

Intelligence artificielle explicable pour le cancer du sein : Une approche visuelle de raisonnement à partir de cas

Jean-Baptiste Lamy, Boomadevi Sekar, Gilles Guezennec, Jacques Bouaud, Brigitte Séroussi

Présentation de l'article suivant :

Explainable artificial intelligence for breast cancer: a visual case-based reasoning approach.
Artificial Intelligence in Medicine 2019



This work was funded by the European Union's Horizon 2020 research and innovation program under grant agreement No 690238.



Université Paris 13, 93017
Bobigny
Sorbonne Universités
INSERM UMRS 1142

Artificial Intelligence In Medicine 94 (2019) 42–53



ELSEVIER

Contents lists available at [ScienceDirect](#)

Artificial Intelligence In Medicine

journal homepage: www.elsevier.com/locate/artmed



Explainable artificial intelligence for breast cancer: A visual case-based reasoning approach

Jean-Baptiste Lamy^{a,*}, Boomadevi Sekar^b, Gilles Guezennec^a, Jacques Bouaud^{a,c},
Brigitte Séroussi^{a,d}

^a LIMICS, Université Paris 13, Sorbonne Universités, INSERM UMRS 1142, 93017 Bobigny, France

^b School of Computing and Mathematics, Ulster University, United Kingdom

^c AP-HP, DRCI, Paris, France

^d AP-HP, Hôpital Tenon, Département de Santé Publique, Paris, France

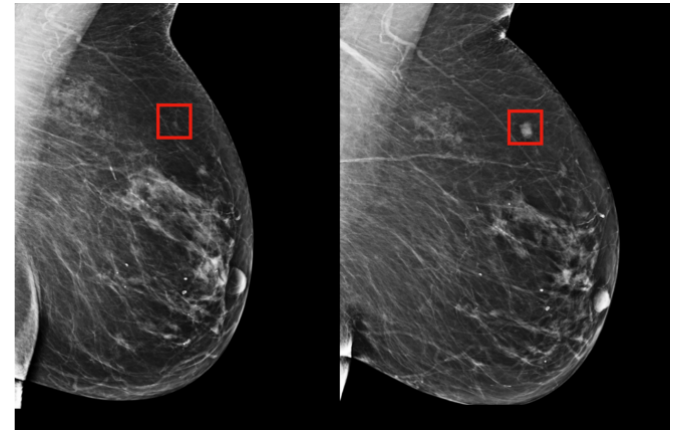


Introduction

- **Le cancer du sein est l'un des plus fréquent chez la femme**
 - ◆ 4 grandes catégories de traitements :
 - Chirurgie, chimiothérapie, thérapie endocrine, radiothérapie

- **L'intelligence artificielle (IA) peut aider au diagnostic**

- ◆ Analyse d'images
- ◆ Explications de l'IA : images annotées



Source: MIT

- **Mais aider à la thérapie est plus complexe !**

- ◆ Le médecin a besoin de comprendre les recommandations de l'IA
 - Dans un temps très limité (3 minutes par patient dans les unités pluridisciplinaires)

- => **IA explicable (*Explainable Artificial Intelligence, XAI*)**

Introduction

🚩 **Projet européen H2020 DESIREE**

◆ **Decision Support and Information Management System for Breast Cancer**



◆ Objectifs:

- Aider à la prise en charge du cancer du sein primitif
- ◆ Conception d'une plate-forme web avec 3 modules d'aide à la décision :
 - Implémentation des guides de bonnes pratiques via une ontologie
 - Apprentissage de règles statistique
 - Raisonnement à partir de cas (RAPC) visuel ←

Introduction

➤ Raisonnement à partir de cas (RAPC, CBR)

◆ Une forme de raisonnement analogique

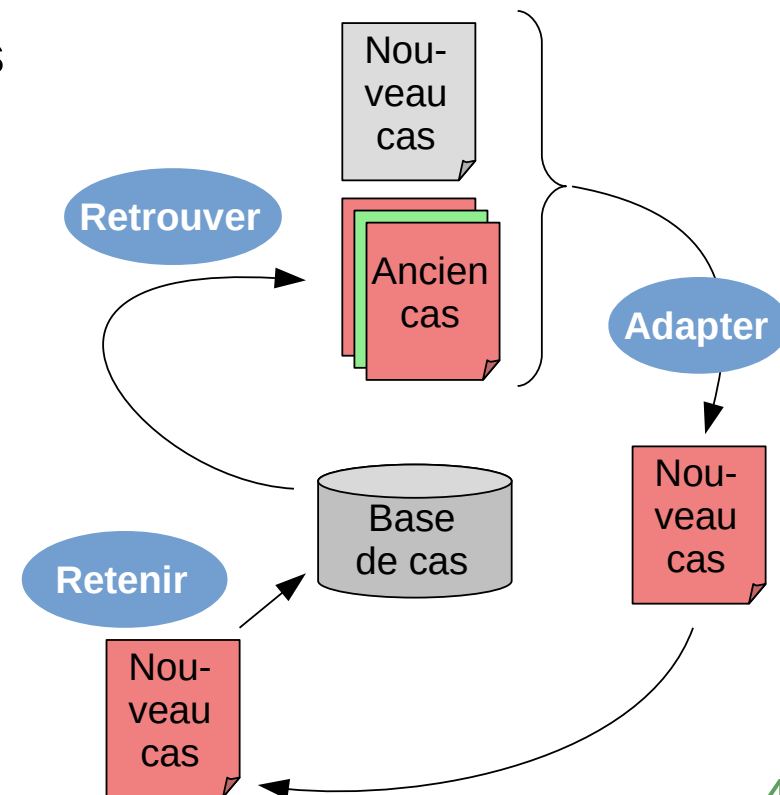
- Exemple type : k plus proche voisin (*k-nearest neighbor*, kNN)

◆ 3 étapes :

- Retrouver les anciens cas similaires dans une base d'anciens cas dont la solution est connue
- Adapter les solutions de ces cas au nouveau cas
- Retenir le nouveau cas dans la base, avec sa solution

◆ Dans le contexte thérapeutique :

- Un cas = un patient
- Une solution = un traitement



Introduction

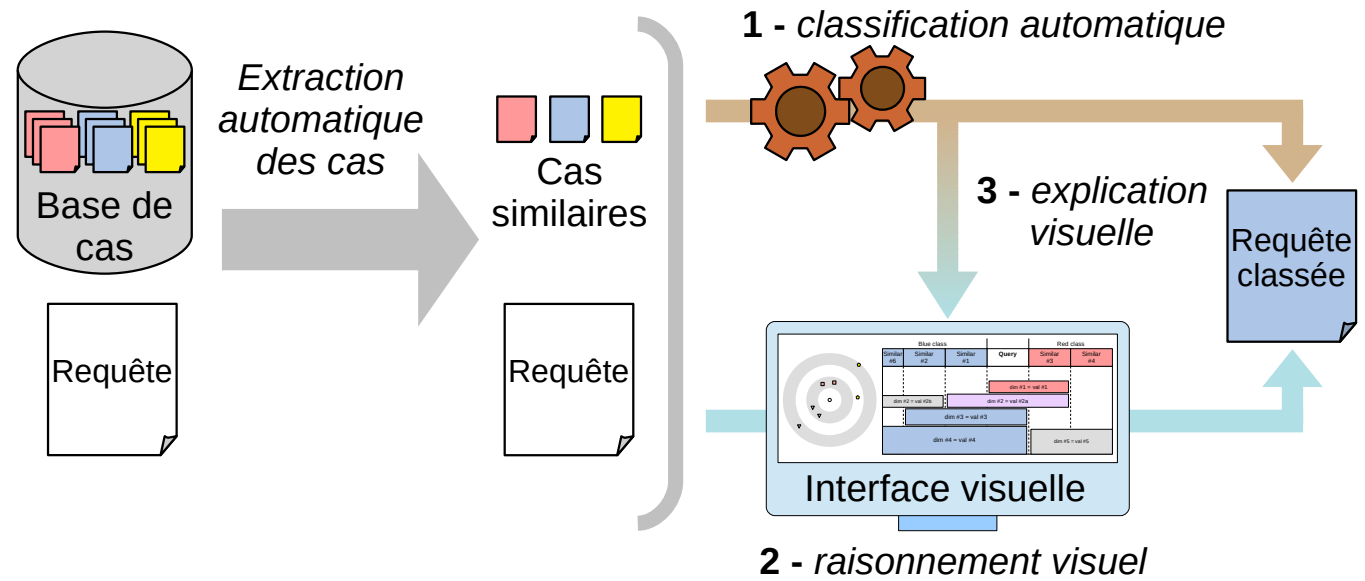
➤ **Raisonnement à partir de cas (RAPC, CBR)**

- ◆ Très intéressant pour produire des explications :
 - Les anciens cas peuvent servir d'explications
 - Ce mode de raisonnement est familier aux cliniciens

➤ => **Une explication peut consister en la présentation de 2-50 cas similaires**

- ◆ Mais 2-50 dossiers patient de cancérologie représentent un volume de données très important !
- ◆ Une solution est l'utilisation de la **visualisation d'information** pour présenter les cas et leur similarité avec le nouveau patient

Une approche automatique ou visuelle

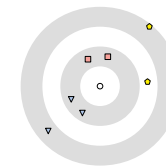


◆ Approche automatique ou visuel

- Traduire visuellement le RAPC

◆ Montre les similarités entre le nouveau cas et les anciens

- Similarités quantitatives
- Similarités qualitatives



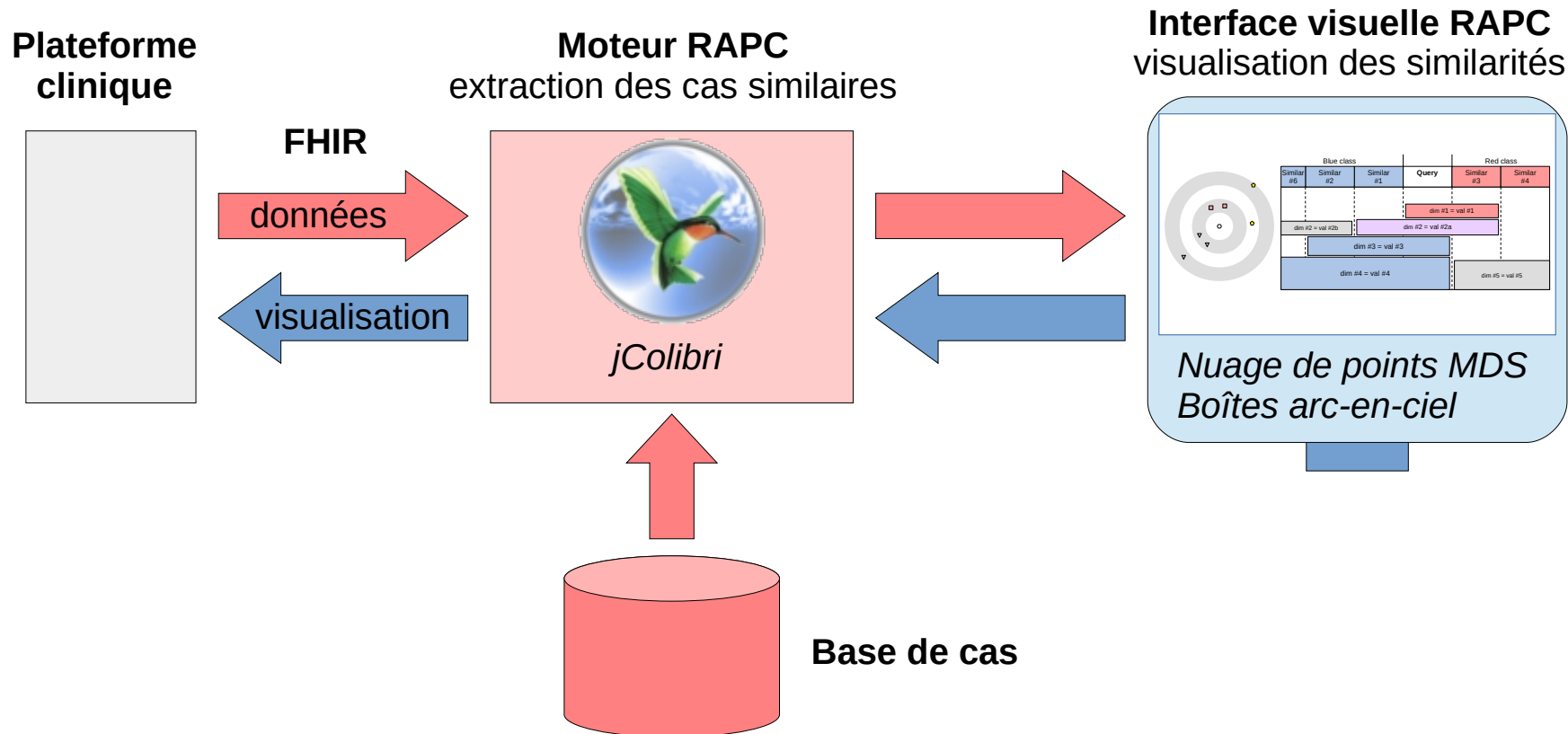
Approche quantitative
Montre les similarités
sous forme de distances

| Blue class | | | Query | Red class | |
|------------|------------|------------|-------|------------------|-----------------|
| Similar #1 | Similar #2 | Similar #3 | | Similar #3 | Similar #4 |
| | | | | dim #1 = val #1 | |
| | | | | dim #2 = val #2a | |
| | | | | dim #3 = val #3 | |
| | | | | dim #4 = val #4 | |
| | | | | | dim #5 = val #5 |

Approche qualitative
Montre les caractéristiques
communes

Architecture

- ◆ La base de cas est une base de données relationnelle
- ◆ Le standard HL7 FHIR est utilisé pour communiquer avec la plateforme clinique
- ◆ Les cas sont retrouvés avec jColibri



Matrice des distances

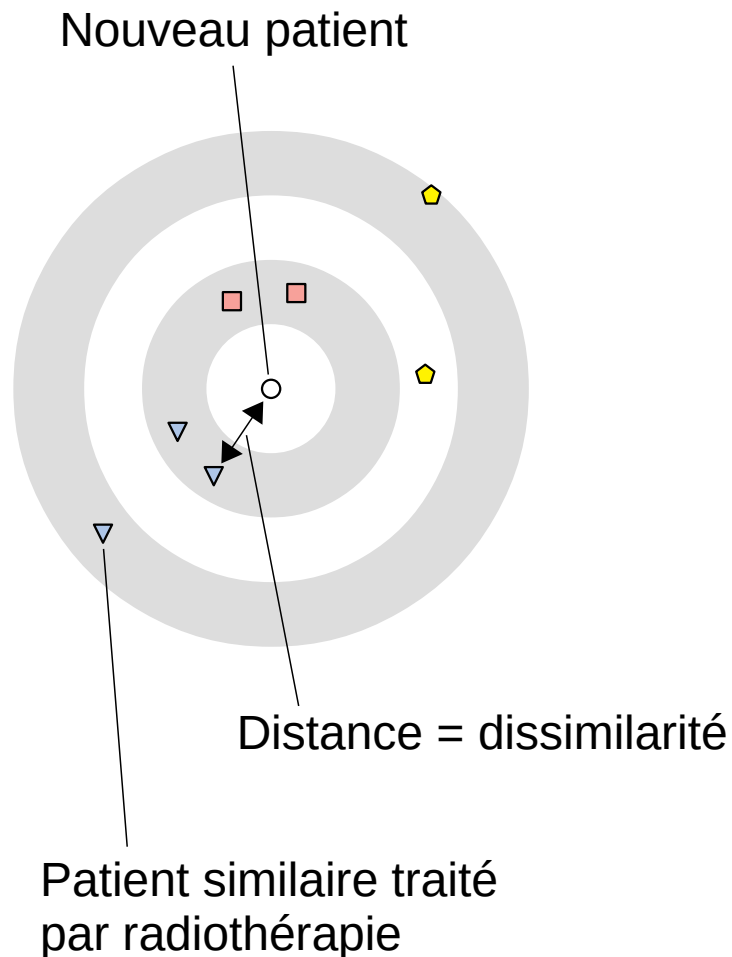
| | Query | Similar #1 | Similar #2 | Similar #3 | Similar #4 | Similar #5 | Similar #6 | Similar #7 |
|--------|----------------|----------------|----------------|----------------|---------------|------------|----------------|------------|
| dim #1 | val #1 | val #1b | val #1c | val #1 | val #1d | val #1e | val #1f | val #1g |
| dim #2 | val #2a | val #2a | val #2b | val #2a | val #2d | val #2e | val #2b | val #2g |
| dim #3 | val #3 | val #3 | val #3 | val #3c | val #3d | val #3e | val #3f | val #3g |
| dim #4 | val #4 | val #4 | val #4 | val #4c | val #4d | val #4e | val #4 | val #4g |
| dim #5 | val #5a | val #5b | val #5c | val #5 | val #5 | val #5e | val #5f | val #5g |
| dim #6 | val #6 | val #6a | val #6b | val #6c | val #6d | val #6e | val #6f | val #6g |
| ... | | | | | | | | |



- ◆ Jcolibri retrouve les cas similaires et calcule la matrice des distances entre cas

| | Query | Similar #1 | Similar #2 | Similar #3 | Similar #4 | Similar #5 | Similar #6 | Similar #7 |
|------------|-------|------------|------------|------------|------------|------------|------------|------------|
| Query | - | | | | | | | |
| Similar #1 | 2.0 | - | | | | | | |
| Similar #2 | 2.1 | 1.5 | - | | | | | |
| Similar #3 | 2.0 | 5.0 | 4.8 | - | | | | |
| Similar #4 | 1.9 | 5.1 | 4.9 | 1.1 | - | | | |
| Similar #5 | 4.5 | 5.2 | 5.2 | 6.0 | 6.1 | - | | |
| Similar #6 | 4,2 | 1.7 | 1.8 | 5.5 | 5.6 | 5.5 | - | |
| Similar #7 | 2.0 | 5.3 | 5.1 | 5.4 | 5.3 | 3.1 | 5.2 | - |

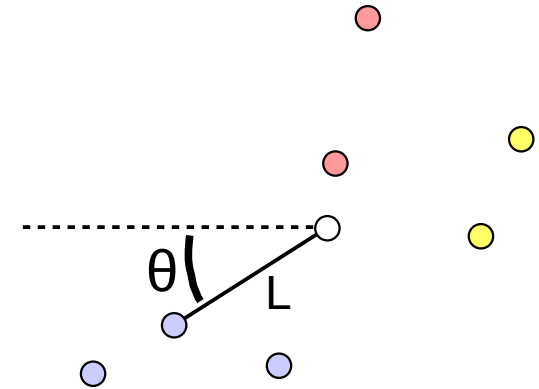
Visualisation des similarités quantitatives



- **Nuage de points : projection en 2D de la matrice des distances**
 - ◆ 1 point = 1 patient
 - ◆ 1 couleur = 1 classe
 - ◆ La cible au fond facilite l'évaluation des distances
- **Méthodes pour tracer le nuage de points**
 - ◆ MDS (multi-Dimensional Scaling), PCA, tSNE,...
- **Ici, 2 types de distances :**
 - ◆ A – Entre le nouveau patient et un patient similaire (**les plus importantes !**)
 - ◆ B – Entre deux patients similaires
- **=> Nous proposons une méthode de MDS en coordonnées polaires**
 - ◆ Préserve les distances du type A au détriment des distances de type B

MDS en coordonnées polaires

- ◆ Origine O = le nouveau patient
- ◆ Chaque patient similaire S est défini par ses coordonnées polaires (L, θ)
- ◆ L figure dans la matrice des distances
- ◆ θ est déterminé en résolvant un problème d'optimisation:
 - Trouver les meilleurs valeurs de θ qui minimisent la fonction de stress :



$$S_p(d) = \sum_{2 < i < j} \frac{(d_{ij} - \delta_{ij})^2}{d_{ij}}$$

Distance sur le nuage de point en 2 dimensions (dépend des valeurs de θ)

Distance vraie dans n dimensions

Le nouveau patient est le numéro 1

Visualisation des similarités qualitatives

Boîtes arc-en-ciel

- ◆ Une technique récente de visualisation d'ensembles
- ◆ 1 patient = 1 élément
- ◆ 1 caractéristique partagée = 1 ensemble
 - Ensemble des patients dont l'âge est supérieur à 60,...
- ◆ Seules les deux classes majoritaires sont conservées
- ◆ Les valeurs numériques sont discrétisées (Minimum Description Length Principle, MDLP)
- ◆ Seules les boîtes ayant la plus haute Information Mutuelle (MI) sont conservées

1)

| | Query | Similar #1 | Similar #2 | Similar #3 | Similar #4 | Similar #5 | Similar #6 | Similar #7 |
|--------|---------|------------|------------|------------|------------|------------|------------|------------|
| dim #1 | val #1 | val #1b | val #1c | val #1 | val #1d | val #1e | val #1f | val #1g |
| dim #2 | val #2a | val #2a | val #2b | val #2a | val #2d | val #2e | val #2b | val #2g |
| dim #3 | val #3 | val #3 | val #3 | val #3c | val #3d | val #3e | val #3f | val #3g |
| dim #4 | val #4 | val #4 | val #4 | val #4c | val #4d | val #4e | val #4 | val #4g |
| dim #5 | val #5a | val #5b | val #5c | val #5 | val #5 | val #5e | val #5f | val #5g |
| dim #6 | val #6 | val #6a | val #6b | val #6c | val #6d | val #6e | val #6f | val #6g |
| ... | | | | | | | | |

2)

| | Similar #6 | Similar #2 | Similar #1 | Query | Similar #3 | Similar #4 |
|--------|------------|------------|------------|---------|------------|------------|
| dim #1 | val #1f | val #1c | val #1b | val #1 | val #1 | val #1d |
| dim #2 | val #2b | val #2b | val #2a | val #2a | val #2a | val #2d |
| dim #3 | val #3f | val #3 | val #3 | val #3 | val #3c | val #3d |
| dim #4 | val #4 | val #4 | val #4 | val #4 | val #4c | val #4d |
| dim #5 | val #5f | val #5c | val #5b | val #5a | val #5 | val #5 |
| dim #6 | val #6f | val #6b | val #6a | val #6 | val #6c | val #6d |
| ... | | | | | | |

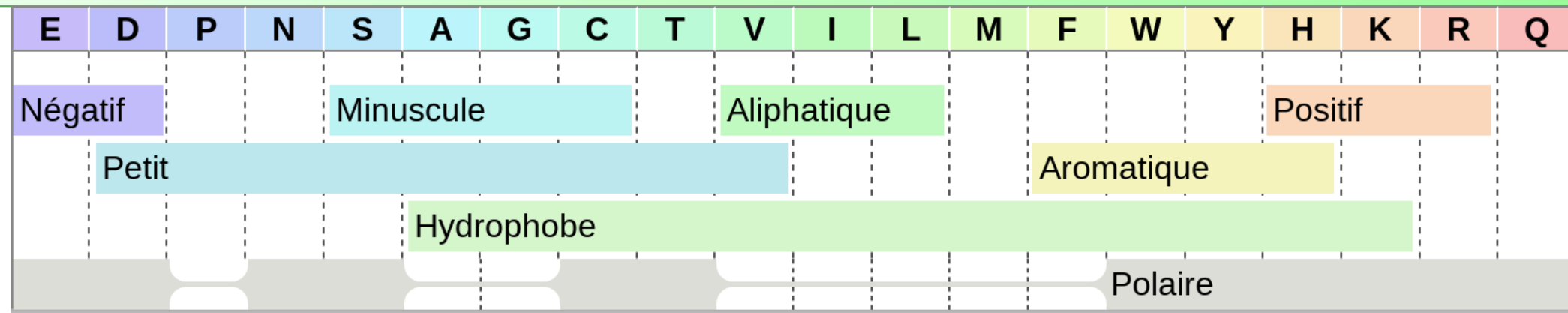
3)

| | Similar #6 | Similar #2 | Similar #1 | Query | Similar #3 | Similar #4 |
|--------|------------------|-----------------|------------------|-----------------|-----------------|------------|
| dim #1 | val #1f | val #1c | val #1b | dim #1 = val #1 | | val #1d |
| dim #2 | dim #2 = val #2b | | dim #2 = val #2a | | | val #2d |
| dim #3 | val #3f | dim #3 = val #3 | | | val #3c | val #3d |
| dim #4 | dim #4 = val #4 | | | | val #4c | val #4d |
| dim #5 | val #5f | val #5c | val #5b | val #5a | dim #5 = val #5 | |
| dim #6 | val #6f | val #6b | val #6a | val #6 | val #6c | val #6d |
| ... | | | | | | |

4)

| | Similar #6 | Similar #2 | Similar #1 | Query | Similar #3 | Similar #4 |
|--|------------------|------------|------------------|-----------------|-----------------|------------|
| | | | | dim #1 = val #1 | | |
| | dim #2 = val #2b | | dim #2 = val #2a | | | |
| | dim #3 = val #3 | | | | | |
| | dim #4 = val #4 | | | | dim #5 = val #5 | |

Les boîtes arc-en-ciel



➤ Une technique récente pour visualiser des ensembles

- éléments => colonnes
- ensembles => boîtes rectangulaires
- Élément non contiguous element in a set => box hole
- Les éléments sont ordonnés par un algorithme métaheuristique (Artificial Feeding Birds) de sorte à placer côte à côte les éléments qui appartiennent aux mêmes ensembles
- Lorsque cela n'est pas possible, des "trous" sont présents dans les boîtes
- Les boîtes sont empilées avec les plus grandes en bas

Traduction algorithmique du raisonnement visuel

function classify(q, X, s, n, m):

q is the query case

X is the case database (we assume that $q \in X$)

s is the dissimilarity measure (a function taking 2 cases and returning their dissimilarity, *e.g.* Euclidean distance)

$n \geq 2$ is the total number of cases considered (query + similar cases)

$m \geq 1$ is the maximum number of boxes selected

For each case i in X , compute $s(q, X_i)$

Let X' be the set of selected cases, X' contains the n elements of X with the lowest dissimilarity $s(q, i)$

We assume that $X'_1 = q$ and X'_2 to X'_n are the similar cases

Let d be the distance matrix between cases in X'

For each case i in X' :

For each case j in X' :

$$d_{i,j} = s(i, j)$$

Let w be the weights of the similar cases

$$w_i = \begin{cases} 1 & , \text{if } s_{max} = s_{min} \\ \frac{s_{max} - s(q, X'_i)}{s_{max} - s_{min}} & , \text{otherwise} \end{cases}$$

with $s_{min} = \min(s(q, X'_i))$ and $s_{max} = \max(s(q, X'_i))$

Let y_1 and y_2 be the two best classes in X' (determined by a majority vote over similar cases, weighted by w_i)

Let X'' be the subset of X' displayed in rainbow boxes

$$X'' = X' \cap (\{q\} \cup y_1 \cup y_2)$$

Let B be the set of candidates boxes (currently empty)

For each dimension Z :

If Z has numeric values:

Discretize Z

For each value v that Z takes in cases X'' :

If q has value v for dimension Z :

Add $Zv = \{x \in X'' \mid x_Z = v\}$ **into** B

For each box Zv in B , compute $MI(Z_v Y) = \sum_{z \in \{Z=v, Z \neq v\}} \sum_{y \in \{y_1, y_2\}} p(z, y) \log \left(\frac{p(z, y)}{p(z)p(y)} \right)$

with $p(y) = \frac{|y|}{|X'' \setminus \{q\}|}$, $p(Z = v) = \frac{|Z_v|}{|X'' \setminus \{q\}|}$, $p(Z \neq v) = \frac{|X'' \setminus Z_v|}{|X'' \setminus \{q\}|}$, $p(Z = v, y) = \frac{|Z_v \cap y|}{|X'' \setminus \{q\}|}$, $p(Z \neq v, y) = \frac{|(X'' \setminus Z_v) \cap y|}{|X'' \setminus \{q\}|}$

Let B' be the set of selected boxes, B' contains the m elements of B with the highest $MI(Z_v Y)$

Compute $S_{y_1} = \sum_{Z_v \in B'} \left(MI(Z_v Y) \times \sum \{w_{2 \leq i \leq n} \mid x_i \in Z_v \cap y_1\} \right)$

Compute $S_{y_2} = \sum_{Z_v \in B'} \left(MI(Z_v Y) \times \sum \{w_{2 \leq i \leq n} \mid x_i \in Z_v \cap y_2\} \right)$

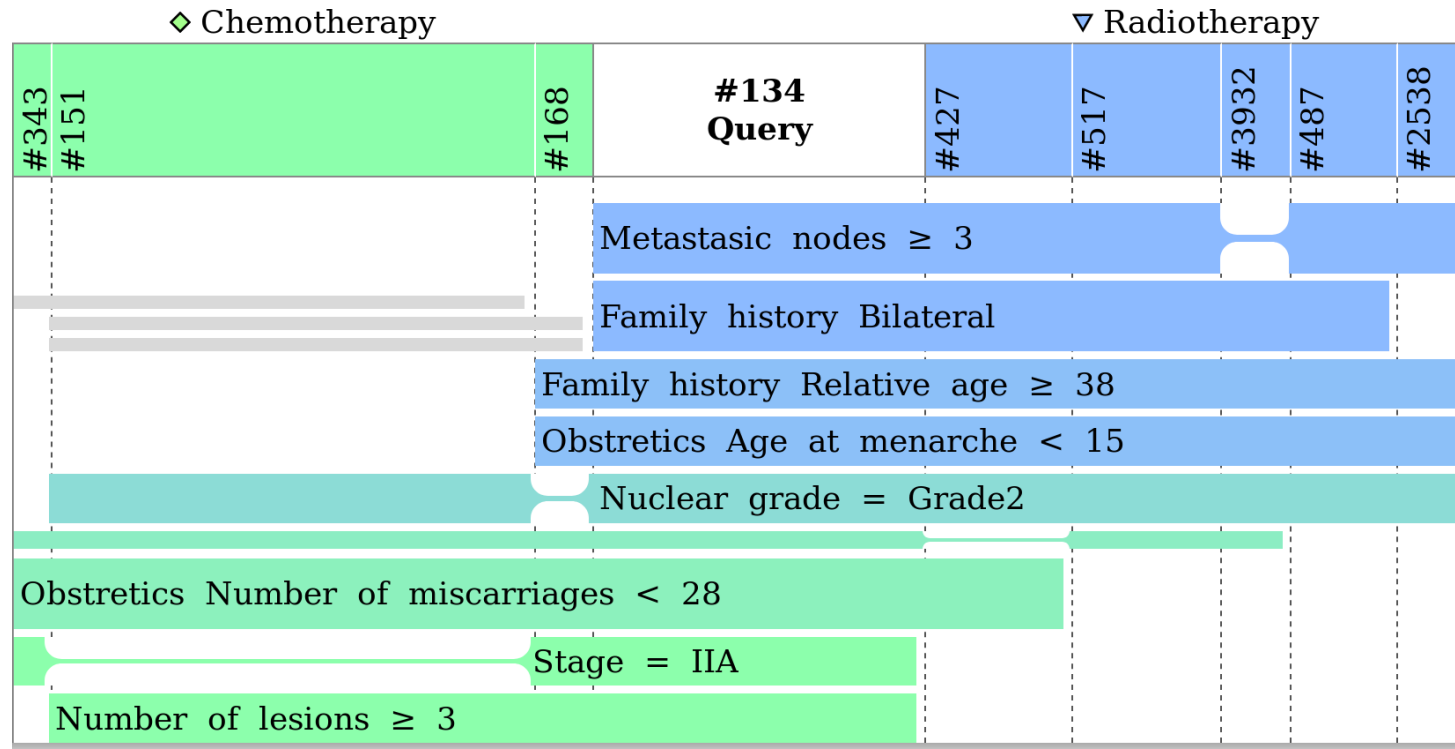
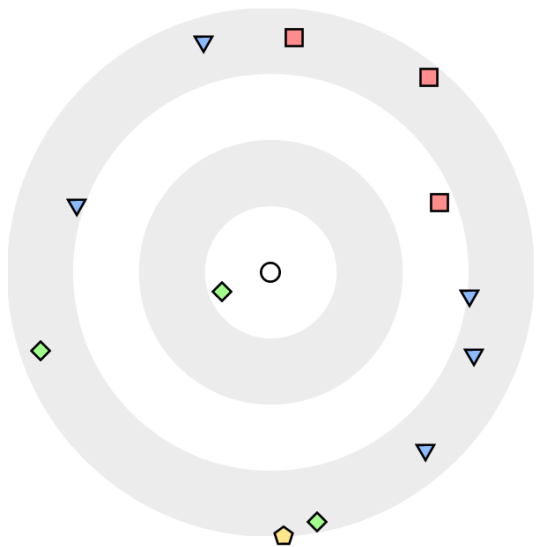
If $S_{y_1} > S_{y_2}$:

return y_1

Else:

return y_2

Interface résultante



- ◆ Les boîtes donnent des arguments en faveur de l'une ou l'autre classes
- ◆ Le traitement recommandé est celui dont la couleur est majoritaire
- ◆ Le clinicien peut choisir une option différente, s'il juge les arguments non pertinents

Discussion et conclusion

- Performances équivalente au kNN, mais explicabilité meilleure
- La visualisation d'ensemble paraît prometteuse pour l'IA explicable
- Limites :
 - ◆ Nécessite une formation des médecins
 - ◆ Les traitements recommandés correspondent à ceux prescrits dans le passé, mais rien ne garanti que ce soient les meilleurs
- Perspectives:
 - ◆ Validation et évaluation cliniques
 - ◆ Adaptation à d'autres domaines
 - ◆ Extension à d'autres méthodes d'IA (*deep learning, boosting, ...*)

References

JB Lamy, B Sekar, G Guezenec, J Bouaud, B Séroussi. Explainable artificial intelligence for breast cancer: a visual case-based reasoning approach. Artificial Intelligence in Medicine 2019

JB Lamy, B Sekar, G Guezenec, J Bouaud, B Séroussi. Hierarchical visual case-based reasoning for supporting breast cancer therapy. International Conference on Advances in Biomedical Engineering (ICABME) 2019

N Choudhury, SA Begum. A survey on case-based reasoning in medicine. International Journal of Advanced Computer Science and Applications (IJACSA) 2016

JM Schoenborn, KD Althoff. Recent Trends in XAI: A Broad Overview on current Approaches, Methodologies and Interactions. Case-Based Reasoning for the Explanation of intelligent systems (XCBR) Workshop 201

JB Lamy. Python et les ontologies. ENI éditions, 2019

