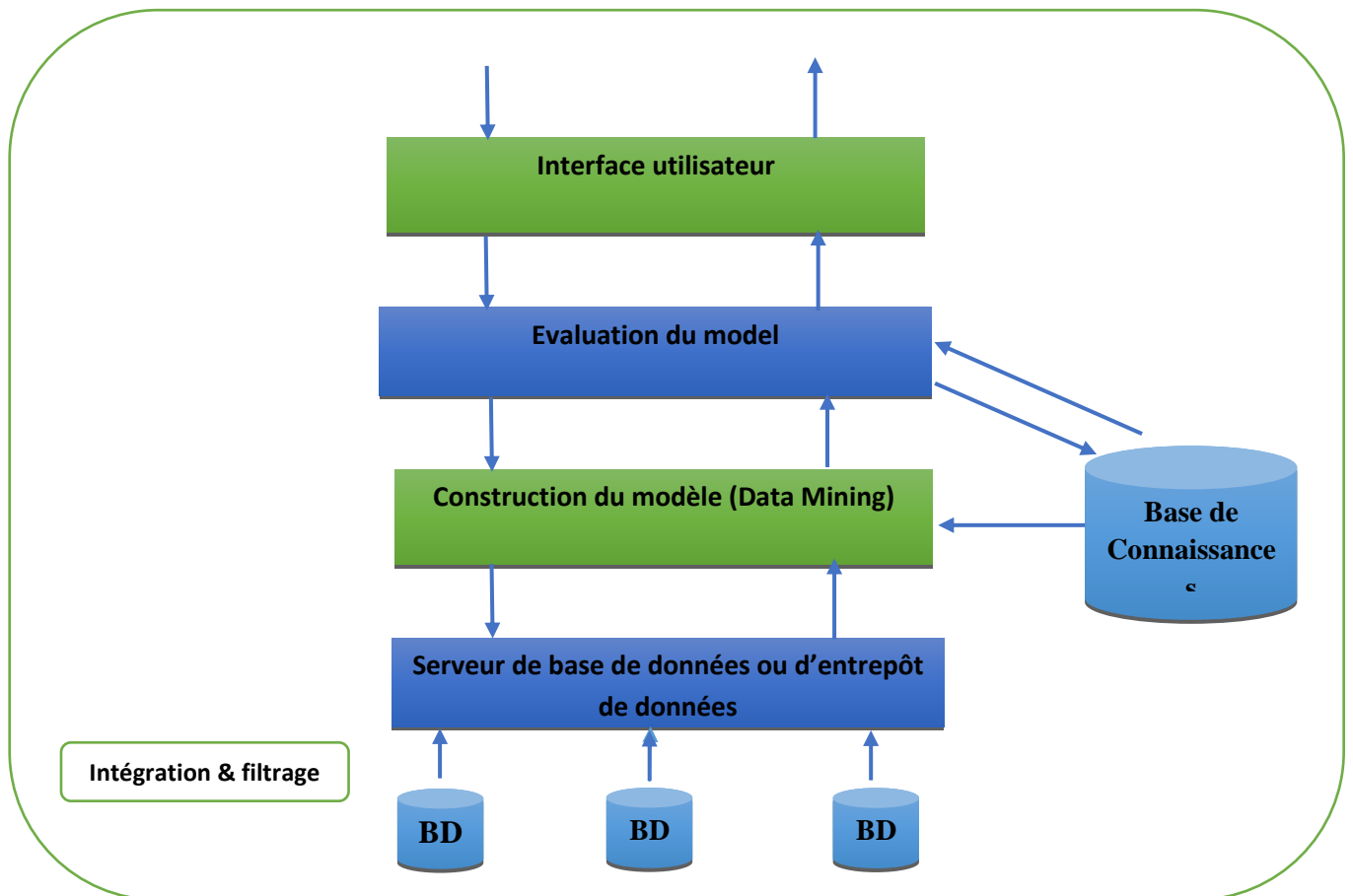


Figure 2: Architecture d'un système type de data mining

3. Le processus Extraction de Connaissances par les Données.

3.1. Définition :

L'Extraction de Connaissances à partir de Données (ECD) désigne tout le cycle de découverte d'informations ou de connaissances dans les bases de données. Il regroupe donc toutes les opérations à effectuer pour extraire de l'information de ces données. [21]

Le data mining est une étape parmi plusieurs qui vise à extraire des connaissances à partir des systèmes d'informations géant, et selon le standard. Il est schématisé comme suit :

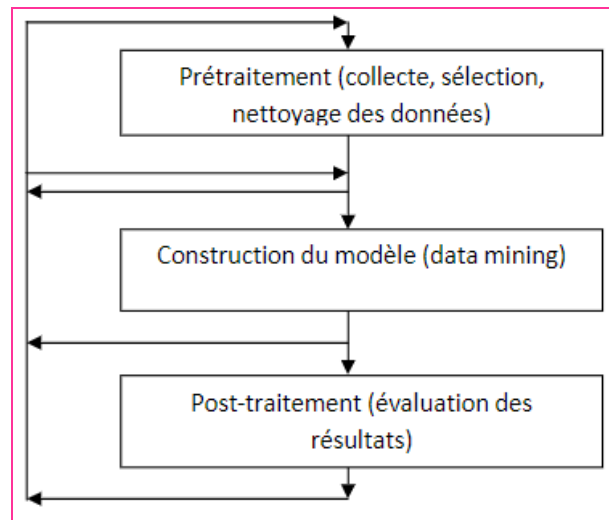


Figure 3 : processus ECD

3.2. Etapes du processus ECD :

L'ECD est un processus homme-machine interactif et itératif, composé de plusieurs phases (Fayyad et al., 1996). Il est communément admis que c'est un processus partiellement automatique, où l'interaction de l'homme est primordiale. Il est itératif, dans le sens où l'utilisateur peut à tout moment revenir à l'une des étapes, afin par exemple de sélectionner ou enrichir les données, intégrer d'autres données, affiner ou changer l'algorithme de fouille, améliorer la présentation ou les mesures d'évaluation des résultats...etc.

L'extraction de règles d'association est un processus constitué de plusieurs phases allant de la collecte, sélection et la préparation des données jusqu'à l'interprétation des résultats, en passant par la phase de recherche des connaissances. [4]

3.2.1. Acquisition des données :

On commence par collecter des données se trouvant dans différents supports, bases de données, datawarhouses et de s'assurer que les données sont bien adaptées à la technique de fouille qui sera utilisée, généralement ses données sont brutes ce qui nécessite une succession de phases de prétraitement.

3.2.2. Prétraitement :

Dans cette étape nous nous intéressons au traitement des données collectées en commençant par la sélection des données objets de l'étude et pertinentes à l'analyse qui seront déterminées et arrêtées.

Une sélection optimale des données nécessite souvent l'aide d'experts du domaine pour déterminer les facteurs, ou variables, les plus aptes à décrire la problématique.

Ensuite une phase de nettoyage est nécessaire pour éliminer les données inutiles tels que: les doublons, les erreurs de saisie, l'intégrité de données et le problème des valeurs manquantes.

A l'étape d'intégration les sources de données hétérogènes seront combinées dans une seule structure.

3.2.3. Construction du model :

C'est la principale phase dans le processus ECD, c'est là où le choix des techniques et algorithmes se fait, pour la fouille des données et l'extraction des concepts, règles ou modèles d'aide à la décision.

3.2.4. Evaluation des résultats :

Cette phase, dite d'évaluation ou de validation, a pour objectif de mesurer l'intérêt des modèles extraits.

Deux approches sont communément utilisées dans la validation : une fondée sur des mesures statistiques et une deuxième par expertise (Gilleron et al., 2000). Ces approches dépendent de la nature de la tâche à mener et du problème considéré. Exemple :

Pour les problèmes de segmentation la validation est essentiellement du ressort de l'expert, qui juge de la qualité et la pertinence des groupes constitués par le système.

Et afin d'évaluer une procédure de classification, les données en entrée sont, généralement, divisées en trois ensembles : un ensemble d'apprentissage, un deuxième de test et un troisième servant à la validation. La matrice de confusion¹ (Gardarin, 1999) et la validation croisée² (Hand et al., 2001) sont d'autres techniques de validation couramment utilisées. [22]

4. Méthodes de classifications associatives :

4.1. Classification et classement :

Dans un classement, les classes sont préalablement connues donc on connaît préalablement à quelle classe l'individu appartient.

Contrairement à la classification, les classes ne sont pas connues au préalable, on ne sait pas à l'avance à quelle classe un individu appartient. La classification se fait en fonction de la population entière.

4.2. Principe de la classification :

Une classe est un ensemble d'éléments qui sont semblables entre eux et qui sont dissemblables à ceux d'autres classes. Classifier consistera à maximiser les similarités des éléments qui sont dans la même classe et à minimiser les similarités de ces éléments avec ceux des autres classes. Inversement, on peut dire que classifier consiste à minimiser la variation intra-classe et à maximiser la variation inter-classe.

4.3. Les types de la classification :

4.3.1. La classification supervisée :

La classification supervisée est basée sur le fait que les classes sont a priori existante, c'est-à-dire l'ensemble de données est déjà classés qu'on appelle « ensemble d'apprentissage », et qu'on l'utilise comme base pour classer le reste des données. [1]

Parmi les méthodes de classification supervisées les plus populaires, on peut citer :

- Les arbres de décision.
- Les réseaux de neurones.
- Les algorithmes génétiques.

4.3.2. La classification non supervisée (clustering) :

Contrairement à la classification (méthodes supervisées), on ne possède pas a priori des connaissances sur les classes prédéfinies des éléments. Donc La division des objets dans les différents groupes (clusters) se procède en se basant sur le calcul de similarité entre les éléments. Alors que l'objectif des méthodes du Clustering est de grouper des éléments proches dans un même groupe de manière à ce que deux données d'un même groupe soient le plus similaires possible et que deux éléments de deux groupes différents soient le plus dissemblables possible [Hartigans, 1975].

Mathématiquement, on a un ensemble X de N données décrites chacune par leurs P attributs. Donc Le Clustering consiste à créer une partition ou une décomposition de cet ensemble en sous parties (clusters) telle que : * Les données appartenant au même groupe se

ressemblent, * Les données appartenant à deux groupes différents soient peu ressemblantes.
[2]

Parmi les méthodes de classification non supervisées les plus populaires, on peut citer :

- Clustering.
- K-means.
- Les règles d'association.

5. Définitions :

Afin de rechercher et d'extraire les règles d'associations on doit définir quelques termes :

5.1. Base de données formelle

Une base de données formelle est la donnée d'un triple (O, P, R) où:

- O : est un ensemble fini d'objets.
- P : est un ensemble fini de propriétés.
- R : est une relation sur $O \times P$ qui permet d'indiquer si un objet $x_i / i = 1..n$ a une propriété p_j avec $j = 1..m$ (noté $x R p$) ou non.

Considérons par exemple la base de données formelle suivante :

| R | a | b | C | d | E |
|-------|---|---|---|---|---|
| x_1 | 1 | 0 | 1 | 1 | 0 |
| x_2 | 0 | 1 | 1 | 0 | 1 |
| x_3 | 1 | 1 | 1 | 0 | 1 |
| x_4 | 0 | 1 | 0 | 0 | 1 |
| x_5 | 1 | 1 | 1 | 0 | 1 |
| x_6 | 0 | 1 | 1 | 0 | 1 |

Où :

$$O = [x_1, x_2, x_3, x_4, x_5, x_6]$$

- $P = [a, b, c, d, e]$
- $x_i R p_j$ si et seulement si la ligne de x et la colonne de p se croisent sur un 1 (et pas sur un 0), par exemple :

- $x_1 R a, x_1 R c$ et $x_1 R d$.

5.2. Motif

Un motif d'une base de données formelle (O, P, R) est un sous-ensemble de P . [8]

L'ensemble de tous les motifs d'une base est donc l'ensemble des parties de P , note 2^P .

On dira qu'un objet $x_i \in O$ possède un motif m si, $\forall p_j \in m, x_i R p_j$.

Pour la base de données en exemple, on a donc :

0=1 Motif de taille 0 : \emptyset .

$C_5^1 = 5$ Motifs de taille 1 : $\{a\}, \{b\}, \{c\}, \{d\}, \{e\}$, qu'on notera, pour simplifier, a, b, c, d et e.

2=10 Motifs de taille 2 : ab, ac, ad, ae, bc, bd, be, cd, ce et de.

3=10 Motifs de taille 3 : abc, abd, abe, acd, ace, ade, bcd, bce, bde, et cde.

4=5 Motifs de taille 4 : abcd, abce, abde, acde et bcde.

$C_5^5 = 1$ Motif de taille 5 : abcde.

Dans cette base, x_1 possède les motifs : $\emptyset, a, c, d, ac, ad, cd$ et acd .

$|p|$ motifs, on va chercher ceux qui apparaissent fréquemment. Pour cela, on va introduire les notions de connexion de Galois et de support d'un motif. [8]

5.3. Itemset

Un Itemset désigne un ensemble d'objets ou d'articles [5][7].

5.4. Itemset

Un K-Itemset est un Itemset de k Items [5][7].

5.5. Support d'un Itemset

Soit A un Itemset de n éléments, dans une base de données transactionnelle D, le support de A est le nombre de transactions dans D incluant A divisé par le nombre total des transactions de D [5][7] :

$$\frac{|D_A|}{|D|}$$

5.6. Support d'une règle d'association

Dans une base de données D , le support d'une règle d'association $A \Rightarrow B$ est le nombre de transactions qui contiennent A et B divisé par le nombre total des transactions [5][7] :

$$\frac{|A \cap B|}{|D|}$$

5.7. Confiance

La confiance d'une règle d'association $A \Rightarrow B$ est le rapport entre le nombre de transactions de D contenant $A \cup B$, et le nombre de transactions de D contenant A . C'est-à-dire que [5][7] :

$$\frac{|A \cap B|}{|A|}$$

Cette formule est équivalente à :

$$\frac{\text{support}(A \cup B)}{\text{support}(A)}$$

5.8. Itemset Fréquent

Un Itemset A est fréquent si et seulement si son support est supérieur à un support minimum défini par l'utilisateur [5][7].

5.9. Transaction

On considère un ensemble $E = \{E_1, \dots, E_n\}$ de n éléments (items) distincts. On appelle une transaction T le sous ensemble E' inclus dans E . Dans une base de données D , chaque transaction est identifiée par une clé unique [5][7].

6. Les motifs fréquents :

C'est un ensemble d'éléments présents dans un nombre « suffisamment grand » de lignes d'une base de données. Le cadre mathématique qui suit est classique, et formalise celui défini par R. Agrawal, T. Imielinski et A. Swami en 1993 [AGR 93] et raffiné par Agrawal et R. Srikant l'année suivante [AGR 94].

7. Les règles d'association :

7.1. Définition 1 :

Les règles d'association, étant une méthode d'apprentissage non supervisé, permettent d'extraire à partir d'un ensemble de transactions, un ensemble de règles qui expriment une possibilité d'association entre différents attributs. [3]

Agrawal et Al (1993) ont été les premiers à évoquer le problème des règles d'association dans la littérature scientifique, en introduisant l'algorithme AIS. Les auteurs ont également introduit les concepts de base de l'extraction des règles à savoir la notion de support minimum et de confiance minimum. Ces deux mesures statistiques sont utilisées comme critères pour éviter de considérer les règles non significatives.

7.2. Définition 2 :

Soit **A** et **B** deux sous-ensembles d'éléments qui appartiennent à un ensemble **M**. On appelle $A \Rightarrow B$ la règle d'association ou la relation entre A et B, signifie que si A se trouve dans une transaction, B peut apparaître dans cette même transaction. L'évaluation de cette relation se base principalement sur le calcul de deux paramètres qu'on appelle le support et la confiance [5][7].

8. Algorithme Apriori :

En 1994, Agrawal et Al ont présenté l'algorithme Apriori, un algorithme d'extraction de règles d'association qui permet de trouver plusieurs items dans la conséquence de la règle d'association trouvée. [5]

L'algorithme Apriori proposé par Agrawal est un algorithme de base qui permet d'extraire des motifs fréquents dans une base ayant plusieurs milliers d'attributs et plusieurs millions d'enregistrements. L'idée est d'effectuer une extraction par niveaux selon le principe suivant [8]:

- On commence par chercher les motifs fréquents de longueur 1 ;
- On combine ces motifs pour obtenir des motifs de longueur 2 et on ne garde que les fréquents parmi eux ;
- On combine ces motifs pour obtenir des motifs de longueur 3 et on ne garde que les fréquents parmi eux ;
- ... continuer jusqu'à la longueur maximale.

Cette approche s'appuie sur les deux principes fondamentaux suivants (qui reposent sur la décroissance du support) :

- Tout sous-motif d'un motif fréquent est fréquent.
- Tout sur-motif d'un motif non fréquent est non fréquent.

L'algorithme de référence basé sur cette approche est l'algorithme Apriori. Le pseudo code suivant décrit l'extraction de motifs fréquents selon ce principe :

Algorithme Apriori

Entrées : Base de données de transactions D , Seuil de support minimum σ

Sorties : Ensemble des motifs fréquents

$i \leftarrow 1$

$C_i \leftarrow$ Ensemble des motifs de taille 1 (un seul motif)

Tantque $C_i \neq \emptyset$ faire

Calculer le Support de chaque motif $m \in C_i$ dans la base

$i \mid \text{support}(m) \geq \sigma \}$

$i+1$ Toutes les combinaisons possibles des motifs de F_i de taille $i + 1$

$i \leftarrow i + 1$

Fin Tantque

i

9. Conclusion :

Dans ce chapitre on a commencé par définir brièvement le data mining et ses principales tâches, le processus ECD et ses étapes, la classification et ses deux types supervisée, et la classification non supervisée, on a mis l'accent sur quelques définitions de termes et notions, on a proposé quelques définitions de termes utilisés dans l' ECD , les motifs fréquent, les règles d'association et finalement on a proposé comme algorithme Apriori qui est l'algorithme de référence pour la recherche des règles d'associations.

Chapitre 2

Cas d'étude

1. Introduction :

Commencé dès les premières semaines de la vie fœtale, le développement anatomique de la glande mammaire suit les étapes hormonales de la vie de la femme : puberté, grossesse, allaitement, sevrage, ménopause.

2. Anatomie descriptive et fonctionnelle

2.1. Situation du sein :

Les seins occupent la partie antéro-supérieure du thorax, de part et d'autre du sternum en avant des muscles pectoraux, en regard de l'espace compris entre la 3ème et la 7ème côte, le mamelon se situant au niveau de la 9ème vertèbre dorsale. En position debout, sous l'influence de son propre poids, le sein tombe légèrement, ce qui crée le sillon infra-mammaire entre la moitié inférieure du sein et le thorax. [9]

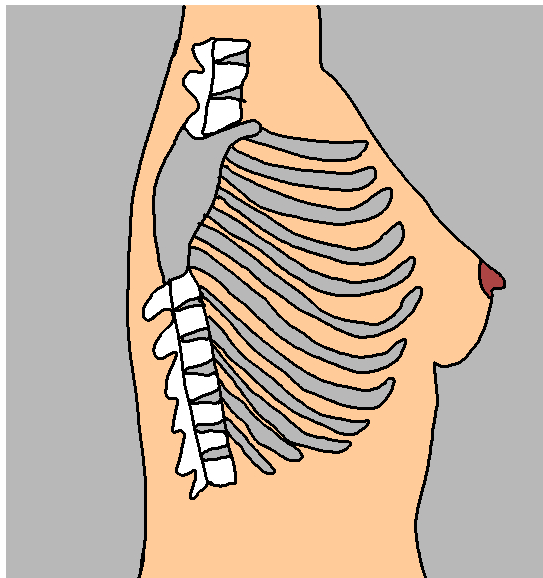


Figure 4 : Situation du sein

L'angle pariéto-mamelonnaire permet d'évaluer le cas échéant le degré de ptose : Sommet = le sillon intra-mammaire

1 côté = ligne joignant le mamelon au sillon

L'autre côté = plan thoracique

Normalement cet angle est de 100 à 110° sur femme debout, Dans les ptoses importantes, il peut atteindre 5°.

2.2. La forme :

Elle est semi-sphérique chez les femmes européennes et asiatiques, plutôt conique chez les femmes africaines.

La taille est d'environ 12 cm en hauteur, 10 cm de largeur et 5 à 6 cm d'épaisseur. Les 2 mamelons sont distants d'environ 20 cm.

A la naissance : les seins ne mesurent que 8 à 10 mm de diamètre.

Entre 9 et 10 ans : surélévation du mamelon suivi de l'élargissement de l'aréole.

Vers 13 ans : bombement de l'aire mammaire et pigmentation de l'aréole.

Vers 18 ans : le sein prend la forme sphérique celle de l'adulte.

Les seins sont fréquemment asymétriques.

2.3. Le poids :

Le poids du sein varie selon la morphologie de la femme, la grossesse et la lactation : de 200 g chez la jeune fille, il peut atteindre 500 g chez la femme allaitante et 900 g dans certains cas.

A la naissance : les seins pèsent chacun 30 à 60 cg.

2.4. La consistance :

La consistance est irrégulière, en particulier lors de la grossesse et de l'allaitement. En comprimant le sein contre la paroi thoracique, la consistance est plus homogène.

A la ménopause : La glande mammaire s'atrophie mais le volume du sein ne diminue pas toujours, compensé par l'augmentation des tissus graisseux.

3. Structure du sein :

De structure complexe composée essentiellement d'un tissu adipeux (graisse) plus au moins important, tissus conjonctif, la glande mammaire, plaque aréolo mamelonnaire.

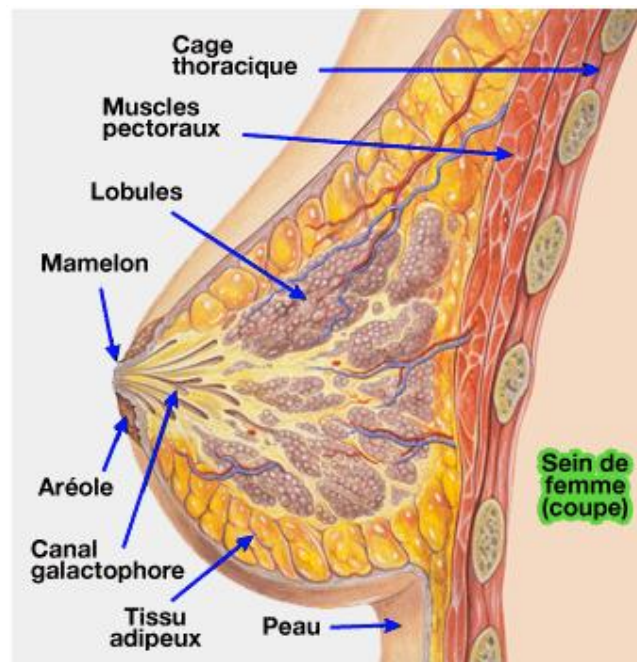


Figure 7 : Structure du sein

3.1. Le tissu adipeux et conjonctif :

Etroitement liée au tissu glandulaire, la quantité de tissu adipeux est en grande partie responsable du volume et la forme des seins.

3.2. La peau et la plaque aréolo-mamelonnaire :

Le revêtement cutané est épais en périphérie et s'amincit au voisinage de l'aréole.

3.3. Le mamelon :

Est cylindrique, conique et pigmenté, mesure 10 mm de hauteur et de largeur.

Les variations de forme du mamelon sont nombreuses :

- Mamelon plat ou court
- Mamelon invaginé
- Mamelon ombiliqué
- + Variations de largeur et de volume

Rôle : A la surface du mamelon, on trouve les orifices d'abouchement (les pores) des canaux galactophores, ces derniers acheminent le lait du lieu de production (acini) jusqu'au mamelon.

3.4. L'aréole :

Est un disque cutané (peau fine) de 15 à 30 mm de diamètre, pigmentée, richement

vascularisé, contient plusieurs terminaisons nerveuses, des muscles lisses et un système abondant de vaisseaux lymphatiques, centré par le mamelon.

Sa surface est irrégulière, constituée de petites saillies appelées les tubercules de Morgagni.

Les tubercules de Montgomery : ce sont des glandes sébacées (1-2 mm) qui, pendant la grossesse sont plus volumineuses et plus nombreuses.

Rôle : sert de repère pour l'enfant lors de la succion grâce à leurs :

Couleur : pigmentation

Chaleur : vascularisation

Et odeur : hormonale

Le mamelon et l'aréole forment une unité appelée la plaque aréolo mamelonnaire.

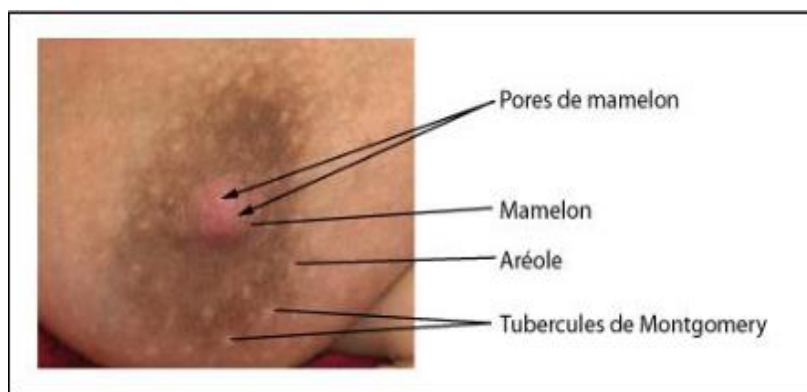


Figure 8 : la plaque aréolo mamelonnaire

4. La glande mammaire :

Dans chaque sein, la glande mammaire est une masse de densité variable, discoïde aplatie d'avant en arrière, de contour irrégulier. Elle est organisée en grappe de 10 à 20 lobes. Chacun est composé de 20 à 40 lobules. La mesure d'un lobule est entre 1.5 et 3 cm de longueur et entre 1 et 2 cm de diamètre. Et chaque lobule contient 10 à 100 alvéoles, chaque lobule mesure entre 0.1 et 1 mm de diamètre.

Chaque sein contient 3 000 à 100 000 alvéoles qui sont les unités sécrétoires.

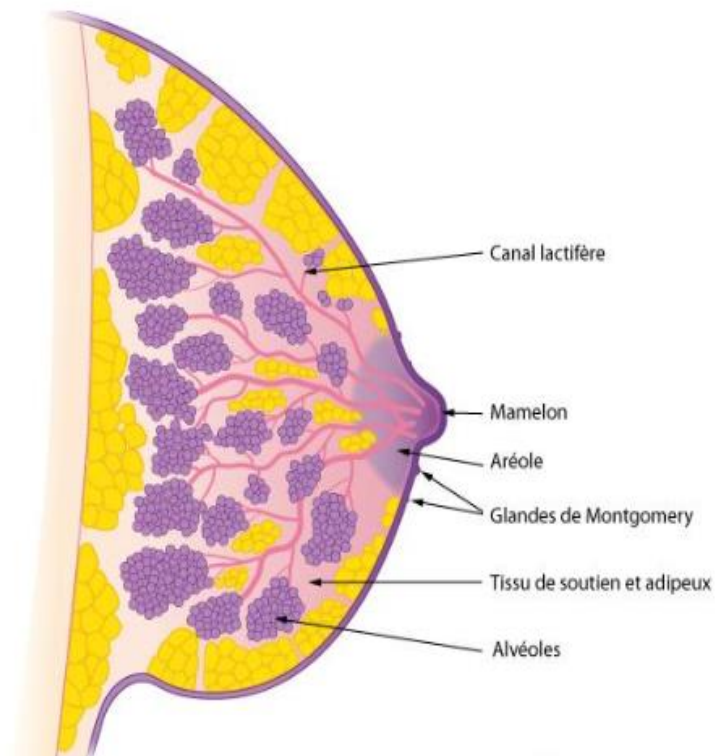


Figure 9 : la glande mammaire

- L'unité fonctionnelle de base est **l'acinus ou alvéole**. L'alvéole est une cavité arrondie en forme de cul de sac qui constitue la partie sécrétrice de lait.
- Chaque acinus se draine par un canal intra lobulaire, galactophore ou alvéolaire ou canal de troisième ordre, ou conduit lactifère.
- Les acini et les canaux intralobulaires forment un **lobule** qui se draine par un **canal interlobulaire** (canal galactophore de **deuxième ordre**)
- Plusieurs lobules se réunissent pour former un **lobe glandulaire** qui se draine par un **canal galactophore de premier ordre** de 2mm en moyenne.

Les canaux galactophores s'élargissent pour former les sinus lactifères de 1cm de diamètre et de 2mm en repos puis se rétrécissent et s'abouchent à la surface du mamelon par un pore de 0.4 à 0.7 mm de diamètre.

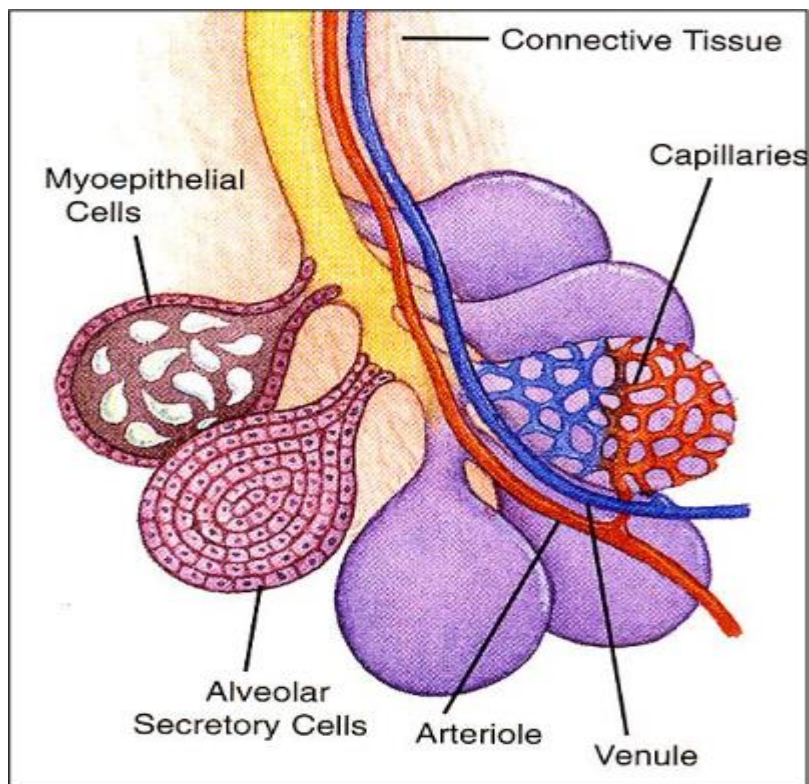


Figure 10 : l'alvéole

La glande mammaire se développe et fonctionne sous l'influence des hormones sexuelles fabriquées par les ovaires. Ces hormones sont de deux types :

Les œstrogènes, qui permettent notamment le développement des seins au moment de la puberté et jouent un rôle important tout au long de la grossesse (assouplissement des tissus, augmentation du volume sanguin nécessaire à l'alimentation du bébé, etc.) ;

La progestérone qui joue notamment un rôle dans la différenciation des cellules du sein et sur le cycle menstruel, en préparant par exemple l'utérus à une éventuelle grossesse (densification et développement de la vascularisation la muqueuse de l'utérus).

4.1. Vascularisation artérielle et veineuse du sein :

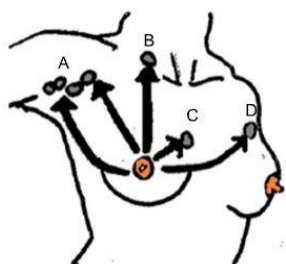
Le sein est parcouru de vaisseaux sanguins et de vaisseaux lymphatiques. Les ganglions et les vaisseaux lymphatiques composent le système lymphatique qui aide notamment à combattre les infections.

Les ganglions lymphatiques du sein sont principalement situés :

- Au niveau de l'aisselle (ganglions axillaires) ;

- Au-dessus de la clavicule (ganglions sus-claviculaires); sous la clavicule (ganglions sous-claviculaires ou infra-claviculaires) ;
- A l'intérieur du thorax, autour du sternum (ganglions mammaires internes).

4.2. Réseaux lymphatiques



- A- groupe axillaire homo-latéral
- B- groupe supra-claviculaire
- C- groupe thoracique interne
- D- groupe axillaire contro-latéral

Figure 11 : réseau lymphatiques

4.2.1. Le réseau lymphatique cutané :

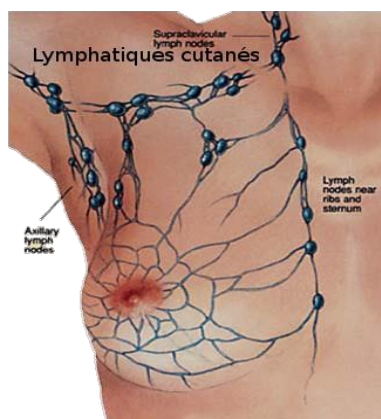


Figure 12 : Réseau lymphatiques cutané

Il existe un double réseau :

- Le plexus superficiel ou dermique
- Le plexus profond ou sous dermique. Ce réseau communique avec celui des territoires voisins.

Il y a deux types de collecteurs :

- Les collecteurs principaux se dirigent vers les ganglions axillaires
- Les collecteurs accessoires se dirigent vers les voie sus claviculaire, la voie mammaire interne et vers le sein opposé.

4.2.2. Le réseau lymphatique glandulaire :

Il existe un réseau superficiel et un réseau profond anastomosés. Ils se drainent vers deux types de collecteurs : certains suivent les galactophores se jettent dans le plexus sous aréolaire et d'autres quittent la glande par sa périphérie.

Les collecteurs se drainent vers les nœuds axillaires et nœuds mammaires internes.

4.2.3. Les ganglions axillaires

Il existe 5 groupes : inférieur, mammaire externe, scapulaire, central, sous claviculaire, 10 à 40 ganglions. Ils forment une masse continue. Il est important de comprendre le système lymphatique parce que c'est un des chemins par lequel le cancer du sein peut disséminer. Les vaisseaux lymphatiques transportent un liquide transparent nommé « la lymphe ». Les cellules du cancer du sein peuvent entrer dans les vaisseaux lymphatiques et se développer dans les ganglions lymphatiques. Ce sont des noyaux de drainage lymphatique sur le trajet des vaisseaux lymphatiques. La majorité des vaisseaux lymphatiques du sein sont connectées avec les ganglions lymphatiques qui se trouvent au niveau de l'aisselle dans la poitrine et au-dessous de la clavicule. [9]

5. Histoire et définition du cancer

Connu depuis au moins 1600 avant JC, le cancer est une pathologie redoutée depuis longtemps. Ce sont d'ailleurs les prêtres d'Esculape et Hippocrate qui l'ont nommée ainsi, le mot cancer dérivant du grec **Karkinos** et du latin **Cancros** qui signifient crabe.

Ce terme de cancer fait penser à la bête carapacée qui plante ses pattes et ses pinces dans la chair vive, provoquant des douleurs rongeantes et lancinantes comparables à celles provoquées par le cancer.

Une seconde version de l'origine étymologique de ce terme se base plus sur l'aspect du cancer, qui lorsqu'il est très évolué et que les veines courent autour de la tumeur devient comparable au crabe et à ces pattes. [10]

5.1. Qu'est-ce qu'un cancer ?

Le terme de « cancer » recouvre un large éventail de maladies différentes qui présentent néanmoins certaines caractéristiques communes :

- Des cellules normales au départ se multiplient de façon incontrôlée ; elles prolifèrent et se

transforment en cellules cancéreuses.

- Les cellules cancéreuses envahissent le tissu sain ; elles le compriment et le détruisent.
- Certaines de ces cellules peuvent se détacher de l'endroit où elles se sont constituées et donner naissance à des foyers secondaires dans d'autres parties du corps : les métastases.

Le terme de « cancer » désigne donc la **prolifération** de cellules qui se multiplient de façon incontrôlée et qui envahissent le tissu sain. Dans le langage courant, on parle aussi de tumeur. Il faut toutefois distinguer les tumeurs bénignes, qui ne mettent pas la vie en danger, des tumeurs malignes, qui peuvent menacer l'existence.

Les tumeurs malignes sont aussi appelées néoplasmes, ce qui signifie « nouvelle formation ».

Il existe plus de deux cents types de cancers différents. On distingue les tumeurs solides, qui se développent à partir des cellules d'un organe et qui forment une masse ou un nodule, et les cancers qui prennent naissance dans le système sanguin ou lymphatique (les leucémies, par ex.). Ceux-ci peuvent notamment se traduire par un gonflement des ganglions lymphatiques, mais aussi par des modifications de la formule sanguine.

Les tumeurs malignes solides qui se forment à partir de tissus superficiels comme la peau, les muqueuses ou les glandes sont appelées carcinomes. Le cancer du sein se classe dans cette catégorie. Les carcinomes représentent la grande majorité des tumeurs malignes.

Les tumeurs malignes solides qui se développent dans le tissu conjonctif, le tissu adipeux, les cartilages, les muscles, les os ou les vaisseaux sont qualifiées de sarcomes.

Exemple : carcinome mammaire

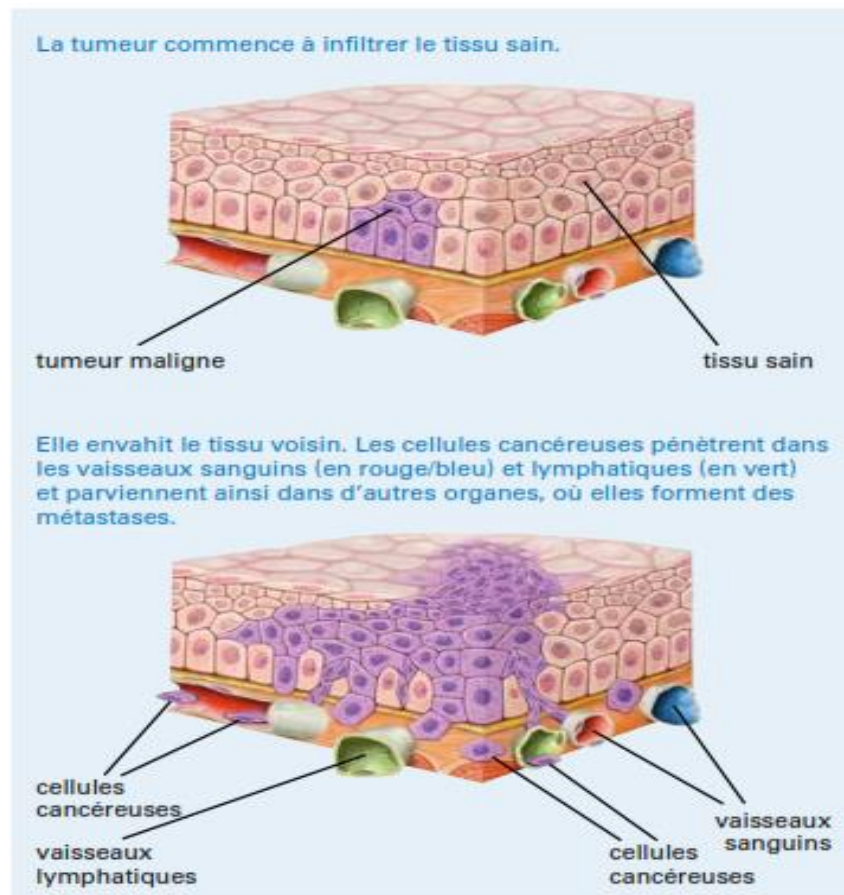


Figure 13 : Formation d'une tumeur

5.2. Le cancer du sein :

Le cancer du sein apparaît lorsque la matière génétique est endommagée ou modifiée dans les différentes cellules somatiques, initialement saines, de la glande mammaire. Les causes de la modification pathologique de la substance génétique de ces cellules demeurent généralement incertaines. Au fil des années, voire même des décennies, se développent à partir de telles cellules endommagées des cellules cancéreuses qui n'assument plus les fonctions qui lui étaient originellement assignées. Les cellules dégénèrent, échappent à la surveillance normale du système immunitaire de l'organisme et se multiplient de manière incontrôlée, elles finissent par donner naissance à une tumeur. Quand celle-ci est maligne, on parle de cancer du sein, ou carcinome mammaire.

Les cellules cancéreuses ne respectent plus aucune limite entre les tissus ou les organes et se propagent dans le tissu environnant pour le détruire. Par ailleurs, il arrive que quelques cellules cancéreuses se détachent de la tumeur et soient transportées par voie lymphatique ou

sanguine dans d'autres organes, où elles peuvent se développer en tumeurs filles, appelées métastases.

5.3. Les différentes formes de cancer du sein :

Le cancer du sein peut apparaître à différents endroits dans le sein. Le plus souvent, il se développe dans les canaux galactophores (carcinome canalaire), ils représentent environ 80% des cancers du sein ;

Ou dans les lobules (carcinome lobulaire), ils constituent 5 à 15 % de tous les cancers du sein. Il existe encore d'autres types de cancer du sein, qui sont plus rares.

5.3.1. In situ ou invasif :

Certaines lésions ne sont pas encore des cancers à proprement parler, mais des stades précurseurs du cancer, ou pré cancéreuses. Elles peuvent, après des années bien souvent, se transformer en cancer (invasif), mais ce n'est pas systématiquement le cas. On parle de tumeurs in situ, c'est à-dire limitées à l'endroit où elles se sont formées.

Les carcinomes invasifs sont souvent bien localisés (in situ) dans un premier temps, mais ils finissent par se propager au tissu voisin. Les cellules cancéreuses des carcinomes invasifs peuvent parvenir dans d'autres régions du corps en passant dans le système lymphatique ou sanguin et y former des métastases au stade précoce déjà.

5.3.2. Tumeurs bénignes et tumeurs malignes :

En grandissant, les tumeurs bénignes compriment le tissu sain, mais elles ne l'envahissent pas et ne forment pas non plus de métastases. Suivant leur localisation, elles peuvent cependant provoquer des troubles importants en écrasant le tissu normal ou en entraînant le rétrécissement d'un canal comme un nerf ou un vaisseau sanguin.

5.4. Corrélation entre l'anatomie et les images mammographiques :

L'image mammographique est le résultat d'atténuation d'un faisceau de rayons X traversant les différents tissus mammaires. L'atténuation de ce faisceau dépend essentiellement de la composition des tissus traversés. En effet, la graisse est considérée comme une zone radio transparente vu qu'elle a une densité physique très légère. De ce fait elle apparaît très sombre sur un cliché mammographique. En revanche, les zones radio opaques apparaissent claires et

correspondent au tissu fibroglandulaire et au calcium qui est le composant essentiel des lésions mammaires.

La radio opacité et l'aspect sur le cliché mammographique. En rassemblant les informations concernant l'anatomie et la radio transparence, on peut confirmer que l'aspect général d'une mammographie est sombre alors que les zones contenant des microcalcifications ou des masses (composées de calcium) sont plus claires.

5.5. Les anomalies de la plaque areolo-mamelonnaire :

La détection d'anomalies de la région aréolo mamelonnaire peut être difficile en raison de l'anatomie complexe de cette région. Une bonne compréhension de son anatomie, des processus bénins et malins qui la touchent, de même que les images spécifiques à chacun sont nécessaires pour porter le diagnostic approprié et, si nécessaire, déterminer de l'intervention à effectuer.

Les termes "**rétraction**" et "**inversion**" du mamelon sont souvent utilisés de façon interchangeable : cela est inexact. Le terme "**inversion**" (1) devrait être utilisé quand seule une fente est repliée vers l'intérieur alors que le terme "**rétraction**" (2) s'applique dans les cas où le mamelon entier ou en parties emble attiré vers l'intérieur du sein, parfois même sous la surface du sein.

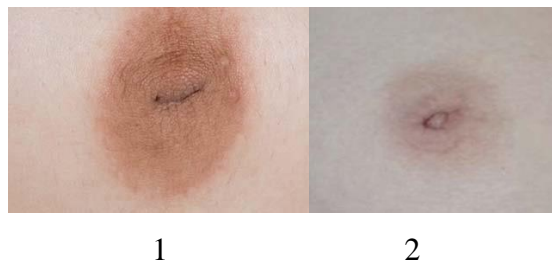


Figure 14 : anomalie de la plaque aréolomamelonnaire

L'inversion et la rétraction peuvent être tous les deux d'origine congénitale ou acquise, unilatérale ou bilatérale. En général, une **inversion** en fente centrale et symétrique indique un processus bénin alors qu'une rétraction de tout le mamelon avec distorsion de l'aréole doit faire évoquer un processus malin.

5.5.1. Changements cutanés :

Des processus bénins et malins peuvent produire des changements visibles à la peau du mamelon incluant érythème, gale, croûte, fissures, vésicules ou une combinaison de ces

derniers. Ces changements peuvent être accompagnés de **douleur** ou **prurit**.

Le diagnostic correct est important en raison de la possibilité d'un néoplasie mammaire. Une analyse cytologique normale d'un écoulement peut être erronée, et l'imagerie peut être normale. L'histoire clinique peut être utile, mais l'analyse histologique d'un spécimen de **biopsie** est souvent nécessaire pour un diagnostic précis.

5.5.2. Masse :

Les masses du mamelon et des tissus sous-aérolaires sont communément bénignes.

D'où l'intérêt d'un examen clinique minutieux complété par d'autres examens complémentaires.

5.5.3. Ecoulement mamelonnaire :

L'écoulement mamelonnaire est un symptôme fréquent.

L'analyse minutieuse de l'aspect de l'écoulement et la meilleure connaissance de l'histoire clinique permet d'identifier la nature de l'écoulement et avoir une meilleure décision thérapeutique

5.6. Nodule mammaire :

C'est le symptôme le plus fréquent du cancer du sein, qui est une néoformation au sein de la glande mammaire, un examen clinique minutieux avec un bon interrogatoire permet de mieux orienter notre enquête étiologique et avoir le diagnostic.

6. Les déclencheurs de cancer du sein :

La modification du noyau de la cellule qui entraîne la multiplication incontrôlée de celle-ci est la cause fondamentale de tous les cancers. La cause de ce dérèglement n'est pas connue. En ce qui concerne plus particulièrement le cancer du sein, des études scientifiques ont cherché à savoir si certaines caractéristiques propres à l'individu où des comportements étaient plus fréquemment observés chez des femmes ayant développé un cancer du sein que chez les autres femmes. On appelle ces caractéristiques des facteurs de risque

Un « facteur de risque » est défini par tout facteur dont la présence induit une augmentation de la probabilité d'apparition de la maladie. Ainsi, éviter ou éradiquer de tels facteurs participe à la prévention primaire.

Parmi les facteurs de risque, on peut distinguer les facteurs « extrinsèques » (ou exogènes) et les facteurs « intrinsèques » (ou endogènes).

Les facteurs de risque « extrinsèques » ne sont pas liés directement à l'individu mais à son

environnement (par exemple la pollution, le tabagisme passif, une activité professionnelle particulière).

Les facteurs de risque « intrinsèques » sont propres à l'individu (par exemple son hérédité, son âge, son sexe, son comportement...). On peut parfois essayer de lutter contre ces facteurs (modifier des comportements par exemple), mais le plus souvent on ne peut que les constater ; la connaissance de ces facteurs permet cependant de définir des populations dites « à risque » qui peuvent être dépistées précocement, on réalise alors une action de prévention secondaire. [14]

Le cancer du sein est une **maladie multifactorielle**. Cela signifie que plusieurs facteurs influents sur le risque de sa survenue. On parle de facteurs de risque.

On connaît aujourd'hui un certain nombre de facteurs de risque du cancer du sein même s'il existe encore des incertitudes quant à l'implication et au poids de plusieurs de ces facteurs.

Une personne qui possède un ou plusieurs facteurs de risque peut ne jamais développer de cancer. Inversement, il est possible qu'une personne n'ayant aucun facteur de risque soit atteinte de ce cancer. [11]

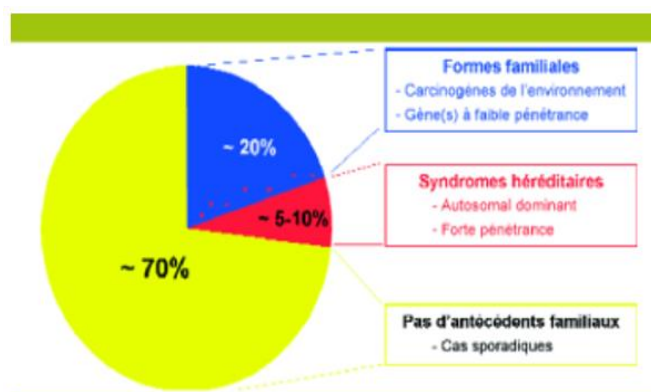


Figure15 : Prévalence de l'atteinte familiale chez les femmes ayant développé un CS

- **Les facteurs de risque lié à l'âge** : En effet, près de 80% des cancers du sein se développent après 50 ans.
- **Les facteurs de risque liés à certains antécédents médicaux personnels et familiaux.**
- **Les facteurs de risque liés aux modes de vie** : tels que la consommation d'alcool et de tabac, un surpoids ou encore pas ou peu d'activité physique peuvent favoriser l'apparition d'un cancer du sein. [11]

Il est possible de les classer en fonction de leur importance : [12]

Risque principal

Risque important

Risque d'importance mineure ou modérée

Risque dont l'importance reste à déterminer ou à clarifier

6.1. Facteurs de risque :

6.1.1. L'âge :

Le risque d'avoir un cancer du sein augmente avec l'âge même s'il peut atteindre des femmes à des âges très différents.

Le risque de cancer du sein chez les jeunes femmes est peu élevé.

En France environ 10% des cas de cancer du sein se manifestent chez les femmes âgées de moins de 35 ans et près de 20% avant 50 ans. [11]

En effet, l'incidence augmente avec l'âge à partir de 45 ans, pour être maximale entre 65-74 ans.

En Algérie Près de 40 % des cancers féminins sont représentés par le cancer du sein. Son incidence continue son ascension et se rapproche inexorablement de l'incidence mondiale (80 nouveaux cas pour 100 000 femmes).[13]

L'âge moyen des sujets atteints de cancer du sein est de 52.5 ± 12.1 ans et un âge médian de 49 ans. L'incidence du cancer du sein augmente notablement dès 35-39 ans et enregistre deux pics à 45-49 ans et 50-44 ans et diminue au-delà de 74 ans. [13]

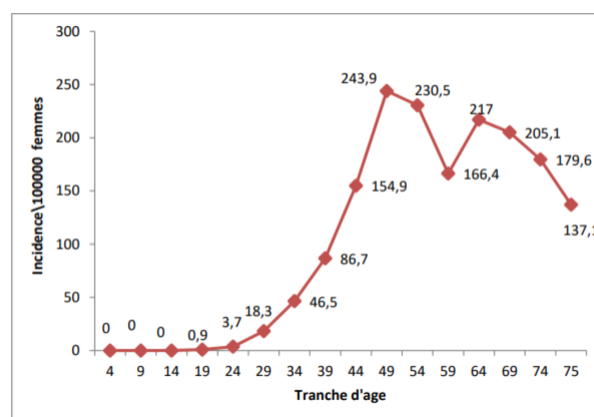


Figure 16 : Répartition des incidences du cancer du sein par tranches d'âges chez les femmes à Alger 2016 (source : INSP)

6.1.2. Antécédents personnels : on à

6.1.2.1. Antécédents de cancer du sein : En plus du risque de récurrence du cancer au niveau du sein traité, une femme qui a eu un cancer du sein a un risque 3 à 4 fois plus élevé de développer un nouveau cancer du sein qu'une femme du même âge. Ce risque justifie un suivi régulier et prolongé.

Le risque de développer un cancer du sein infiltrant est 8 à 10 fois plus importantes chez les femmes qui ont déjà eu un cancer canalaire in situ (CCIS) ou un cancer lobulaire in situ (CLIS). Par ailleurs, la présence d'un CLIS fait augmenter le risque de cancer dans les deux seins. [11]

6.1.2.2. Affections bénignes du sein : Les lésions pré-malignes ou bénignes du sein se caractérisent par des changements histologiques, associés à un risque augmenté de transformation maligne.

Depuis l'introduction de la mammographie de dépistage, on détecte de plus en plus ces modifications histologiques et donc on estime et on prévient le risque de leur transformation maligne. On peut identifier parmi ces lésions bénignes : [11]

6.1.2.3. L'hyperplasie canalaire simple : L'hyperplasie canalaire simple se caractérise par une prolifération intracanaire de cellules épithéliales, entourées de cellules myoépithéliales et d'une membrane basale intacte).

Elle peut se voir à tout âge, mais elle est moins fréquente après 60 ans.

Elle n'augmente pas le risque absolu de cancer chez les femmes de moins de 55 ans.

L'hyperplasie canalaire simple ne nécessite donc pas de traitement ni de surveillance particulière. [11]

6.1.2.4. L'hyperplasie canalaire atypique : L'hyperplasie canalaire atypique correspond soit à une lésion possédant quelques-unes des caractéristiques du carcinome canalaire in situ.

Le risque de développer un cancer du sein est **quatre à cinq** fois plus élevé que dans la population normale. Et donc elle nécessite un traitement spécifique (exérèse

chirurgicale de la lésion et à une analyse complète de la pièce, afin d'exclure la présence d'un carcinome). [11]

6.1.2.5. Cicatrice radiaire : La cicatrice radiaire se présente comme une image d'allure maligne à la mammographie, alors que cette affection est de nature bénigne en soi.

Une surveillance radiologique est nécessaire. En revanche la présence d'une hyperplasie canalaire atypique ou d'un carcinome canalaire in situ impose l'exérèse chirurgicale. [11]

6.1.2.6. L'adénose sclérosante : L'adénose sclérosante ne montre aucune atypie cytologique. En absence d'une hyperplasie atypique ou d'un carcinome par de multiples biopsies, une surveillance clinique, et radiologique est indispensable. [11]

6.1.2.7. Les atypies épithéliales planes : Ce terme descriptif fait référence à une métaplasie cylindrique, appelée aussi changement cylindrique ou hyperplasie cylindrique.

L'attitude que nous proposons face à une atypie épithéliale plane est une surveillance clinique, et radiologique complétée par une exérèse chirurgicale si présence d'une hyperplasie canalaire atypique ou un carcinome canalaire in situ associé. [11]

6.1.2.8. Les lésions papillaires : Ces lésions se manifestent cliniquement par un écoulement séreux et sanglant.

Histologiquement, le papillome bénin se caractérise par une prolifération intracanaulaire de cellules épithéliales

Le risque non négligeable de sous-estimation d'une lésion papillaire maligne nous fait préconiser une exérèse chirurgicale. [11]

6.1.2.9. Exposition à des radiations médicales : Une irradiation du thorax peut augmenter le risque de cancer du sein. Le niveau de risque est en rapport avec la dose totale reçue et l'âge de la femme.

Les femmes jeunes (avant 30 ans) qui ont eu des radiothérapies répétées du thorax ou un traitement par irradiation (radiothérapie) sur le thorax pour traiter un autre cancer (comme le lymphome hodgkinien par exemple) ont un risque plus élevé de cancer du sein. [11]

6.1.3. Antécédent familiaux :

Près de 20 à 30% des cancers du sein se manifestent chez des femmes ayant des antécédents familiaux de cancers dont des cancers du sein, par exemple plusieurs cas de cancer du sein dans la même famille.

Parfois on ne sait pas exactement si cette disposition familiale est due au hasard, à un mode de vie que des membres de la famille ont en commun, à un facteur héréditaire qui a été transmis des parents à leurs enfants par les gènes ou bien à une association de ces éléments.

De manière générale, les études montrent que :

- Lorsque c'est une parente au premier degré (la mère, la sœur ou la fille) qui a déjà eu un cancer du sein, en particulier si le diagnostic a été posé à un âge jeune (avant 50 ans) avant la ménopause, le risque d'être atteint de ce type de cancer est environ deux fois plus élevé ;
- Lorsque ce sont des parentes au second degré (comme une grand-mère, une tante ou une nièce, d'un côté ou l'autre de la famille) qui ont déjà été atteintes du cancer du sein, ce risque augmente légèrement. [11]

Ces études précisent par ailleurs que le risque de développer un cancer du sein augmente d'autant plus que :

- Une parente au premier degré a été atteinte d'un cancer dans les deux seins (cancer du sein bilatéral) avant la ménopause ;
- Deux membres de la famille ou plus ont été atteintes d'un cancer du côlon ou de l'ovaire ;

- Une parente a été atteinte d'un cancer du sein et d'un cancer de l'ovaire ou de plusieurs cancers différents ;
- Un parent de sexe masculin a été atteint du cancer du sein. [11]

6.1.4. Prédisposition génétique :

Lorsque plusieurs personnes d'une même famille sont atteintes du même cancer, il peut s'agir d'un cancer héréditaire dû à une anomalie au niveau d'un gène (anomalie génétique) qui se transmet d'une génération à une autre. Cette anomalie est encore appelée mutation génétique.

Seule une petite partie des cancers du sein, 5 à 10%, sont héréditaires, c'est-à-dire attribuable à une mutation génétique (qu'elle soit identifiée ou non).

La recherche a permis d'identifier un certain nombre de mutations génétiques favorisant la survenue de cancers du sein. Le plus souvent, celles-ci portent sur des gènes appelés BRCA1 (pour BReast Cancer 1 : gène 1 du cancer du sein) et le BRCA2 (pour BReast Cancer 2 : gène 2 du cancer du sein). Etre porteur d'une mutation sur l'un de ces gènes ne se traduit pas systématiquement par l'apparition d'un cancer, mais augmente le risque d'en développer un. C'est ce que l'on appelle une prédisposition génétique. [11]

6.1.5. Les facteurs de risque liés aux modes de vie :

6.1.5.1. La consommation de tabac : La consommation de tabac est associée à une augmentation du risque de plusieurs cancers dont le cancer du sein.

De récentes études ont montré que des femmes exposées au tabagisme passif (dont l'entourage consomme du tabac) ont un risque de cancer du sein inférieur à celles qui sont exposées au tabagisme actif (qui consomme elle-même du tabac) mais tout de même plus élevé que le risque de femmes jamais exposées au tabac. (Afsset, Inserm, 2008). [11]

6.1.5.2. La consommation d'alcool : La consommation de boissons alcoolisées est associée à une augmentation du risque de plusieurs cancers dont le cancer du sein. Elle augmenterait les taux d'œstrogène qui joue lui-même un rôle important dans le développement des cellules du cancer du sein.

L'augmentation de risque de cancers du sein est significative dès une consommation moyenne d'un verre par jour. Des études de référence attribuent 17% des cancers du sein à la consommation d'alcool régulière, même modérée. [11]

6.1.5.3. Le surpoids/ L'obésité : Le surpoids (IMC compris entre 25 et 29,9) et l'obésité (IMC de 30 ou plus) augmentent le risque de cancer du sein chez la femme ménopausée.

A l'inverse, l'activité physique est associée à une diminution de risque de cancer du sein après la ménopause. [11]

7. Conclusion :

Plusieurs facteurs de risque du cancer du sein ont été clairement mis en évidence. Toutefois, pour la majorité des femmes présentant un cancer du sein, il n'est pas possible d'identifier des facteurs de risque particuliers (IARC, 2008 ; Lacey et al. 2009).

Ont dit que le cancer du sein est une association de facteurs ou déclencheurs, pour cela il serait caractéristique de notre modèle.

Chapitre 3

Conception et mise en œuvre

1. Introduction :

Pour le développement de tout système, la phase de conception est primordiale pour cela, dans ce chapitre on va exposer la conception globale et la conception détaillé de notre système.

Dans le système qu'on a proposé, nous nous somme focalisé sur une étape importante dans le processus du datamining celle de la prédiction des nouveaux modèles, pour notre modèle c'est l'évaluation des critères prédictifs pour une population de Femme ayant un risque modéré de cancer du sein « la classification par les motifs fréquents ».

2. La conception du système

Notre système est globalement composé de cinq composants au niveau du back office.

2.1. La conception générale du système

Le schéma suivant représente l'architecture globale du système proposé (Figure17) :

- Acquisition des données cibles.
- Prétraitement.
- Classification.
- Visualisation.

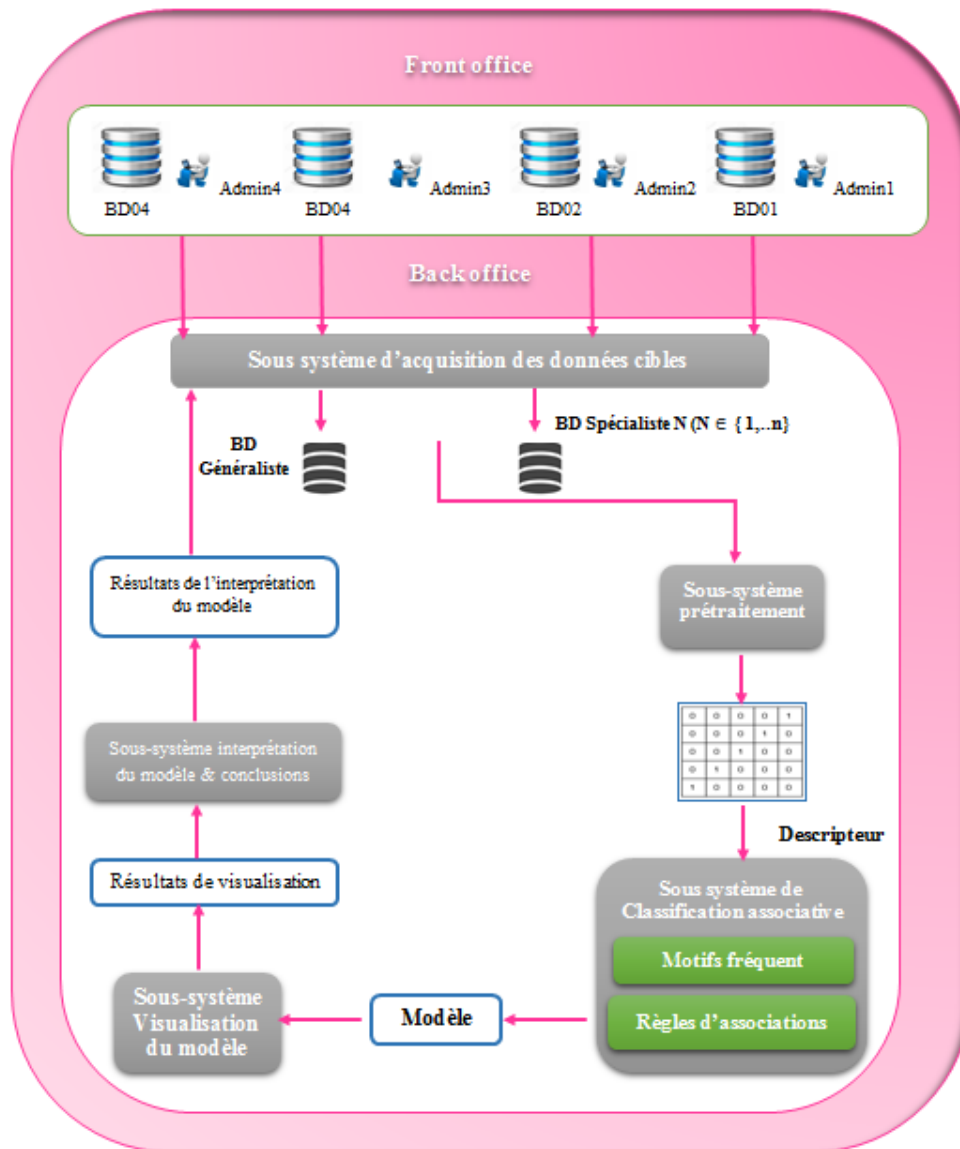


Figure 17 : l'architecture générale du système.

2.2. Conception Détaillée :

Dans cette partie nous allons détailler les différents composants du système en précisant les entrées, les sorties et les traitements de chaque sous-système.

2.2.1. Sous système d'acquisition des données cibles :

2.2.1.1. Acquisition des informations acteurs :

Le sous-système d'acquisition des données acteurs contient les identifications concernant les médecins.

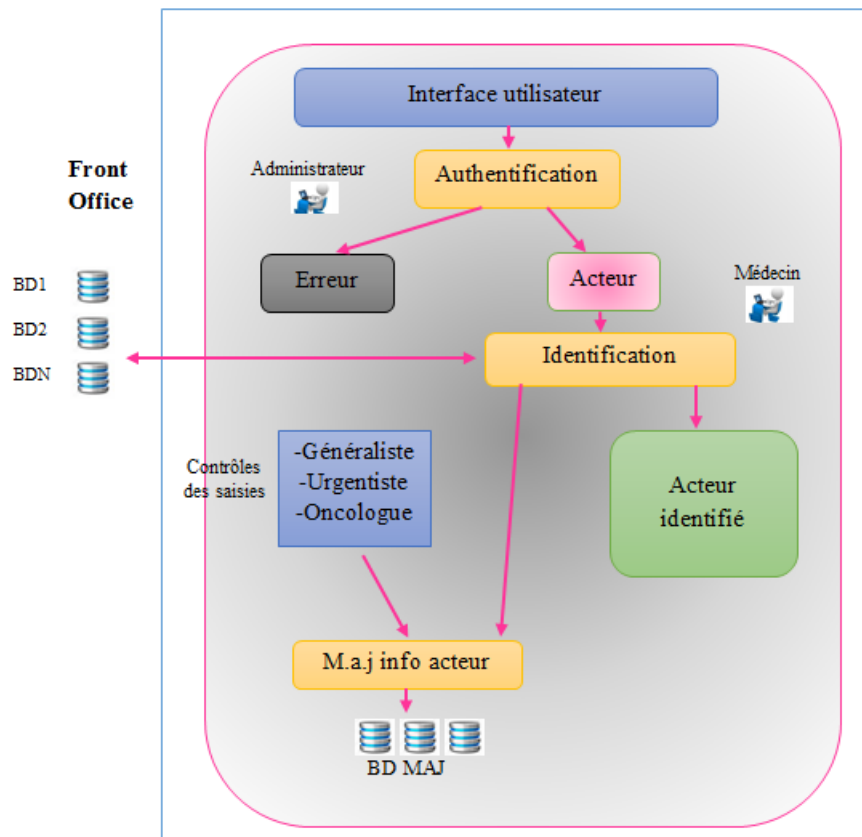


Figure18 : unité d'acquisition des informations des acteurs

2.2.1.2. Acquisition des informations des patients :

Le consultant représente l'unité fondamentale dans le processus de dépistage et l'acquisition des données pour notre système. La figure suivante représente l'unité fonctionnelle identification du consultant(e).

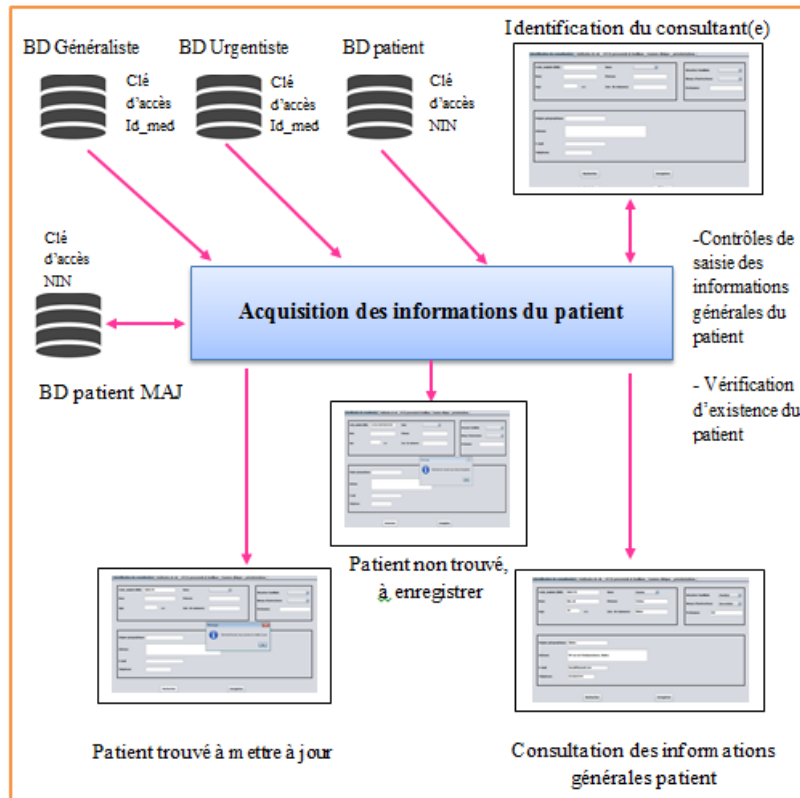


Figure19 : unité fonctionnelle identification du consultant(e).

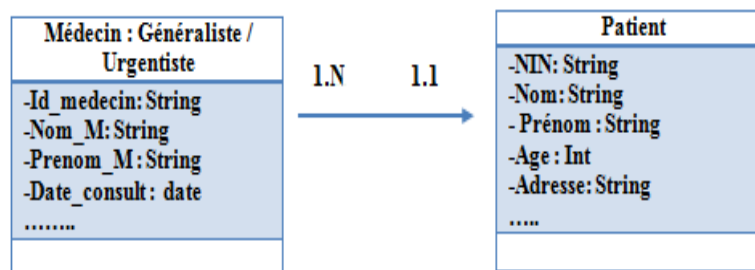


Figure20 : diagramme de structure d'acquisition des informations des patients.

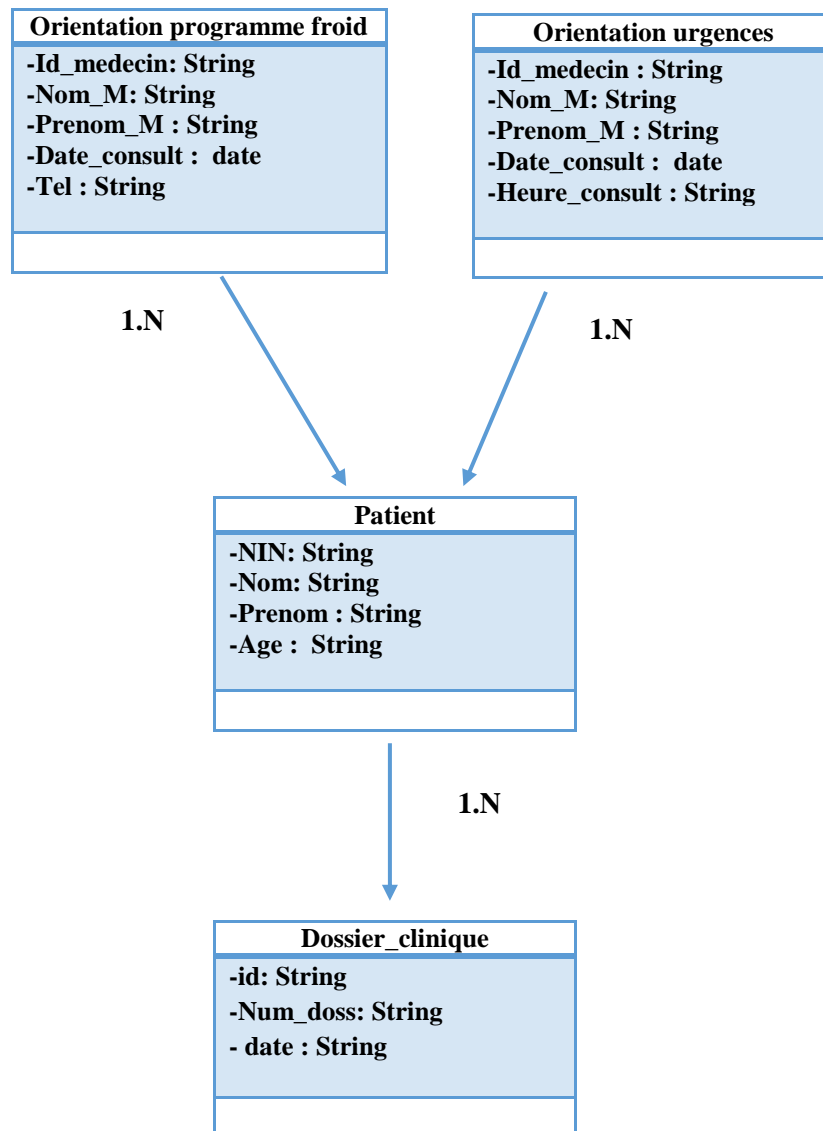


Figure21 : diagramme de structure du processus de consultation.

2.2.1.3. Acquisition des informations des habitudes de vie du consultant :

Cette unité est utilisée pour l'acquisition des informations des habitudes de vie du consultant (tabac, alcool, alimentation, activités physique). La figure suivante représente l'unité fonctionnelle habitudes de vie.

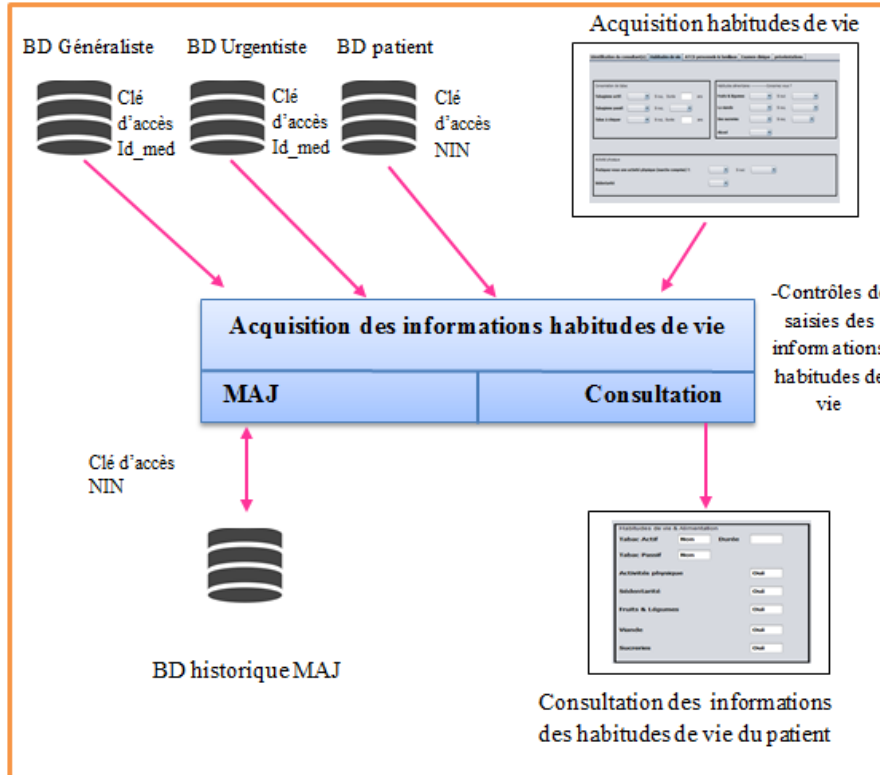


Figure22 : unité fonctionnelle habitudes de vie.

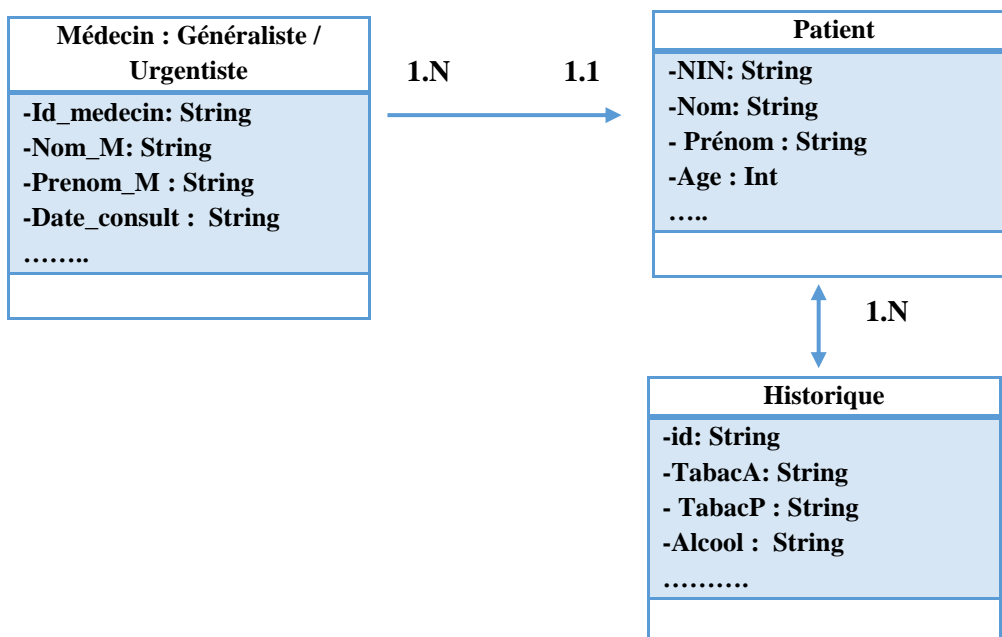


Figure 23 : diagramme de structure habitudes de vie patient.

2.2.1.4. Acquisition des informations des antécédents du consultant :

Cette unité a pour but l'acquisition des informations des antécédents pathologique, physiologie et familiaux du consultant. La figure suivante représente l'unité fonctionnelle antécédents du consultant.

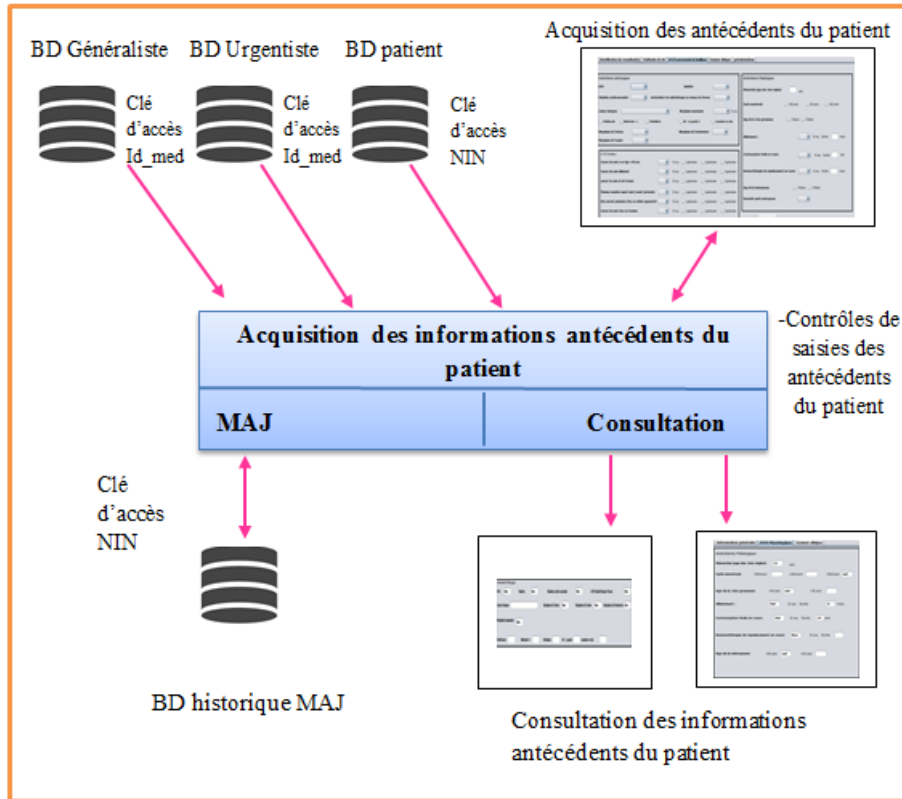


Figure 24 : unité fonctionnelle antécédents personnelle et familiaux.

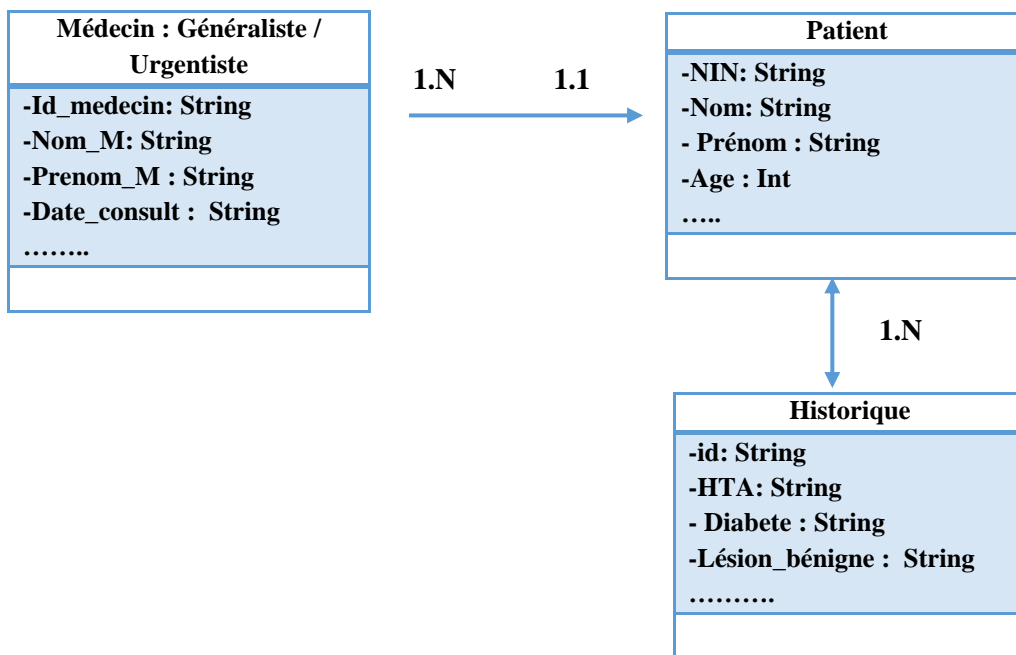


Figure 25 : diagramme de structure antécédents patient.

2.2.1.5. Acquisition des informations de l'examen clinique du consultant :

Dans cette unité les informations du consultant de l'examen clinique réalisé au début par un médecin généraliste ou bien un médecin urgentiste sont saisies. La figure suivante représente l'unité fonctionnelle examens clinique.

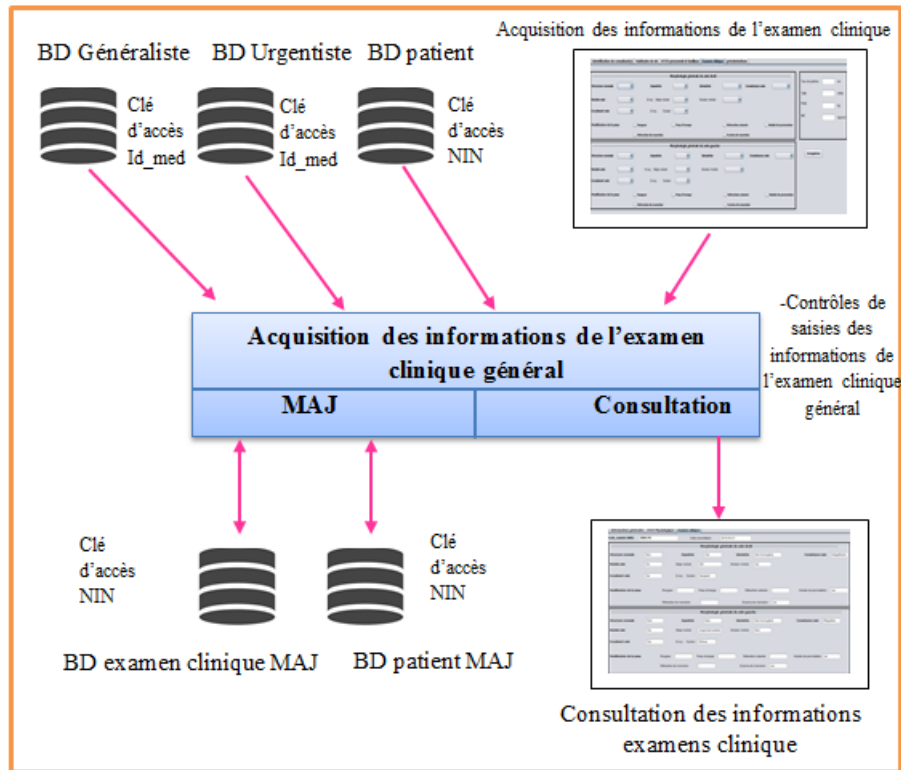


Figure 26 : unité fonctionnelle examens clinique.

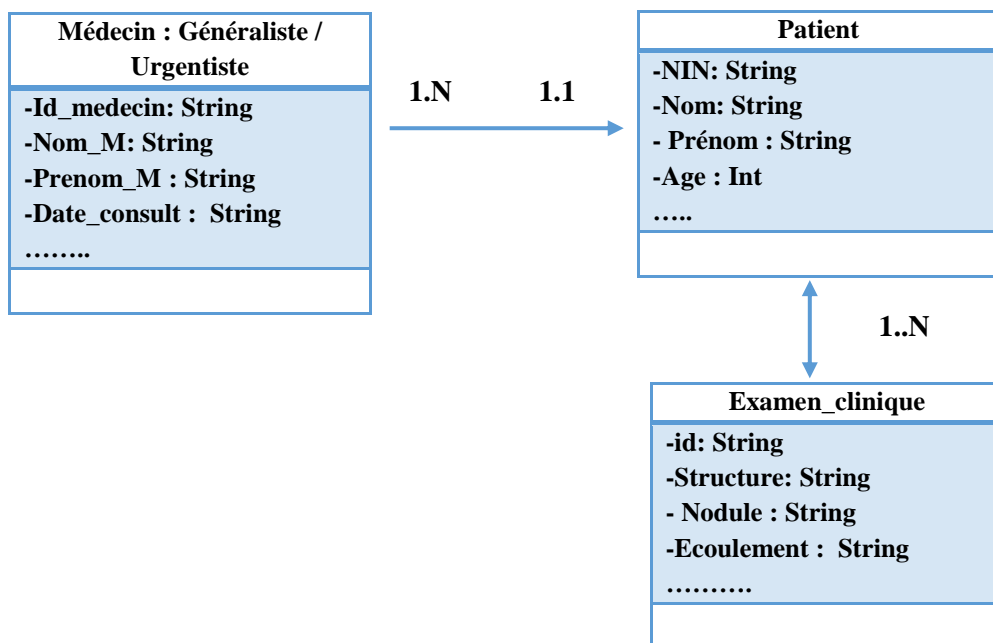


Figure 27 : diagramme de structure examens clinique.

2.2.1.6. Acquisition des informations de pré-orientations du consultant :

Lors de l'orientation du consultant vers un médecin oncologue ou vers un service commun (radio, laboratoire...etc.), c'est dans cette unité ou les informations du consultant et de l'orientation (numéro orientation, nom et prénom oncologue...etc.) sont saisies. La figure suivante représente l'unité fonctionnelle pré-orientations.

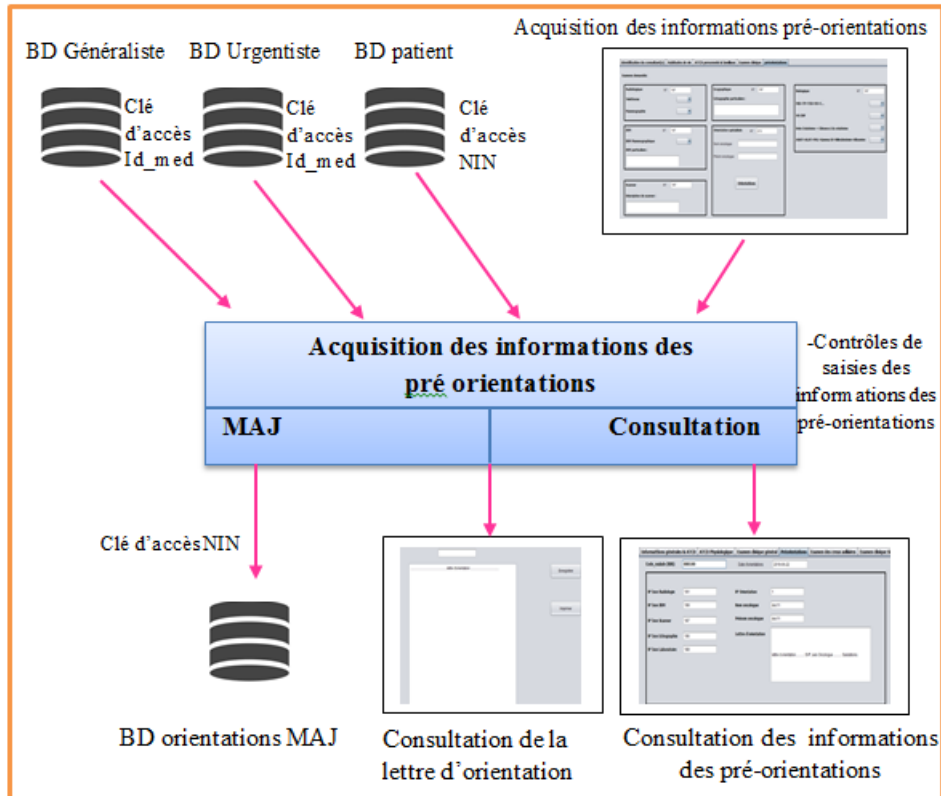


Figure 28 : unité fonctionnelle pré-orientations.

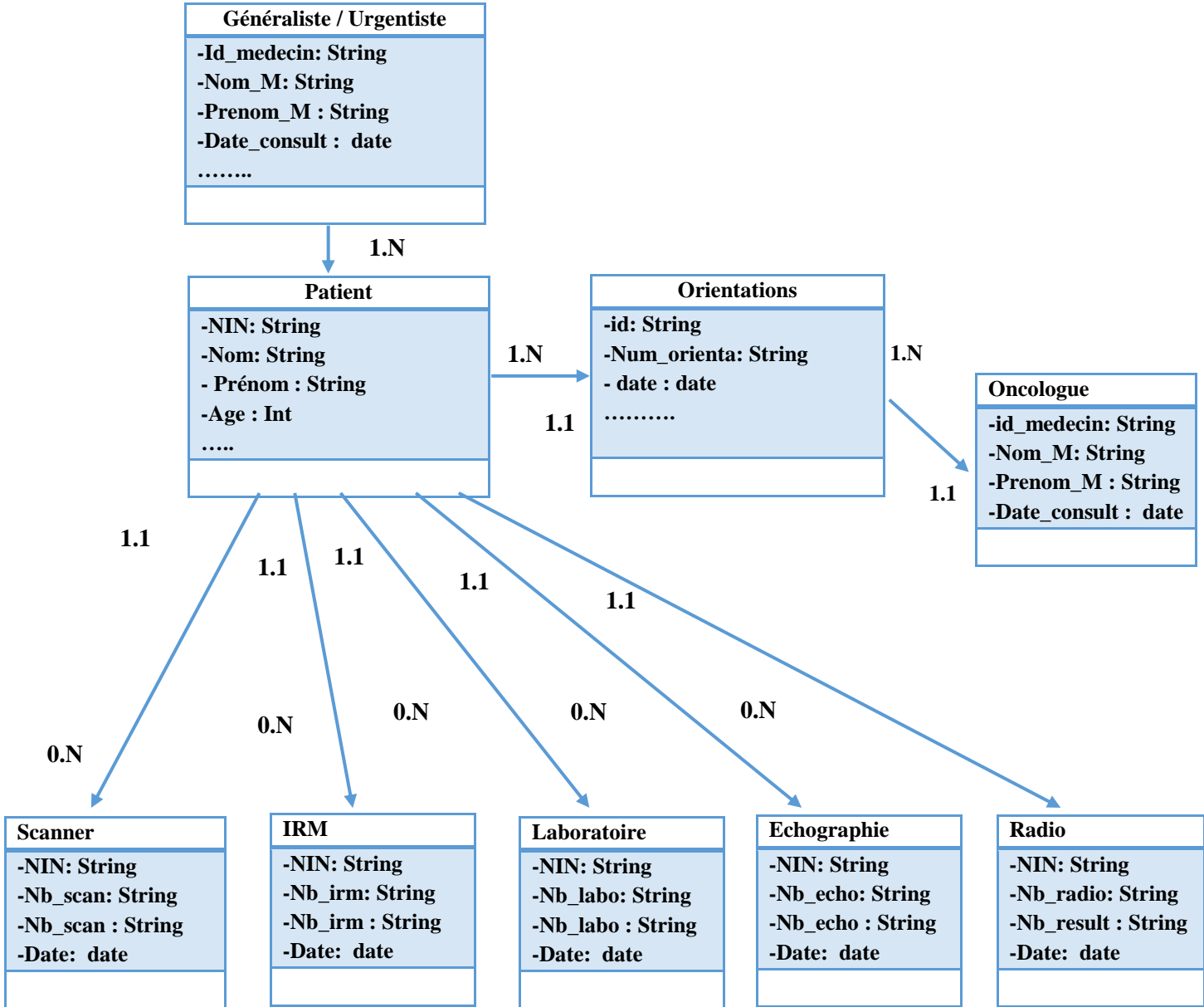


Figure 29 : diagramme de structure pré-orientations.

2.2.1.7. Acquisition des informations des examens des creux axillaires :

L'acquisition des informations des examens des creux axillaires (adénopathie) faits par l'oncologue sont saisis dans cette unité. La figure suivante représente l'unité fonctionnelle examens des creux axillaires.

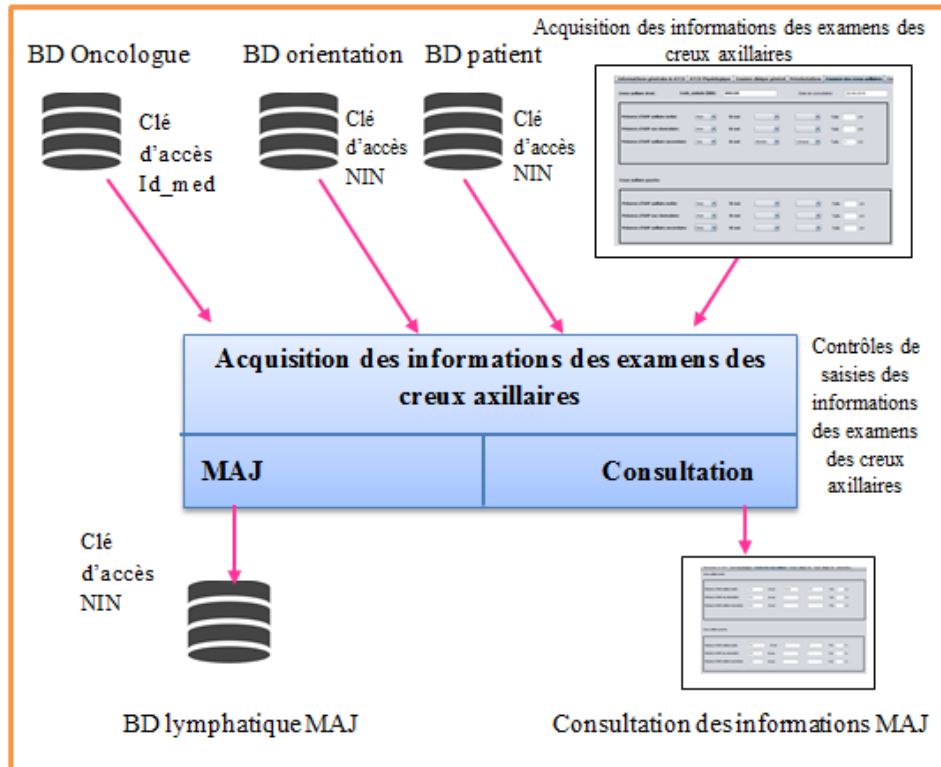


Figure 30 : unité fonctionnelle examens des creux axillaires.

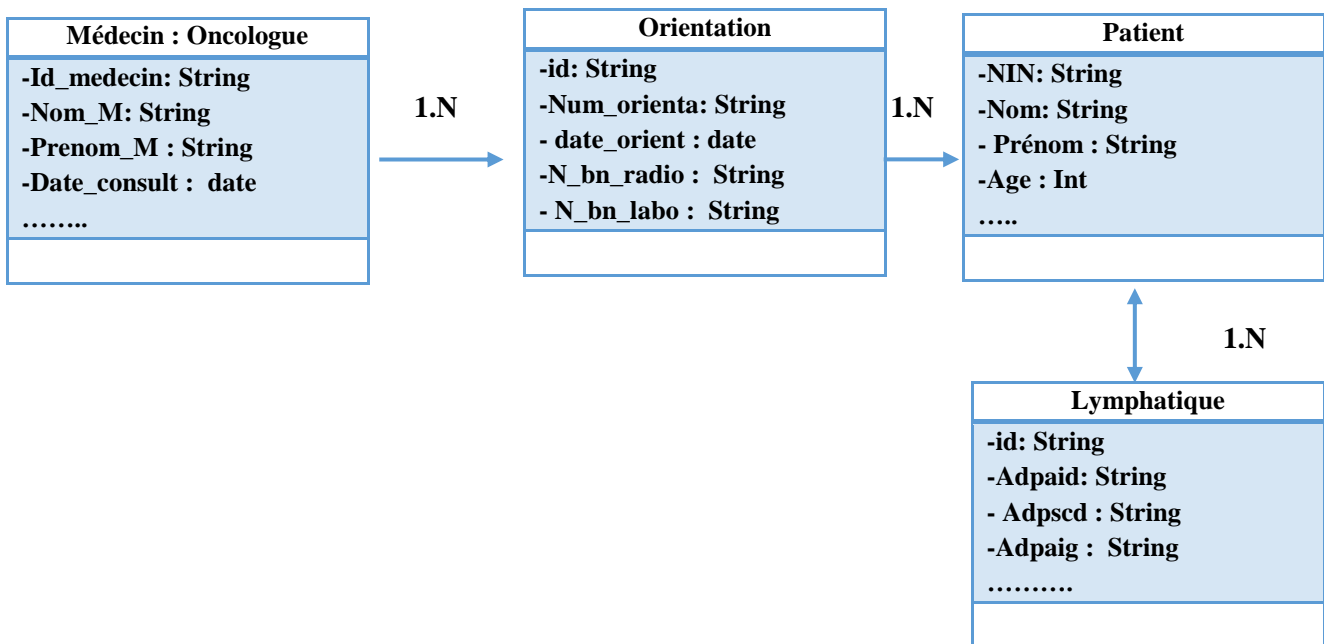


Figure 31 : diagramme de structure examen des creux axillaires.

2.2.1.8. Acquisition des informations des examens sein droit :

Cette unité sert à saisir les informations des examens et diagnostics faits par l'oncologue sur le sein droit du patient. La figure suivante représente l'unité fonctionnelle examens clinique sein droit.

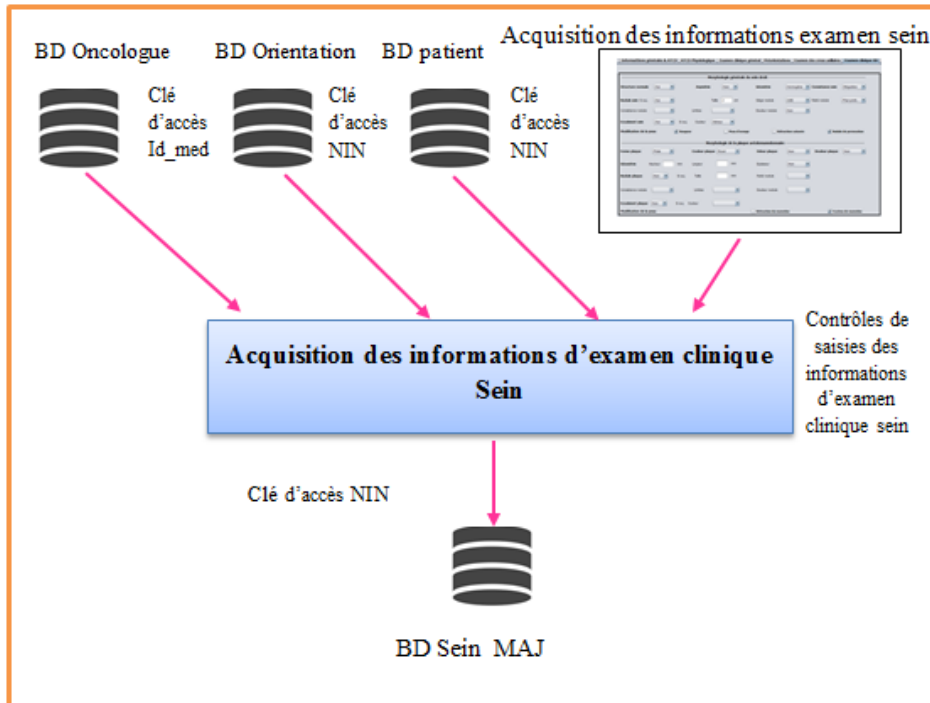


Figure 32 : unité fonctionnelle examen clinique sein.

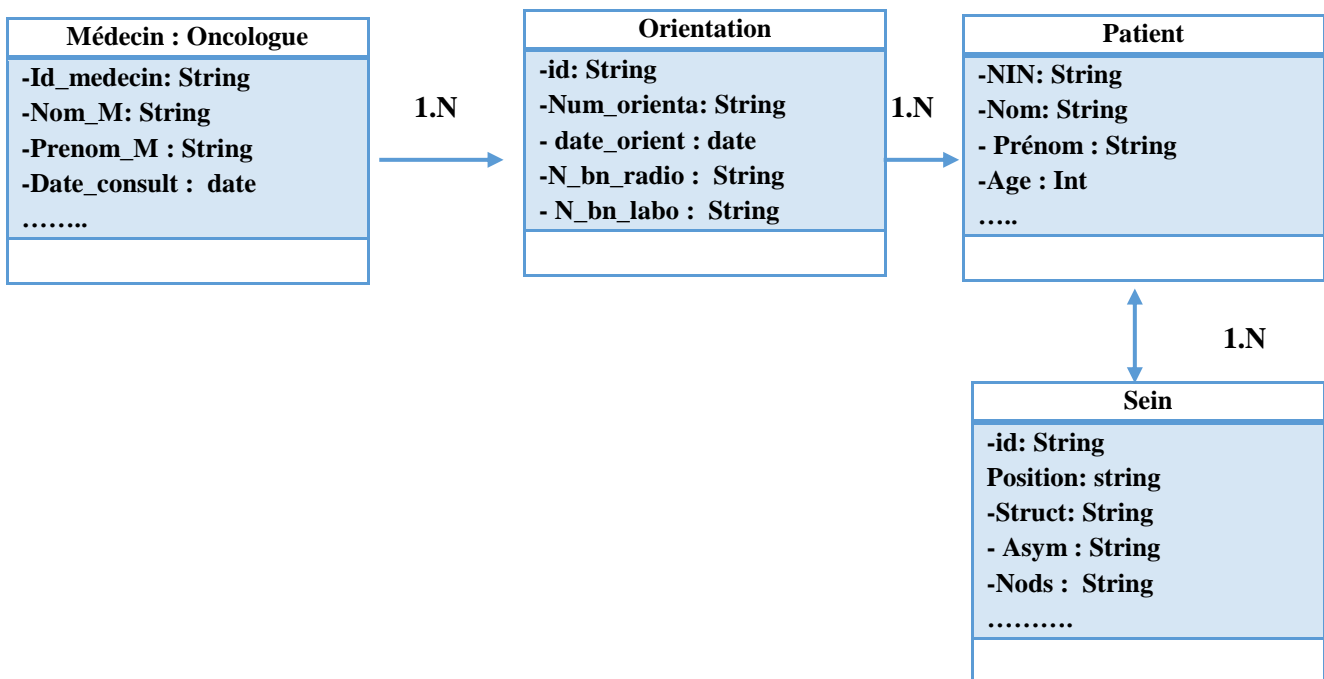


Figure 33 : diagramme de structure examen clinique du sein.

2.2.1.9. Acquisition des informations des orientations :

Après avoir être examiné par un médecin oncologue le patient peut être orienté vers un autre médecin spécialiste ou vers le service commun (radio, laboratoire, anapath...etc), c'est dans cette unité ou les informations du consultant et de l'orientation (numéro orientation, lettre...etc) sont saisies. La figure suivante représente l'unité fonctionnelle post-orientations.

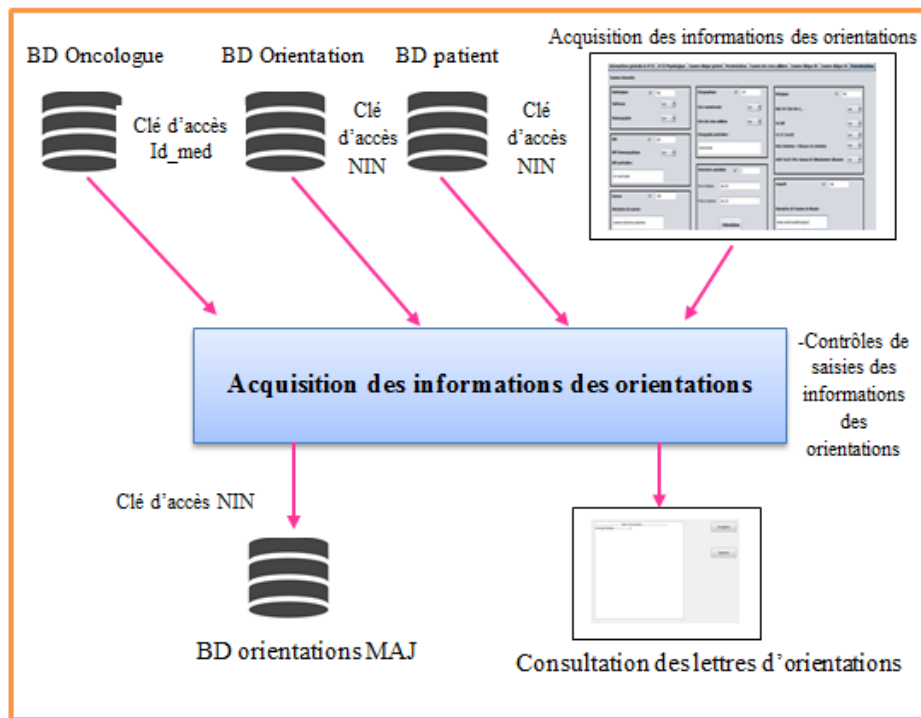


Figure 34 : unité fonctionnelle des orientations.

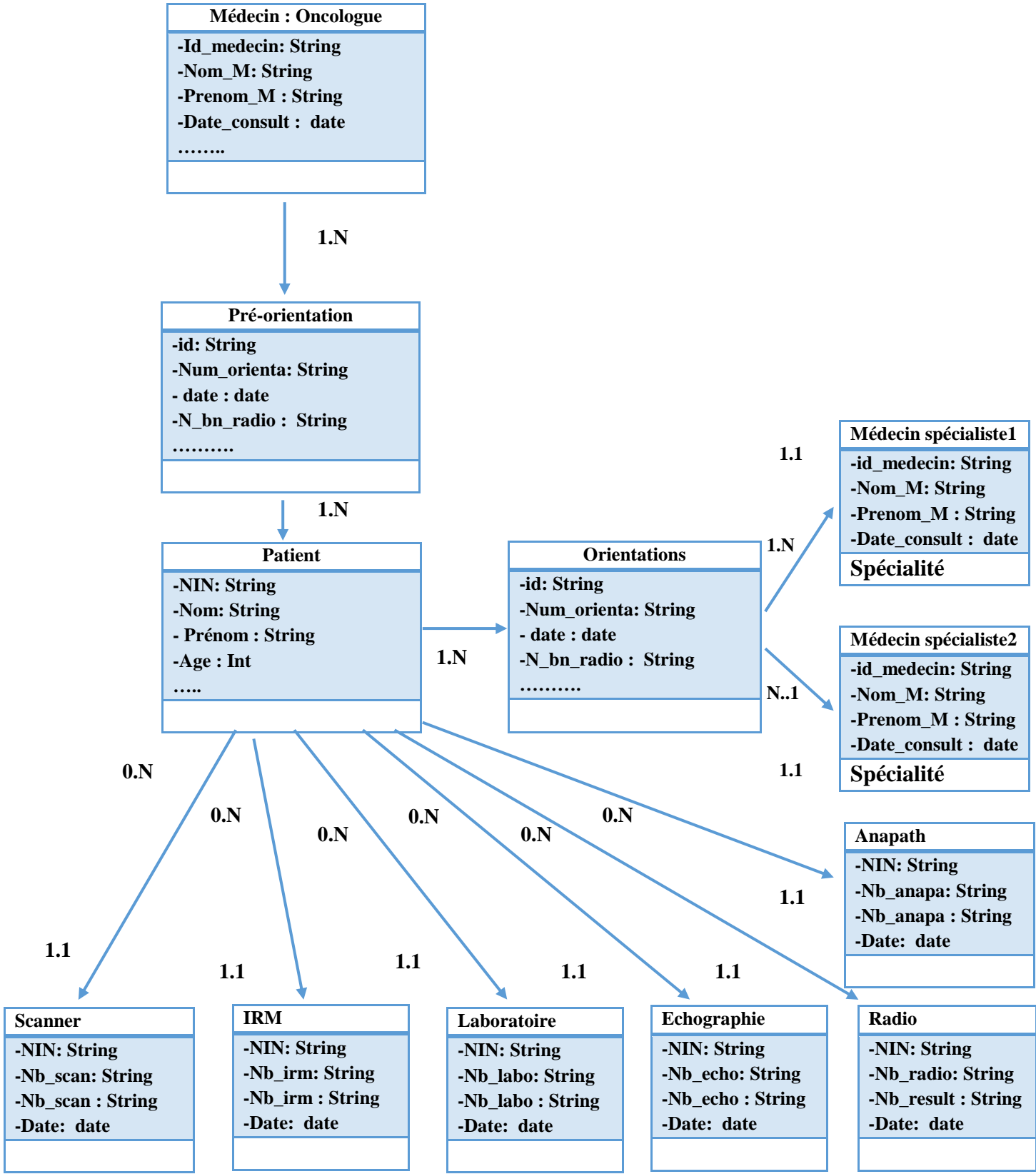


Figure 35 : diagramme de structure post-orientations.

2.2.2. Le sous-système du Prétraitement

Dans l'étape de prétraitement on fait l'extraction des motifs qui caractérisent chaque patient en termes de facteurs de risques, et on élimine tous bruits ou informations inutiles.

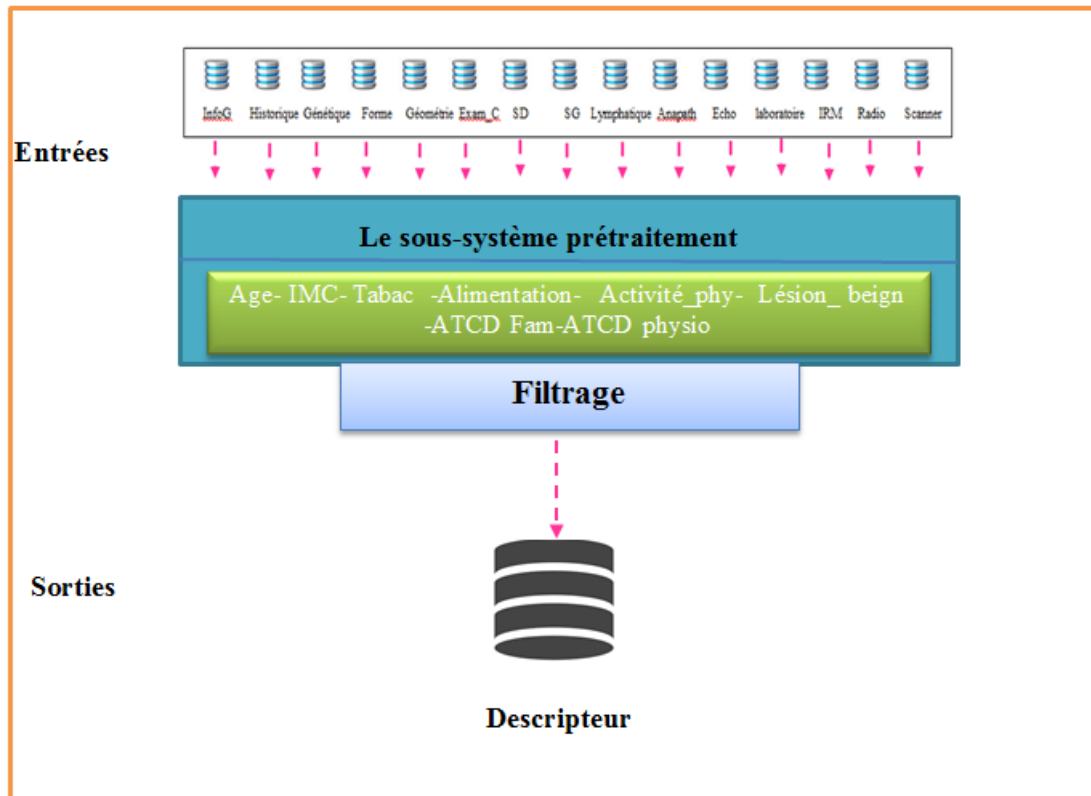


Figure 36 : Architecture du sous-système de prétraitement

2.2.2.1. Les entrées du Sous système

L'ensemble des entrées de ce sous-système représente les informations collectées lors des acquisitions des données à travers les différentes unités.

2.2.2.2. La sortie du Sous système

La donnée sortante du sous-système Prétraitement représente un Descripteur. Ce descripteur est une conception d'une table bidimensionnelle dont ces valeurs sont binaires.

Chaque ligne de la table représente un individu et chaque colonne représente un facteur de risque ' motif '.

2.2.3. La classification

Dans ce module on procède à l'extraction des facteurs fréquents et les règles solides depuis notre descripteur et on introduisant l'algorithme Apriori, avec l'entrée du seuil de confiance et un seuil de support. On aura en sortie notre modèle pour la prise de décision.

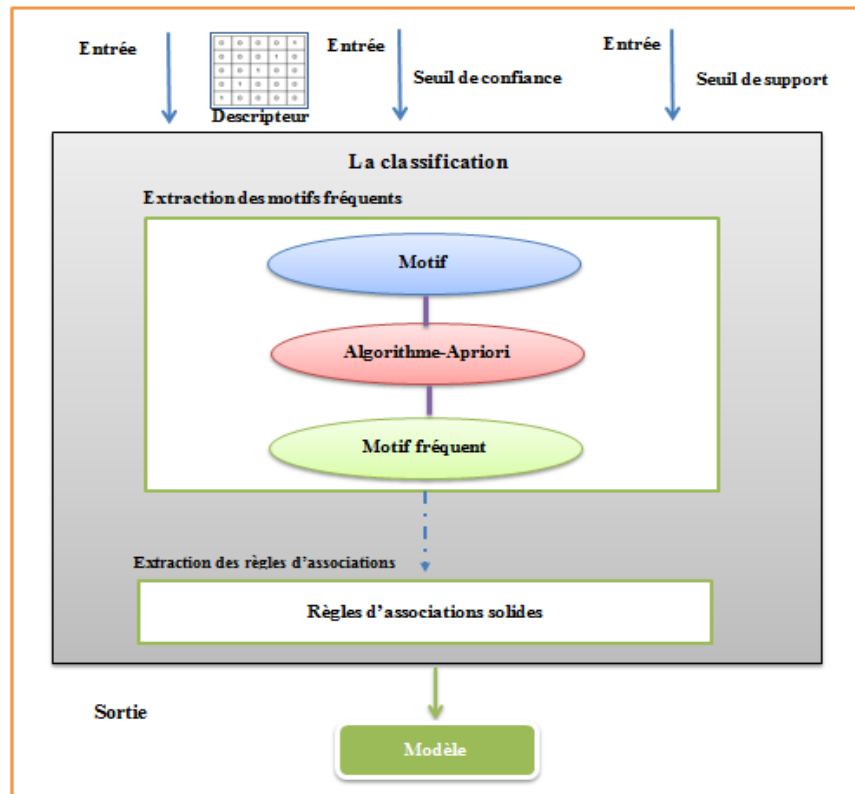


Figure 37 : Architecture du sous-système de la classification.

Le schéma suivant représente le processus de l'extraction des facteurs (motifs) fréquents.

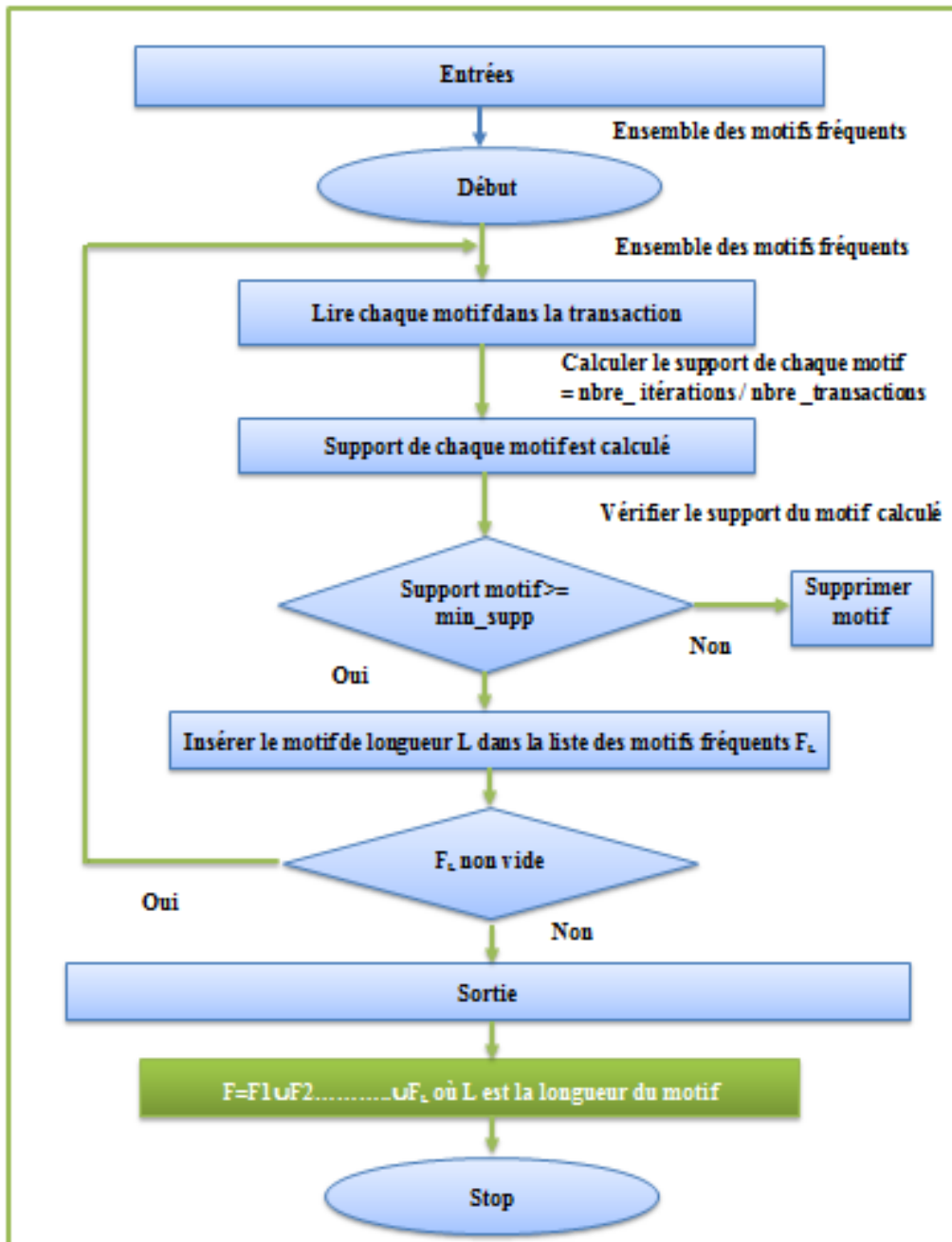


Figure 38 : Le processus de l'extraction des motifs fréquents.

Le schéma suivant représente le processus de l'extraction des règles solides

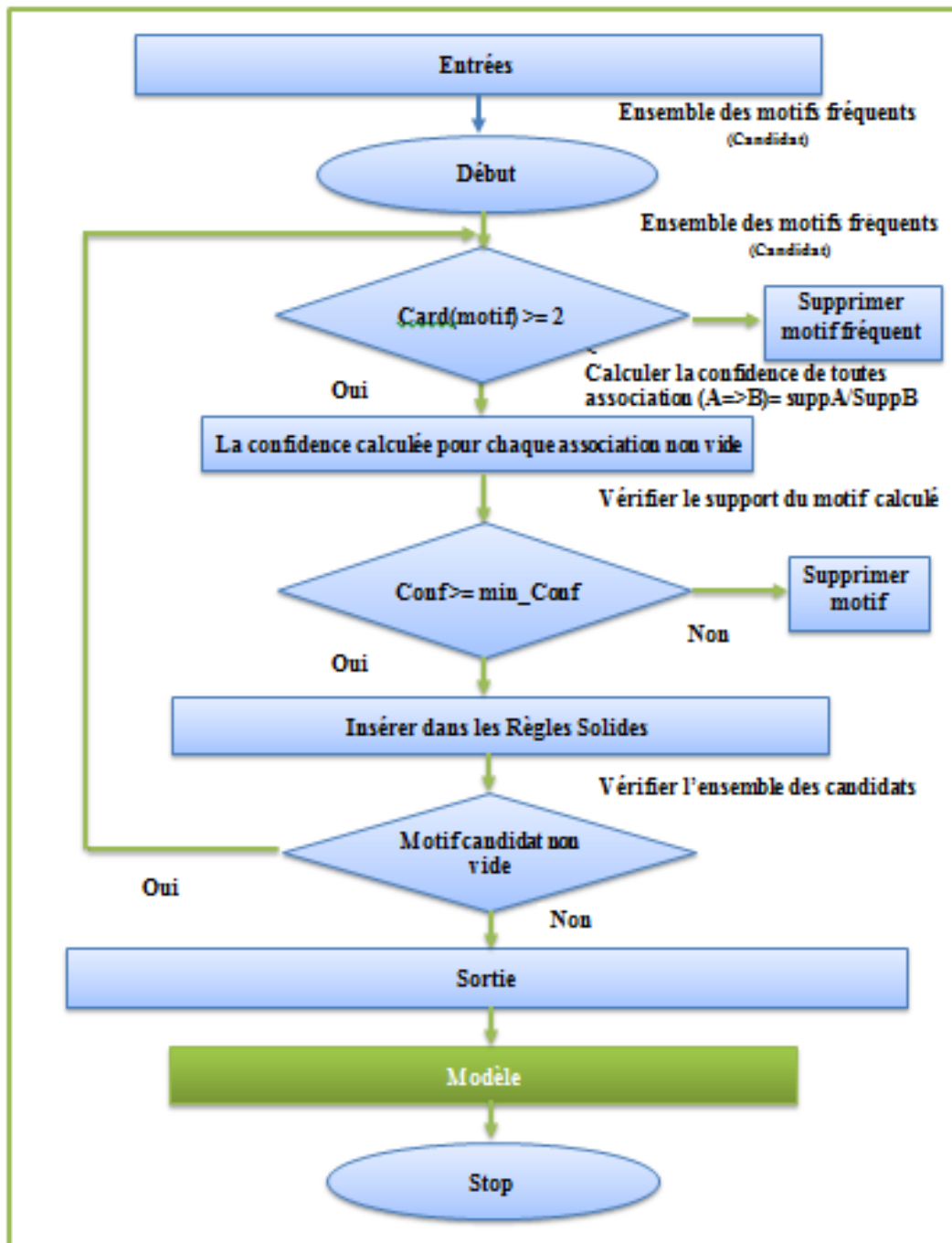


Figure 39 : Le processus de l'extraction des Règles Solides.

2.2.4. Le sous-système de la visualisation

Ce module sert à visualiser les résultats pour estimer la qualité, mesurer l'utilité du modèle prédit par des graphes.

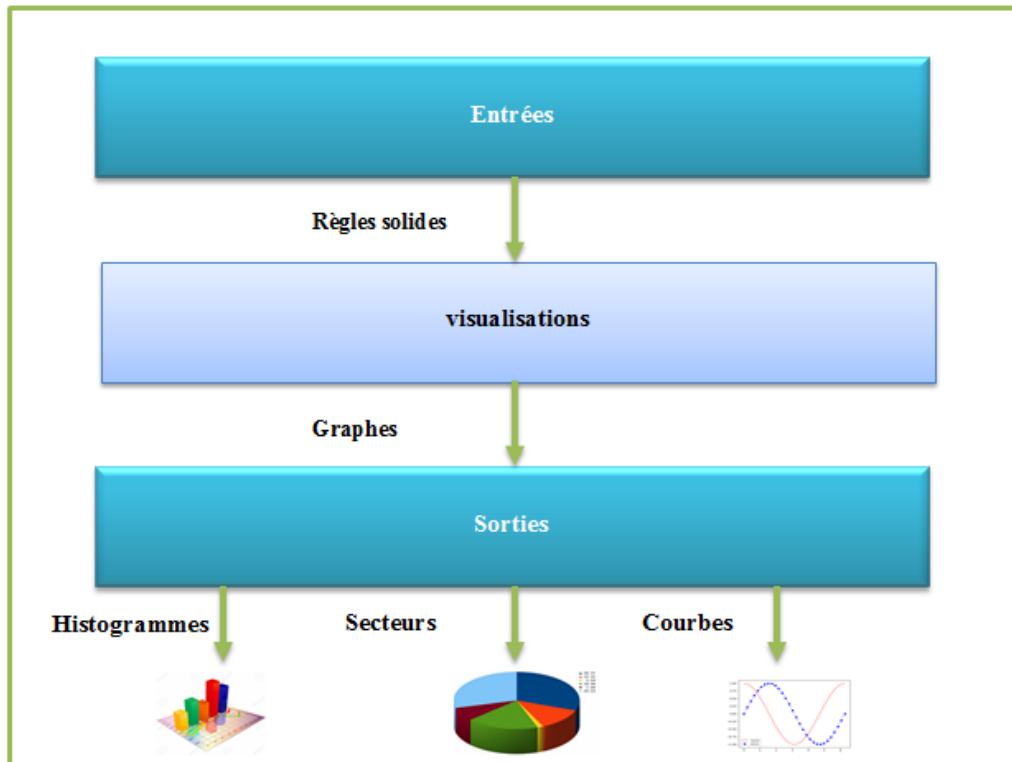


Figure 40: L'architecture du sous-système de visualisation.

3. Conclusion

La conception nous a permis de modéliser notre système de prédiction des déclencheurs du cancer du sein chez la femme par les motifs fréquents. Nous avons commencé par une architecture globale du système proposé, et nous l'avons suivi par une conception détaillée de ses différentes composantes.

Notre système est composé en premier lieu d'un module d'acquisition des données cibles, en deuxième lieu un système de prétraitement permettant de générer la table bidirectionnelle appelée Descripteur.

Ensuite un système de la classification associative pour extraire les déclencheurs (motifs) fréquents et les règles solides pour donner naissance à notre modèle.

Enfin la mise en œuvre d'un système de visualisation des résultats des prédictions.

Chapitre 4

L'implémentation et la validation du
système

1. Introduction :

Dans ce chapitre, nous allons décrire la mise en œuvre des différents sous système de notre modèle proposé. On commence par la description de l'environnement matériel/logiciel, le choix du langage, les outils de programmation, le code des différents sous-systèmes et on termine par les tests et la visualisation des résultats. L'environnement de développement choisit est le NETBEANS.

2. L'environnement matériel/logiciel :

2.1. Matériel

On a développé notre système sur un ordinateur portable ACER avec les caractéristiques suivantes :

- Modèle : TravelMate 5744
- Processeur : Intel ® Core ™ i3 CPU M 380 @ 2.53GFz, 2533 MHz
- Mémoire : 4 Go RAM et 500 Go Disque
- Ecran/Poids : 15.6 pouces et 2.4 Kg

2.2. Logiciel

2.2.1. Système d'exploitation utilisé

On a un utilisé un système d'exploitation Microsoft Windows 7 professionnel Service Pack 1, à 32Bits sur notre machine.

2.2.2. Langage de Programmation utilisé

Pour l'implantation du notre Système on a choisi le langage de programmation orienté objet « **JAVA** » développée par Sun Microsystems. Dans un des premiers papiers sur le langage **JAVA**, SUN le décrit comme suit : « Java : a simple, object-oriented, distributed, robust, secure, architecture neutral, portable, high-performance, multithreaded, and dynamic language ». [15]

```

private void récapActionPerformed(java.awt.event.ActionEvent evt) {
    // Récapitulatif

    String Id = null, Nom = null, Prenom = null;
    Id = id1.getText();
    Nom = nom1.getText();
    Prenom = prenom1.getText();
    Date actuelle = new Date();
    DateFormat dateFormat = new SimpleDateFormat("dd-MM-yyyy");
    String date = dateFormat.format(actuelle);
    String dc = date;
    date_consult2.setText(dc);
    id2.setText(Id);
    id3.setText(Id);

    try {
        String req2 = "select * from historique where id='" + Id + "'";
        rt2 = st.executeQuery(req2);
        if (rt2.next()) {
            //JOptionPane.showMessageDialog(null, "element trouver");
            tabaa.setText(rt2.getString(1));
            dureetaba.setText(rt2.getString(2));
        }
    }
}

```

Figure 41 : exemple code java.

2.2.3. L'environnement utilisé

Le **NetBeans** est un environnement de développement intégré (EDI), placé en open source. En plus de Java, **NetBeans** permet la prise en charge native de divers langages tels le C, le C++, le JavaScript, le XML, le Groovy, le PHP et le HTML. Il offre toutes les facilités d'un IDE moderne (éditeur en couleurs, projets multi-langage, éditeur graphique d'interfaces et de pages Web). **NetBeans** est disponible sous Windows, Linux, Solaris

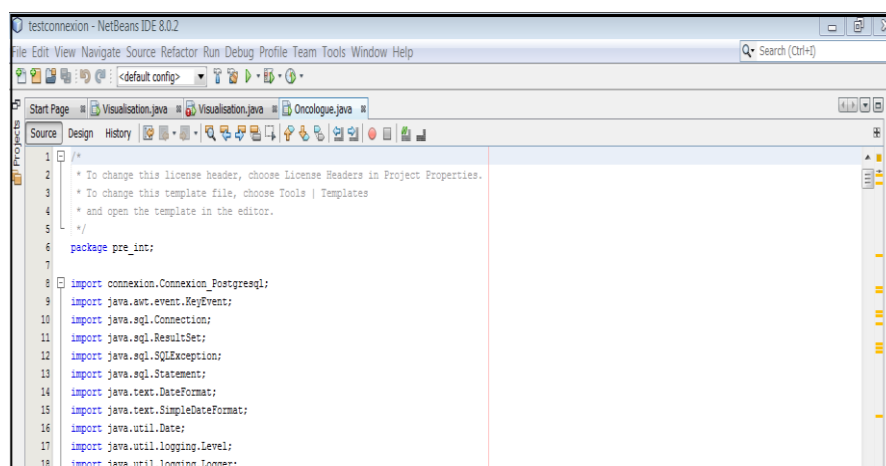


Figure 42 : Fenêtre Netbeans.

2.2.4. Les SGBDS utilisé

Afin de gérer les bases de données utilisées dans notre système on a utilisé :

- ❖ **PostgreSQL** : est un des principaux SGBD-Relationnelles) du marché, il est libre et gratuit, parmi ses avantages : déploiement illimité, excellent support, fiabilité et stabilité légendaires, conçu pour une grande capacité, outils graphiques de modélisation et d'administration, très bonne compatibilité SQL. Le **PostgreSQL** est compatible avec presque toutes les plates-formes (linux, Windows, Solaris...etc.) [16].

4. L'implémentation :

Notre système interactif comprend :

- Sous-système d'acquisition des données cibles
- Sous-système de prétraitement
- Sous-système de la classification par les motifs fréquents.
- Sous-système de la visualisation (test et validation)

4.1. Sous-système d'acquisition des données cibles

Ce sous-système permet l'acquisition de toutes les données nécessaires pour l'ensemble des patients.

- ✓ « **ConnexionPostgreSQL.java** »

Cette classe java permet la connexion dynamique à la base de données.

La figure suivante représente le code de cette classe :

```
public class Connexion_Postgresql {
    public static Connection c2;
    public static Connection Connexion_Postgresql() {
        try {

            String url = "jdbc:postgresql://localhost:5432/DB";
            String user = "postgres";
            String psw = "root";
            Class.forName("org.postgresql.Driver");

            c2 = DriverManager.getConnection(url, user, psw);
            // st = c.createStatement();
            JOptionPane.showMessageDialog(null, "Connexion avec succes");
        } catch (ClassNotFoundException | SQLException ex) {
            JOptionPane.showMessageDialog(null, "Erreur:" + ex.toString());
            c2=null;
        }
        return c2;
    }
}
```

Figure 43 : Connexion PostgreSql.

✓ **La liaison entre les données**

Cette dernière est gérée par les requêtes SQL.

Les figures suivantes représentent les classes « Acteurs », « Action acteurs » le code source de la méthode « Acteurs () » et « Action_acteurs () » et leurs unités :

- ✓ La classe gestion des acteurs : donne le choix d'accès d'un médecin (généraliste, urgentiste ou oncologue) pour l'identification.

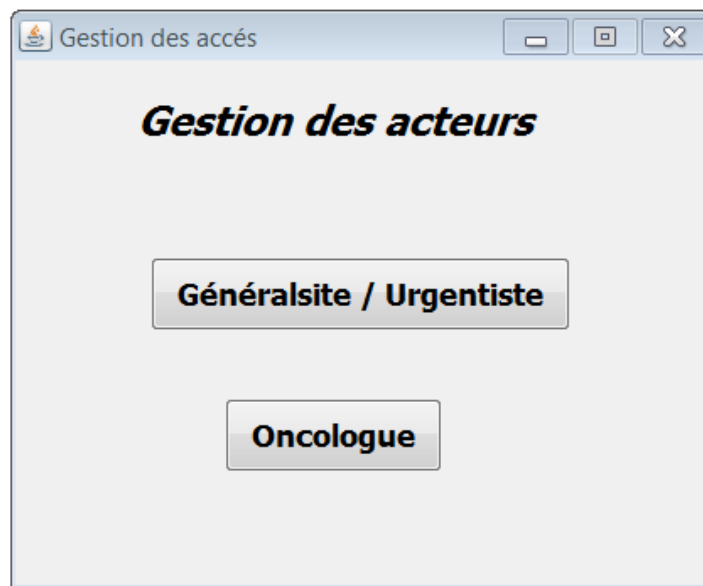


Figure 44 : la classe Acteurs ().

```

public Acteurs() {
    initComponents();
}

/**
 * This method is called from within the constructor to initialize the form.
 * WARNING: Do NOT modify this code. The content of this method is always
 * regenerated by the Form Editor.
 */
@SuppressWarnings("unchecked")
Generated Code

private void jButton1KeyPressed(java.awt.event.KeyEvent evt) {

}

private void jButton1ActionPerformed(java.awt.event.ActionEvent evt) {
    action_acteur p = new action_acteur();
    login1 p = new login1();
    p.setVisible(true);
    dispose();
}

private void jButton3ActionPerformed(java.awt.event.ActionEvent evt) {
    login2 p = new login2();
    p.setVisible(true);
    dispose();
}

```

Figure 45 : code source de la classe Acteurs ().

- ✓ La classe login : permet l'identification des médecins par nom d'utilisateur et mot de passe.

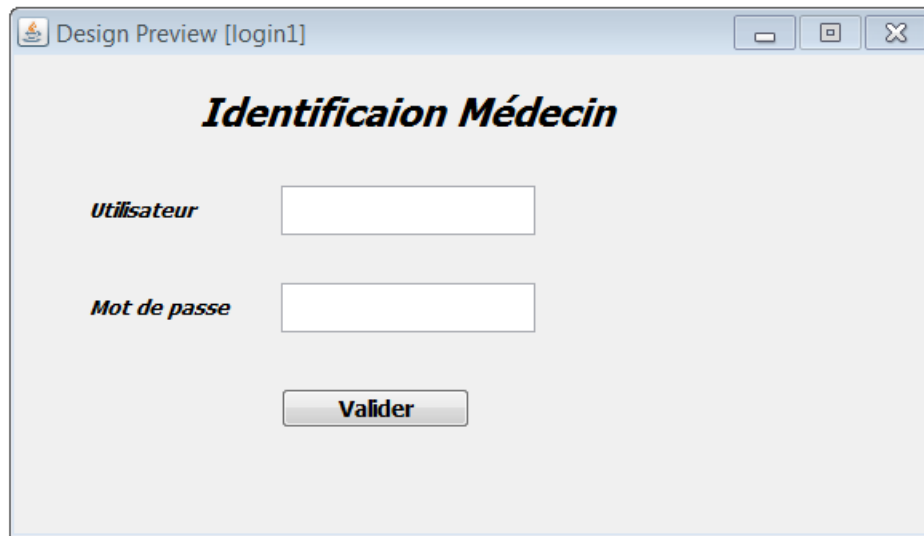


Figure 46 : la classe login ().

- ✓ L'unité Identification patient de la classe 'Action_acteurs' permet l'acquisition des données générales du patient, pour la création du dossier médical électronique.

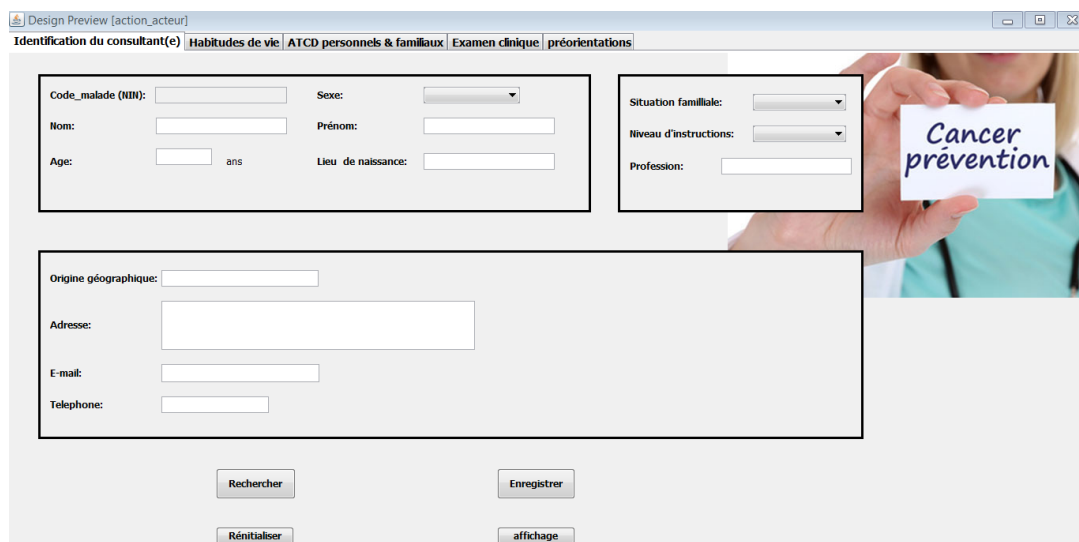


Figure 47 : unité Identification patient classe Action_acteurs avant acquisition de données.

```

private void enregistrer3ActionPerformed(java.awt.event.ActionEvent evt) {
// Historique
String Id = null, Hta = null, Mcardio = null, Surpoid = null, Diabete = null, Co = null, Atcdrth = null, Neopm = null, Multif = null,
int i, j;
Id = id.getText();
Hta = jComboBoxhta.getSelectedItemAt().toString();
Mcardio = jComboBoxcardio.getSelectedItemAt().toString();
Diabete = jComboBoxdiab.getSelectedItemAt().toString();
Co = jComboBoxco.getSelectedItemAt().toString();
Atcdrth = jComboBoxrth.getSelectedItemAt().toString();
Neopm = jComboBoxnm.getSelectedItemAt().toString();
Menarchie = jTextField3.getText();
Allait = jComboBoxallait.getSelectedItemAt().toString();
Allaitduree = jTextField22.getText();
Contracepo = jComboBoxcontro.getSelectedItemAt().toString();
Contracepduree = jTextField27.getText();
Hormono = jComboBoxhr.getSelectedItemAt().toString();
Hormonoduree = jTextField28.getText();
Neopute = jComboBoxnut.getSelectedItemAt().toString();
Neopendom = jComboBoxnendo.getSelectedItemAt().toString();
Lesionb = jComboBoxlesbegn.getSelectedItemAt().toString();
Surpoid = jComboBoxsurpoids.getSelectedItemAt().toString();
}

```

Figure 48 : code source Action_acteurs().

- ✓ L'unité habitude de vie, sert à acquérir les informations des habitudes de vie du patient.

Figure 49 : unité habitudes de vie avant acquisition de données.

- ✓ L'unité ATCD personnel et familiaux : permet d'acquérir les informations antécédents personnel et familiaux du patient.

Figure 50 : unité antécédents personnel et familiaux avant acquisition de données

- ✓ L'unité Identifications générales permet la consultation des informations généraux, habitudes de vie, antécédents pathologique du patient déjà enregistré.

Figure 51 : unité informations générales, habitudes de vie, antécédent après acquisition de données.

- ✓ L'unité examen clinique assiste le médecin généraliste ou urgentiste à enregistrer les informations de l'examen clinique.

Figure 52 : unité examen_clinique avant acquisition de données.

- ✓ L'unité examen clinique en mode consultation, permet la consultation des informations des examens clinique passés.

Figure 53 : unité examen_clinique après acquisition de données.

- ✓ L'unité pré-orientations permet au médecin généraliste ou urgentiste à faire leurs orientations munies d'une lettre et les enregistrer sur le dossier médicale électronique du patient.

Figure 54 : unité pré-orientation avant acquisition de données.

- ✓ L'unité pré-orientations en mode consultation permet au médecin oncologue de visualiser les orientations et la lettre du patient établie par le médecin généraliste ou urgentiste.

Figure 55 : unité pré-orientation après acquisition de données.

- ✓ L'unité examen des creux axillaires permet au médecin oncologue d'enregistrer les informations de l'examen du patient.

Figure 56 : unité examen des creux axillaires avant acquisition de données.

- ✓ L'unité examen des creux axillaires en mode consultation permet au médecin oncologue de visualiser les examens des creux axillaires déjà enregistré.

Figure 57 : unité examen des creux axillaires après acquisition de données.

- ✓ L'unité examen clinique sein droit identique à l'unité du sein gauche permet au médecin oncologue d'enregistrer les informations de l'examen spécialisé des seins.

The screenshot shows a web-based medical form with the following sections:

- Morphologie générale du sein droit:**
 - Structure normale (dropdown), Asymétrie (dropdown), Géométrie (dropdown), Consistance sein (dropdown)
 - Nodule sein: Si oui (dropdown), Taille (text input, cm), Siège nodule (dropdown), Fixité nodule (dropdown)
 - Consistance nodule (dropdown), Limites (dropdown), Douleur nodule (dropdown)
 - Écoulement sein (dropdown), Si oui (dropdown), Couleur (dropdown)
 - Modification de la peau: Rougeur, Peau d'orange, Rétraction cutanée, Nodule de perméation
- Morphologie de la plaque aréolomamelonnaire:**
 - Forme plaque (dropdown), Couleur plaque (dropdown), Odeur plaque (dropdown), Douleur plaque (dropdown)
 - Géométrie: Hauteur (text input, mm), Largeur (text input, mm), Épaisseur (dropdown)
 - Nodule plaque (dropdown), Si oui (dropdown), Taille (text input, mm), Fixité nodule (dropdown)
 - Consistance nodule (dropdown), Limites (dropdown), Douleur nodule (dropdown)
 - Écoulement plaque (dropdown), Si oui (dropdown), Couleur (dropdown)
 - Modification de la peau: Rétraction du mamelon, Eczéma de mamelon

An 'Enregistrer' button is located in the top right corner of the form area.

Figure 58 : unité examen clinique sein droit / identique sein gauche.

L'unité postorientations permet au médecin oncologue d'établir des orientations vers les services communs, ou vers un autre confrère muni d'une lettre d'orientation.

The screenshot shows the 'Postorientations' tab with the following layout:

- Examens demandés:**
 - Radiologique (N°: 192):** Telethorax (dropdown), Mammographie (dropdown)
 - IRM (N°: 200):** IRM Mammographique (dropdown), IRM particulière: (text input)
 - Scanner (N°: 188):** Description de scanner: (text input)
 - Echographique (N°: 197):** Echo mamelonnaire (dropdown), Echo des creux axillaires (dropdown), Echographie particulière: (text input)
 - Orientation spécialisée (N°: 2):** Nom médecin (text input), Prénom médecin (text input), Orientations (button)
 - Biologique (N°: 190):** FHS+TP+TCK+VS+C... (dropdown), VS CRP (dropdown), CA 15-3 & ACE (dropdown), Urée Créatinine + Clérence à la créatinine (dropdown), ASAT+ALAT+PAL+Gamma Gt+Bilirubinémie+Albumine (dropdown)
 - Anapath (N°: 189):** Description de l'examen de Biopsie: (text input)

Figure 59 : unité post-orientations.

4.2. Sous-système de prétraitement

L'implémentation du sous-système de prétraitement de notre conception se fait par la programmation d'une classe java, « pretraitement .java » représente le composant interne

du sous-système, toutes les bases de données représentent ses entrées et « descripteur.java » représente sa sortie.

4.2.1. La sortie du sous-système

➤ Descripteurs.java

Ce descripteur se présente sous forme d'une table binaire dont les lignes représentent l'ensemble des individus dit « patients », et les colonnes représentent les facteurs fréquents de la maladie.

Le résultat des motifs extraits, sauvegardés dans une base de données « Btransac.txt ». Les figures suivantes représentent le composant « descripteur », ainsi le code source de la méthode descripteur ().

Figure 60 : descripteur avant l'exécution.

| Id | age | sedent | surpoidm... | alcool | tabac | fruitleg | viande | sucré | actvphy | menarchie | age1gross | allait | dureeallait | contracepo | menopo | hormono |
|--------|-----|--------|-------------|--------|-------|----------|--------|-------|---------|-----------|-----------|--------|-------------|------------|--------|---------|
| 000001 | 1 | | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 000003 | 1 | | | 0 | 0 | 0 | 0 | 1 | 0 | 0 | | 1 | 0 | 0 | 0 | 0 |
| 000048 | 1 | | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | | 0 | 1 | 0 | 0 | 0 |
| 000029 | 1 | | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | | 0 | 1 | 1 | 0 | 0 |
| 000052 | 1 | | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | | 0 | 1 | 0 | 0 | 1 |
| 000030 | 1 | | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | | 0 | 1 | 1 | 0 | 0 |
| 000061 | 0 | | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | | 0 | 1 | 1 | 0 | 0 |
| 000035 | 0 | | | 0 | 0 | 0 | 0 | 1 | 0 | 0 | | 0 | 0 | 1 | 0 | 0 |
| 000043 | 1 | | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | | 0 | 0 | 1 | 0 | 0 |
| 000014 | 1 | | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | | 1 | 0 | 0 | 0 | 0 |
| 000010 | 0 | | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | | 0 | 1 | 1 | 0 | 0 |
| 000012 | 0 | | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | | 0 | 1 | 0 | 0 | 0 |
| 000009 | 1 | | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | | 0 | 1 | 1 | 0 | 0 |
| 000011 | 1 | | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | | 0 | 1 | 1 | 0 | 0 |
| 000008 | 0 | | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | | 0 | 1 | 1 | 0 | 0 |
| 000013 | 1 | | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | | 0 | 0 | 1 | 0 | 0 |
| 000007 | 1 | | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | 0 | 1 | 1 | 0 | 0 |
| 000038 | 0 | | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | | 0 | 1 | 1 | 0 | 0 |
| 000031 | 1 | | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | | 1 | 1 | 0 | 0 | 0 |

Figure 61 : descripteur après exécution.

```

private void jButton2ActionPerformed(java.awt.event.ActionEvent evt) {
    jButton2.setCursor(Cursor.getPredefinedCursor(Cursor.WAIT_CURSOR));
    BufferedWriter write = null;
    clear2();

    try {
        write = new BufferedWriter(new FileWriter(new File("BTransac.txt")));
        try {

            String req = "select id, age , sedent,surpoidmeno,alcool,tabac,fruitleg,viande,
rt1 = st1.executeQuery(req);
d.setRowCount(0);

            while (rt1.next()) {
                String ligne = "";
                String Id = rt1.getString("id");
                JOptionPane.showMessageDialog(null, Id);
                String age = rt1.getString("age");
                String sedent = rt1.getString("sedent");
                String surpoidmeno = rt1.getString("surpoidmeno");
                String alcool = rt1.getString("alcool");
                String tabac = rt1.getString("tabac");
                String fruitleg = rt1.getString("fruitleg");
                String viande = rt1.getString("viande");
                String sucre = rt1.getString("sucre");
                String activphy = rt1.getString("activphy");
                String menarchie = rt1.getString("menarchie");
                String ageigross = rt1.getString("ageigross");

```

Figure 62 : Code source méthode descripteur.

4.3. Sous-système de la classification

Le sous-système de la classification a été programmé dans la classe java Apriori().

➤ AlgorithmeApriori.java

Après avoir appliqué l'algorithme Apriori sur la base de données « Btransac.txt » on extrait l'ensemble des règles solides, qui sont sauvegardés dans une table « reglesolide » dans la base de données « DB ».

Les figures suivantes représentent le sous-système de la « Classification», ainsi le code source de l'Algorithme Apriori.

Pour Extraire des Règles Solides Veuillez :

Entrez la Valeur du Support Minimum:

0.1

Entrez la valeur du Confiance Minimum:

0.1

Extraire_Règles_Solides

Figure 63 : L'acquisition des valeurs support & confiance.

| Motifs Fréquent: | | | | Les Règles Solides | |
|-------------------|-----------|---------|----------|--------------------|-----------|
| ItemSet | Fréquence | Support | Longueur | Les Règles Solides | confiance |
| dureeallait | 113 | 0.62 | 1 | age sucre=>imc | 0.68 |
| contracepo | 110 | 0.61 | 1 | age imc=>sucre | 1.07 |
| age | 135 | 0.75 | 1 | age=>imc sucre | 1.3 |
| imc | 123 | 0.68 | 1 | imc sucre=>age | 0.76 |
| sucre | 175 | 0.97 | 1 | imc=>age sucre | 1.47 |
| dureeallait su... | 109 | 0.6 | 2 | sucre=>age imc | 0.92 |
| contracepo su... | 106 | 0.58 | 2 | | |
| age imc | 95 | 0.52 | 2 | | |
| age sucre | 130 | 0.72 | 2 | | |
| imc sucre | 120 | 0.66 | 2 | | |
| age imc sucre | 92 | 0.51 | 3 | | |

Visualisation Graphique

Figure 64 : L'extraction des motifs fréquents et les Règles Solides.

4.4. Les tests et la visualisation des résultats

Après avoir présenté les principaux modules de notre système prédictif des facteurs de risque dans du cancer du sein chez la femme par les motifs fréquents, nous essayons dans cette phase de présenter les résultats obtenus.

La fenêtre principale de notre système s'affiche comme suit :

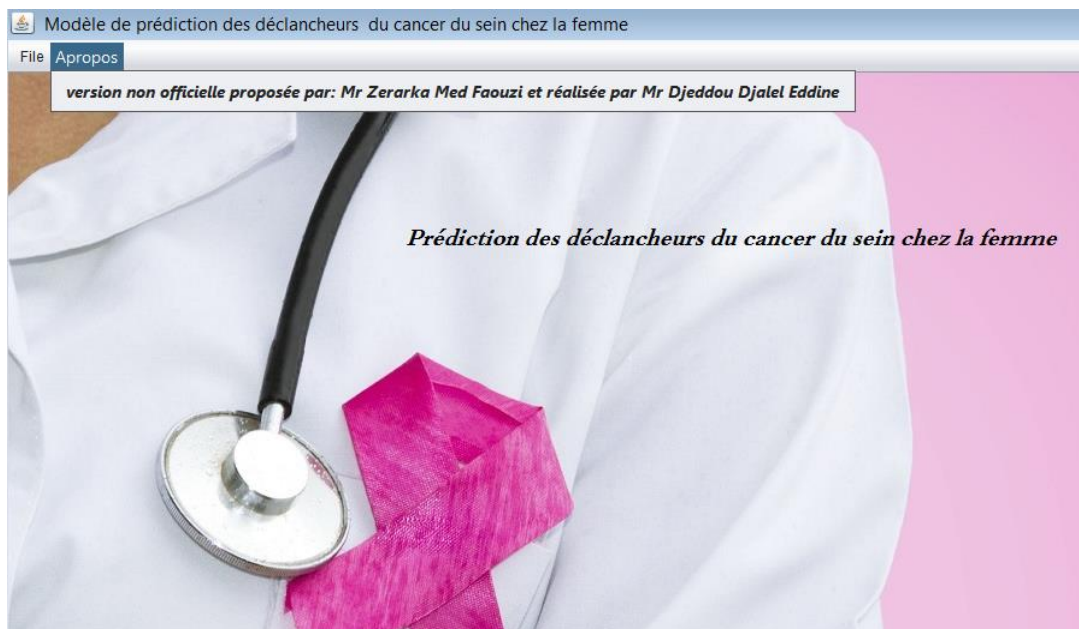


Figure 65 : L'interface principale du système proposé.

Notre système cyclique et diviser en trois parties :

4.4.1. La présentation du sous-système d'acquisition des données cible

- ✓ En entré on à l'acquisition des identificateurs de médecins et toutes les informations du patient collectées lors de la consultation.

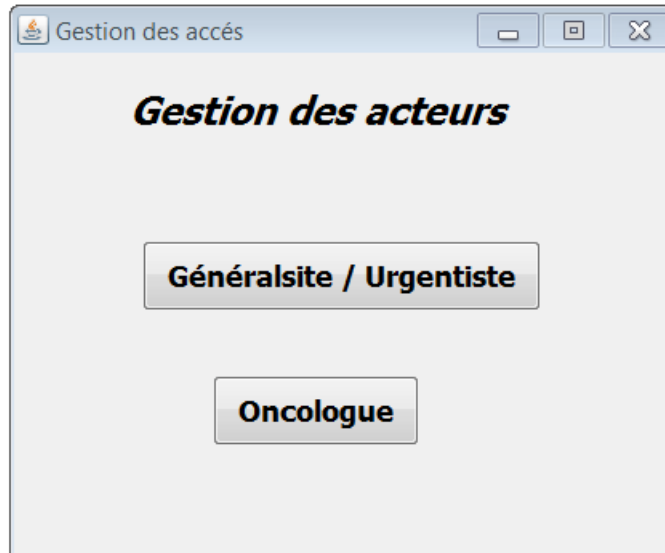


Figure 66 : sous système d'acquisition des données cibles médecin.

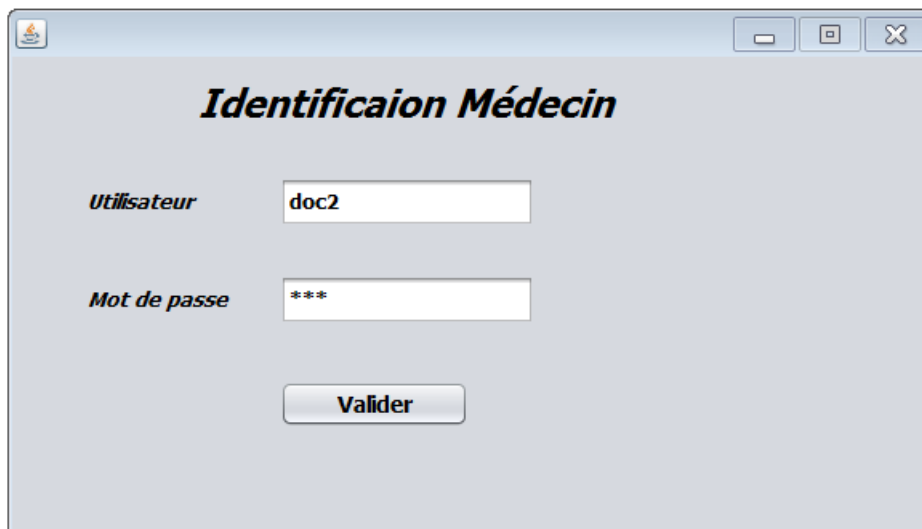


Figure 67 : Identification des médecins.

Figure 47 : unité d'identification patient classe Action_acteur avant acquisition.

- ✓ A la sortie on a des unités d'affichage de toutes les informations du patient collectées lors de la consultation.

Figure 51 : unité informations général, habitudes de vie, antécédent après acquisition de données.

4.4.2. La présentation du sous-système Prétraitement

Dans cette phase les données acquis dans les base de donnée son sélectionné et classés en classe de facteurs de risque. A la sortie on aura une table binaire.

- ✓ Interface du sous-système de prétraitement.



Figure 68 : sous système de prétraitement.

- ✓ La sortie du sous-système de prétraitement, la base de données binaire.

| Id | age | sedent | surpoidme... | alcool | tabac | fruitleg | viande | sucres | activphy | menarchie | age1gross | allaît | dureeallaît | contracepo | menopo | hormono |
|--------|-----|--------|--------------|--------|-------|----------|--------|--------|----------|-----------|-----------|--------|-------------|------------|--------|---------|
| 000001 | 1 | | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 000003 | 1 | | | 0 | 0 | 0 | 0 | 1 | 0 | 0 | | 1 | | | | |
| 000048 | 1 | | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | | 0 | 1 | 0 | 0 | |
| 000029 | 1 | | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | | 0 | 1 | 1 | 0 | |
| 000052 | 1 | | | 0 | 0 | 0 | 0 | 1 | 0 | 0 | | 0 | 1 | 0 | | 1 |
| 000030 | 1 | | | 0 | 0 | 0 | 0 | 1 | 1 | 0 | | 0 | 1 | 1 | 0 | |
| 000051 | 0 | | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | | 0 | 1 | 1 | 0 | |
| 000035 | 0 | | | 0 | 0 | 0 | 0 | 1 | 0 | 0 | | 0 | 0 | 1 | | |
| 000043 | 1 | | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | | 0 | 0 | 1 | 0 | |
| 000014 | 1 | | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | | 1 | | 0 | 0 | |
| 000010 | 0 | | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | | 0 | 1 | 1 | 0 | |
| 000012 | 0 | | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | | 0 | 1 | 0 | | |
| 000009 | 1 | | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | | 0 | 1 | 1 | 0 | |
| 000011 | 1 | | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | | 0 | 1 | 1 | 0 | |
| 000008 | 0 | | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | | 0 | 1 | 1 | 0 | |
| 000013 | 1 | | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | | 0 | 0 | 1 | 0 | |
| 000007 | 1 | | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | 0 | 1 | 1 | | |
| 000038 | 0 | | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | | 0 | 1 | 1 | | |
| 000031 | 1 | | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | | 1 | | 0 | | |

Figure 61 : descripteur après exécution.

4.4.3. Le sous-système de la classification

Dans cette étape nous avons fait l'extraction des motifs fréquents par l'implémentation de l'algorithme apriori.

- ✓ Interface de saisie des valeurs support et confiance.

Figure 63 : L'acquisition des valeurs support & confiance.

- ✓ Après exécution de l'algorithme apriori, l'affichage des facteurs fréquents et les règles solides.

| ItemSet | Fréquence | Support | Longueur |
|-------------------|-----------|---------|----------|
| dureeallait | 113 | 0.62 | 1 |
| contracepo | 110 | 0.61 | 1 |
| age | 135 | 0.75 | 1 |
| imc | 123 | 0.68 | 1 |
| sucre | 175 | 0.97 | 1 |
| dureeallait su... | 109 | 0.6 | 2 |
| contracepo su... | 106 | 0.58 | 2 |
| age imc | 95 | 0.52 | 2 |
| age sucre | 130 | 0.72 | 2 |
| imc sucre | 120 | 0.66 | 2 |
| age imc sucre | 92 | 0.51 | 3 |

| Les Règles Solides | confiance |
|--------------------|-----------|
| age sucre=>imc | 0.68 |
| age imc=>sucre | 1.07 |
| age=>imc sucre | 1.3 |
| imc sucre=>age | 0.76 |
| imc=>age sucre | 1.47 |
| sucre=>age imc | 0.92 |

Figure 64 : L'extraction des motifs fréquents et les Règles Solides.

4.4.4. Le sous-système de la visualisation du modèle

L'étape de visualisation permet de donner une lecture graphique aux nouveaux modèles.

- ✓ Interface de choix de visualisation.

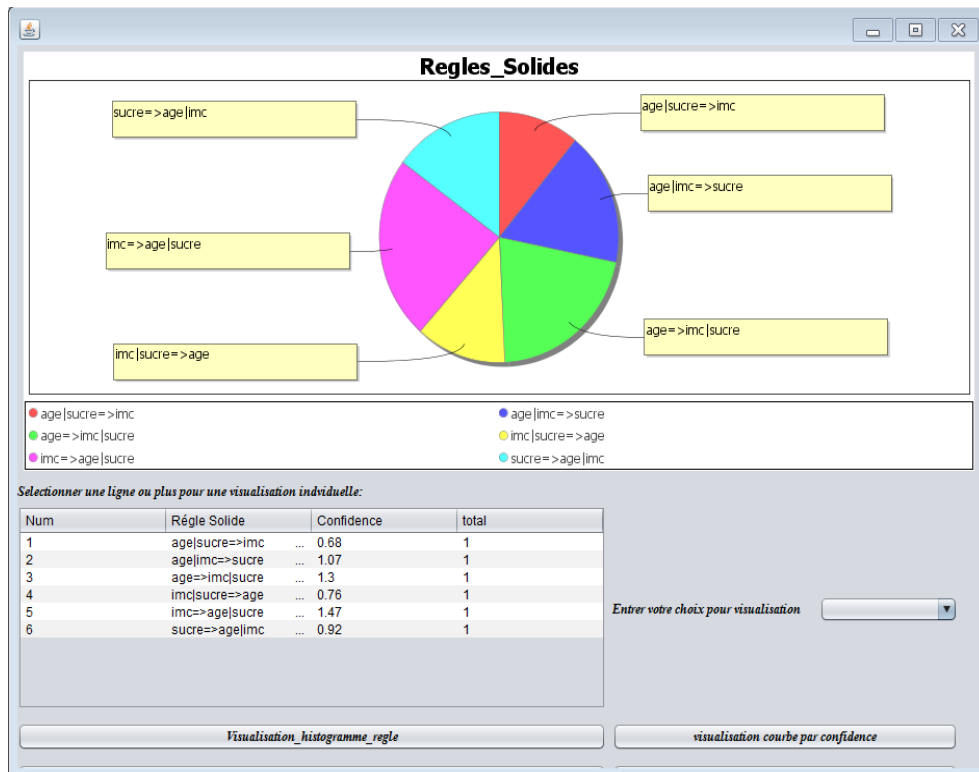


Figure 69 : le sous-système de la visualisation.

- ✓ Les figures suivantes représentent les visualisations par histogramme et courbe, suite aux choix de la sélection des règles ou intervalle de confiance.

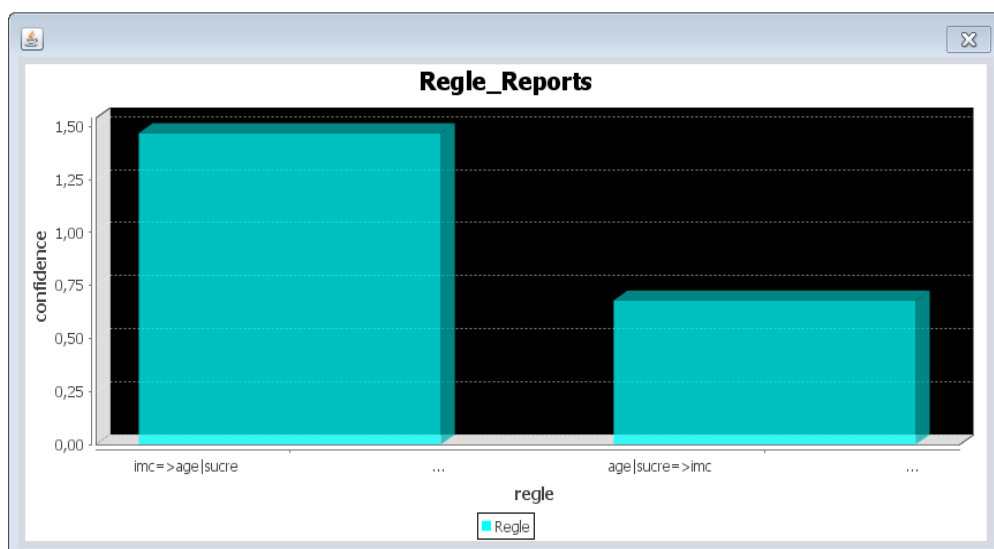


Figure 70 : Histogramme des règles.

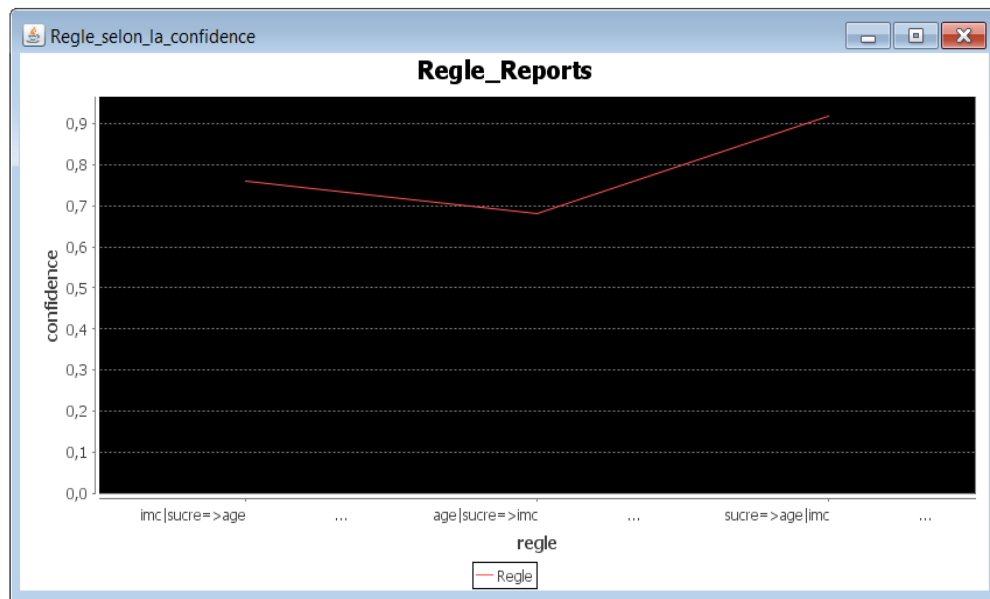


Figure 71 : Courbe des règles.

5. Conclusion

Dans ce chapitre, Nous avons représenté l'implémentation de notre système de prédiction des facteurs de risque du cancer du sein chez la femme, par la technique associative et, l'extraction des motifs fréquent et les règles solides.

Initialement on a présenté l'implémentation des différents modules composant notre système, Ensuite on a présenté un test de notre réalisation dans le but de validation par simulation du système,

Finalement nous avons essayé de visualiser les résultats obtenus avec notre système.

Conclusion générale

Conclusion et perspectives

Au terme de ce mémoire, nous procédons à un récapitulatif du travail effectué, rappelons que notre travail consiste à développer un modèle prédictif des facteurs de risque (déclencheurs) dans le développement du cancer du sein chez la femme.

Dans un premier temps, nous nous sommes attelés, à définir les différents concepts liés au data mining et au processus ECD, les méthodes de classifications associative pour finir par notre algorithme de sélection Apriori adapté au contexte du sujet.

Dans un second temps, nous avons procédé aux définitions de notre cas d'études le cancer du sein chez la femme, en concluant qu'il faut se focaliser sur les facteurs de risque comme éléments déclencheurs du cancer du sein.

La troisième partie de ce mémoire était consacré à la conception du système qui est basé sur l'algorithme Apriori et l'extraction des règles solides permettant de prédire un nouveau modèle prédictif des facteurs de risque liés au cancer du sein, où on a entamé cette phase par une analyse du processus de l'extraction des connaissances à partir des données saisies dans la base d'entraînement. Le système proposé comprend quatre sous-systèmes à savoir : le sous-système d'acquisition des données cibles, le sous-système du prétraitement, le sous-système de la classification par les motifs fréquents et enfin le sous-système de visualisation de résultats du modèle.

Le premier sous-système sert à l'acquisition des données cibles des patients, le deuxième prévoit un **prétraitement** sur les données cibles pour avoir uniquement les données utiles et nécessaire qui décrivent les facteurs de risque et à homogénéiser ces données pour pouvoir les disposer dans une table bidirectionnelle appelé **descripteur**, la classification associative par les motifs fréquent a été adapté pour l'estimation d'un nouveau modèle de prédiction dans le but d'aider le médecin à prendre les mesures nécessaires pour prédire le risque de la maladie.

Le modèle a donné des résultats de tests compatibles avec les dernières recherches effectuées actuellement en termes de facteurs de risque où les chercheurs montrent clairement que l'incidence élevé des principaux cancers essentiellement le cancer du sein précoce avant 50 ans, est en majeure partie causé par le mode de vie (l'obésité, la mauvaise alimentation et démunissions activité physique). Et qu'on peut prévenir en adoptant de saines habitudes de vie, maintien d'un poids normal, alimentation riche en végétaux et activité physique régulière.

La quatrième partie de ce mémoire était réservée à l'implémentation de notre modèle par les motifs fréquents ainsi que les différentes validations des différentes étapes de notre système.

Ce travail n'a pas été exempt d'obstacles, en effet nous avons été confrontés à l'obstacle des sources de données pour la base d'entraînement, où les formulaires des dépistages déjà faits dans nos services hospitaliers étaient incomplets et manquaient de facteurs pris dans notre étude.

Comme perspective, il serait intéressant d' :

- ✓ Étudier d'autres facteurs qui ne sont pas considérés à risque pour générer un nouveau modèle adapté à la population de sa propre région.
- ✓ Adapter le questionnaire lié au dépistage aux facteurs recherchés.
- ✓ Créer une base de données centralisée contenant tous les patients du pays.

Références Bibliographiques

Référence :

- [1] <http://depot-e.uqtr.ca/1201/1/030110265.pdf>
- [2] <http://dSPACE.univ-tlemcen.dz/bitstream/112/1045/4/Memoire.pdf>
- [3] <https://studylibfr.com/doc/4918004/rapport-de-stage-de-fin-d-%C3%A9tude>
- [4] https://hal.archives-ouvertes.fr/file/index/docid/467753/filename/Extraction_de_bases_pour_les_regles_d_association_a_partir_des_itemsets_fermes_frequents_Pasquier_INFORSID_2000.pdf
- [5] <http://biblos.hec.ca/biblio/memoires/m2006no62.pdf>
- [6] https://hal.archives-ouvertes.fr/file/index/docid/467760/filename/PASCAL_Un_algorithme_d_extraction_des_motifs_frequents_Bastide_et_al._TSI_2002.pdf
- [7] La recherche de motifs fréquents : une méthode de fouille de données : <http://bruno.maitresdumonde.com/optinfo/ads-scie/2007.pdf>
- [8] http://abdelhamid-djeffal.net/web_documents/polycopefda.pdf
- [9] <http://campus.cerimes.fr/maieutique/UE-obstetrique/glandemammaire/site/html/cours.pdf>
- [10] <https://tel.archives-ouvertes.fr/tel-00910388>
- [11] <https://www.e-cancer.fr/Patients-et-proches/Les-cancers/Cancer-du-sein/Facteurs-de-risque>
- [12] http://acces.ens-lyon.fr/acces/thematiques/sante/epidemiologie/cancersein/risq_cancer_sein
- [13] <http://www.insp.dz/index.php/Non-categorise/registre-des-tumeurs-d-alger.html>
- [14] <http://campus.cerimes.fr/gynecologie-et-obstetrique/enseignement/item139/site/html/2.html>
- [15] https://books.google.dz/books?id=1P6MbaXSNYwC&pg=PT219&lpg=PT219&dq=definition+de+oracle+8&source=bl&ots=6VNMAPs0uS&sig=ke4EGVOGQbhKHFix62AgzFFW1Xs&hl=fr&sa=X&ved=0ahUKEwjJ5oeV_sTbAhVQkRQKHxeMCdcQ6AEIjQEwCA#v=onepage&q=definition%20de%20oracle%208&f=false
- [16] définition de postgresql : https://www.postgresql.org/message-id/attachment/6011/presentation_erwan_fs.html
- [17] Amsalhem P, Eteessami R, Morère J-F. Epidémiologie, facteurs de risque, dépistage. Part of the series Oncologie pratique. 2007:P 11–22. [Google Scholar]

