

# RAPPORT INTERMÉDIAIRE

## Détection et caractérisation de callot sanguin grâce au IRMs

*Detection and characterization of Clot with MRIs*

Écrit par:

Philippe Chuzel

Sous la supervision de :

Ali Mansour



**ENSTA**  
Bretagne

Option Systèmes Perception Information Décision

# Contents

<b>I État de l'art</b>	<b>6</b>
<b>1 Qu'est ce que la thrombose veineuse profonde</b>	<b>7</b>
1.1 terminologie . . . . .	7
1.2 Causes et conséquence . . . . .	7
<b>2 Quels moyens de détection?</b>	<b>9</b>
2.1 Elastométrie . . . . .	9
2.2 Échographie . . . . .	10
2.3 Computed tomography angiography / Angiographie . . . . .	10
2.4 Scanner . . . . .	11
2.5 IRM . . . . .	11
<b>3 Option suivie</b>	<b>12</b>
3.1 Obtention de courbe de rehaussement à partir des IRMs de perfusion. . . . .	13
3.2 Carte de paramètre pharmacocinétique. . . . .	14
3.3 Classification des courbes de rehaussement à partir de la classification spectrale. . . . .	15
<b>II Classification spectrale</b>	<b>19</b>
<b>4 Intérêt de la classification spectrale.</b>	<b>20</b>
<b>5 Rappel sur la théorie des graphes.</b>	<b>22</b>
5.1 Définitions de base. . . . .	22
5.2 Choix du laplacien de la matrice. . . . .	22
<b>6 Construction de la matrice de voisinage.</b>	<b>23</b>
6.1 Choix du type de voisinage. . . . .	23
6.2 Choix de la distance et de la similitude. . . . .	23
6.3 Choix du paramètre de dispersion. . . . .	24
<b>7 Quels algorithmes de classification spectrale?</b>	<b>26</b>
<b>8 Résultats de l'algorithme et interprétations.</b>	<b>29</b>
<b>III Application de la classification spectrale sur des données réelles</b>	<b>33</b>
<b>9 Réalisation de la chaîne de traitement.</b>	<b>34</b>
<b>10 IMRs T2 du crane CHRU Brest Cavale Blanche.</b>	<b>36</b>

<b>11 IMRs T1 de la prostate de l'INSERM de Lille.</b>	<b>37</b>
11.1 Types de courbe rencontrées dans le cadre du cancer de la prostate . . . . .	37
11.2 Résultats de l'algorithme . . . . .	38
<b>IV La problématique du nombre de classe</b>	<b>40</b>
<b>12 Choix automatique du nombre de classe.</b>	<b>41</b>
12.1 Algorithme 1 : Analyse des valeurs propres et le n-gap . . . . .	41
12.2 Algorithme 2 : Minimisation par la norme de Frobenius . . . . .	42
12.3 Algorithme 2 : Minimisation par rotation . . . . .	43
<b>13 Application de ces algorithmes sur des données réelles</b>	<b>44</b>
<b>V Annexes</b>	<b>45</b>

# **Abstract**

# **Introduction**

# **Part I**

## **État de l'art**

# Chapter 1

## Qu'est ce que la thrombose veineuse profonde

### 1.1 terminologie

On appelle thrombose le fait qu'un caillot sanguin ou thrombus se développe à l'intérieur d'un vaisseau sanguin. Il peut être soit artériel, soit veineux. Dans les deux cas, ce caillot peut obstruer le vaisseau et donc empêcher la circulation du sang.

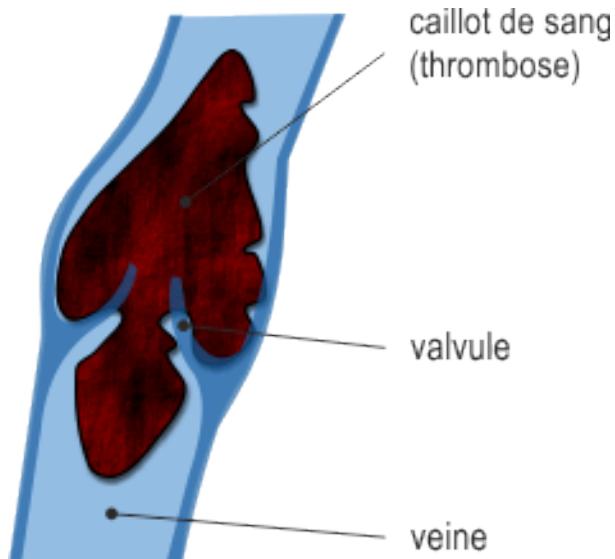


Figure 1.1: Exemple illustrant la thrombose.

La phlébite correspond au cas de la thrombose veineux et plus précisément pour une veine profonde et est consécutif à l'inflammation de la veine et au risque de déplacement du caillot. D'où le terme de thrombose veineux profond DVT.

### 1.2 Causes et conséquence

Parmi les principales causes, nous pouvons énumérer:

- Une lésion dans une veine qui entraîne une coagulation au niveau du tissu sanguin (blessure...).
- Certains paramètres génétiques qui peuvent favoriser le processus de thrombose.
- Les périodes prolongées dans une même position (alitement, accouchement, voyage long en avion...).

- Des précédentes expériences traumatisantes pur le corps (cancer, opération...)
- Influence de certaines substances sur la paroi des artères (tabac, médicaments...)
- Facteurs favorisant la formation de thrombus (obésité, diabète, hypercholestérolémie...)

Une fois le caillot formé, il peut se produire différentes chose:

- Il peut de lui même se désagréger et disparaître naturellement.
- Il peut se calcifier ou changer de structure ce qui peut amener à la formation de pelote fibreuse.
- Il peut se détacher et atteindre un des vaisseaux pulmonaires. Ceci se traduit par une embolie pulmonaire qui s'avère très dangereuse.

Dans tous les cas, la présence de ce caillots dans le système circulatoire sanguin d'un patient peut mener à des perturbations graves du flux sanguin avec l'apparition d'oedème ou d'ulcère pour la zone inférieur du corps que nous allons étudier.

# Chapter 2

## Quels moyens de détection?

Nous allons donner ici une liste succincte des méthodes existantes qui pourraient s'avérer utiles pour la caractérisation des caillots sanguins.

### 2.1 Elastométrie

L'élastométrie est une méthode d'acquisition d'image médicale non invasive très pratique car très facile à mettre en œuvre, peu coûteuse en temps et en argent. Elle consiste à recréer une représentation d'une zone du corps à partir de l'élasticité du tissu biologique [9].

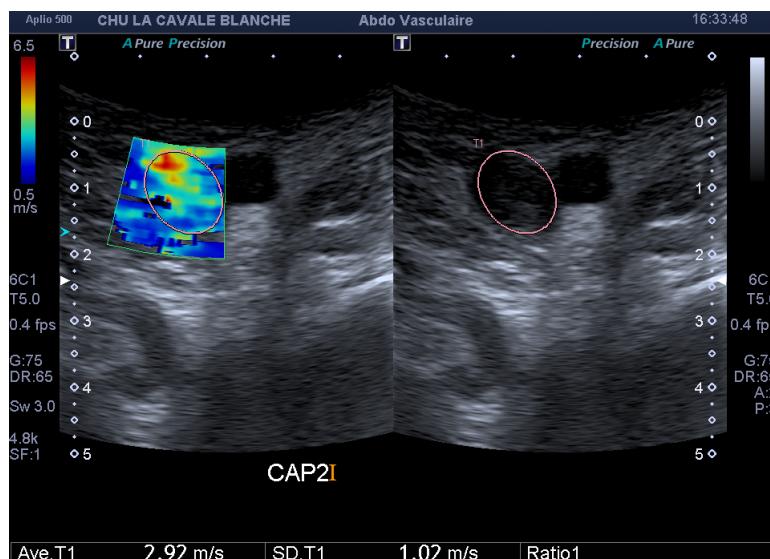


Figure 2.1: Exemple d'image d'élastométrie.

En l'état, nous pouvons donner les résultats suivants:

- Il est toujours nécessaire de définir un protocole unique à tous les patients pour faire en sorte que les images obtenues par ce procédé soient comparables.
- Une nouvelle sonde sera disponible dans les prochaines semaines et une partie des images vont être probablement refaites.
- Bien que la mise en œuvre soit facile, les images obtenues sont de médiocre qualité en terme de résolution.

Pour l'instant, un doctorant cherche des pistes dans cette piste pour la détection et la caractérisation des thromboses veineuses sanguin.

## 2.2 Échographie

L'échographie est une technique d'imagerie qui consiste à utiliser des ondes ultrasons. Là encore, en fonction des propriétés du milieu, certains tissus ou parties du corps apparaîtront en noir, gris ou blanc.



Figure 2.2: Exemple d'image échographique pour l'obstétrerie.

Cette méthode présente les mêmes avantages. Elle est peu chère, facile à mettre en œuvre et non invasive pour le patient. Néanmoins, elle présente également des images de très faible qualité en terme de résolution.

Cette piste est également explorée pour ce projet.

## 2.3 Computed tomography angiography / Angiographie

L'angiographie est une technique d'imagerie qui est invasive. Bien qu'elle présente plusieurs intérêts, notamment car elle consiste à faire une imagerie des vaisseaux sanguins par rayon X (veines ou artères), il y a plusieurs effets secondaires (vomissement, vertiges, nausées, réactions allergiques, saignement important...) ou contre-indications (Grossesse, tension artérielle trop basse...). Cette piste ne sera pas étudiée ici.

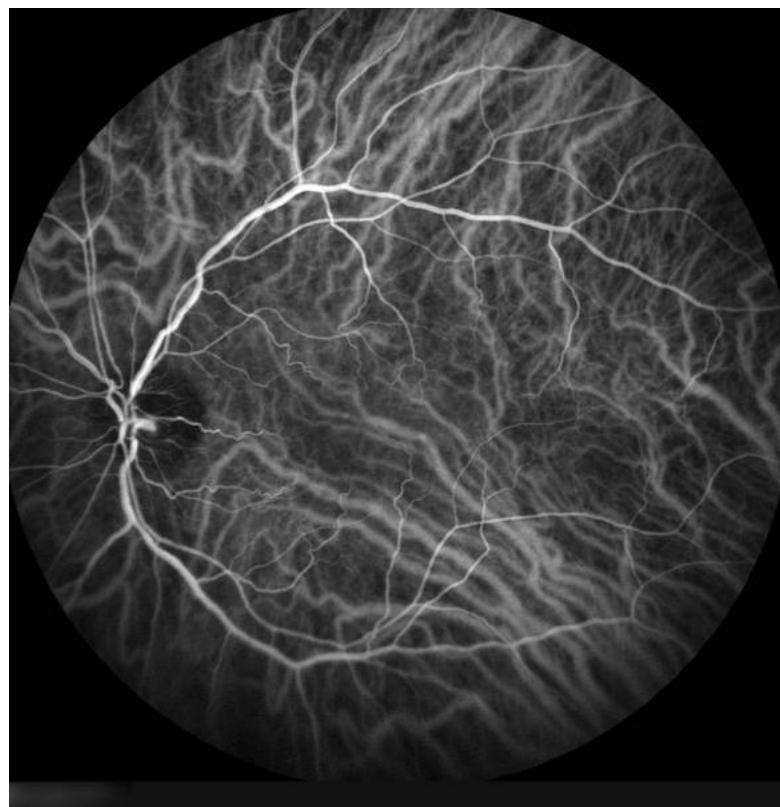


Figure 2.3: Exemple d'angiographie pour l'œil.

## 2.4 Scanner

**Rencontre avec un médecin prévu jeudi 21 avril**

## 2.5 IRM

**Rencontre avec un médecin prévu jeudi 21 avril**

# Chapter 3

## Option suivie

Dans le cadre de ce PFE, nous allons nous focaliser sur deux méthodes qui sont développées dans la thèse de M. Thatare [13] dans le cadre de l'aide au diagnostic du cancer de la prostate. La première va aboutir à la réalisation de carte de paramètre pharmacocinétique alors que l'autre va aboutir à faire une classification directement sur toutes les données.

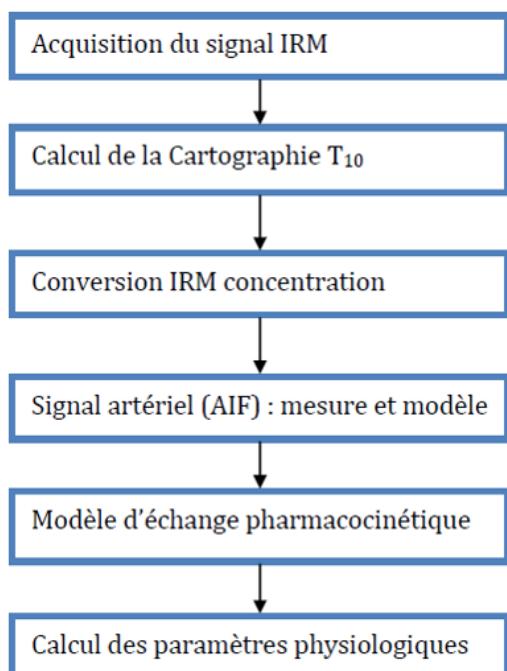


Figure 3.1: Méthode thèse Tartare.

### 3.1 Obtention de courbe de rehaussement à partir des IRMs de perfusion.

La stratégie qui a été développée dans cette thèse peut être résumer en quelques lignes. Ils ont proposé d'utiliser les IRMs de perfusion, ici l'IRM T1 ou DCE-MRI, afin de décrire le fonctionnement des tissus de la zone étudiée en fonction d'une certaine caractéristique. Ici, la stratégie adoptée est de décrire la vascularisation des différents zones de la prostate car les zones tumorales ont tendance à avoir une vascularisation plus importante et donc cela va se traduire un hypo ou hyper signal sur l'IRM.

Le principe de l'IRM de perfusion développée ici est d'injecter un produit de contraste, ici le galonium, qui est visible par IRM et de voir comment se produit sa diffusion au cours du temps dans les différentes zones du corps. À terme, nous obtenons un ensemble d'image traduisant l'évolution de la concentration du produit dans la zone étudiée au cours du temps. A partir de ceci, nous pouvons déterminer à partir de ces images la courbe d'intensité au cours du temps de chaque pixel de l'IRM. L'idée est donc de voir le comportement d'une zone du corps avant pendant et après injection de PdC.

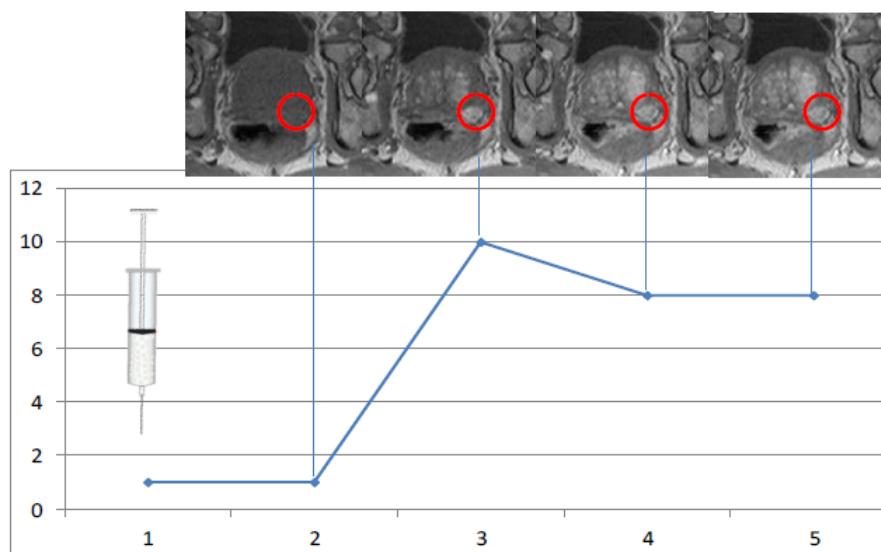


Figure 3.2: Exemple courbe de rehaussement.

Plusieurs méthodes d'analyse s'offrent à nous à partir d'ici:

- Analyse qualitative: Méthode d'analyse qui consiste à regarder la forme de la courbe et en déduire "à l'œil" l'état du tissu.
- Analyse semi-quantitative: Méthode plus rigoureuse que la précédente qui consiste à calculer le temps de pic de la courbe, le temps d'absorption et de dilution (wash-in et wash-out)... afin d'en déduire des caractéristiques sur le tissu.
- Analyse quantitative: Méthode d'analyse beaucoup plus contraignante car elle impose de proposer un modèle qui peut caractériser les échanges entre cellules afin de traduire l'évolution de la concentration de PdC dans le tissu. Néanmoins, avec une telle méthode, nous pouvons obtenir des cartes de paramètres pharmacocinétiques qui traduisent les zones homogènes dans la zone étudiée.

La méthode qui a été choisie est la méthode quantitative car elle permet d'obtenir des résultats particulièrement pertinents pour réaliser de l'aide au diagnostique.

Néanmoins, un gros problème reste encore à résoudre. Nous avons ici que des courbes d'intensité en fonction du temps et non pas de concentration en fonction du temps. Il faut donc encore trouver un moyen de trouver une transformation entre ces deux courbes.

Pour ce faire, il faut déterminer deux paramètres:

- L'AIF: Fonction d'entrée artérielle. Cette fonction fait référence à la concentration de PdC dans le vaisseau alimentant le tissu étudié. Il faut donc détecter quelle est l'artère alimentant le tissu, l'isoler et estimer cette fonction à partir de la courbe de rehaussement de tous les points appartenant à la veine.
- T10: Temps de relaxation longitudinal au temps zéro. Cette constante de temps est fortement liée au temps de réponse du tissu à la perturbation engendrée par l'injection de produit de contraste.

Avec ces deux paramètres, il est possible de passer de la courbe de intensité au cours du temps à la courbe concentration au cours du temps. Ces courbes sont ensuite approximées par des expressions théoriques qui correspondent à des modèles. On peut donc en déduire quels sont les valeurs des différents paramètres qui permettent d'interpoler ces courbes et ainsi réaliser des cartes de paramètre pharmacocinétique. L'autre possibilité est de faire une classification sur les différentes courbes de concentration directement.

## 3.2 Carte de paramètre pharmacocinétique.

A partir du travail qui a été réalisé au-dessus, il nous est possible de faire des cartes montrant les variations des valeurs des paramètres pharmacocinétiques. L'idée est de récupérer les valeurs des paramètres qui permettent d'interpoler les courbes de concentration en fonctions du temps et de voir quelles sont les zones de la prostate qui présentent les même propriété.

Le modèle utilisé dans cette thèse est appelé modèle de Toft et on parvient au final à des cartes comme ci-dessous 3.3.

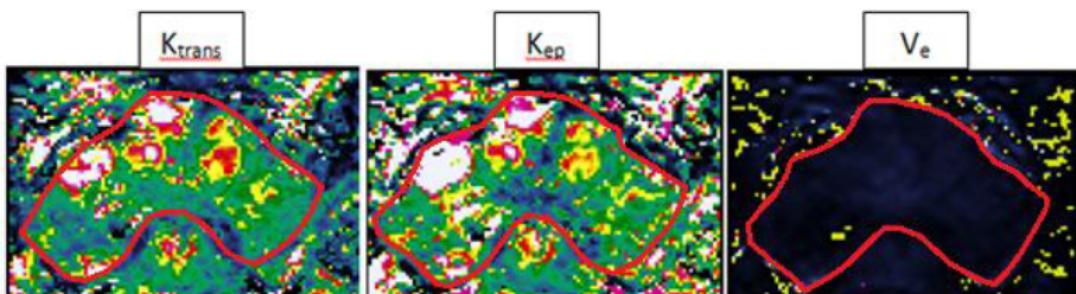


Figure 3.3: Carte de paramètres pharmacocinétiques.

La figure qui suit 3.4 montre un exemple d'application particulièrement parlant. L'IRM de gauche représente l'IRM de perfusion sous angle T1, l'IRM de droite représente l'IRM de perfusion sous angle T2 et la carte à droite représente les différentes lésions de la prostate après analyse par un médecin. Au final, La carte obtenue par le paramètre  $K_{trans}$  permet d'identifier de façon très claire une des lésions de la prostate du patient.

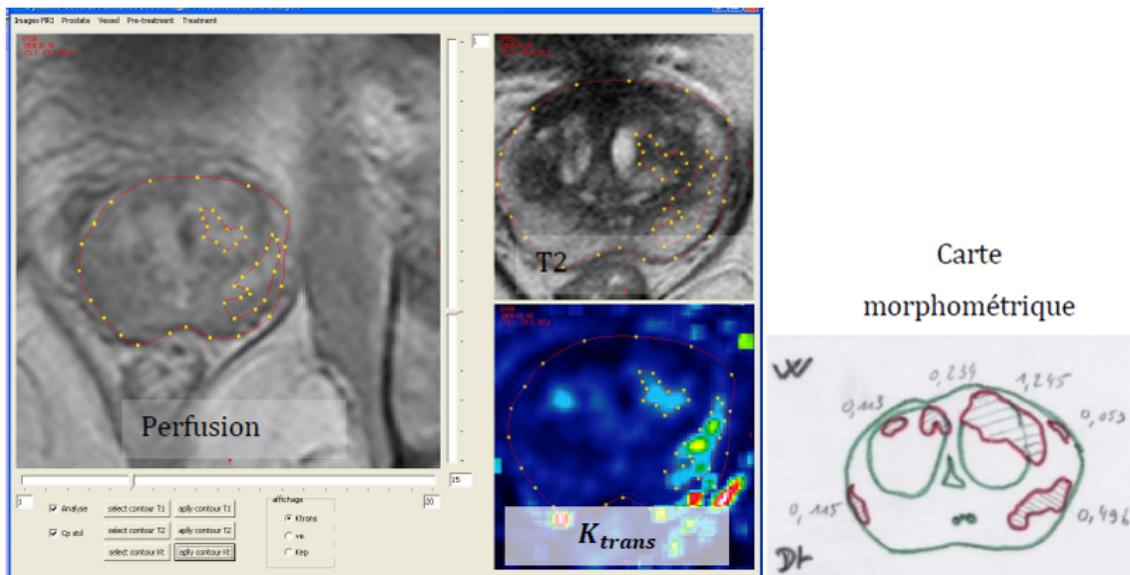


Figure 3.4: Exemple de carte de paramètres pharmacocinetiques.

### 3.3 Classification des courbes de rehaussement à partir de la classification spectrale.

Autre méthodologie développée pour réaliser des programmes d'aide au diagnostic consiste à réaliser une classification ou des clusters des différentes courbes d'intensité obtenues afin d'isoler les courbes présentant le même comportement entre elles. Ainsi, on obtient une nouvelle carte segmentée qui indique les régions qui présentent les mêmes caractéristiques, ici la vascularisation.

La figure qui suit montre la méthodologie appliquée ici 3.5.

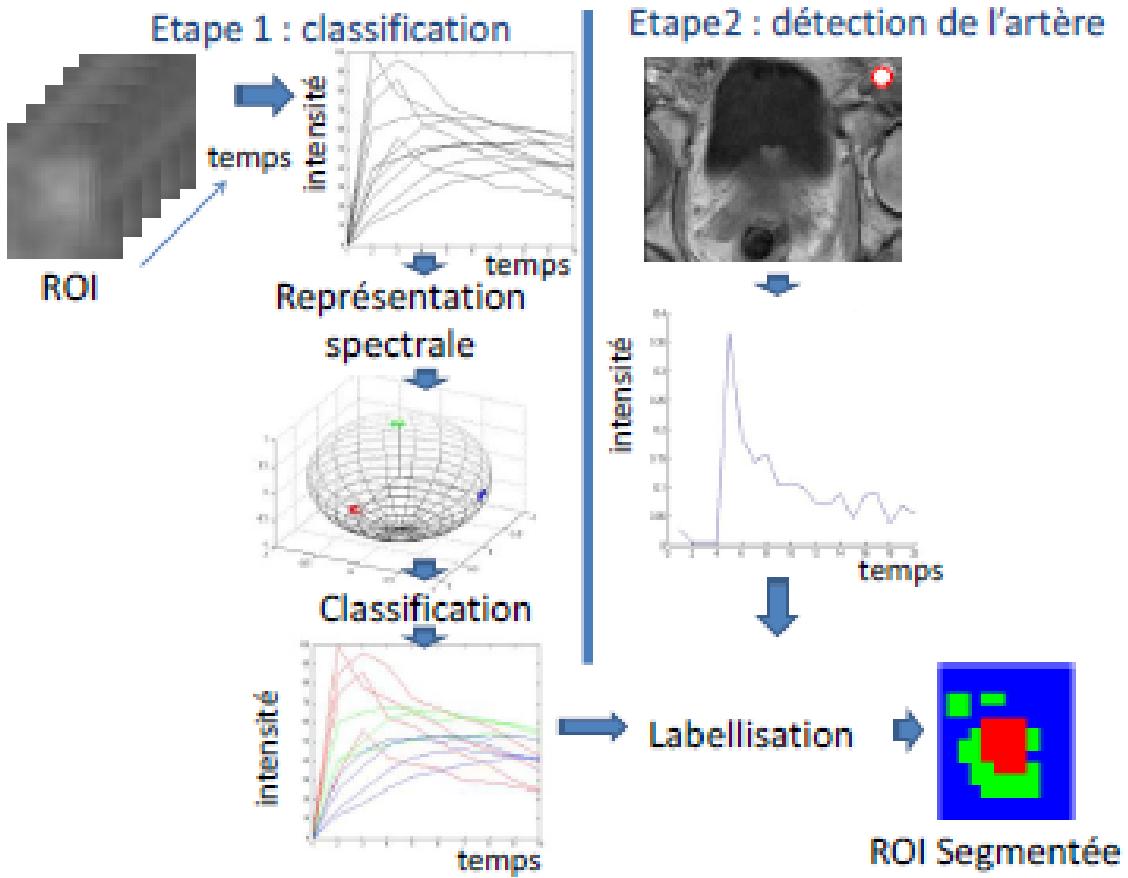


Figure 3.5: Méthodologie suivie pour la classification.

La présente thèse utilise la classification spectrale car elle est particulièrement adaptée aux types d'entrées (signaux temporels) que nous avons ici. La partie suivante va détailler en profondeur cette théorie. Dans le cadre de cette thèse, il réalise une classification des différents signaux en trois classes et il parvienne à obtenir une image segmentée où les zones tumorales sont clairement identifiables.

La figure qui suit 3.6 montre le résultats de la classification sur des données simulées. Comme on peut le voir, le programme est très performant.

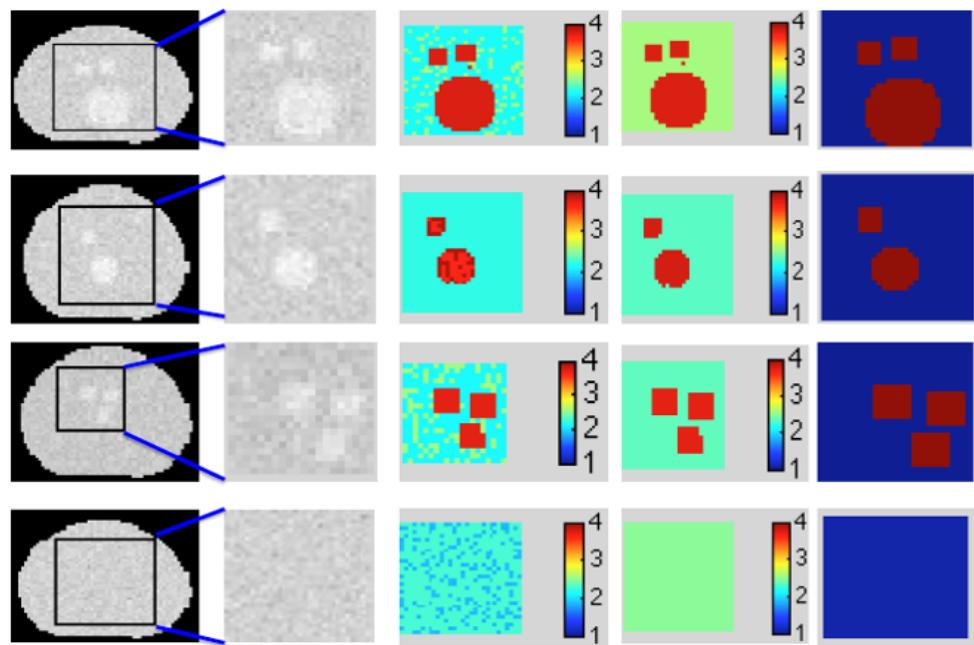


Figure 3.6: Dans l'ordre, - prostate avec présence ou non de tumeur - zoom sur la région d'intérêt - résultat de l'algorithme de classification - fusion de l'information en un résultat zone saine / zone tumorale - table de vérité.

Dans le cadre de la lutte contre le cancer prostatique, le précédent programme permettait d'avoir une segmentation d'image automatique assez performant mais qui montrait des très bons résultats lorsqu'il était couplé à d'autres méthodes d'aide au diagnostique.

# Conclusion

La partie que nous venons de voir présente beaucoup de piste intéressante pour les objectif que nous poursuivons. Néanmoins, cette méthodologie ne pourra pas être réutilisé tel quel pour le thrombose veineuse profonde. Néanmoins, plusieurs étapes comme l'extraction de signaux temporelle à partir d'IRMs de perfusion, la classification spectrale ou l'extraction de paramètres physiologiques peuvent s'avérer particulièrement pertinent tant pour ce projet que pour d'autre enjeux médicaux (cancer du sein...).

## **Part II**

# **Classification spectrale**

# Chapter 4

## Intérêt de la classification spectrale.

Il existe une très large bibliographie sur la classification non supervisée. Néanmoins, comme le montre la figure suivante, certaines données bien que étant facilement classifiable "à l'œil nu", un algorithme du K-means ne parvient pas à aboutir au résultat escompté.

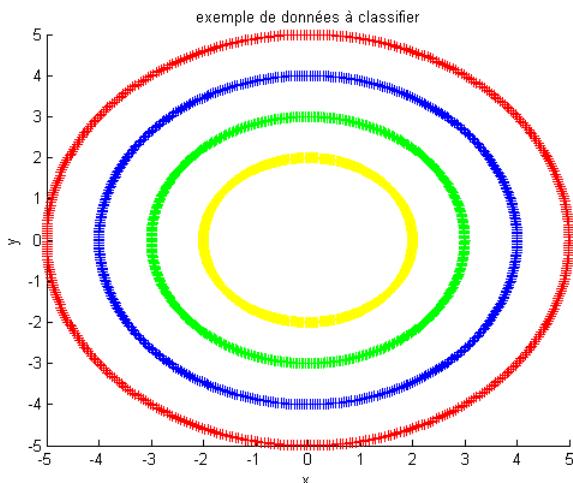


Figure 4.1: Ensemble de données de départ. Les données sont générées sur un cercle dont le rayon varie selon la classe

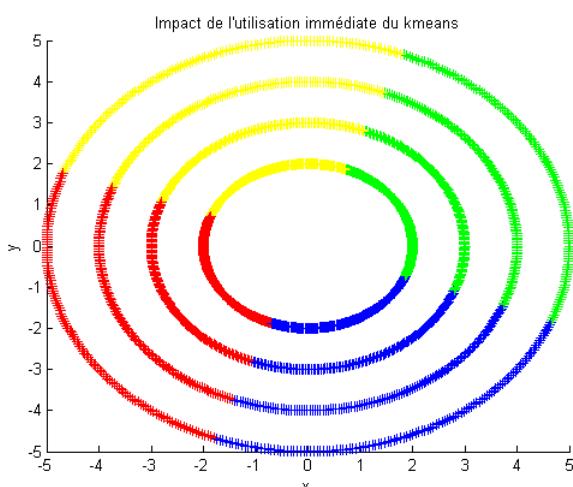


Figure 4.2: Application du Kmeans sur cet ensemble de données. Les données ne sont pas correctement classifiées.

Dans notre cas d'étude, nous devons étudier des signaux temporels dont on ne connaît a priori aucunes informations discriminantes. Il est donc nécessaire de parvenir à trouver des paramètres sur lesquelles la classification puisse s'appuyer. L'idée de la classification spectrale va justement d'extraire les valeurs et vecteurs propres d'une certaine matrice afin de classifier les données. Pour ce faire, la classification spectrale propose de créer un graphe à partir des données et d'extraire les informations pertinentes de la matrice de voisinage.

# Chapter 5

## Rappel sur la théorie des graphes.

### 5.1 Définitions de base.

Un graphe est constitué du couple  $(V, E, W)$ , où  $V$  correspond à ensemble des nœuds du graphe,  $E$  correspond à l'ensemble des couple  $(v_i, v_j)$  symbolisant le lien entre deux nœuds et  $W$  correspond à la matrice de voisinage. Par définition, le coefficient  $W_{ij}$  prend comme valeur le poids liant le nœud  $v_i$  au nœud  $v_j$ . Il s'en suit que la matrice  $W$  est symétrique. Nous verrons plus tard comment nous pouvons construire une tel matrice.

Par la suite, nous définissons la matrice  $D$ , diagonale, qui contient les degrés des nœuds du graphe. Le degré d'un nœud est défini par  $d_i = \sum_{v_j \in V} W_{ij}$ .

### 5.2 Choix du laplacien de la matrice.

Il existe trois définition différentes de laplacien d'une matrice. Si on appelle  $D$  la matrice de degré du graphe et  $W$  la matrice de voisinage du graphe, on aboutie aux définitions suivantes.

**Matrice Laplacienne standard non normalisée** Ce laplacien est donné par la relation:

$$L = D - W \tag{5.1}$$

**Matrice Laplacienne normalisée non symétrique** Ce laplacien est donné par la relation:

$$L = I - D^{-1}W \tag{5.2}$$

**Matrice Laplacienne normalisée symétrique** Ce laplacien est donné par la relation:

$$L = I - D^{-1/2}WD^{-1/2} \tag{5.3}$$

Ces trois matrices présentent les mêmes propriétés:

- Ces matrices présentent la valeur propre 0 une seule fois pour le vecteur propre  $I$  ou  $D^{1/2}I$ .
- Ces matrices sont réelles semi-définies positives.

Par conséquent, quand nous chercherons à trouver les  $k$  plus petites valeurs propres de cette matrice, il nous sera possible très facilement de les récupérer.

# Chapter 6

## Construction de la matrice de voisinage.

Il y a trois choix à faire pour définir la matrice de voisinage d'un graphe.

### 6.1 Choix du type de voisinage.

Pour un ensemble de nœud donné, il y a plusieurs manière de créer les liens qui vont compléter l'ensemble (E,V,W).

**Graphe des epsilon voisin:** On considère ici que deux nœuds ont un lien si la distance entre eux est inférieur à  $\epsilon$ .

**Graphe des k-plus proche voisins:** On considère ici que tous les nœuds ont k voisins et qu'ils sont les k-plus proche du nœud considéré.

**Graphe entièrement connecté:** On considère que tous les sommets sont reliés entre eux.

Ici, nous avons implémenté le graphe entièrement connecté car il nous donne de très bons résultats et qu'il est très simple à mettre en place.

### 6.2 Choix de la distance et de la similitude.

Dans la littérature de la classification, il est nécessaire de choisir une distance et une fonction de similarité.

Plusieurs choix s'offre à nous pour les distances:

- **Distance de Manhattan :**  $\sum_i |x_i - y_i|$
- **Distance d'euclidienne :**  $(\sum_i (x_i - y_i)^2)^{(1/2)}$
- **Distance de Minkowski :**  $(\sum_i |x_i - y_i|^p)^{(1/p)}$

La distance euclidienne a ici été choisie.

La fonction de similitude doit être une fonction qui va associer une valeur comprise entre 0 et 1 à deux sommet du graphe:

$$f(x_i, x_j) \Rightarrow [0; 1] \quad (6.1)$$

La littérature donne plusieurs types de fonction de similitude mais deux sont à retenir.

- La similitude euclidienne qui est définie par la fonction

$$f(x_i, x_j) = \frac{1}{1 + \frac{d^2(x_i, x_j)}{\sigma^2}} \quad (6.2)$$

- La similitude gaussienne qui est définie par la fonction

$$f(x_i, x_j) = \exp\left(-\frac{d^2(x_i, x_j)}{2\sigma^2}\right) \quad (6.3)$$

Dans les deux cas, le coefficient  $\sigma$  est un paramètre qui traduit la dispersion des données. Nous avons ici choisi la similitude gaussienne car elle permet de traduire de manière plus efficace la similitude des données.

### 6.3 Choix du paramètre de dispersion.

Le gros problème pour les données que nous aurons à traiter est que le paramètre  $\sigma$  devra être déterminé de manière automatique en fonction des données d'entrées. Ce paramètre va traduire la similitude ou non des différents éléments entre eux et donc va conditionner tous les résultats de l'algorithme.

Pour ce faire, une solution assez élégante a été proposée [19] et consiste à définir le paramètre  $\sigma$  comme le produit  $\sigma_i\sigma_j$ , où ces deux coefficient dépendent du voisinage de chaque point.

Ainsi, on définit le coefficient par  $\sigma_i = d(x_i, x_r)$  où  $x_r$  correspond au r-ième voisin le plus proche de  $x_i$ . Aussi étrange que cela puisse être, après beaucoup de test, on constate comme dans le document [13] que les meilleurs résultats sont obtenus pour  $r = 7$  ou  $r = 9$ .

## Conclusion

Au final, nous arrivons aux matrices d'affinité suivantes. Les données qui ont été utilisées sont issues d'une IRMs.

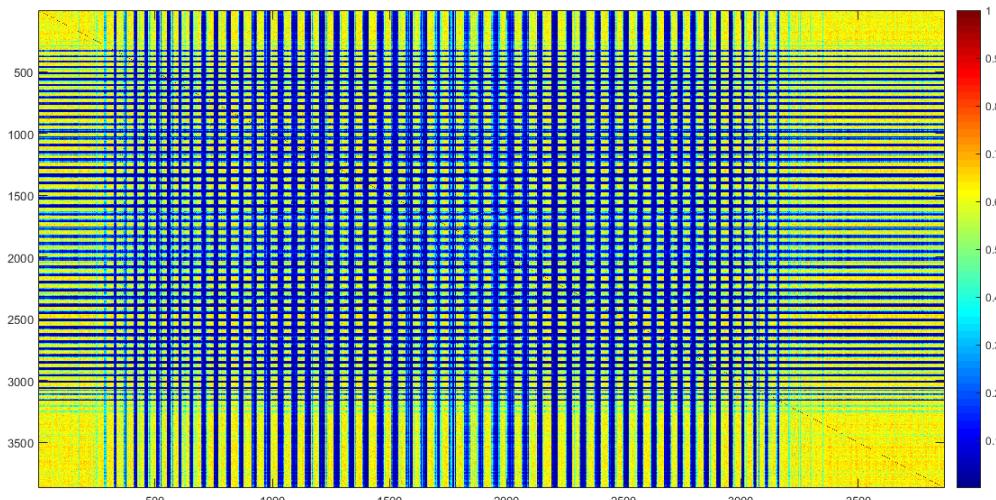


Figure 6.1: Exemple de matrice de similitude.

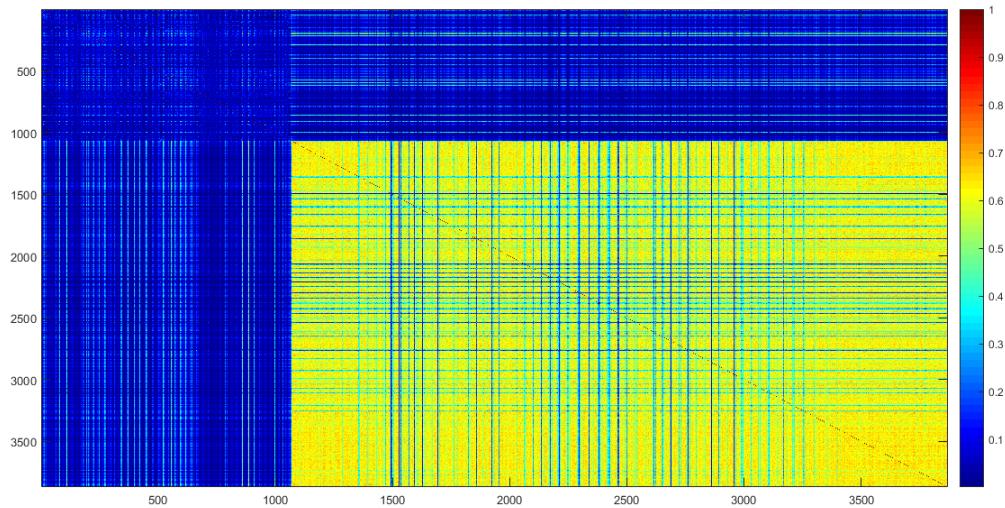


Figure 6.2: Matrice de similitude après utilisation de l'algorithme de classification spectrale.

# Chapter 7

## Quels algorithmes de classification spectrale?

Les algorithmes de classification spectrale sont répartis en deux classes:

- La première classe qui va faire une séparation en deux groupes de manière récursive à partir de la seconde plus petite valeur propre de la matrice laplacienne issue du graphe jusqu'à avoir k clusters.
- La seconde consiste à faire une classification sur les k plus petits vecteurs propres de la matrice laplacienne issue du graphe grâce à un algorithme plus simple comme le K-means.

Dans le cadre de ce projet, nous avons principalement étudié la seconde classe d'algorithme de classification et particulièrement celui développé par Jordan and Weiss. Au total, nous avons implémenté 4 algorithmes différents de classification spectrale. Les détails théoriques de ces algorithmes peuvent être trouver dans ces références [15] et [12].

Les trois algorithmes correspondent à la seconde classe d'algorithme:

---

### Algorithm 1 Unnormalized spectral clustering

---

#### Inputs

- 1: Matrice  $W$  matrice de voisinage
- 2: Entier  $k$  nombre de classe

#### Output

- 1: Vecteur  $Ver$  table de vérité calculée.

#### Algorithm

- 1:  $L \leftarrow$  Laplacien standard de la matrice  $W$
  - 2:  $VectP \leftarrow$  matrice contenant les vecteurs propres des k plus petites valeurs propres non nulles de  $L$
  - 3:  $Ver \leftarrow$  résultat de l'algorithme du Kmeans sur la matrice VectP en k clusters.
  - 4: **return**  $Ver$
-

---

**Algorithm 2** Normalized spectral clustering, Shi and Malik

---

**Inputs**

- 1: Matrice  $W$  matrice de voisinage
- 2: Entier  $k$  nombre de classe

**Output**

- 1: Vecteur  $Ver$  table de vérité calculée.

**Algorithm**

- 1:  $L \leftarrow$  Laplacien asymétrique de la matrice  $W$
  - 2:  $VectP \leftarrow$  matrice contenant les vecteurs propres des  $k$  plus petites valeurs propres non nulles de  $L$
  - 3:  $Ver \leftarrow$  résultatat de l'algorithme du Kmeans sur la matrice  $VectP$  en  $k$  clusters.
  - 4: **return**  $Ver$
- 

---

**Algorithm 3** Normalized spectral clustering, Jordan and Weiss

---

**Inputs**

- 1: Matrice  $W$  matrice de voisinage
- 2: Entier  $k$  nombre de classe

**Output**

- 1: Vecteur  $Ver$  table de vérité calculée.

**Algorithm**

- 1:  $L \leftarrow$  Laplacien symétrique de la matrice  $W$
  - 2:  $VectP \leftarrow$  matrice contenant les vecteurs propres des  $k$  plus petites valeurs propres non nulles de  $L$
  - 3: normaliser toutes les lignes de la matrice  $VectP$
  - 4:  $Ver \leftarrow$  résultatat de l'algorithme du Kmeans sur la matrice  $VectP$  en  $k$  clusters.
  - 5: **return**  $Ver$
- 

Le dernier algorithme appartient à la première classe d'algorithme.

---

**Algorithm 4** Normalized spectral recursive clustering, Shi and Malik

---

**Inputs**

- 1: Matrice  $W$  matrice de voisinage
- 2: Entier  $k$  nombre de classe

**Output**

- 1: Vecteur  $Ver$  table de vérité calculée.

**Algorithm**

- 1:  $L \leftarrow$  Laplacien asymétrique de la matrice  $W$
  - 2:  $VectP \leftarrow$  seconde plus petite valeur propre de  $L$
  - 3: Partitionner le graphe en deux sous-ensemble selon les valeurs de  $VectP$
  - 4: Répéter l'algorithme jusqu'à avoir  $k$  clusters.
  - 5:  $Ver \leftarrow$  résultat de l'algorithme.
  - 6: **return**  $Ver$
-

# Chapter 8

## Résultats de l'algorithme et interprétations.

Nous avons implémenté les différents algorithmes que nous avons précédemment présenté et voici les différents résultats sur l'ensemble de données que nous avons utilisé dans le chapitre 4.

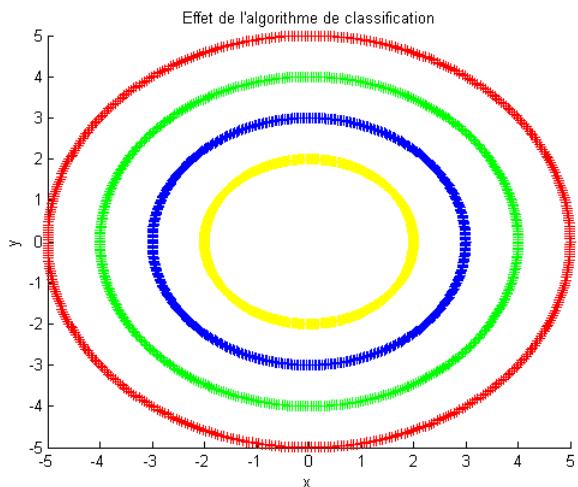


Figure 8.1: Application du premier algorithme sur les premières données. Les données sont correctement classifiées.

On peut voir que le résultat obtenu est bien celui espéré et que c'est une solution parfaitement viable pour agréger des données dans des clusters bien que les liens entre elles ne soient pas identifiable immédiatement. Les valeurs propres extraites du laplacien et qui sont utilisées pour la classification sont données ci-dessous.

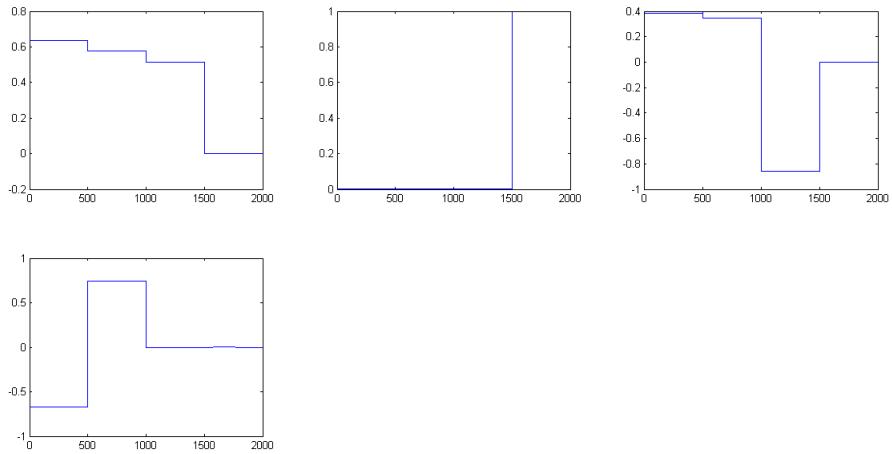


Figure 8.2: Vecteurs propres extraits du laplacien.

On peut clairement voir comment les K-means sur ces données extraites permettent des discriminer les quatre classes. Nous l'avons testé sur un ensemble de test et les figures suivantes montrent les résultats que nous avons obtenu.

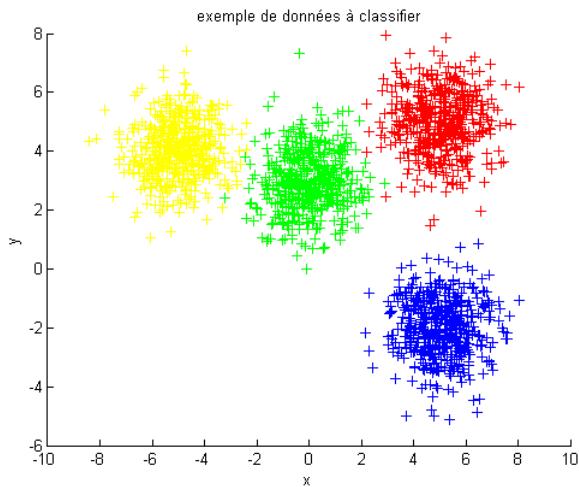


Figure 8.3: Second ensemble de données.

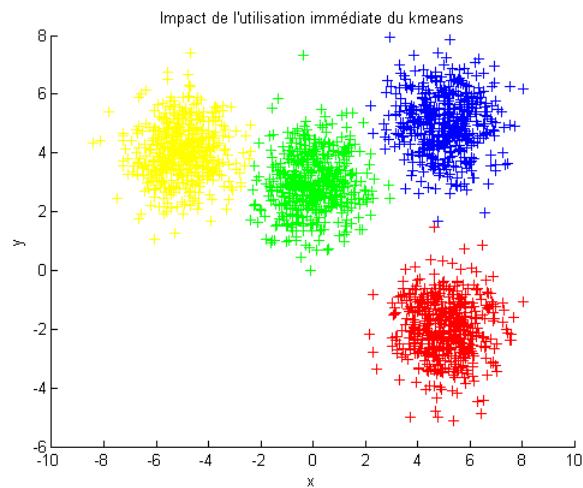


Figure 8.4: Application du Kmeans sur le second ensemble de données.

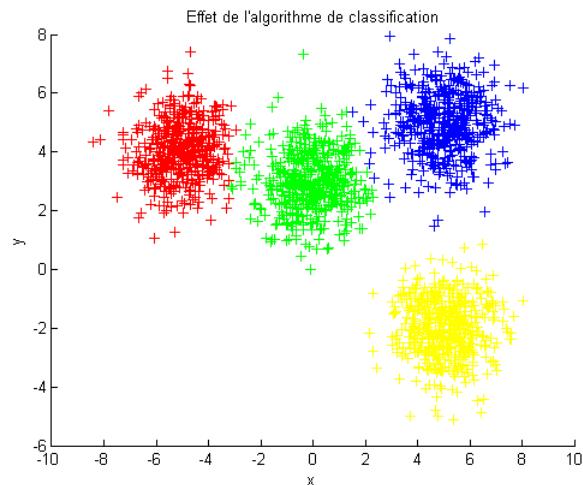


Figure 8.5: Application du premier algorithme sur le second ensemble de données.

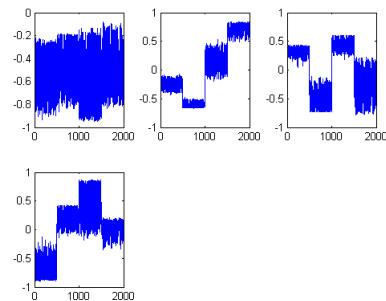


Figure 8.6: Vecteurs propres extraits du laplacien.

Là encore, le résultat est très satisfaisant. Néanmoins, on peut clairement voir que les distributions aléatoires ont un impact sur la régularité des vecteurs propres.

# Conclusion

Au terme de cette étude, nous pouvons extraire ces différents résultats:

- Tout d'abord, la classification spectrale est particulièrement adaptée pour les problèmes que les algorithmes traditionnels comme le K-means ne peuvent résoudre. On arrive à se placer facilement dans un nouvel espace qui permet d'identifier clairement les éléments appartenant à une même classe.
- La mise en place des différents algorithme est au final relativement simple et peut couteuse en temps de calcul.
- Elle est particulièrement adaptée pour le problème que nous étudions.

Néanmoins, il y a plusieurs contraintes qui faut prendre en compte:

- Le résultat final est fortement impacté par l'utilisation du K-means à la fin de l'algorithme de classification spectrale. Il arrive que l'algorithme ne converge pas vers les solutions souhaitées en fonction de l'initialisation.
- Il y a un certain impact du bruit sur les vecteurs propres qui présentent des irrégularités comme le montre la figure [8.6](#).

## **Part III**

# **Application de la classification spectrale sur des données réelles**

# Chapter 9

## Réalisation de la chaîne de traitement.

Je tiens ici à remercier le HIA Clermont Tonnerre de Brest et le CHRU de Brest pour m'avoir fourni les différentes IRMs de perfusion dont je vais vous montrer les résultats.

Quelque soit l'IRM considérée, la chaîne de traitement restera la même. Il faudra choisir l'IRM d'un patient, stocker les images IRMs par coupe, choisir une coupe puis une région d'intérêt et enfin appliquer l'algorithme sur les éléments sélectionnés.

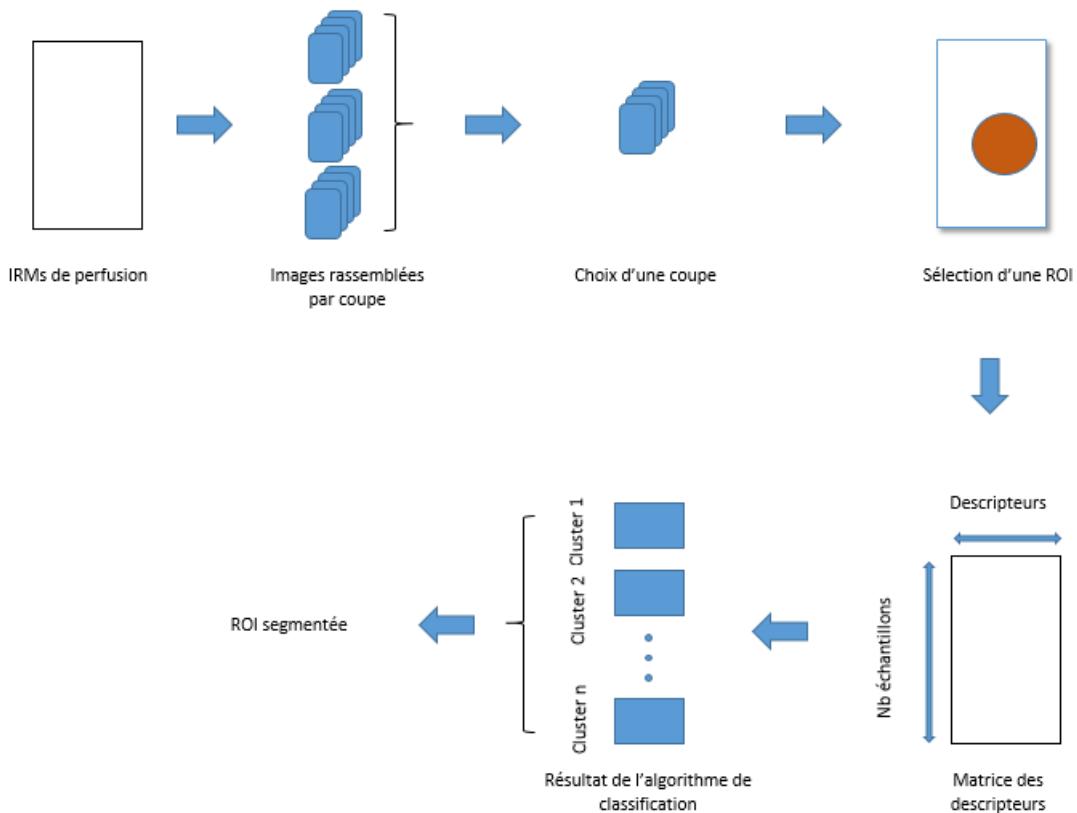


Figure 9.1: Chaîne de traitement développée dans le cadre du projet.

La sélection de la région d'intérêt se fait assez facilement avec Matlab comme le montre la figure suivante. L'idée est de prendre une zone et une coupe qui contiennent la zone présentant une pathologie.

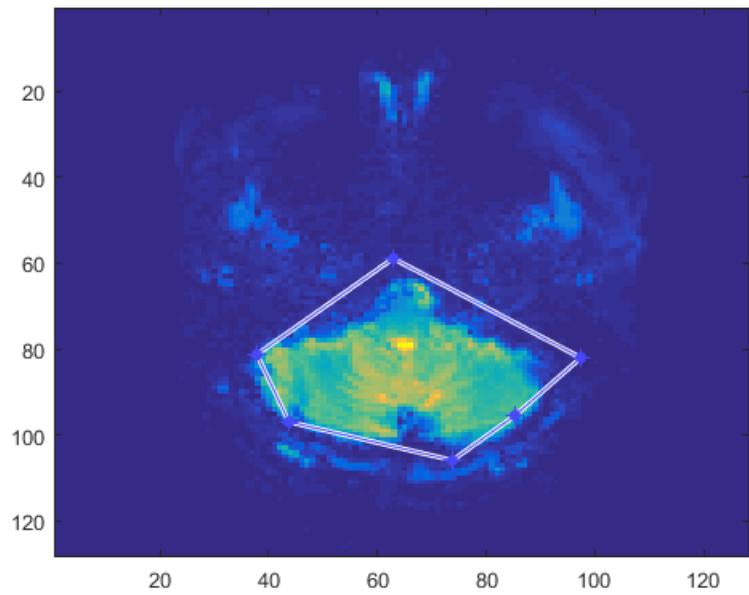


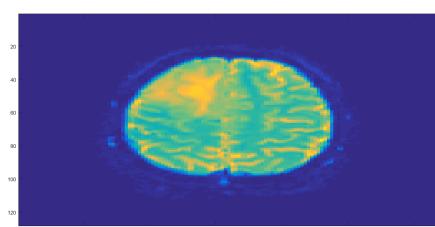
Figure 9.2: Choix d'une ROI avec Matlab.

# Chapter 10

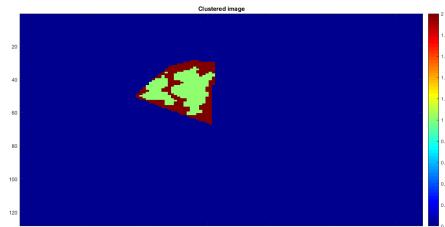
## IMRs T2 du crane CHRU Brest Cavale Blanche.

Pendant ce PFE, j'ai pu récupérer une dizaine d'IRM de perfusion de cerveau présentant des lésions. La liste ci-dessous donne une liste des lésions considérées:

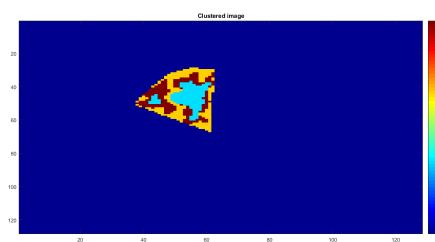
- Patient 2: Cancer visible sur les coupes 13 à 22.
- Patient 6: \* visible sur les coupes 5 à 9.
- Patient 9: Lésion post-opératoire visible sur les coupes 15 à 19.



(a) Image original.



(b) 2 classes.



(c) 3 classes.

Figure 10.1: Résultats de notre algorithme sur le patient 2 avec 2 et 3 classes.

# Chapter 11

## IMRs T1 de la prostate de l'INSEERM de Lille.

### 11.1 Types de courbe rencontrées dans le cadre du cancer de la prostate

Une partie du travail qui est présenter ici peut être retrouver dans le lien suivant [1]. Ce document présente les différentes courbes caractéristiques de la prostate. Les courbes d'intensité présentent toutes un hyper signal en IRM T1 mais les courbes issus de tissus cancéreux présentent un pic d'absorption beaucoup plus marqué avec un temps d'assimilation très caractéristique alors que les tissus sain présentent une courbe d'absorption beaucoup plus lissée. La figure suivant issu de [1] donne une illustration de ces courbes caractéristique.

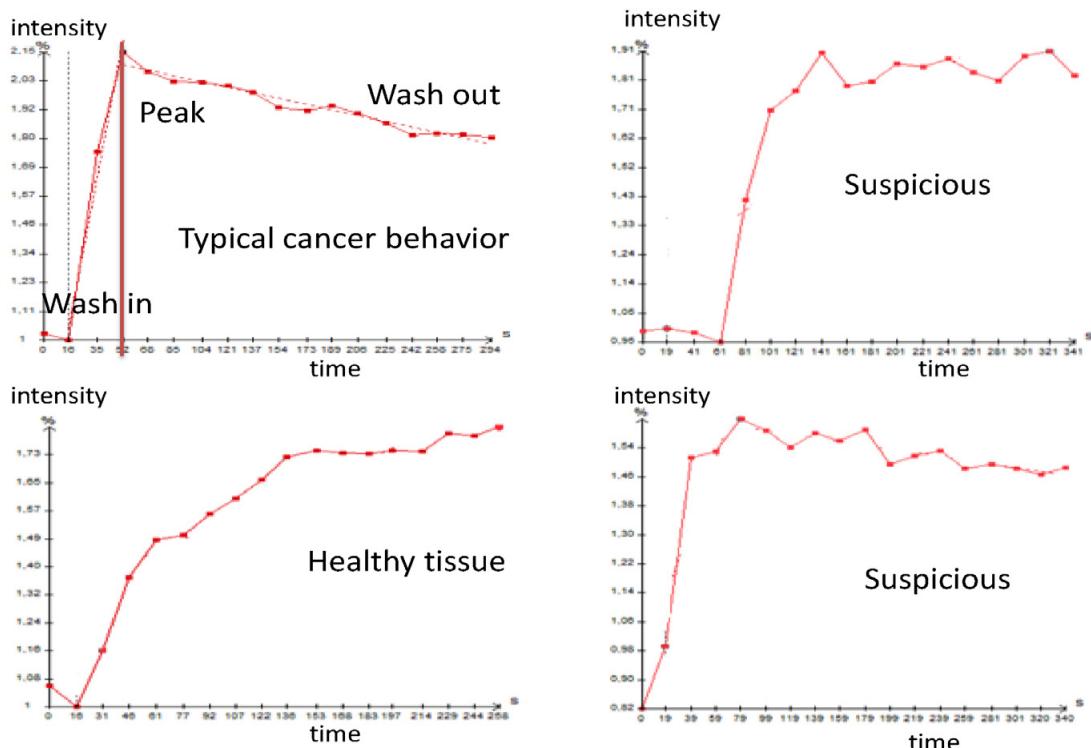


Figure 11.1: Courbes caractéristiques des tissus sain et cancéreux de la prostate.

Néanmoins, certaines courbes sont difficilement à classifier. L'intérêt de notre algorithme va donc être de pouvoir distinguer aisément les tissus sains et les tissus cancéreux et d'estimer avec une autre

classe supplémentaire les cas d'incertitude.

**NOTA BENE:** Dans le document précédent, il construit le graphe de similitude à partir de la similitude entre la courbe d'intensité du voxel considéré et de l'AIF. Chose qui n'est pas fait ici.

## 11.2 Résultats de l'algorithme

Pour appliquer nos algorithmes, il y a un petit pré-traitement à faire. Pour discriminer les différentes courbes, il faut surtout voir le phénomène de diffusion au sein du tissu et normaliser les courbes afin de faire ressortir le comportement du tissu.

Tous les courbes sont donc remise à l'échelle en appliquant la formule suivante:

$$x(t) \leftarrow \frac{x(t) - x_{min}}{x_{max} - x_{min}} \quad (11.1)$$

Tous les signaux temporels sont donc placés dans une matrice comme expliqué dans la figure 9.1. Au final, en appliquant nos algorithmes à une IRM de perfusion de la prostate, nous arrivons au résultat suivant :

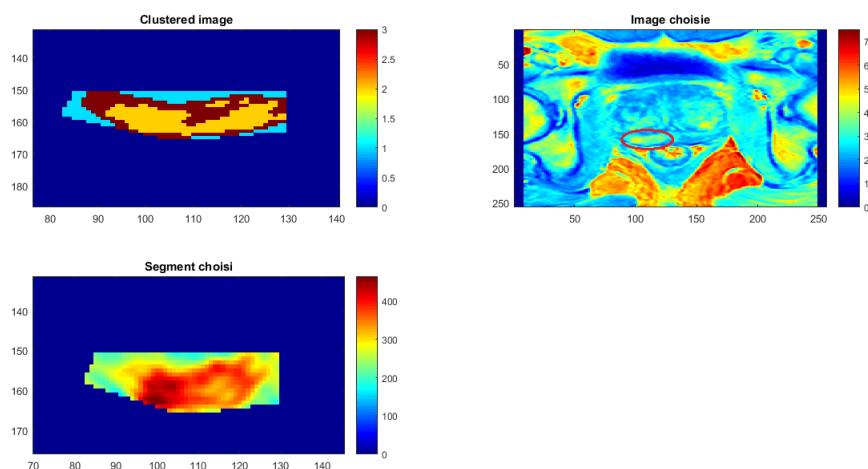


Figure 11.2: Coupes sélectionnées. En haut à gauche le résultat de l'algorithme de classification. En haut à droit, image de la coupe sélectionnée avec la ROI représentée par l'ellipse rouge. En bas à gauche, ROI en fausses couleurs.

La figure suivante montre comment les courbes sont réparties selon les différents clusters.

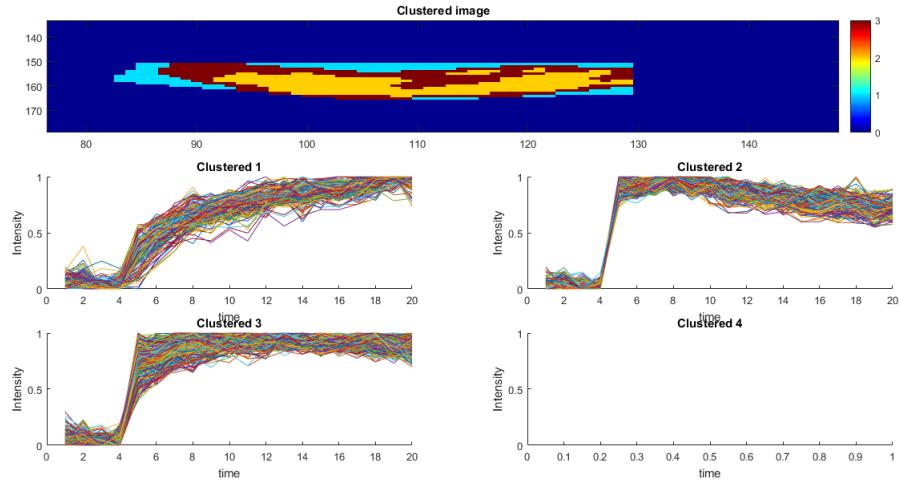


Figure 11.3: Résultat de l'algorithme sur l'IRM avec les courbes associées aux clusters.

Nous avons ici lancé une simulation avec 3 classes. Nous pouvons interpréter ces résultats de la façon suivante:

- Le cluster 1 rassemble les courbes qui sont saines en accord avec les informations issus de la figure 11.1
- Le cluster 2 lui rassemble les courbes des tissus tumoraux qui présentent un temps d'absorption très court et un pic très marqué.
- Le cluster 3 lui représentent des courbes qui sont à mi chemin entre les deux autres courbes caractéristiques. Ici, l'algorithme comme l'humain va avoir du mal à trancher sur l'appartenance de ces tissus au cluster sain ou tumoral.

Ici, on arrive à un résultat en partie semblable au médecin qui a à analyser ces images. On arrive à identifier de façon claire la zone tumorale et la zone litigieuse qui est assez souvent traitement de la même manière.

Néanmoins, il y a quelques remarques à faire:

- Il faut parfois choisir une bonne représentation des données de départ afin que l'algorithme converge vers une solution satisfaisante.
- Il est nécessaire de choisir une ROI qui est composé d'une zone homogène et de préférence de taille pas trop importante. En effet, l'algorithme ne converge pas si il y a trop de point sélectionné.
- Il y a un grand enjeu sur la détermination automatique du nombre de classe mais ceci sera plus longtemps abordé dans la partie 4.

## **Part IV**

### **La problématique du nombre de classe**

# Chapter 12

## Choix automatique du nombre de classe.

### 12.1 Algorithme 1 : Analyse des valeurs propres et le n-gap

Dans le cas d'une classification idéale, la matrice laplacienne  $L$  peut être mise sous la forme d'une matrice par bloc où tout les blocs non-diagonaux sont nuls.

$$L = \begin{bmatrix} x^{(1)} & 0 & \cdots & 0 \\ 0 & x^{(2)} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & x^{(n)} \end{bmatrix}$$

Cela traduit le fait que les blocs ou clusters sont homogènes et ne présentent pas de similitude entre eux. Cependant, vu comment la matrice  $L$  est construite, la valeur propre 0 va avoir une multiplicité égale au nombre de bloc.

Une première méthode consiste donc à afficher les  $n$  premières valeurs propres de  $L$  et de "compter" le nombre de fois que la valeur propre 0 apparaît. Dans le cas des données simulées ci-dessous:

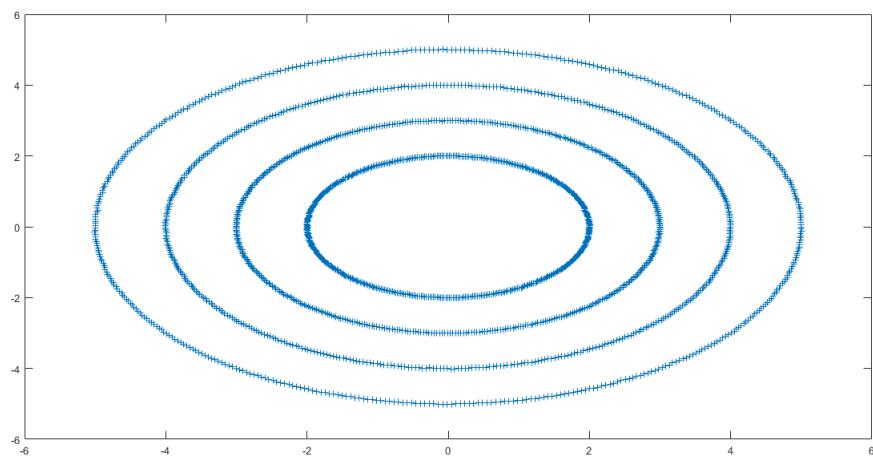


Figure 12.1: Exemple simple de données à classifier. Le nombre de classe est 4.

Nous arrivons aux valeurs propres suivantes:

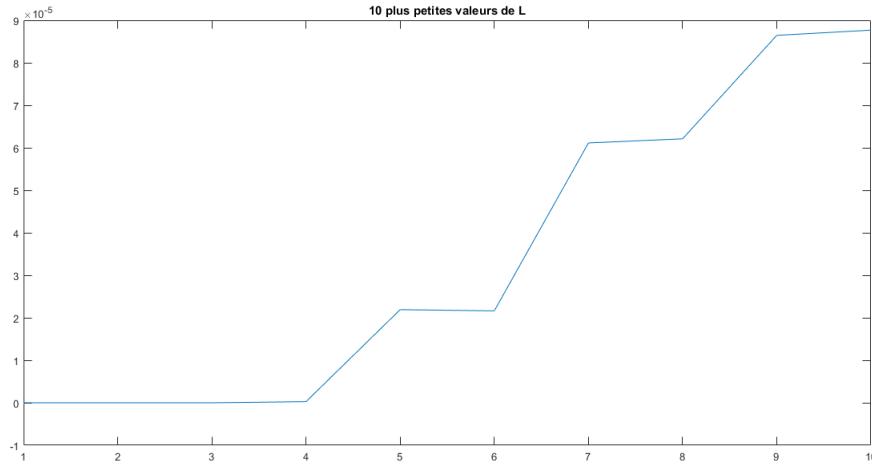


Figure 12.2: 10 plus petites valeurs propres de L. On observe bien que on a un saut à partir de n=5. Le nombre optimal de classe est 4.

Une autre méthode consiste à calculer la différence entre deux valeurs propres successives et voir à partir de quand un gap significatif se produit afin d'identifier le nombre de classe optimale. Néanmoins, ces deux méthodes sont hautement discutables de par leur manque de rigueur.

## 12.2 Algorithme 2 : Minimisation par la norme de Frobenius

Une autre méthode consiste à voir pour quel nombre de classe la similitude entre classe est la plus faible. On définit la norme de Frobenius d'une matrice par la formule suivante:

$$\|L\|_F = \sum_{i=1}^{N_i} \sum_{j=1}^{N_j} |L_{ij}|$$

L'idée va donc être de calculer le volume des blocs diagonaux et de calculer une fonction coût qui dépend du nombre de classe.

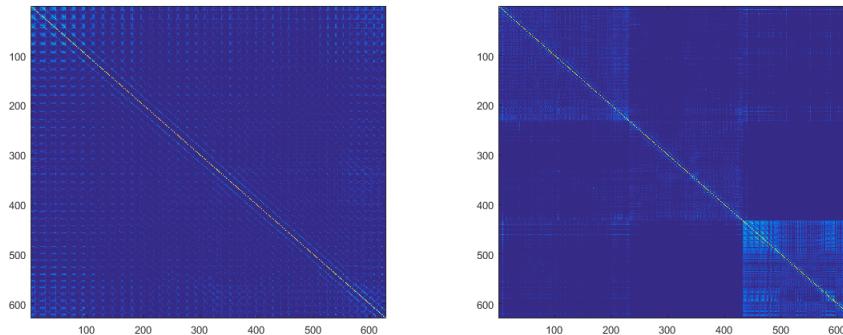


Figure 12.3: Exemple de matrice laplacienne avant et après utilisation de l'algorithme de classification spectrale en trois classes et ré-agencement. On observe bien l'apparition de blocs.

On définit donc le ratio de Frobenius entre deux blocs de la matrice L par:

$$r_{ij} = \frac{\|L^{ij}\|_F}{\|L^{ii}\|_F}$$

On va donc regarder les valeurs de  $r_{ij}$  et voir pour quel valeur de  $n$  on atteint un minimum. Le but de cet algorithme va consister à trouver la valeur de  $n$  qui va minimiser la fonction :

$$\nu_k = \frac{2}{k(k-1)} \sum_{i=1, j=i+1}^k (r_{ij})$$

Il faut donc regarder pour plusieurs valeurs de n quel est le nombre de classe qui minimise cette fonction.

### 12.3 Algorithme 2 : Minimisation par rotation

Le dernier algorithme consiste à utiliser la même approche que précédemment mais

# **Chapter 13**

## **Application de ces algorithmes sur des données réelles**

# **Part V**

## **Annexes**

# Anonymisation des images DICOM

Pour anonymiser les données, j'utilise le programme "Santé Dicom Editor". Il suffit d'ouvrir le dossier contenant les images et aller dans l'onglet Edit → Anonymise current file/series. Il existe des fonctions Matlab pour faire le même travail mais certains champs sensibles étaient malgré tous intacts. Voir [http://www.santesoft.com/win/sante\\_dicom\\_editor/sante\\_dicom\\_editor.html](http://www.santesoft.com/win/sante_dicom_editor/sante_dicom_editor.html).



Figure 1: Programme pour l'anonymisation des données.

# Visualisation des données en 3D et 4D

Un gros problème lorsque nous recevons les images Dicom est que nous ne pouvons voir de façon très claire la localisation de la tumeur. J'ai donc réutilisé et modifié des codes Matlab disponibles sur internet afin d'avoir une représentation 3D et 4D. Ils sont disponibles sur les liens suivant:

<http://fr.mathworks.com/matlabcentral/fileexchange/41465-4d-volume-visualization>  
<http://www.mathworks.com/matlabcentral/fileexchange/37268-3d-volume-visualization>

Et sur le github du PFE :

<https://github.com/chuzelph-ENSTA-Bretagne/ThromboseIRM.git>

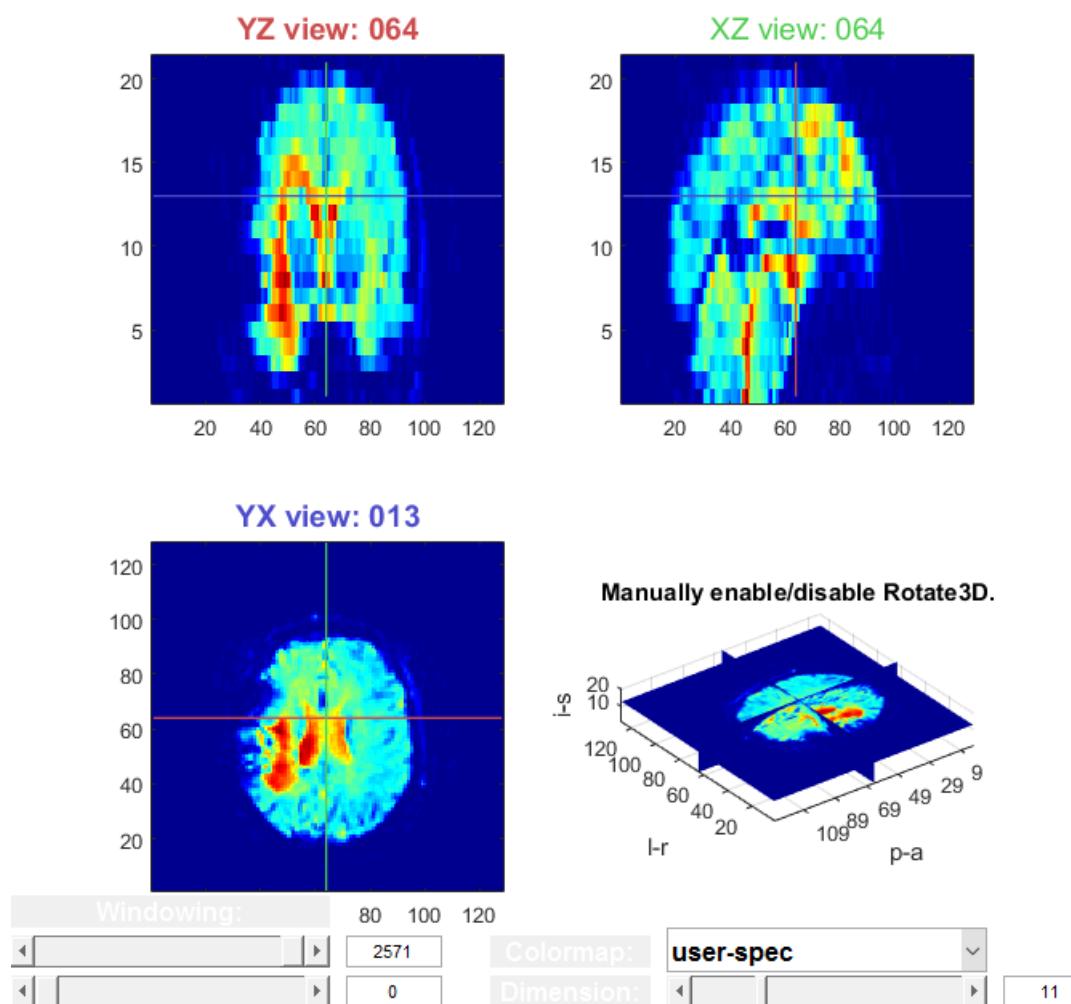


Figure 2: Programme Matlab pour la visualisation de données en 4D.

# Bibliography

- [1] Kunio Doi. Computer-aided diagnosis in medical imaging: historical review, current status and future potential. *Computerized medical imaging and graphics*, 31(4):198–211, 2007.
- [2] Ahmed Idbaih, Monique Boukobza, Isabelle Crassard, Raphaël Porcher, Marie-Germaine Bousser, and Hugues Chabriat. Mri of clot in cerebral venous thrombosis high diagnostic value of susceptibility-weighted images. *Stroke*, 37(4):991–995, 2006.
- [3] Jun Jiao, Xuan Mo, and Chen Shen. Image clustering via sparse representation. In *Advances in Multimedia Modeling*, pages 761–766. Springer, 2010.
- [4] David S Liebeskind, Nerves Sanossian, William H Yong, Sidney Starkman, Michael P Tsang, Antonio L Moya, David D Zheng, Anna M Abolian, Doojin Kim, Latisha K Ali, et al. Ct and mri early vessel signs reflect clot composition in acute stroke. *Stroke*, 42(5):1237–1243, 2011.
- [5] Dario Livio Longo, Walter Dastrù, Lorena Consolino, Miklos Espak, Maddalena Arigoni, Federica Cavallo, and Silvio Aime. Cluster analysis of quantitative parametric maps from dce-mri: application in evaluating heterogeneity of tumor response to antiangiogenic treatment. *Magnetic resonance imaging*, 33(6):725–736, 2015.
- [6] Marina Meila. Comparing clusterings. 2002.
- [7] Sandrine Mouysset. *Contributions à l'étude de la classification spectrale et applications*. PhD thesis, Institut National Polytechnique de Toulouse-INPT, 2010.
- [8] Andrew Y Ng, Michael I Jordan, Yair Weiss, et al. On spectral clustering: Analysis and an algorithm. *Advances in neural information processing systems*, 2:849–856, 2002.
- [9] Jonathan Ophir, Ignacio Cespedes, Hm Ponnekanti, Y Yazdi, and Xin Li. Elastography: a quantitative method for imaging the elasticity of biological tissues. *Ultrasonic imaging*, 13(2):111–134, 1991.
- [10] Cedric Schmitt, Anis Hadj Henni, and Guy Cloutier. Characterization of blood clot viscoelasticity by dynamic ultrasound elastography and modeling of the rheological behavior. *Journal of biomechanics*, 44(4):622–629, 2011.
- [11] Cédric Schmitt, Emmanuel Montagnon, Anis Hadj Henni, Shijie Qi, and Guy Cloutier. Shear wave induced resonance elastography of venous thrombi: A proof-of-concept. *Medical Imaging, IEEE Transactions on*, 32(3):565–577, 2013.
- [12] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(8):888–905, 2000.
- [13] Guillaume TARTARE. *Contribution à l'analyse de l'IRM dynamique pour l'aide au diagnostic du cancer de la prostate*. PhD thesis, ENSTA Bretagne, 2014.

- [14] Sebastian Temme, Christoph Grapentin, Christine Quast, Christoph Jacoby, Maria Grandoch, Zhaoping Ding, Christoph Owenier, Friederike Mayenfels, Jens W Fischer, Rolf Schubert, et al. Noninvasive imaging of early venous thrombosis by 19f magnetic resonance imaging with targeted perfluorocarbon nanoemulsions. *Circulation*, 131(16):1405–1414, 2015.
- [15] Ulrike Von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 17(4):395–416, 2007.
- [16] Sen Wu, Xiaodong Feng, and Wenjun Zhou. Spectral clustering of high-dimensional data exploiting sparse representation vectors. *Neurocomputing*, 135:229–239, 2014.
- [17] Shuicheng Yan and Huan Wang. Semi-supervised learning by sparse representation. In *SDM*, pages 792–801. SIAM, 2009.
- [18] Allen Y Yang, Zihan Zhou, Arvind Ganesh Balasubramanian, S Shankar Sastry, and Yi Ma. Fast-minimization algorithms for robust face recognition. *Image Processing, IEEE Transactions on*, 22(8):3234–3246, 2013.
- [19] Lihi Zelnik-Manor and Pietro Perona. Self-tuning spectral clustering. In *Advances in neural information processing systems*, pages 1601–1608, 2004.