Intégration hiérarchique de données d'imagerie cardiaque par apprentissage automatique

Benoit Freiche

Université Claude Bernard Lyon 1

THÈSE de DOCTORAT DE L'UNIVERSITÉ CLAUDE BERNARD LYON 1

Ecole Doctorale nº 160
Électronique, Électrotechnique, et Automatique

Discipline : Traitement du Signal et de l'Image

Soutenue publiquement le 07/12/2023, par :
Benoît Freiche

Intégration hiérarchique de données d'imagerie cardiaque par apprentissage automatique

---

Devant le jury composé de :

LORENZI Marco CR, INRIA	Rapporteur
RUAN Su PU, Université de Rouen	Rapporteure
ALLASSONNIERE Stéphanie PU, Université Paris Cité	Examinatrice
BASARAB Adrian PU, Université Lyon 1	Président
MATEUS Diana PU, Ecole Centrale Nantes	Examinatrice
CLARYSSE Patrick DR CNRS, INSA Lyon	Directeur de thèse
DUCHATEAU Nicolas MCU, Université Lyon 1	Examinateur (co-encadrant)
CROISILLE Pierre PU-PH, Centre Hospitalier Universitaire de Saint-Etienne	Invité

Table des matières

Résumé

L'imagerie médicale fournit des informations fines jusqu'à l'échelle du pixel, sous la forme de données images ou de multiples descripteurs de haute dimension extraits de ces images, que les médecins exploitent pour le diagnostic et le suivi quantitatif des patients. Dans le cadre de cette thèse, nous souhaitons améliorer la caractérisation des lésions myocardiques dans une population de patients ayant subi un infarctus aigü et suivis par imagerie par résonance magnétique (IRM). L'objectif de nos travaux est d'exploiter plusieurs descripteurs de haute dimension provenant des images IRM pour mieux caractériser la pathologie et son évolution. À l'heure actuelle, les méthodes d'analyse computationnelle mélangent tous ces types de descripteurs simultanément et sans distinction, stratégie qui montre rapidement ses limites. En effet, les descripteurs peuvent être partiellement corrélés entre eux et/ou avec la pathologie, et représenter une quantité conséquente d'information mais difficile à correctement analyser (données de haute dimension et hétérogènes).

Cette stratégie contraste avec la pratique clinique où les médecins, en se basant sur leur expérience, sont capables d'ordonner ces différentes informations et de les traiter de façon progressive selon une approche hiérarchique. Dans cette thèse, nous nous inspirons de cette approche pour proposer une analyse par niveaux hiérarchiques qui incorpore progressivement les différentes informations extraites des images. Nous démontrons sa pertinence pour caractériser l'infarctus aigü du myocarde dans une population de patients à l'aide d'apprentissage de représentation non supervisé.

Nos contributions méthodologiques principales concernent la prise en compte et l'exploitation du lien hiérarchique dans les données, dans le cadre de l'apprentissage de représentations. Une première contribution évalue le potentiel d'une hierarchie à deux niveaux dans le cadre de l'apprentissage de variétés. Une deuxième contribution propose une formulation générique de la hiérarchie en se basant sur une méthode probabiliste, les processus Gaussiens à variable latente. Du point de vue applicatif, nous démontrons la pertinence de ces approches pour l'exploitation directe du contenu d'images IRM à rehaussement précoce et tardif, sous-exploitées par les médecins, et pour la caractérisation de lésions du myocarde complexes de par leur forme et leur taille réduite (infarctus et lésions de reperfusion).

Mots clés : Apprentissage de représentation, fusion d'informations, réduction de dimension, imagerie cardiaque, infarctus aigü du myocarde.

Publications

Articles de revues internationales

Hierarchical data integration with Gaussian processes : application to the characterization of cardiac ischemia-reperfusion patterns.
Freiche B, Bernardino G, Clarysse P, Duchateau N.
Under review at IEEE Transactions on Medical Imaging, 2023.

Articles de conférences internationales

Characterizing myocardial ischemia and reperfusion patterns with hierarchical manifold learning.
Freiche B, Clarysse P, Viallon M, Croisille P, Duchateau N.
Proc. Statistical Atlases and Computational Models of the Heart (STACOM), MICCAI'21 Workshop, LNCS 2022 ;13131 :66-74.

Résumés de conférences internationales

Hierarchical manifold learning for the interpretation of multi-level data - Application to cardiac imaging.
Freiche B, Clarysse P, Viallon M, Croisille P, Duchateau N.
Medical Image Analysis and Artificial Intelligence (MAI), Sino-French workshop 2021.

Chapitre 1

Introduction

Ce premier chapitre introduit la notion de compréhension d'une pathologie à l'aide de l'imagerie médicale. Cette introduction traite notamment des méthodes d'imagerie de routine pour l'étude de la fonction cardiaque, et détaille le cas particulier de l'infarctus aigü du myocarde, qui constitue l'application centrale de cette thèse. Elle pose aussi des bases pour l'exploitation des données issues de l'imagerie médicale grâce à l'apprentissage statistique, en exposant les principales problématiques et limites actuelles, en particulier dans le cadre de multiples modalités de données. Ce sont certaines de ces limites que cette thèse cherche à adresser.

1.1 Contexte clinique

1.1.1 Rôle de l'imagerie dans la caractérisation des pathologies cardiaques

Les causes des pathologies sont nombreuses et diverses, et elles peuvent être bénignes comme mortelles. Pour les pathologies complexes touchant des organes dont les fonctions sont vitales pour le bon fonctionnement du corps (le cerveau, le cœur, les poumons...), il est impératif de trouver des traitements adaptés et efficaces, et ce dès les premiers symptômes. Le choix d'un traitement adéquat requiert une compréhension profonde de la pathologie en question.

Il est d'abord important de caractériser le mieux et le plus tôt possible les différents stades d'une maladie. On parle alors de diagnostic précoce. Par exemple, d'après *Ryan et al.* en 1996 [1], 3.5% des patients traités par thrombolyse (traitement médicamenteux) dans la première heure suivant l'apparition des symptômes d'un infarctus du myocarde étaient sauvés, contre 1.6% si le traitement était donné dans les 7 à 12 heures suivant les symptômes. Par ailleurs, il est essentiel d'appréhender l'évolution de ces maladies au cours du temps, et notamment d'anticiper et de stratifier les risques pour les patients (en d'autres termes existe-t-il des sous-groupes identifiables avec des risques de mortalité,

hospitalisation, etc. différents ?) [2]. Les conséquences directes d'une meilleure caractérisation des pathologies se trouvent au niveau de l'amélioration de la prise en charge des patients, et sur la réduction des coûts humains et/ou matériels associés à cette pathologie [3].

Les données médicales sur lesquelles on peut s'appuyer pour caractériser une pathologie sont d'abord des —nombreux— indicateurs scalaires (tension, rythme cardiaque, analyses biologiques. etc.) mesurés lors d'une visite et éventuellement suivis dans le temps. Ces indicateurs sont relativement faciles à intégrer dans une analyse computationnelle, notamment du point de vue des arbres de décision cliniques. Cependant, les pathologies peuvent mettre en jeu des mécanismes complexes insuffisamment caractérisés par ces indicateurs simples, et une analyse plus fine est nécessaire, notamment par imagerie.

Dans le cadre de cette thèse, nous nous concentrons sur le cœur et une pathologie associée à forte prévalence, l'infarctus du myocarde, étudiée par le biais de l'imagerie cardiaque. Cette imagerie permet l'observation des structures anatomiques (les différentes cavités cardiaques, les parois du myocarde, les valves) et de certains paramètres fonctionnels associés (flux sanguins, fraction d'éjection, déformation myocardique, etc.).

Un cas propice à l'utilisation de l'imagerie cardiaque est la détermination de la cause de symptômes qui peuvent être provoqués par des anomalies cardiaques, tels que les douleurs thoraciques et l'essoufflement [4]. Selon les recommandations officielles de la Société Européenne de Cardiologie (ESC) [5], le premier contrôle effectué en pratique lors de l'accueil d'un patient avec une suspicion d'infarctus du myocarde est l'électrocardiogramme (ECG). Cette technique consiste à mesurer les signaux électriques cardiaques par le biais d'électrodes placées sur le torse du patient. L'ECG a une forme saine de référence, et l'on peut déduire en cas de motifs pathologiques certaines spécificités de la maladie développée par le patient [6]. L'ECG permet notamment de confirmer la suspicion d'infarctus en cas par exemple d'élévation du segment ST (voir Figure 1.1). On parle alors d'infarctus avec élévation du segment ST, ou STEMI. Cependant, bien que très utile au diagnostic et à la prise en charge rapide des patients, l'ECG n'est pas aussi complet que l'imagerie et ne permet pas un examen assez approfondi, notamment pour caractériser l'étendue des lésions et leurs conséquences sur la fonction cardiaque.

Les recommandations cliniques prescrivent ainsi au moins une échocardiographie à l'admission du patient, puis une échocardiographie complète dans les semaines qui suivent. Selon les patients, il peut être également recommandé de réaliser une acquisition d'image à l'effort quand l'état du patient est stable [9]. L'IRM est quand à elle utilisée pour le diagnostic et le suivi de l'infarctus, afin d'obtenir des mesures anatomiques plus précises [9], car les images ont une meilleure résolution spatiale que l'échocardiographie, et permettent de mieux caractériser les lésions et les tissus (notamment via les acquisitions de réhaussement). La suite de cette section donne un aperçu de ces deux techniques d'imagerie, et discute plus précisément des acquisitions et mesures pertinentes dans le cadre de l'infarctus du myocarde.

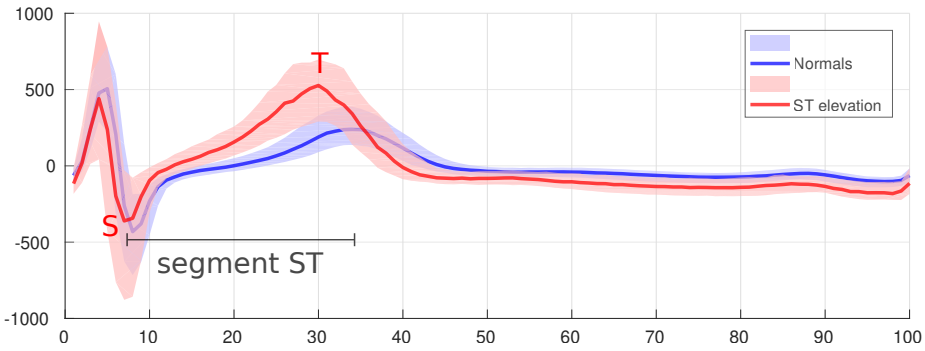


FIGURE 1.1 – Variabilité dans les signaux ECG tirés de la base de données CPSC2018 [7] utilisée pour le challenge Physionet en 2020 [8]. Cette figure représente la médiane et les 1er et 3ème quartiles du signal ECG (dérivation V3) pour les individus ayant un signal normal en bleu (N=917) et les individus présentant une élévation du segment ST en rouge (N=182). Figure réalisée par N. Duchateau.

Échographie (ultrasons) : L'échocardiographie est une technique d'imagerie médicale sûre, non invasive et rapide à mettre en place, qui est utilisée pour visualiser les structures du cœur et leur fonction pour une variété de pathologies et symptômes [10]. Elle utilise des ondes sonores à haute fréquence, ou ultrasons (US), pour créer des images en temps réel des organes, des tissus et des structures de l'intérieur du corps. Les US sont émis par une sonde d'échographie placée en contact avec la peau. Les ondes transmises sont ensuite réfléchies par les différents tissus et organes puis renvoyées sous forme d'écho à la sonde. La vitesse de propagation des ondes dans le corps dépend du type de tissu ou de matière traversée. Leur temps de retour à la sonde permet d'en estimer la nature et donc de produire une image. Les informations recueillies sont ensuite converties en images en quasi temps réel, et visualisées sur un écran. Ces images sont de différents types en fonction de la méthode utilisée pour la visualisation. On peut par exemple citer les acquisitions de mode B ("B" pour *Brightness*), le Doppler tissulaire, et le Doppler de flux, respectivement utilisées en clinique pour visualiser les parois du myocarde, leur vitesse selon la direction de l'émission ultrasonore, et la vitesse des flux sanguins au niveau des valves. Plus récemment et principalement dans le cadre de la recherche, le suivi de *speckle* (en anglais *speckle tracking echocardiography*, STE) permet également la quantification du mouvement et de la déformation (*strain*) localement en chaque point du myocarde, à partir de séquences d'images de mode B.

L'échocardiographie possède une bonne résolution temporelle, qui dépend de la méthode utilisée : elle est typiquement de 40 à 80 images par seconde pour des séquences de mode B 2D, et de 100 à 140 images par seconde pour l'échocardiographie Doppler Couleur [10]. Les images ont cependant une résolution spatiale limitée, la présence des motifs de "*speckles*" et la présence d'artéfacts limitant la

visualisation précise des contours endocardique et epicardique, notamment dans certaines régions du myocarde. Afin de mieux standardiser les acquisitions et les analyses, le cœur est imagé selon différents axes en fonction de l'orientation de la sonde échographique. Par exemple, dans le cas d'une acquisition apicale, cela donne accès à 3 types de vues distribuées sur la circonférence du coeur et permettant de visualiser les différentes chambres et valves cardiaques, comme le montre la Figure 1.2.

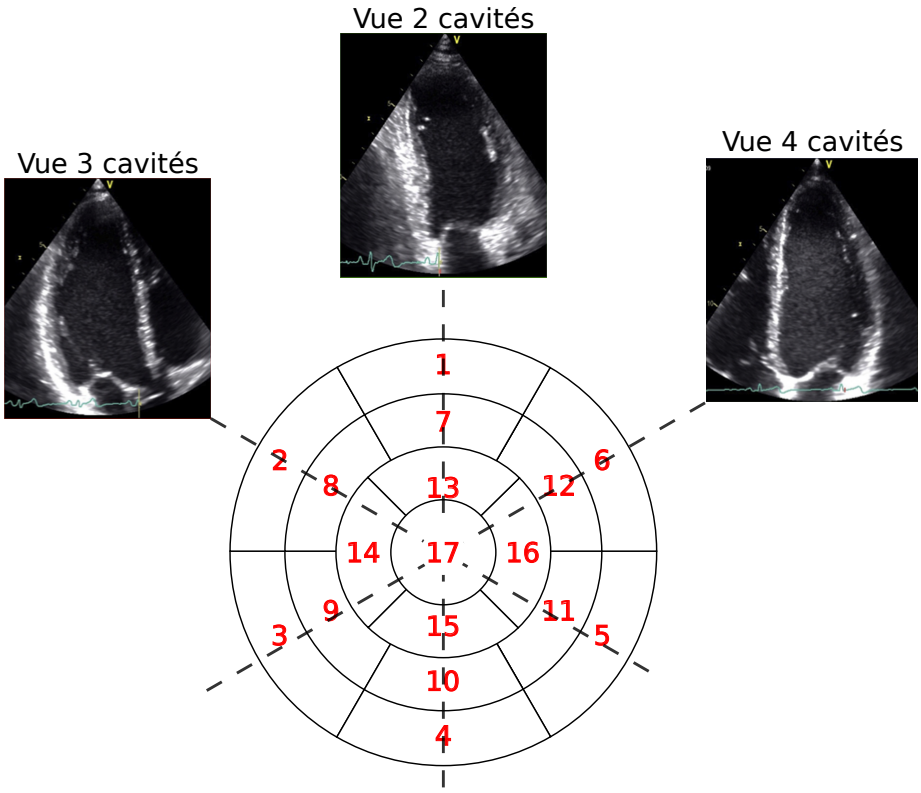


FIGURE 1.2 – Trois vues apicales permettent de couvrir l'ensemble des segments du ventricule gauche, numérotés de 1 à 17 selon la convention de l'American Heart Association [11]. Image adaptée de [12]

Imagerie par Résonnance Magnétique (IRM) : L'IRM est une technique d'imagerie médicale qui repose sur l'utilisation d'un champ magnétique intense et d'ondes radio pour produire des images détaillées du corps humain. Le principe de l'IRM repose sur le moment magnétique intrinsèque des atomes qui composent les tissus corporels, qui les rend sensibles à un champ magnétique. Dans un imageur IRM, le patient est soumis à un champ magnétique intense fixe. Les spins des atomes d'hydrogène présents dans l'eau que contient

son corps s'alignent sur ce champ magnétique. On modifie ensuite ce champ magnétique, et une antenne capte les émissions des protons lors de la relaxation (retour en position initiale après qu'on ait stoppé la modification du champ magnétique). Les émissions des protons diffèrent en fonction du type de tissu dans lequel se trouve le proton, ce qui permet de différentier les structures et de produire une image anatomique de haute qualité. Par rapport à l'échocardiographie, elle permet une visualisation plus précise des structures anatomiques grâce à un meilleure résolution spatiale. Les acquisitions dynamiques (séquences d'images "cine") permettent également d'étudier la fonction cardiaque, malgré une résolution temporelle plus faible.

Certaines acquisitions IRM permettent de caractériser la viabilité des tissus (voir Figure 1.3). C'est le cas des acquisations de réhaussement précoce et tardif ("early" et "late" Gadolinium enhancement / EGE et LGE, respectivement), qui consistent à injecter au patient un produit de contraste (le gadolinium), qui mettra en évidence des zones avec un défaut de perfusion (l'infarctus, visible par un hyper-réhaussement du signal sur les images LGE, et l'obstruction microvasculaire (MVO), visible par un affaissement de signal sur les images EGE et LGE). Ces acquisitions constituent le standard en IRM clinique pour déterminer la taille de l'infarctus [13]. Ce sont les modalités qui seront principalement utilisées dans cette thèse.

D'autres types d'acquisition permettent de caractériser la nature des tissus : par exemple, les images pondérées en T1 et T2 semblent permettre de visualiser des zones de myocarde lésées, l'oedème et la fibrose. Des exemples de données LGE, et pondérées en T1 pré et post injection de l'agent de contraste sont présentées sur la Figure 1.4.

1.1.2 Cœur et infarctus du myocarde

1.1.2.1 Fonctionnement et alimentation du cœur

Le cœur est un muscle qui permet l'oxygénation des organes du corps humain en propulsant le sang, vecteur de l'oxygène, via une action de pompe, à travers l'organisme. Le cœur est constitué de 4 cavités, représentées sur la Figure 1.5-(a) : les deux ventricules, droit et gauche (VD et VG), et les deux oreillettes, droite et gauche également (OD et OG). Les oreillettes réceptionnent le sang avant de remplir les ventricules, qui se contractent pour propulser le sang. Le ventricule droit a pour fonction d'envoyer le sang non-oxygéné vers les poumons afin qu'il se recharge en oxygène, alors que le ventricule gauche propulse le sang oxygéné vers les organes du corps humain via l'aorte. Le cœur étant lui-même constitué d'un muscle (le myocarde), il est alimenté en sang via des artères, appelées artères coronaires (droite et gauche, voir Figure 1.5-(b)). L'artère coronaire gauche se divise en deux, l'artère descendante antérieur gauche (LAD) et l'artère circonflexe (LCX), toutes deux vascularisant principalement la partie supérieure du ventricule gauche. L'artère coronaire droite irrigue quand à elle le ventricule droit et la partie inférieure du ventricule gauche, comme le montrent les Figures 1.5 et 1.6.

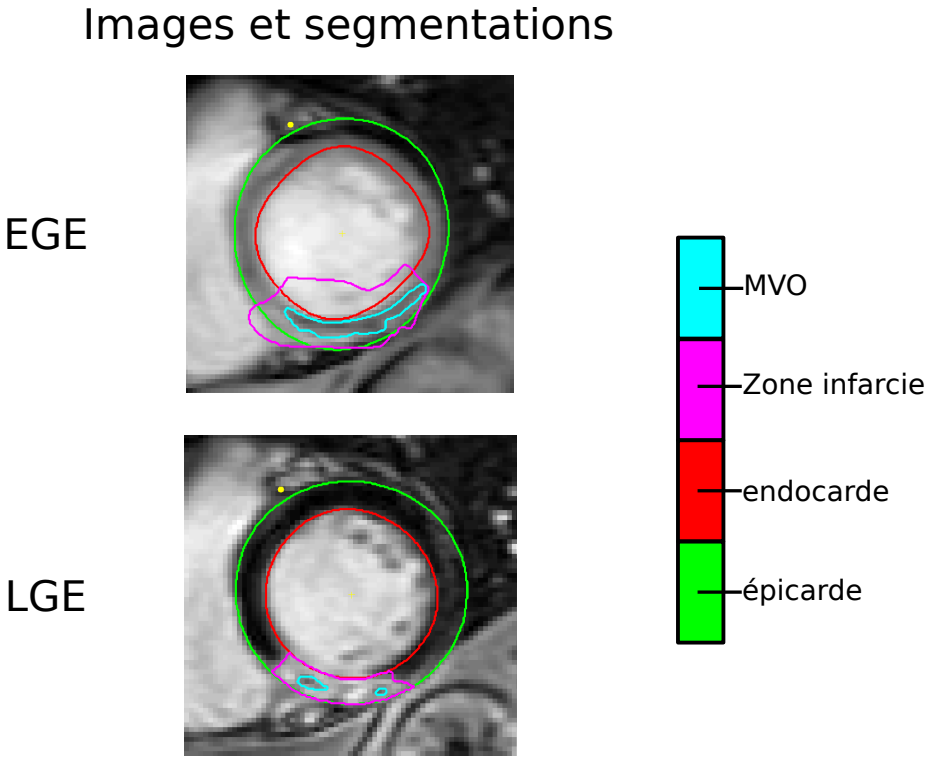


FIGURE 1.3 – Exemple d'acquisitions EGE et LGE sur le même patient, issu de la cohorte MIMI [14]. L'image EGE est acquise entre 4 et 10 minutes après l'injection du produit de contraste, contre entre 10 et 20 minutes pour l'image LGE. La première ligne montre la donnée EGE, et la segmentation du MVO associé, ainsi que la zone infarcie. Cette zone n'est pas détectable sur la donnée EGE, elle est donc ici estimée grossièrement. La deuxième ligne montre la donnée LGE, avec cette fois également la segmentation précise de l'infarctus. La donnée EGE donne une estimation plus fiable du MVO mais ne permet pas d'estimer la zone infarcie, tandis que la donnée LGE permet d'estimer la zone infarcie mais donne une estimation du MVO moins robuste.

1.1.2.2 Infarctus du myocarde

Les pathologies cardiaques affectent la capacité de pompage du coeur, perturbant sa fonction et provoquant des problèmes d'oxygénation du corps humain. Les maladies cardiovasculaires sont la première cause de décès prématuré dans le monde, et la seconde en France (après le cancer) [2]. En France, environ une personne toutes les 4 minutes est victime d'un accident cardiovasculaire [3]. Cette thèse s'intéresse en poarticulier à l'infarctus aigü du myocarde, qui est la ma-

2. Source : Ministère Français de la Santé, 2022
3. Source : CépiDC-Inserm et solidarités-santé.gouv, 2016

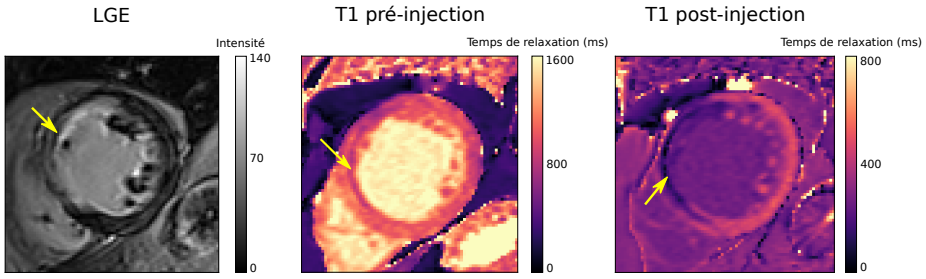


FIGURE 1.4 – Exemples d'acquisitions LGE, pondérée T1 pré-injection et post-injection. Ces images proviennent du même patient, et ont été acquises 12 mois après l'infarctus (ici, dans le territoire de l'artère descendante gauche, indiqué par une flèche jaune). Elles proviennent de la base de données HIBISCUS-STEMI [15] et m'ont été fournies par R. Deleat-Besson, doctorant à CREATIS.

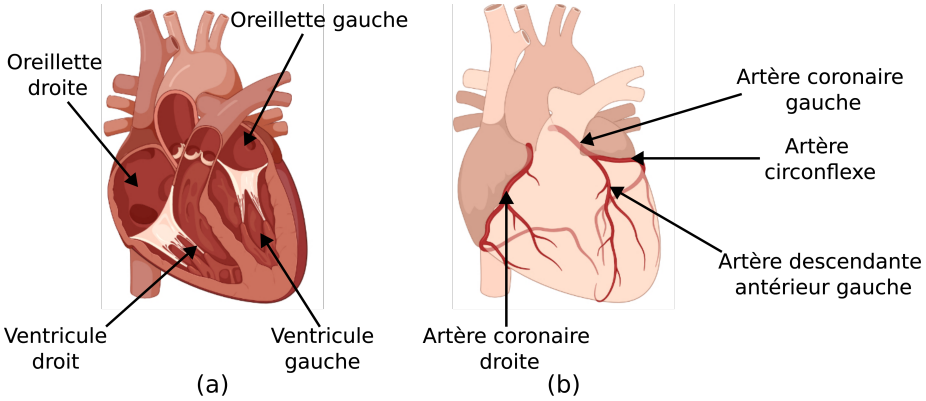


FIGURE 1.5 – (a) Coupe du cœur représentant les ventricules droit et gauche et les oreillettes droite et gauche. (b) Vue extérieure indiquant les artères coronaires et leur subdivisions principales. (Schémas créés à partir de BioRender [1].)

ladie cardiovasculaire la plus fréquente. En 2019, on dénombre en France 80000 infarctus du myocarde, dont 12000 provoquent le décès du patient [4].

Infarctus aigü du myocarde : L'infarctus du myocarde correspond à la nécrose d'une partie du muscle cardiaque. Cette maladie est provoquée par la sténose (ou rétrécissement du diamètre) plus ou moins prononcée d'une des artères coronaires. La sténose coronarienne est engendrée par la formation de plaques d'athérome, dans lesquelles des cellules inflammatoires et des lipides se réorganisent avec d'autres éléments, pour conduire à une modification locale de l'aspect et de la nature de la paroi vasculaire. Ce dépôt graisseux augmente progressivement, ralentissant le débit en sang à l'intérieur de l'artère coronaire

4. Source : Fédération Française de Cardiologie, 2019

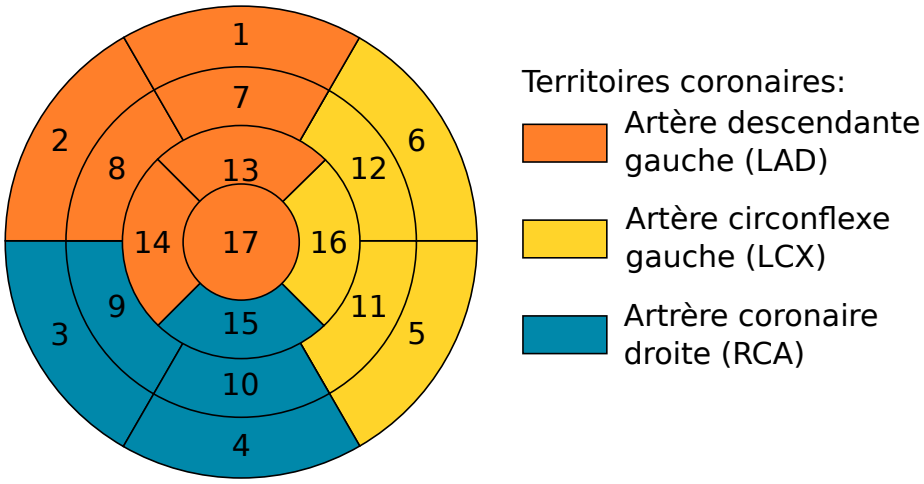


FIGURE 1.6 – Représentation de référence de l'American Heart Association (AHA) en 17 segments [11], indiquant les différents segments du cœur sous forme de vue aplatie ("Bull's eye"). Les couleurs indiquent quelle est l'artère coronaire irriguant chaque segment.

concernée, et donc l'apport en oxygène dans la région irriguée par cette artère. Cette occlusion mène dans un premier temps à l'ischémie partielle du muscle cardiaque : les cellules myocardiques manquent de sang donc d'oxygène. Ce manque d'oxygène fragilise les cellules et peut, sans prise en charge, mener à leur nécrose. L'infarctus du myocarde est d'abord une pathologie aigüe : il évolue rapidement et soudainement (à l'opposé, les cancers par exemple sont des maladies chroniques, qui ont des évolutions lentes et longues). La mort des cellules du myocarde impacte la capacité de pompe du cœur et son activité électrique, jusqu'au décès du patient, d'où la nécessité de traiter l'infarctus le plus rapidement possible (10% des gens atteints d'infarctus meurent dans l'heure qui suit [5]).

Traitement de l'infarctus et lésions de reperfusion : À l'heure actuelle, on soigne majoritairement l'infarctus du myocarde par angioplastie si le délai d'intervention est court. L'occlusion est traitée en gonflant un ballon pour écraser l'amas graisseux et redonner son volume à l'artère. S'y ajoute potentiellement la pose d'un endoprothèse métallique, le "stent", pour s'assurer que l'artère conserve sa forme dans le futur (voir Figure 1.7). On complète cette opération par la prise de médicaments contre l'agrégation de plaquettes. Comme l'artère est n'est plus obstruée, le sang afflue de nouveau dans le coeur : c'est la reperfusion cardiaque. Néanmoins, si cette reperfusion permet de stopper l'ischémie, certaines cellules ne supportent pas bien ce brusque afflux sanguin, ce qui peut provoquer des blessures de reperfusion de gravité variable (œdème réversible,

5. Source : Assurance Maladie, 2019

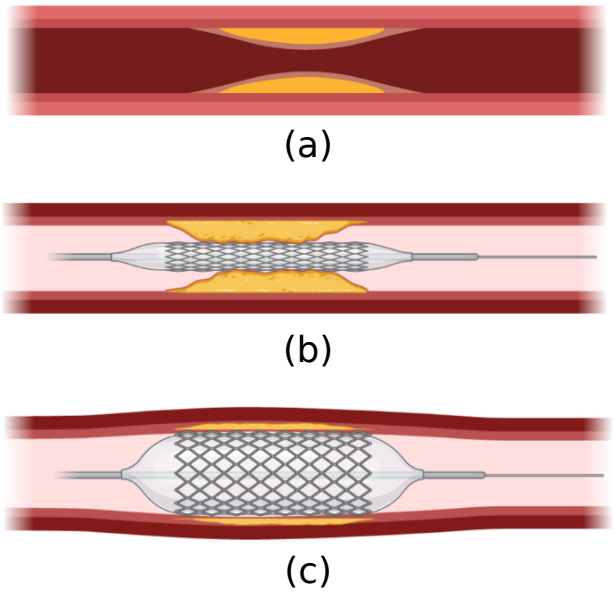


FIGURE 1.7 – Principe de l'angioplastie avec pose d'un stent : (a) artère sténosée ; (b) ballon (avec un stent) avant qu'il soit gonflé ; (c) artère une fois le ballon gonflé. (Schémas créés à partir de BioRender [6])

obstruction ou destruction des capillaires, etc.) [16]. Les lésions de reperfusion se caractérisent par un phénomène de non-reflux sanguin dans certaines zones du myocarde avec notamment des zones d'obstruction microvasculaire (MVO), situées à l'intérieur de la zone infarcie (voir Figure 1.3), comme nous le verrons plus en détail dans la suite de cette thèse.

Remodelage : Dans les mois qui suivent un infarctus du myocarde, et en particulier si celui-ci n'a pas été traité, le coeur s'adapte morphologiquement aux changements provoqués par la présence de la zone infarcie, jusqu'à être incapable de remplir a minima sa fonction. On observe alors des changements structurels, certains ayant pour objectif de compenser en partie la capacité de pompage cardiaque, d'autres reflétant directement la nécrose des tissus : c'est le remodelage. Ce phénomène est d'abord local et se caractérise par un amincissement du myocarde et une dilatation dans la zone infarcie, ainsi que par l'épaississement des segments voisins ; puis global avec une dilatation sphérique de la cavité [17] (voir Figure 1.8).

1.1.2.3 Utilisation des données par les cliniciens

Pour mieux comprendre et soigner ces pathologies, il est primordial d'exploiter au maximum l'information présente dans les données médicales collectées sur les patients. À l'heure actuelle, l'IRM de réhaussement tardif ("Late Gadolinium

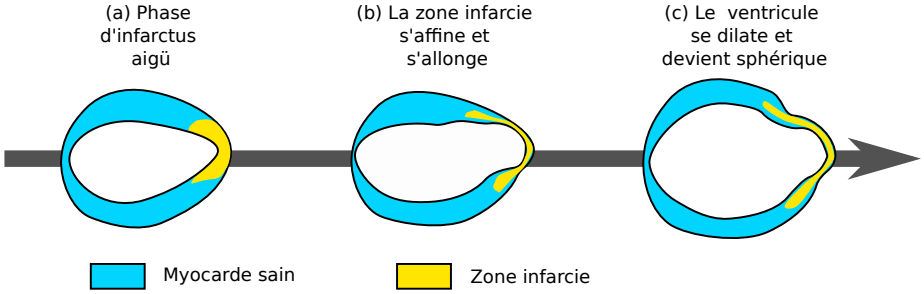


FIGURE 1.8 – Les différentes étapes du remodelage post infarctus aigü du myocarde (a). La paroi ventriculaire s'affine dans la zone infarcie (b), puis le ventricule se dilate (c). Figure inspirée de *Konstam et al.* [17]

Enhancement" / LGE) est la modalité de choix pour caractériser la forme et l'ampleur de l'infarctus, en particulier via des mesures scalaires : la taille de l'infarctus, sa transmuralité (pourcentage d'infarctus traversant le myocarde) et son étendue le long de l'endocarde (*Endocardial surface area*) [18]. Comme commenté dans la Section 1.1.1, d'autres types d'acquisition IRM peuvent être utilisés, notamment les acquisitions IRM pondérées T1 et T2 pour caractériser le contenu des tissus. En outre, certains motifs de l'ECG sont connus et utilisés pour déterminer de potentiels risques liés à la pathologie et déterminer le type d'intervention chirurgicale nécessaire [6]. Enfin, l'imagerie dynamique (IRM cine et échocardiographie) permet le calcul d'indicateurs globaux tels que la fraction d'éjection (pourcentage de sang présent dans la cavité ventriculaire gauche expulsé à chaque pulsation, en pratique estimé par la différence relative de volumes entre la fin de systole et la fin de diastole) ou régionaux/locaux tels que la déformation myocardique (strain) [19]. Ces indicateurs, bien qu'indispensables à la prise en charge des patients en pratique, sont encore trop limités pour rendre compte de la complexité des mécanismes d'ischémie-reperfusion à l'œuvre. Il est donc nécessaire de trouver de nouveaux moyens d'exploiter ces données, en intégrant pleinement le contenu des données image au-delà de ces indicateurs scalaires utilisés actuellement en clinique.

L'exploitation des images médicales représente cependant un défi pour de nombreuses raisons. Premièrement, ces images et les descripteurs qui en découlent sont en haute dimension, ce qui complexifie leur exploitation directe : il est difficile d'avoir une vision d'ensemble des données, et le traitement demande de grandes capacités de calcul. Ensuite, la quantité de données n'est pas toujours suffisante pour bien représenter la diversité des individus et les variations des pathologies. De plus, la qualité des données est très variable d'un centre médical à un autre et dépend de l'expérience du praticien et de la diversité des protocoles cliniques utilisés. Cette problématique est renforcée par la variabilité des morphotypes et rend la comparaison entre les différentes images difficile, même si les protocoles d'imagerie sont identiques. Il arrive également que les données ou

que leurs étiquettes soient absentes ou incomplètes. Toutes ces difficultés forcent l'information à être recoupée entre différents examens ou différentes vues pour avoir une vision d'ensemble d'un cas clinique. Les médecins recoupent ces informations en se servant de leur expérience et de leurs connaissances, ce qui reste subjectif et difficilement transposable à l'échelle de grandes populations. Les outils d'analyse computationnelle (et en particulier ceux basés sur l'apprentissage statistique) ont ainsi un fort potentiel pour fournir un soutien quantitatif pour l'analyse statistique de données d'imagerie, complexes et de différents types.

1.2 Contexte methodologique :

Un des outils disponibles pour la caractérisation des pathologies est l'analyse statistique de données médicales au sein d'une population d'individus pathologiques ou sains. Cette analyse est réalisée à partir de descripteurs complexes extraits des images ; ce sont des données de haute dimension, dont on désire préserver les propriétés. Il est pertinent de considérer que les images médicales, tout comme un grand nombre de descripteurs de haute dimension extraits de ces images, appartiennent à une variété non linéaire. Prenons l'exemple des deux images "jouets" d'infarctus représentées en Figure 1.9 : ces deux motifs d'infarctus sont identiques, mais tournés différemment. Si l'on calcule la moyenne linéaire des deux images, on obtient une image qui ne représente pas un motif plausible d'infarctus. La moyenne attendue est en fait un motif d'infarctus similaire aux deux précédents, positionné entre les deux infarctus (a) et (b) de la Figure 1.9, qu'on ne peut obtenir que de façon non linéaire.

La variété non linéaire sous-jacente est généralement inconnue, mais il est possible de l'estimer en utilisant des méthodes d'apprentissage de variétés, qui seront discutées dans le Chapitre 2.

Pour obtenir une caractérisation plus complète de la population, il est pertinent d'avoir recours à plusieurs descripteurs de haute dimension. Ces descripteurs montrent en effet des aspects différents et complémentaires des données. Cependant, ils sont en général de types hétérogènes (données tabulaires, images, maillages, etc.) ce qui complique leur analyse. Une solution est de fournir une représentation commune à l'aide d'un apprentissage multi-modal ou multi-descripteurs, par exemple à l'aide des méthodes d'alignement variétés ou de fusion de données qui seront discutées au Chapitre 2 de ce document.

La variabilité des données et des problématiques cliniques est importante, et le type d'analyse à effectuer est dicté par l'application clinique. Si l'on veut proposer un diagnostic ou un pronostic, l'analyse sera en général supervisée. On peut également analyser les données de façon non-supervisée, comme dans l'application de cette thèse, si l'on cherche à identifier des sous-groupes, à analyser la variabilité d'une population, étudier des tendances ou détecter des anomalies. D'autres défis sont propres à l'accés aux données cliniques et nécessitent d'adapter les algorithmes et méthodes statistiques utilisées. Ces défis propres à l'application sont relatifs à la quantité de données disponibles (et donc à la

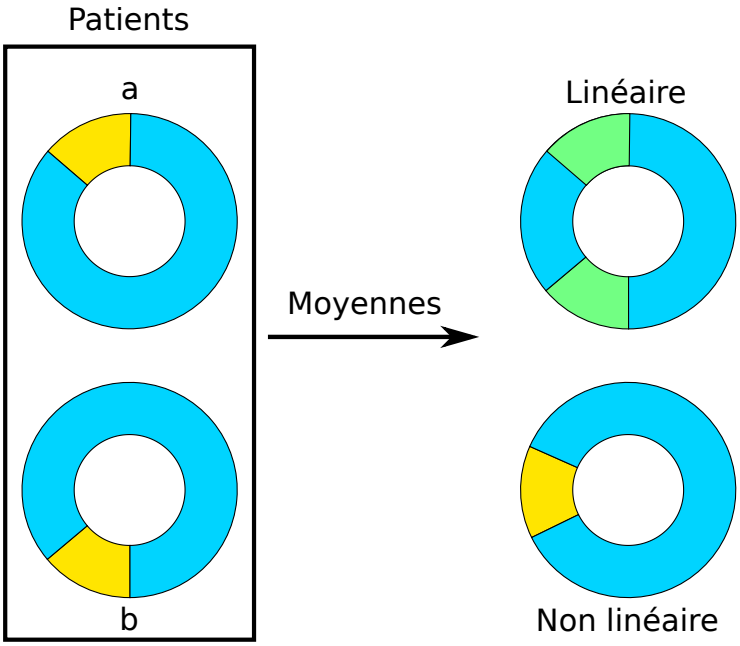


FIGURE 1.9 – Illustration de la nécessité d'appliquer des méthodes non linéaires quand on traite des images médicales. Les représentations (a) et (b) schématisent des segmentations d'infarctus du myocarde (myocarde sain en bleu clair, segmentation de l'infarctus en jaune). Ces deux infarctus sont tournés différemment, et leur moyenne linéaire n'est pas elle-même un motif d'infarctus, contrairement à sa version non linéaire.

variabilité de motifs pathologiques à disposition), à la qualité variable des données (acquisition dans des centres différents) et aux données manquantes pour diverses raisons qui impactent l'analyse de population. Ces problématiques requièrent une adaptation spécifique à chaque problème.

1.2.1 Réduction de dimensions non supervisée

Bien qu'une grande majorité de méthodes sont actuellement développées dans le cas supervisé, cette stratégie n'est parfois pas viable pour caractériser une population selon l'application clinique visée. D'abord, les labels ne sont pas forcément fiables (comme dans le cas de pathologies représentant un continuum depuis la normalité, par exemple l'insuffisance cardiaque à fraction d'éjection préservée [20]). Par ailleurs, ceux-ci ne sont pas toujours accessibles (par exemple, la survie du patient pour des données relativement récentes, ou lorsque la pathologie est insuffisamment cernée par les cliniciens, comme c'est le cas pour les mécanismes d'ischémie-reperfusion que nous étudions dans cette thèse).

Dans cette thèse, nous nous concentrons sur le développement de méthodes

d'apprentissage non supervisé pour analyser des données d'IRM cardiaque issues de patients avec infarctus aigü du myocarde. Du point de vue clinique, l'objectif premier est d'obtenir une représentation simplifiée mais informative des données permettant d'analyser la variabilité des lésions (complexes) au sein d'une population ou de sous-groupes de patients. L'objectif à moyen terme est d'utiliser cette représentation pour faire de la stratification de risque au sein d'une population, et de caractériser l'évolution des patients dans le cadre d'études longitudinales.

Nous construisons pour cela sur les méthodes de réduction de dimensions, issues de l'apprentissage non supervisé. L'hypothèse est que les données de haute dimension appartiennent à un espace non linéaire gouverné par un nombre restreint de dimensions, dont il est possible d'estimer les principales dimensions. Cette représentation simplifiée (l'espace latent) doit permettre à la fois une meilleure visualisation des données de la population, mais également doit être pertinente statistiquement pour pouvoir quantifier les différences entre patients. Dans notre cas, les données d'IRM cardiaque majoritairement utilisées dans le cadre de cette thèse (acquisitions de réhaussement tardif) sont constituées pour chaque patient (après rééchantillonnage / alignement sur une référence commune) d'une pile de 21 coupes, chaque coupe étant une image 2D de 80 pixels par 80 pixels. Cela conduit à un total de 134400 dimensions par patient et par descripteur, que l'on cherche à réduire.

Si l'objectif est d'identifier plusieurs groupes distincts dans le jeu de données (notamment en vue de quantifier le risque associé à chaque sous-groupe), il est pertinent de faire appel aux méthodes de regroupement ("clustering" en anglais). Néanmoins, ces méthodes ne sont pas directement applicables en très haute dimension et nécessitent l'estimation d'un espace de dimension réduite, ce que permet l'apprentissage de représentation.

Au-delà du simple regroupement de sujets, la réduction de dimensions doit pouvoir permettre de situer les patients les uns par rapport aux autres (voir Figure 1.10). Plus précisément, l'apprentissage de variétés (*manifold learning*) [21, 22] vise à estimer un espace latent dans lequel les sujets sont distribués selon une métrique préalablement définie, permettant leur analyse statistique quantitative. Par exemple, l'algorithme Isomap [23] construit une représentation réduite à partir des distances géodésiques entre paires de points de données sur un graphe d'adjacence. Disposer de telles représentations est extrêmement intéressant pour quantifier les anomalies par rapport à une population de référence, ou quantifier l'évolution d'un patient.

Ces méthodes ont été élaborées pour analyser des descripteurs complexes. Cependant, en pratique, ces descripteurs sont nombreux et hétérogènes, et les concepts décrits jusqu'alors ne sont donc pas complètement adaptés.

1.2.2 Intégration de données hétérogènes

Comme commenté dans la sous-Section 1.1.2.3, un deuxième défi majeur pour notre analyse est d'exploiter plusieurs modalités d'imagerie, ou plusieurs descripteurs issus de l'imagerie, de types hétérogènes. Les images de réhausse-

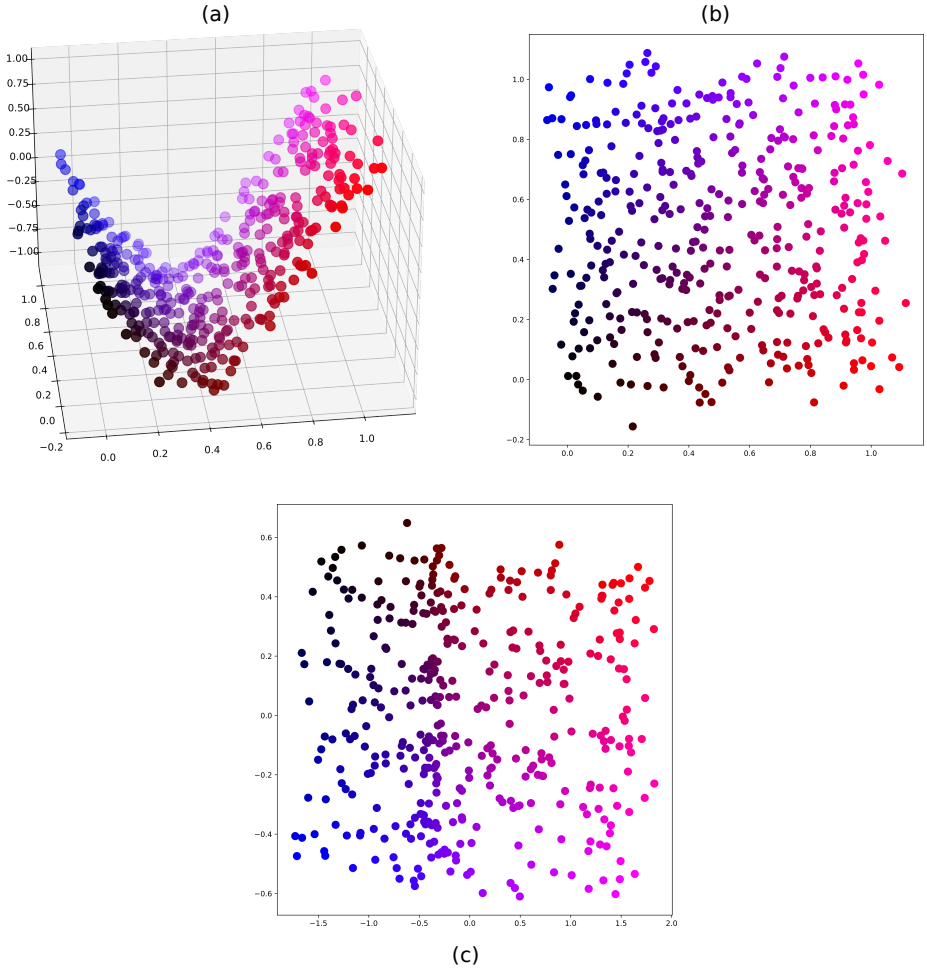


FIGURE 1.10 – Illustration du concept d'apprentissage de représentations. À partir de l'espace de haute dimension (a) (ici 3 dimensions), on cherche à extraire la vérité terrain (2 dimensions) (b) en estimant un espace latent de faible dimension (ici 2 dimensions) (c). L'espace latent a ici été estimé avec l'algorithme Isomap (avec 5 voisins) qui sera présenté plus en détail en Section 2.2.2. Le code couleur a été créé spécifiquement pour mieux examiner l'arrangement des échantillons dans les différents espaces. L'espace latent estimé (c) doit bien sûr se rapprocher le plus possible de la distribution réelle (b), ce que l'on observe ici à une rotation 2D près.

ment tardif invitent par exemple à analyser la segmentation de la zone infarcie, mais aussi celle du MVO, qui correspondent à des motifs binaires disponibles en chaque point du myocarde. Il est également intéressant d'exploiter directement l'image, qui contient des textures et des motifs potentiellement pertinents et

non disponibles dans la segmentation, également disponibles en chaque point du myocarde. Ces descripteurs sont complémentaires car ils décrivent des aspects différents de l'infarctus du myocarde : dans notre exemple, pour avoir une vue complète de l'obstruction microvasculaire, il est également important de regarder quelle est la zone infarcie et son étendue. D'où le besoin essentiel d'algorithmes capables d'exploiter conjointement ces différents descripteurs : dans le cadre de l'apprentissage de représentations, ces algorithmes sont regroupés sous les méthodes de fusion de données, dont nous donnons une description plus avancée en Section 2.3.3.

En plus des segmentations ou du contenu image disponible en chaque pixel des acquisitions LGE, nous pourrions tout à fait intégrer des contenus issus d'autres acquisitions : par exemple, les images pondérées en T1 ou T2 dont les valeurs en chaque pixel ne peuvent pas forcément être concaténées ou agglomérées avec les informations du LGE, ou des données de dimensionalités ou types différents comme les caractéristiques cliniques des patients (age, sexe, indice de masse corporelle, traitement, etc.) qui sont des scalaires, ou les signaux ECG. Le déséquilibre induit par ces descripteurs se manifeste de deux façons principales :

1. Dans la manière de combiner ce descripteur avec d'autres (la concaténation n'étant pas pertinente dans de nombreux cas),

2. Dans la quantité d'information présente dans chaque descripteur, et donc son influence relative sur le résultat global ; les descripteurs pouvant par ailleurs être partiellement redondants.

À l'heure actuelle, la majorité des algorithmes spécifiques au traitement de plusieurs descripteurs (et allant au-delà d'une concaténation ou d'un traitement multi-canaux) ciblent la fusion de ces informations dans une représentation commune, typiquement un espace latent dans lequel un point représente un compromis entre les informations issues des différents descripteurs utilisés en entrée. Cette stratégie est néanmoins limitée sur plusieurs points. D'abord, elle fait l'hypothèse qu'il existe une représentation qui à elle seule puisse représenter plusieurs données d'entrée très diverses, potentiellement redondantes ou pas du tout porteuses de la même information. Ensuite, cette fusion se fait en considérant toutes les données en même temps, ce qui est sous-optimal du point de vue calculatoire (rapidité et robustesse notamment), et ne tient absolument pas compte des connaissances a priori que l'on pourrait posséder sur les données.

Pour reprendre l'exemple de l'infarctus du myocarde, nous savons que la segmentation de la zone infarcie du myocarde sur les images LGE donne la position et l'étendue de l'infarctus, qui est primordial pour la bonne compréhension de la donnée MVO, bien plus difficile à exploiter sans ce support. Il est regrettable que les méthodes de fusion existantes ne permettent pas la prise en compte d'une hiérarchie dans les descripteurs utilisés, alors que cette connaissance est clé dans le raisonnement des médecins en pratique clinique (voir Section 1.1.2.3).

1.3 Objectifs de cette thèse

Pour remédier à cela, nous proposons dans cette thèse de prendre en compte le lien hiérarchique entre les modalités d'imagerie cardiaque au sein d'un algorithme d'apprentissage de représentation. L'exploitation de ce lien, inspiré du raisonnement des cliniciens, devrait permettre de bien mieux intégrer des données hétérogènes et partiellement redondantes, tout en facilitant l'intégration de descripteurs complexes.

Plus concrètement, dans le cadre de cette thèse, j'ai exploité le lien hiérarchique entre différentes modalités d'imagerie cardiaque pour la caractérisation de populations de patients atteints d'infarctus aigü du myocarde. À partir des verrous méthodologiques décrits précédemment, trois objectifs majeurs peuvent être dégagés pour cette thèse :

1. Développer de nouvelles méthodes pour l'intégration hiérarchique de données multiples, hétérogènes et de haute dimension,
2. Converger vers une formulation générique et généralisable,
3. Évaluer son apport pour la caractérisation des lésions d'ischémie-reperfusion chez des patients avec infarctus aigü du myocarde suivis par imagerie.

La suite de ce document présente une synthèse des principales méthodes d'apprentissage de représentation pertinentes dans le cadre de cette thèse (Chapitre 2), les complète par une introduction détaillée aux modèles de processus Gaussiens à variables latentes (Chapitre 3), et détaille les deux contributions méthodologiques principales de cette thèse (Chapitres 5 et 6) pour considérer une hiérarchie dans le cadre de l'analyse de données multiples de haute dimension. Les données utilisées aux Chapitres 5 et 6 sont décrites au Chapitre 4.

Chapitre 2

État de l'art : Apprentissage de représentations

Le chapitre précédent a introduit le besoin d'une meilleure exploitation des données d'imagerie cardiaque dans leur ensemble, en particulier dans le cadre de l'infarctus du myocarde. La dimensionalité de ces données est une des problématiques principales pour les traiter de façon pertinente (au minimum $128 \times 128 \times 21 \approx 340000$ pixels pour couvrir un cœur en multi-coupes par exemple). La quantité d'échantillons nécessaires pour correctement représenter un tel espace de données augmente drastiquement avec le nombre de dimensions (c'est ce qu'on appelle la malédiction de la dimensionalité : *curse of dimensionality*). Un nombre aussi élevé de dimensions peut également poser des problèmes de mémoire et de capacité de calcul. La réduction de dimensions vise à pallier ce problème en estimant un espace de faible dimension représentant au mieux les données initiales, qu'on appelle espace latent. Cet espace latent permet de :

1. Clarifier la lecture des données de haute dimension en en réduisant la redondance et en permettant même une visualisation,
2. Définir un cadre mathématique permettant de réaliser des statistiques entre individus ou entre sous-groupes d'individus,
3. Dégager des motifs permettant une meilleure interprétabilité des tendances au sein d'une population,
4. Combiner plusieurs descripteurs dans une même représentation latente.

2.1 Réduction de dimensions

Dans le cadre de l'apprentissage de représentations [24], il existe un grand nombre de méthodes permettant la réduction de dimensions [21, 22]. Certaines sont supervisées, d'autres non. Pour cette thèse, nous privilégions la réduction de dimensions non supervisée. En effet, dans notre cas applicatif, les étiquettes ne sont pas disponibles. De façon plus générale, même quand les étiquettes sont

accessibles, elles ne sont pas forcément fiables et pourraient induire un biais dans la représentation qu'on cherche à produire. En apprenant de façon non supervisée, on s'affranchit donc de ces biais et on effectue des statistiques sans a priori. Cette propriété est intéressante, car notre but dans un premier temps est de produire une représentation visant à améliorer la compréhension de la pathologie ciblée : on ne cherche pas directement à produire un diagnostic ou un pronostic. Une autre raison de produire des représentations non supervisées est liée à la nature de ce que l'on cherche à décrire. La transition d'un état sain à un état pathologique est un évènement continu, alors que les étiquettes sont catégoriques (généralement binaires). On vise ici à caractériser une pathologie et ses évolutions, donc à produire une représentation continue, pour laquelle des étiquettes discontinues ne sont pas pertinentes.

Ce chapitre présente d'abord une première famille de méthodes pour la réduction de dimensions non supervisée, l'apprentissage de variétés, qui propose un cadre permettant l'exploitation de distances statistiques entre échantillons, y compris dans l'espace latent. La suite du chapitre complète cette présentation avec d'autres méthodes pour le calcul d'un espace latent à partir de données en haute dimension, comme les auto-encodeurs (réseaux de neurones). La fin de ce chapitre traite des extensions de ces méthodes à plusieurs descripteurs.

Le jeu de données CelebA : Cette synthèse sur la réduction de dimensions non supervisée s'appuie sur un jeu de données réel non médical avec lequel j'ai travaillé en début de thèse : la base d'images CelebA [25]. CelebA est un jeu de données d'images de dimensions 178×218 pixels, représentant des visages de célébrités. Il donne accès également à des points d'intérêts, qui sont les positions des centres des deux yeux, du nez, et les positions des coins de la bouche. Le jeu de données contient également 40 labels portant sur les caractéristiques des images, tels que la couleur des cheveux, le sexe de la personne, son expression faciale (sourire ou non), etc. L'avantage de ce jeu de données dans notre contexte méthodologique est la présence de points d'intérêts : ces points d'intérêts sont un premier niveau naturel de hiérarchie. Le jeu de données complet est constitué de plus de 200000 individus aux apparences très différentes : un apprentissage de représentations lancé directement sur les images avec une métrique simple sera très fortement conditionné par l'apparence globale de l'image (dont le niveau de gris et l'apparence du fond) et représentera moins le contenu lié aux visages. Nous avions observé ce phénomène sur le jeu de données complet, ce qui nous a poussé à travailler sur une portion plus spécifique du jeu de données pour limiter cet effet. Dans un premier temps, j'ai ainsi sélectionné un sous-ensemble de ces données pour mieux focaliser l'analyse sur des visages cohérents et donc sur leur apparence : plus de 4700 images représentant majoritairement des femmes aux cheveux de teinte foncée avec un fond d'image clair. Cette sélection a été effectuée par projections successives sur un espace latent appris avec l'algorithme *Diffusion maps*, qui sera présenté plus tard dans ce chapitre. Quelques exemples d'images de CelebA et des points d'intérêts correspondants sont présentés sur la Figure 2.1.

FIGURE 2.1 – Exemples de visages de célébrités tirés du jeu de données CelebA. Les points d'intérêts (position des yeux, des coins de la bouche et du nez) sont représentés en orange. Ces images sont issues d'une sélection de 4714 images, et comprennent majoritairement des femmes aux cheveux bruns sur fond clair. Il existe néanmoins quelques exceptions : certains fonds ne sont clairs que sur les bords, certaines images représentent des hommes. Il peut également y avoir des écritures (couvertures de magazines), des accessoires (chapeaux, lunettes) susceptibles de perturber l'estimation d'espaces latents dans le cas où l'on utilise une métrique simple pour comparer les échantillons.

2.2 Apprentissage de représentations

Pour effectuer la réduction de dimensions, on choisit de se placer dans un cadre théorique précis : l'apprentissage de variétés mathématiques. Une variété mathématique est un espace topologique localement Euclidien. Par exemple, la terre peut être considérée comme une variété : sa surface globale est courbe mais localement plane. L'apprentissage de variétés vise à estimer, à partir des échantillons disponibles, un espace latent de faible dimension associé à la variété

dont sont issus ces échantillons. Il est intéressant que les propriétés de l'espace latent et/ou la façon dont il a été construit permettent de calculer des distances simples entre échantillons (idéalement, des distances Euclidiennes), associées à des distances pertinentes le long de la variété (par exemple, distance géodésique ou distance de diffusion). Cela permet notamment de caractériser la population étudiée (par exemple, au travers de l'estimation de la moyenne et de la variabilité), de quantifier les différences entre individus ou l'évolution d'un sujet, et d'utiliser cette représentation comme entrée pour d'autres modèles (par exemple pour de la classification ou du clustering).

Notations utilisées : Dans la suite de ce chapitre, nous notons

$$\mathbf{Y} = [\mathbf{y}_1, \cdots, \mathbf{y}_N] \in \mathbb{R}^{N \times D}$$

l'ensemble des observations disponibles, N correspondant au nombre d'échantillons (souvent, le nombre de patients dans le jeu de données) et D à la dimensionalité des données (par exemple le nombre de pixels/voxels dans chaque images), que l'on cherche à réduire. La réduction de dimensions vise à estimer un espace latent de dimensionalité d, donc les variables latentes $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_N] \in \mathbb{R}^{N \times d}$, avec $d \ll D$.

Dans le cas où nous avons M types d'observations (par exemple, des images issues de plusieurs modalités, ou différents types de descripteurs extraits des images), nous notons

$$\mathbf{Y}^{(m)} = [\mathbf{y}_1^{(m)}, \cdots, \mathbf{y}_N^{(m)}] \in \mathbb{R}^{N \times D}, \qquad m \in [1, M]$$

ces observations. De même, dans le cas où l'on calcule un espace latent par type d'observation, les variables latentes associées aux observations $\mathbf{Y}^{(m)}$ seront notées $\mathbf{X}^{(m)} = [\mathbf{x}_1^{(m)}, \cdots, \mathbf{x}_N^{(m)}] \in \mathbb{R}^{N \times d}$, avec $d \ll D$.

2.2.1 Apprentissage de variétés linéaire

Ce sous-ensemble de méthodes consiste à trouver la combinaison linéaire des dimensions d'entrée permettant de représenter au mieux la population étudiée, selon un critère donné. L'analyse en composantes principales (ACP), une des méthodes linéaires les plus répandues, considère la covariance dans les données de haute dimension, et estime ainsi les principales directions de variabilité dans les données. Plus précisément, la méthode consiste à diagonaliser la matrice de covariance des données d'entrée, les vecteurs propres associés aux d plus grandes valeurs propres indiquant les principales dimensions à retenir.

En pratique, le nombre d'échantillons étant souvent restreint par rapport à la dimensionalité des données, on considère plutôt la matrice de covariance associée à la transposée des échantillons (donc de dimensionalité moindre), dont les vecteurs propres et valeurs propres sont reliés à ceux du problème original [26].

Ces méthodes linéaires sont intéressantes et répandues car elles sont simples à mettre en œuvre et qu'elles donnent accès facilement et rapidement à un

résumé des données en peu de dimensions. Par ailleurs, le passage des données de haute dimension vers l'espace latent de faible dimension (et vice-versa) s'effectue simplement par un changement de base linéaire, en utilisant les vecteur propres estimés. Ces méthodes sont cependant vite limitées : par exemple, une ACP sur des images n'est pas capable de retrouver la variété non linéaire sous-jacente et biaise ainsi l'analyse statistique de ces images, comme discuté au Chapitre 1 (voir Figure 1.9). Il est donc nécessaire dans les cas réels d'utiliser des méthodes d'apprentissage de variétés non linéaires.

2.2.2 Apprentissage de variétés non linéaire

Pour l'analyse d'images médicales, comme introduit ci-dessus, il est très souvent nécessaire d'utiliser des méthodes non linéaires. Les différentes méthodes d'apprentissage de variétés non linéaires ont de nombreux points communs, et peuvent être regroupées sous le même paradigme décrit en 2007 par *Yan et al.* [21]. Cette méthodologie repose sur le calcul d'un graphe d'affinité entre les échantillons, qui approxime la variété. Prenons l'exemple de l'algorithme *Laplacian eigenmaps* [27], dont la formulation est très proche du cadre englobant toutes les méthodes [21], et à la base de la première contribution de cette thèse.

Il s'agit tout d'abord de construire un graphe d'affinité entre échantillons : l'affinité est une valeur de 0 à 1 quantifiant la ressemblance entre 2 individus (et donc associée à une distance, la métrique le long de la variété). Par exemple, pour les *Laplacian eigenmaps*, on considère la distance de diffusion (liée à un noyau Gaussien dont la largeur définit l'échelle d'observation). Cette distance est souvent mise à 0 si un échantillon j n'est pas un des K plus proches voisins de l'échantillon i. Du point de vue du graphe, cela signifie que les nœuds i et j ne sont pas liés par une arête. Plus formellement, cette matrice d'affinité éparse $\mathbf{W} = [W_{ij}] \in \mathbb{R}^{N \times N}$ est calculée de la façon suivante :

$$W_{ij} = \begin{cases} \exp\left(-\frac{d(\mathbf{y}_i, \mathbf{y}_j)}{2\sigma^2}\right) & \text{si } j \in \mathcal{N}_K(i), \\ 0 & \text{sinon,} \end{cases} \tag{2.1}$$

où $d(.)$ désigne la distance dans l'espace de haute dimension, σ est un facteur d'échelle fixé par l'utilisateur, et $\mathcal{N}_K(i)$ est le voisinage (comportant K éléments) de l'individu i. La distance $d(.)$ est souvent prise comme la distance Euclidienne entre les images, considérées comme des vecteurs colonnes : $d(\mathbf{y}_i, \mathbf{y}_j) = \|\mathbf{y}_i - \mathbf{y}_j\|^2$.

Ensuite, il s'agit de résoudre :

$$\arg\min_{\mathbf{X}} \sum \|\mathbf{x}_i - \mathbf{x}_j\|^2 W_{ij}, \tag{2.2}$$

sous la contrainte $\sum_i D_{ii} \|\mathbf{x}_i\|^2 = 1$, avec $D_{ii} = \sum_j W_{ij}$.

Intuitivement, si l'affinité W_{ij} entre les échantillons i et j est proche de 1, c'est à dire si ces échantillons sont similaires, $\mathbf{x}_i$ et $\mathbf{x}_j$ seront forcés à être proches dans l'espace latent. dans le cas contraire, rien ne contraint ces deux échantillons à être proches dans l'espace latent. On contraint donc les échantillons proches

dans l'espace d'origine à être également proches dans l'espace latent. La fonctionnelle de l'Équation 2.2 s'écrit également sous forme matricielle :

$$\arg\min_{\mathbf{X}} \mathrm{Tr}(\mathbf{X}^T\mathbf{L}\mathbf{X}), \tag{2.3}$$

sous la contrainte $\mathbf{X}^T\mathbf{D}\mathbf{X} = \mathbf{I}$, avec $\mathbf{L} = \mathbf{D} - \mathbf{W}$ le Laplacien de $\mathbf{W}$, où $\mathbf{D}$ est la matrice diagonale telle que $D_{ii} = \sum_j W_{ij}$.

L'algorithme des *Laplacian eigenmaps* effectue ensuite la décomposition spectrale du Laplacien associé au graphe des échantillons. Cette décomposition spectrale est effectuée en pratique en diagonalisant la matrice $\tilde{\mathbf{W}} = (\mathbf{D})^{-\frac{1}{2}}\mathbf{W}(\mathbf{D})^{-\frac{1}{2}}$, qui est symmétrique, ce qui correspond à travailler avec le Laplacien normalisé.

Pour plus de robustesse aux différences de densité dans la distribution des échantillons le long de la variété, nous utilisons la généralisation des *Laplacian eigenmaps* à travers l'algorithme des *Diffusion maps* [28], [29]). Il s'agit de normaliser en amont de la diagonalisation la matrice $\mathbf{W}$ par l'affinité totale de chaque échantillon avec les autres échantillons, c'est-à-dire utiliser $\frac{W_{ij}}{D_{ii}D_{jj}}$ au lieu de W_{ij}. Nous utilisons la version la plus simple du paramètre de temps de diffusion ($t = 1$), qui correspond à considérer que $\mathbf{W}$ encode la probabilité de passer d'un échantillon i à un échantillon j en une seule étape.

Enfin, l'espace latent est obtenu à partir des vecteurs propres associés aux plus hautes valeurs propres, après avoir enlevé le cas trivial associé à la valeur propre 0.

2.2.3 Comparaison des différentes méthodes sur CelebA

Nous proposons ici une illustration qualitative des méthodes classiques de réduction de dimension présentées précédemment. Nous comparons les espaces latents estimés avec l'ACP, Isomap, *Diffusion Maps* et GP-LVM. Ces espaces ont été appris sur un échantillon de 400 images de visages tirées aléatoirement du jeux de données CelebA [25]. Comme expliqué dans ce chapitre en Section 2.1, nous avons sélectionné des images ayant des caractéristiques similaires pour minimiser la variabilité des motifs dans ce jeu de données. La méthode qui a été choisie pour la sélection des images est non supervisée : nous avons d'abord sélectionné les images possédant des fonds blancs en nous basant sur les valeurs des bords des images, puis cette pré-sélection projetée dans un espace latent calculé par *Diffusion Maps*. Nous avons fixé une valeur de seuil et réduit l'espace de données à 4714 images, dans lequel nous avons échantillonné aléatoirement 400 sujets pour les expériences qui suivent. Les images obtenues représentent principalement des femmes aux cheveux bruns. Il y a cependant quelques exceptions dues à la méthode de sélection des images. Par exemple, la Figure 2.2 montre des individus proches au sens de la distance Euclidienne, métrique qui a été utilisée pour la sélection des données.

Pour assurer une comparaison équitable, les espaces qui suivent ont été calculés avec des paramètres similaires (même nombre de dimensions, facteurs d'échelle ou nombre de voisins égaux suivant les méthodes).

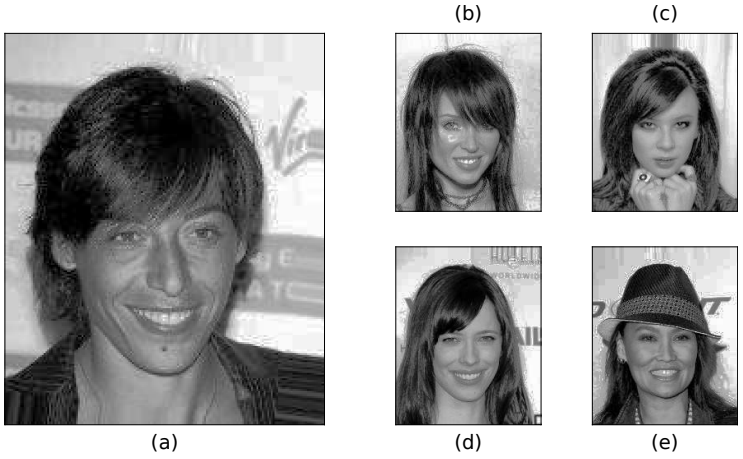


FIGURE 2.2 – (a) Image représentant un homme de la base de données CelebA qui a été sélectionnée avec notre méthode. (b-c-d-e) sont les images les plus proches au sens de la distance Euclidienne. Pour l'image (c), les mains de la femme ont une apparence pixel-à-pixel proche du le cou de l'homme, et la distance Euclidienne est insuffisante pour refléter ces différences (sauf si nous disposions d'un nombre très important d'échantillons). De la même façon pour l'image (e), le chapeau se superpose avec les cheveux de l'homme (a), ce qui n'est pas souhaitable.

Indices d'évaluation liés à CelebA : Par nature, l'apprentissage statistique non supervisé ne possède pas de méthode ou de métrique d'évaluation intrinsèque. Autrement qu'en observant directement l'espace latent, une façon d'apprécier les qualités d'un espace latent est d'établir des indices quantitatifs représentant certaines caractéristiques du jeu de données. CelebA donne accès aux indicateurs des positions (*landmarks*) de la bouche, des yeux et du nez. Ces indicateurs permettent de déterminer deux indices de positions :

1. L'orientation du visage, calculée en récupérant la position relative du nez par rapport aux positions moyennes de l'alignement des yeux et des coins de la bouche,
2. La quantité de "sourire", mesurée en calculant l'espacement horizontal entre les deux coins de la bouche.

Ces informations sont cachées dans les images mais sont importantes. On peut considérer qu'une bonne représentation latente estimée à partir de ces données doit être ordonnée au moins partiellement selon ces informations. De manière plus quantitative, certaines dimensions latentes doivent être partiellement corrélées avec le sourire et l'orientation du visage.

On cherche également à évaluer si les espaces arrivent à restituer des informations plus fines liées directement aux textures et à l'apparence des images. Ici, nous nous focalisons sur la moyenne d'intensité des pixels à l'intérieur de

la zone centrale définie par les positions des yeux et de la bouche. Cette valeur correspond au teint moyen du visage, caractéristique physique que l'on devrait retrouver dans un espace représentant une population de visages. C'est une information spécifique à la donnée image. Cela permet également de s'affranchir en grande partie de l'orientation du visage, mais surtout des teintes des cheveux et du fond de l'image, sur la mesure de distance.

La Figure 2.3 résume les opérations réalisées pour obtenir ces indices.

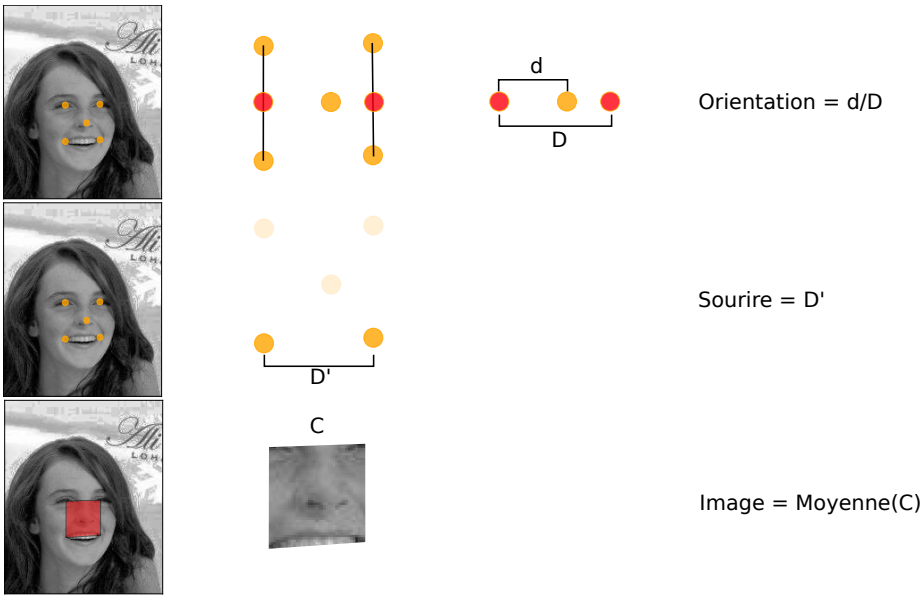


FIGURE 2.3 – Indices choisis pour évaluer l'estimation d'un espace latent à partir du jeu de données CelebA. On calcule l'orientation du visage en regardant la position du nez par rapport à la moyenne des positions des yeux et des coins de la bouche, le sourire en regardant seulement l'écartement des coins de la bouche, et on évalue le teint du visage en récupérant la moyenne de l'intensité de la zone délimitée par les yeux et la bouche.

Observations sur les espaces latents : Le nombre de dimensions choisies pour l'espace latent est important : il doit être minimal mais suffisamment élevé pour encapsuler un maximum d'information. La Figure 2.4 montre l'effet du nombre de dimensions sur la décroissance des valeurs propres pour l'algorithme *Diffusion Maps*. Pour cet exemple, le nombre de dimensions latentes idéal peut être évalué à 10, valeur correspondant à la fin de la plus forte pente de décroissance des valeurs propres. Cela correspond à 82% de la décroissance totale des valeurs propres. Il n'y a cependant pas de consensus sur la meilleure méthode pour choisir le nombre de dimensions latentes.

Nous représentons sur les Figures 2.5 2.6 les deux premières dimensions des

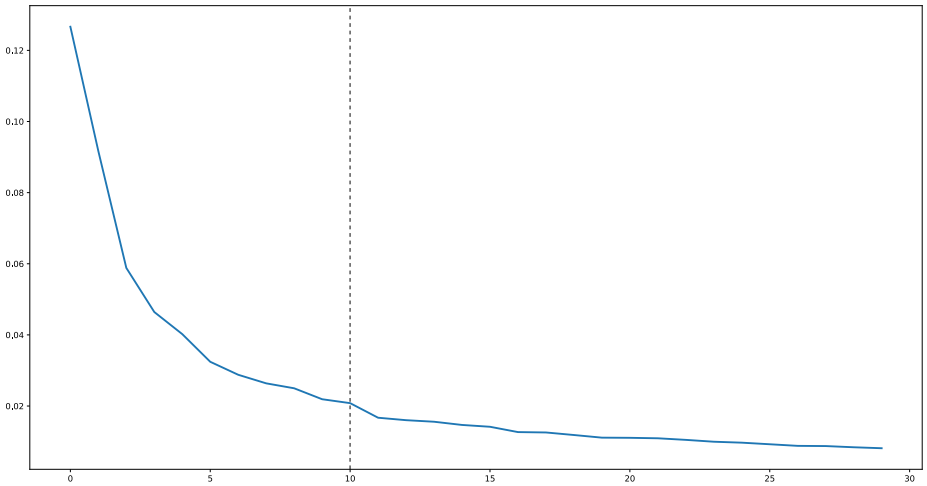


FIGURE 2.4 – Décroissance des valeurs propres pour l'algorithme *Diffusion Maps*.La décroissance est forte sur les 10 premières dimensions, et se stabilise autour de 10 dimensions latentes.

espaces latents calculés par *Diffusion Maps* et le GP-LVM. Ces espaces ont été calculés en 10 dimensions (voir Figure 2.4), sur les 400 images tirées de CelebA. Sur l'espace de la Figure 2.5 calculé avec *Diffusion Maps*, la première dimension encode assez clairement l'orientation du visage : les personnes respectivement à gauche / à droite de l'espace ont majoritairement le visage tourné vers leur gauche / leur droite, tandis que les personnes au centre de l'espace regardent droit devant elles. Des observations similaires peuvent être réalisées sur l'espace de la Figure 2.6 calculé par GP-LVM. La deuxième dimension encode quand à elle le teint. Pour l'espace GP-LVM, les visages du haut ont des teintes claires, tandis que ceux du bas ont des teintes majoritairement foncées. C'est l'inverse en ce qui concerne l'espace calculé par *Diffusion Maps* (le signe des vecteurs propres et donc l'orientation des dimensions n'étant pas déterminé). L'étalement dans l'espace latent paraît donc cohérent à première vue, au moins pour ces deux caractéristiques.

Pour avoir une idée quantitative de la propension de ces méthodes à représenter les données, nous avons calculé les corrélations entre chacun de ces espaces et les indices quantitatifs de CelebA décrits au paragraphe précédent 2.2.3. Ces corrélations sont représentées (en valeur absolue) sur la Figure 2.7, et permettent de confirmer les impressions visuelles observées sur les deux premières dimensions.

Ces espaces ont cependant quelques défauts, pointés sur les Figures 2.5 et 2.6 par les zones colorées. Deux types d'erreurs peuvent être relevées : (i) certaines images ne sont pas placées dans la zone qui encode leur particularité (par exemple un teint de visage clair au milieu de teints foncés) (ii) d'autres sont des *outliers* qui sont pourtant placés dans une zone à forte densité (un homme au

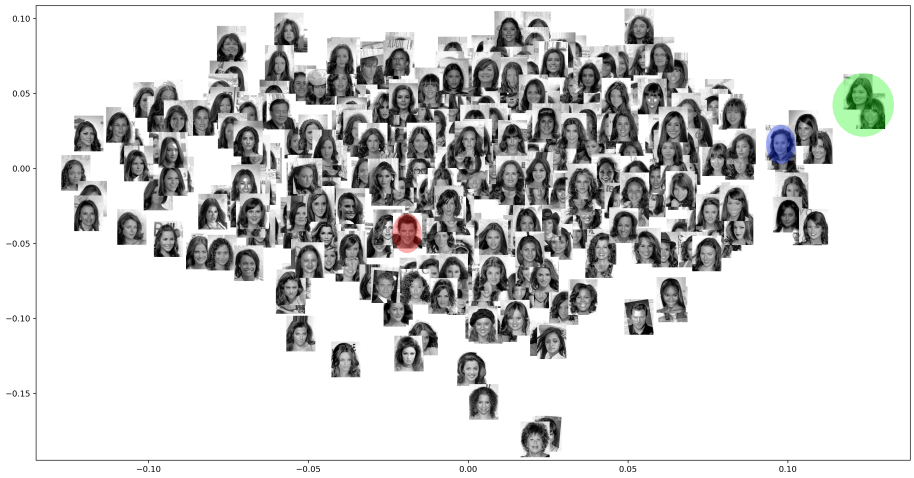


FIGURE 2.5 – Espace latent calculé par *Diffusion Maps*. La zone rouge met en évidence l'image d'un homme aux cheveux courts au milieu de l'espace, alors qu'il devrait être sur les bords. En vert, deux images de femmes sont isolées et considérées comme *outliers* sans raisons apparentes. La zone bleue représente quant à elle une personne qui regarde en face de l'objectif, alors qu'elle est dans une zone où tous ses voisins sont tournés vers leur droite.

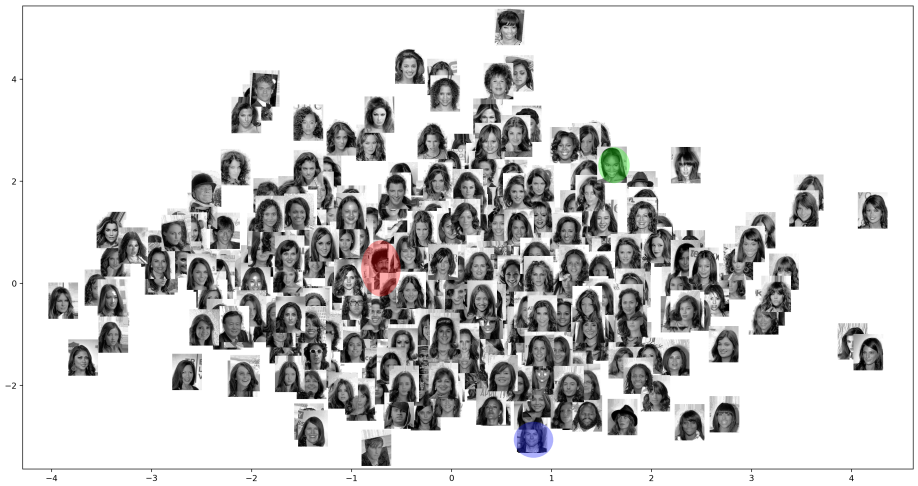


FIGURE 2.6 – Espace latent calculé par GP-LVM. Dans la zone rouge, on aperçoit un homme avec une moustache, en plein milieu de l'espace. La zone verte représente une femme tournée du mauvais côté par rapport à la disposition de l'espace, et l'homme en bleu a le teint pâle, dans une zone qui est supposée encoder les teints les plus foncés du jeu de données.

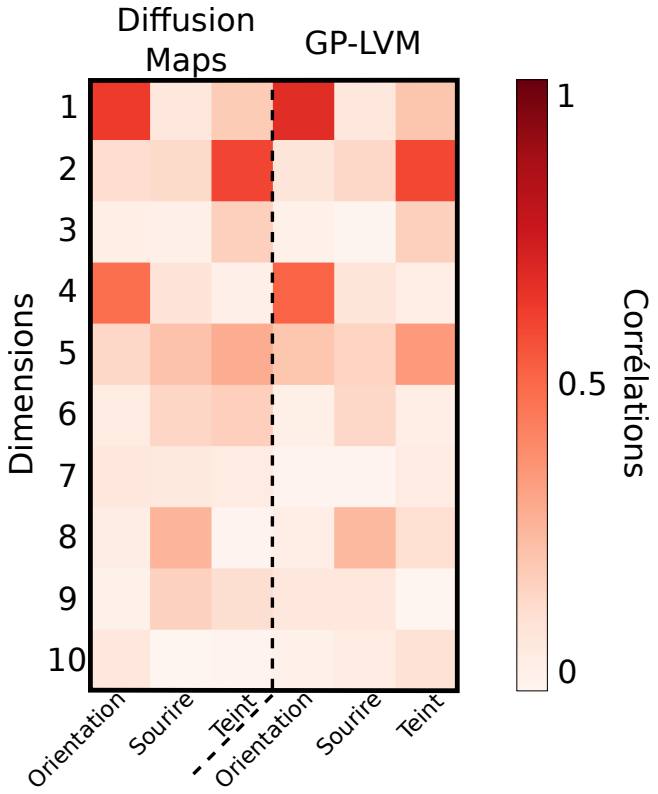


FIGURE 2.7 – Matrice des corrélations entre les espaces produits par *Diffusion Maps* et GP-LVM, et les indices quantitatifs associés à CelebA. Cela permet d'observer que les espaces en question ont des comportements similaires sur ce jeu de données, et qu'ils sont capables d'encoder des caractéristiques propres aux images. On s'aperçoit cependant que l'indice lié au sourire, qui représente une information cachée dans l'image, n'est que très faiblement corrélé aux différents espaces. Cette information n'est donc pas encodée.

milieu de femmes, une personne portant un chapeau avec des personnes n'en portant pas). Ces défauts mettent en évidence les limites de la distance Euclidienne pixel-à-pixel pour comparer les images.

Les espaces monomodaux permettent donc d'exploiter les images mais la faiblesse de la métrique limite leur analyse. Pour une meilleure exploitation des données d'images, on peut citer deux principales pistes d'amélioration :

1. Améliorer la métrique ; par exemple en utilisant des réseaux de neurones convolutionnels (voir Section 2.2.6), pouvant mieux prendre en compte la structure spatiale des pixels et donc les textures de l'image,
2. Incorporer des informations supplémentaires ; par exemple en utilisant les

positions des points d'intérêts, dans un apprentissage hiérarchique. C'est l'idée principale de cette thèse, et un exemple d'application sur CelebA sera développé en sous-Section 3.2.1 dans le Chapitre 3.

2.2.4 Méthodes plus récentes : t-SNE et UMAP

Cette sous-Section présente deux méthodes plus récentes pour la réduction de dimensions, utilisées en pratique principalement pour la visualisation des dernières couches des réseaux de neurones, ou comme une aide à la visualisation d'espaces latents. Ces méthodes sont t-SNE [30] et UMAP [31] introduites respectivement en 2008 et 2018.

t-SNE : t-SNE (*t-Distributed Stochastic Neighbor Embedding*) est une méthode probabiliste et stochastique visant à trouver un espace de faible dimension avec une fonction de coût basée sur la divergence de Kullback-Leibler (mesure de dissimilarité d'une distribution de probabilité à une autre) par rapport à une distribution de Student.

t-SNE calcule la probabilité conditionnelle $p_{i|j}$ entre les vecteurs $\{\mathbf{y}_i\}_{i=1}^N$ de l'espace de départ :

$$p_{i|j} = \frac{\exp(-(\|\mathbf{y}_i - \mathbf{y}_j\|^2/2\sigma_i^2))}{\sum_{k\neq i}\exp(-(\|\mathbf{y}_k - \mathbf{y}_j\|^2/2\sigma_k^2))}, \tag{2.4}$$

avec σ_i un facteur d'échelle, fixé par l'utilisateur, généralement dépendant de la densité de l'espace de départ (i.e. de la distance au plus proches voisin). On calcule ensuite p_{ij}, distribution de probabilité de la similarité entre les échantillons i et j, en symétrisant $p_{i|j}$:

$$p_{ij} = \frac{p_{i|j} + p_{j|i}}{2N}. \tag{2.5}$$

On définit également la probabilité q_{ij} dans l'espace latent par un noyau basé sur la distribution de Student pour le calcul des similarités au sein de l'espace latent :

$$q_{ij} = \frac{(1 + \|\mathbf{x}_i - \mathbf{x}_j\|^2)^{-1}}{\sum_{k\neq l}(1 + \|\mathbf{x}_k - \mathbf{x}_l\|^2)^{-1}} \tag{2.6}$$

t-SNE consiste à faire correspondre au mieux la distribution q (équation 2.6) de l'espace latent à la distribution p (équation 2.5) de l'espace en grande dimension. Pour cela, la fonction de coût de t-SNE est la divergence de Kullback-Leibler entre q et p :

$$D_{KL}(p, q) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}} \tag{2.7}$$

En pratique, on résout t-SNE en initialisant les variables latentes $\{\mathbf{x}_i\}_{i=1}^N$ par une distribution normale multivariée, puis on effectue une descente de gradient jusqu'à convergence.

UMAP : UMAP (*Uniform Manifold Approximation and Projection for Dimension Reduction* [31]) est une méthode plus récente, qui a été pensée pour dépasser certains problèmes spécifiques à t-SNE :

1. UMAP est généralement plus rapide que t-SNE pour les grands ensembles de données. L'algorithme utilise une approximation basée sur des graphes, qui permet de réduire considérablement le temps de calcul tout en préservant la structure globale des données. Cela le rend plus adapté à l'exploration de grandes bases de données.
2. UMAP est plus robuste que t-SNE dans le sens où de petites variations des paramètres ou de l'initialisation ont généralement moins d'impact sur les résultats finaux. Cela facilite la répétabilité des expériences et la comparaison entre différentes exécutions.
3. UMAP est capable de gérer plus facilement des données avec une densité d'échantillons hétérogène. Il utilise une méthode d'échantillonnage basée sur la densité locale qui lui permet de s'adapter aux différentes échelles présentes dans les données.
4. UMAP offre également plus de contrôle sur les paramètres et les hyperparamètres que t-SNE. Il permet de régler plus finement l'équilibre entre la préservation de la structure locale et globale.

Limites de t-SNE et UMAP : t-SNE et UMAP, bien que partageant des propriétés intéressantes et étant des méthodes parmi les plus récentes pour la réduction de dimensions dans le cadre de l'apprentissage non profond, souffrent du même problème : il n'est pas possible d'exploiter quantitativement les distances entre les points dans l'espace latent. Contrairement à t-SNE et UMAP, les méthodes classiques d'apprentissage de variétés offrent des garanties géométriques dans l'espace latent. Par exemple, dans le cas de l'algorithme Isomap, les distances dans l'espace latent correspondent aux distances géodésiques dans l'espace initial. Ces méthodes fournissent donc des mesures de distances qui ont une signification interprétable et cohérente.

2.2.5 Reconstruction

Les méthodes d'apprentissage non supervisées ont par nature un désavantage sur les méthodes supervisées : l'évaluation quantitative de leur performance est bien plus difficile en l'absence de vérité terrain. En effet, si l'on veut comparer deux méthodes de réduction de dimensions, on n'a a priori accès qu'aux données d'origine et aux espaces latent appris. Dans ce contexte, il est en général difficile de quantifier quel espace a incorporé le plus d'informations, et à partir de quelles informations issues de l'espace de départ. Par ailleurs, l'espace latent est difficilement interprétable en tant que tel par l'utilisateur, contrairement à la visualisation des données originales.

Un des leviers à ces problèmes est la reconstruction des données d'origine : en se plaçant en certains points caractéristiques de l'espace latent, on reconstruit la donnée d'entrée correspondante. La reconstruction est un moyen d'exploitation

des espaces latents : si l'on possède une bonne reconstruction, on peut par exemple trouver des motifs pathologiques associés aux différentes dimensions de l'espace latent, ou à des sous-groupes spécifiques. C'est également grâce à la reconstruction que l'on peut interpréter les statistiques non linéaires sur notre jeu de données obtenues grâce à l'espace latent. Par exemple, pour obtenir la moyenne non linéaire d'un jeu de donnée, on calcule la moyenne (par rapport à la distance Euclidienne) dans l'espace latent, et on reconstruit le point associé.

Cependant, les méthodes d'apprentissage de variétés non linéaires classiques (contrairement aux auto-encodeurs qui seront présentés dans la suite de ce chapitre) ne possèdent pas intrinsèquement de méthodes de reconstruction. Nous avons donc utilisé une méthode spécifique pour reconstruire les données à partir de l'espace latent. La méthode que nous avons choisie est une régression à noyaux et plus particulièrement sa version multi-échelles présentée par *Bermanis et al.* en 2013 [32], qui permet de prendre en compte plusieurs niveaux de résolution dans l'espace latent et ainsi d'être plus robuste à la densité non uniforme des échantillons. Nous utilisons l'implémentation décrite dans [33].

Cette reconstruction possède des défauts (par exemple, léger floutage des contours de l'infarctus ou visages) mais permet d'interpréter les principales tendances de l'espace latent du point de vue des données de haute dimension, un point clé pour l'interaction avec les cliniciens.

2.2.6 Auto-encodeurs (AE)

Certaines méthodes à base de réseaux de neurones, en particulier les auto-encodeurs, permettent également de faire de l'apprentissage de représentations [24]. Le fonctionnement de l'auto-encodeur repose sur la réalisation simultanée de deux tâches : la réduction de dimensions et la reconstruction de ses propres données d'entrées [34]. Il comporte 3 parties principales : un encodeur, qui compresse peu à peu l'information jusqu'à un goulot d'étranglement (une couche de neurones de dimension $d \ll D$), une représentation latente de dimension d estimée à partir de cette couche, et enfin un décodeur qui recompose l'information afin de retrouver les données d'origine (voir Figure 2.8).

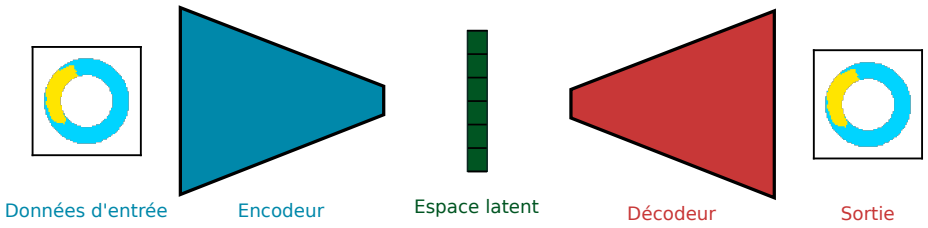


FIGURE 2.8 – Schéma de fonctionnement de l'auto-encodeur. Les données d'entrée (à gauche, de dimension D) sont progressivement encodées dans un vecteur de faible dimension (en vert, de dimension d), puis reconstruites par le décodeur. L'espace latent pour l'ensemble de la population est donc de dimension $N \times d$).

Pour ce type de méthodes, on cherche à maximiser la qualité de la reconstruction tout en diminuant le nombre de dimensions de la représentation latente. Si la reconstruction est bonne avec un nombre de dimensions faible, cela signifie que l'on compresse l'information efficacement.

Auto-encodeurs variationnels : L'auto-encodeur variationnel (*Variational Autoencodeur*, VAE) [35] est une version probabiliste de l'auto-encodeur : on fait l'hypothèse a priori que la distribution des données dans l'espace latent suit une loi normale et, au lieu d'apprendre directement la position des points dans l'espace latent, on apprend une densité de probabilité, encodée par sa moyenne et son écart-type. Les données de haute dimensions sont reconstruites en tirant un échantillon à partir de cette densité de probabilité (en général, la moyenne, correspondant à la valeur la plus probable).

Les AEs, VAEs et leur dérivés ont été largement utilisés ces dernières années pour des applications en imagerie médicale [36,37]. Ils possèdent néanmoins des caractéristiques qui les différencient de l'apprentissage de variétés usuel :

1. Tout d'abord, par leur structure, ces réseaux reconstruisent leurs données d'entrées. Cette reconstruction est essentielle pour l'interprétation de l'espace latent, notamment dans le contexte des applications médicales. Par exemple, pour estimer la moyenne non linéaire d'une population, il suffit de calculer la moyenne linéaire des positions des individus dans l'espace latent, puis de reconstruire grâce au décodeur la donnée qui serait associée à cette coordonnée moyenne. Cependant, la reconstruction des VAEs a tendance à produire des images floues, contrairement à d'autres modèles génératifs comme les GANs, ces derniers ciblant directement une meilleure qualité d'image en sortie. Les méthodes d'apprentissage de variétés, quant à elles, ne possèdent pas de reconstruction intrinsèque ; on a donc besoin dans ce cadre-là d'utiliser une méthode de reconstruction externe (voir sous-Section 2.2.5), souffrant également d'un floutage comparable à celui des VAEs.
2. Les auto-encodeurs présentent également des limites : premièrement, s'agissant de réseaux de neurones profonds, les étapes de la décomposition des images au fil des couches de neurones sont plus difficiles à interpréter. De plus, contrairement aux méthodes d'apprentissage de variétés plus classiques (e.g. ACP, Isomap, *Diffusion maps*), l'encodage de l'espace latent ne garantit des dimensions d'importance différente et que l'on pourrait ordonner. Enfin, si l'espace latent est explicitement contraint pour respecter une distribution statistique comme dans les VAEs, il n'y a pas de garantie que la variété soit correctement "dépliée", c'est-à-dire qu'il soit licite de calculer des distances Euclidiennes dans l'espace latent, et ce malgré des efforts pour mieux contrôler l'équilibre entre encodage et propriétés statistiques comme dans le cas simple des β-VAEs [38] ou de certaines méthodes de désentrelacement [39]. Dans notre cadre, on cherche à expliquer des phénomènes sous-jacents à une population et les mécanismes complexes liés à la pathologie, il est donc nécessaire d'obtenir des espaces

latents dans lesquels ce type d'analyses statistiques est réalisable.

3. Enfin, une autre difficulté liée à l'apprentissage de réseaux profonds est la quantité de données nécessaire à une bonne généralisation par les réseaux. Dans de nombreuses applications cliniques, on possède moins de patients que de variables à exploiter, ce qui rend les réseaux de neurones (et de fait les auto-encodeurs) moins adaptés à la tâche. Cette limite est également valable pour l'apprentissage de variétés décrit en sous-Section 2.2.2.

2.2.7 Transformeurs :

Plus récemment, un ensemble de méthodes d'apprentissage profond basées sur des mécanismes d'auto-attention a vu le jour, les transformeurs, proposés en 2017 par *Vaswani et al.* [40]. Ce type de réseaux est inspiré de méthodes de traitement automatique du langage naturel (*Natural language Processing*, NLP) et permet également de faire de la réduction de dimensions, ainsi que du mélange de données. L'architecture ViT [41] a ensuite été développée pour adapter les transformeurs au traitement d'images. Son principe général est schématisé sur la Figure 2.9 : l'image d'entrée est tout d'abord découpée en patchs de plus petite taille, qui sont transformés en éléments, ou *tokens*, par une projection linéaire. Ces *tokens* sont ensuite encodés par le transformeur ; enfin, un perceptron multicouches les utilise pour la tâche visée (par exemple, de la classification). L'encodeur est composé de L couches d'encodage, dont la composition est illustrée en Figure 2.10 pour le transformeur ViT.

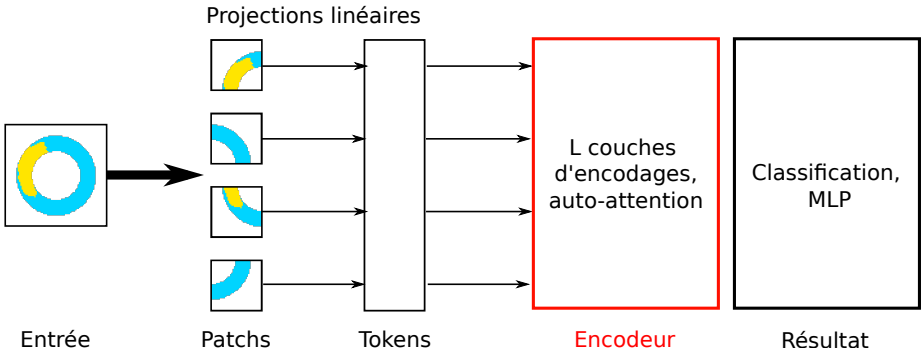


FIGURE 2.9 – Représentation schématique du transformeur ViT. Les données sont segmentées en patchs, puis projetés pour former des *tokens*, qui sont ensuite encodés par un module basé sur de l'auto-attention, puis utilisés pour la tâche visée (par exemple, de la classification).

La spécificité des transformeurs réside dans le module d'auto-attention, représenté sur la Figure 2.11. Pour chaque élément en entrée, le transformeur génère trois matrices : une requête (*query*), une clé (*key*) et une valeur (*value*). Ces matrices sont obtenus en multipliant la séquence d'entrée par trois matrices de poids apprises lors de l'entraînement du modèle. L'étape suivante consiste à

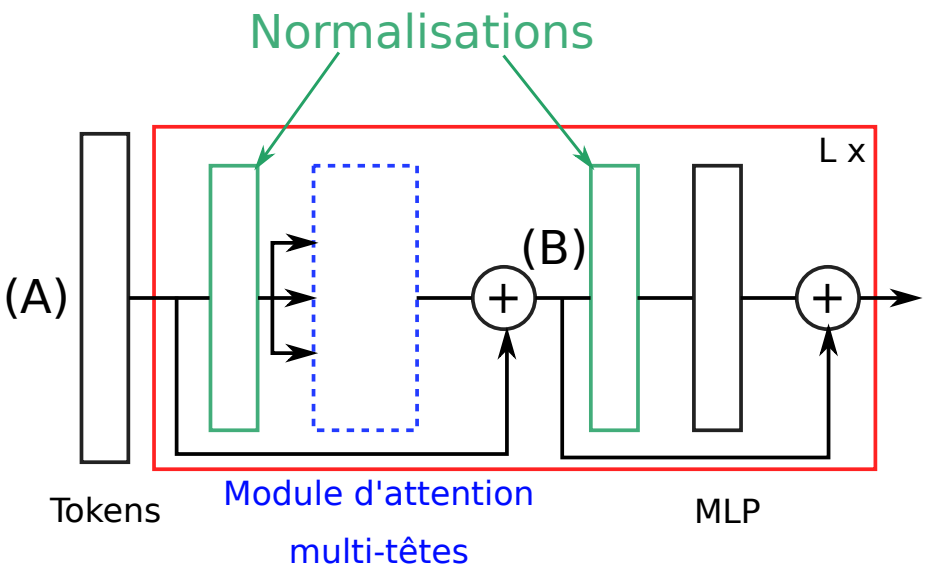


FIGURE 2.10 – Représentation schématique d'une couche d'encodage du transformeur : les *tokens* (A) sont normalisés, puis passent dans le module d'attention (voir Figure 2.11). Ce résultat est ensuite ajouté avec les *tokens* d'entrée, puis passe par une nouvelle couche de normalisation, puis par un perceptron multi-couches (qui introduit de la non linéarité dans le système), avant d'être ajouté avec le résultat (B). Cette suite d'opération est effectuée L fois, L correspondant au nombre de couches choisies par l'utilisateur.

calculer un score d'attention pour chaque paire d'éléments (i,j) dans la séquence, en utilisant la similarité entre la requête de l'élément i et la clé de l'élément j (impliquant le produit scalaire entre ces deux vecteurs). Les poids d'attention sont ensuite normalisés puis utilisés pour pondérer les valeurs correspondantes de chaque élément de la séquence. Ces valeurs pondérées sont sommées pour obtenir une représentation agrégée, qui capture les informations importantes de la séquence en tenant compte des relations entre les éléments. Cette opération est réalisée dans plusieurs modules différents, appelés modules (*head*), dont les résultats sont ajoutés puis projetés linéairement dans un espace de faible dimension.

Ce qui rend les transformeurs intéressants pour le traitement d'images médicales est leur flexibilité et leur capacité à dépasser le contexte local de l'image grâce à l'opération d'auto-attention qui permet de lier des pixels éloignés. Ce sont des outils puissants qui permettent de combiner des informations et de faire de la fusion de données. Cependant, ces algorithmes sont gourmands en ressources, et lorsqu'on les utilise avec en entrée différents types de données, ils souffrent des mêmes limitations que les autres méthodes de fusion de données présentées dans ce Chapitre : ils fusionnent tous les descripteurs en même temps, et ne tiennent donc pas compte des éventuels liens entre les données. De plus, ils

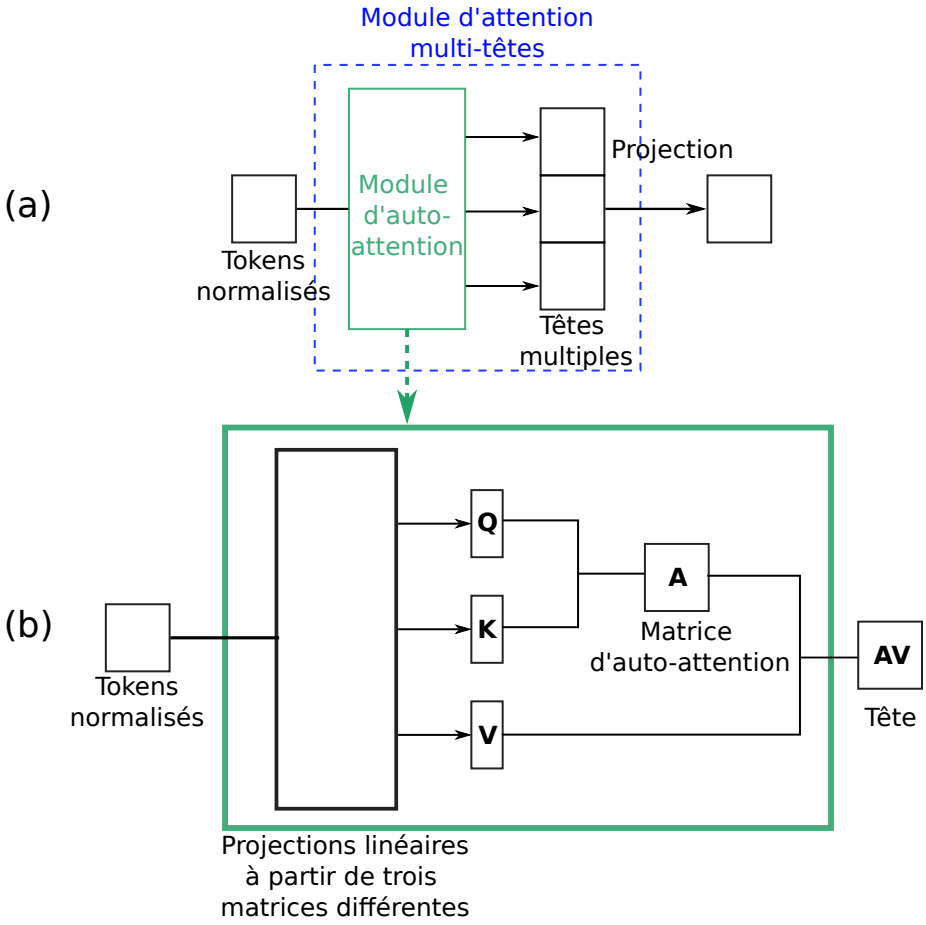


FIGURE 2.11 – Représentation schématique du module d'attention multi-têtes du transformeur (a), et en particulier du module d'auto-attention (b). Les matrices $\mathbf{Q} \in \mathbb{R}^{N+1 \times D_h}$ *Query* et $\mathbf{K} \in \mathbb{R}^{N+1 \times D_h}$ *Key* sont combinés pour former la matrice d'auto-attention $\mathbf{A} \in \mathbb{R}^{N+1 \times N+1}$, qui est multipliée par la *Value* $\mathbf{V} \in \mathbb{R}^{N+1 \times D_h}$ pour former une tête. $D_h = D/k$ avec k le nombre de têtes.

ne produisent pas d'espace latent, ce qui limite les possibilités d'interprétations.

2.2.8 Processus Gaussiens à variables latentes :

Les modèles de processus Gaussiens à variables latentes [42] constituent une approche probabiliste pour la réduction de dimensions, basée sur un modèle de graphe reliant les observations (les données en haute dimension) et les variables latentes. Ils permettent de capturer les relations sous-jacentes dans les données tout en fournissant une représentation compacte et explicative, ainsi qu'une estimation robuste des incertitudes associées aux prédictions. Ces modèles offrent

de plus un cadre flexible pour représenter les interactions entre descripteurs. Ce type de méthodes sera utilisé pour une des contributions principales de cette thèse (Chapitre 6), et sera décrit et testé plus en détail dans le Chapitre 3.

2.3 Apprentissage de représentation multi-descripteurs

Toutes les méthodes décrites dans la Section précédente ont d'abord été conçues pour l'exploitation d'un unique descripteur, entraînant souvent une caractérisation incomplète du système étudié. Les approches multi-descripteurs sont essentielles pour la pratique clinique (dont elles sont partie intégrante, contrairement à leur version computationnelle), car elles amènent à la fois une caractérisation plus précise et plus robuste des pathologies. Dans le cas de l'application que nous développons dans les Chapitres 5 et 6, nous considérons plusieurs types de données d'IRM de réhaussement : les segmentations de l'infarctus (sur des images de LGE), les segmentations des régions d'obstruction microvasculaire / MVO (sur des images de EGE et LGE), et les images en nuances de gris (également EGE et LGE). L'objectif de notre étude est de combiner les informations provenant de ces différents descripteurs pour obtenir une caractérisation plus riche de l'état du cœur des patients.

Pour l'analyse computationelle et notamment pour l'apprentissage de représentations, les difficultés liées à l'analyse conjointe de plusieurs descripteurs extraits des images sont nombreuses :

— Comme évoqué dans l'introduction, ces données peuvent être de types hétérogènes, ce qui rend impossible une concaténation directe en entrée,

— Elles peuvent avoir des dimensionalités différentes donc des influences sur l'espace latent distinctes,

— Elles peuvent également présenter des corrélations partielles à prendre en compte pour limiter la redondance d'information dans l'espace latent.

Toujours dans l'exemple des données étudiées dans les Chapitres 5 et 6 de cette thèse, l'information de texture présente dans les images en niveaux de gris est totalement perdue quand on exploite directement les segmentations, alors qu'elle est potentiellement porteuse d'informations (pour lever les ambiguïtés sur la méthode optimale de segmentation de la lésion [43], [44], mais aussi pour prendre en compte la "zone grise" liée à la récupération du myocarde et au pronostic du patient [45]). Cependant, utiliser seulement les images en niveaux de gris peut s'avérer délicat, car l'information est difficile à exploiter (le MVO a des niveaux de gris parfois similaires au myocarde sain et peut donc affecter une méthode basée sur une métrique simpliste, tout comme la présence d'artéfacts IRM) [46]. L'exploitation conjointe de ces données est susceptible d'amener plus de robustesse dans l'analyse des niveaux de gris ou de petites lésions comme le MVO. Néanmoins, une intégration progressive des données peut être plus intéressante que leur mélange simultané au même niveau, comme nous le développerons plus spécifiquement dans les Chapitres 5 et 6.

2.3.1 Mélanges naïfs : agrégation précoce ou tardive de données

Cette première sous-partie présente deux concepts pour une exploitation conjointe de plusieurs descripteurs : l'agrégation tardive ou précoce des données d'entrée. Soit $\mathbf{Y}^{(k)} \in \mathbb{R}^{N \times A}$ et $\mathbf{Y}^{(l)} \in \mathbb{R}^{N \times B}$ deux descripteurs de dimensionnalités A et B respectivement, connus pour N échantillons, et Φ une fonction de mélange définie de la façon suivante :

$$\Phi(\mathbf{Y}^{(k)}, \mathbf{Y}^{(l)}) :\rightarrow \mathbf{Y}^{(m)} \in \mathbb{R}^{N \times C}, \tag{2.8}$$

avec $C \leq A + B$. L'exemple le plus basique de fonction pour Φ est la **concaténation** ; dans ce cas, $C = A + B$.

Définissons également Ψ une fonction de réduction de dimensions :

$$\Psi(\mathbf{Y}) :\rightarrow \mathbf{X} \in \mathbb{R}^{N \times d}, \tag{2.9}$$

avec $d \ll D$, $\mathbf{Y} \in \mathbb{R}^{N \times D}$.

Étant donné ce contexte, l'opération de fusion précoce consiste à mélanger les données (avec la fonction Φ, puis à réduire la dimension (avec la fonction Ψ), tandis que l'opération de fusion tardive fait l'inverse, comme le montre la Figure 2.12.

Ces deux stratégies permettent d'exploiter toutes les modalités, et constituent une base sur laquelle on peut construire des stratégies d'intégration de données multiples. Néanmoins, elles ne sont pas satisfaisantes en l'état car elles prennent insuffisamment en compte l'hétérogénéité des données, et n'explicitent pas les interactions entre descripteurs. L'apprentissage de représentations fournit des méthodes intéressantes pour aller au-delà de ces limitations : les stratégies d'alignement et de fusion, développées dans les sous-Sections suivantes.

2.3.2 Alignement de variétés

L'alignement consiste à calculer un espace latent par descripteur, tout en contraignant les espaces estimés à être alignés les uns par rapport aux autres.

2.3.2.1 Méthodes linéaires

La méthode des moindres carrés partiels (*Partial Least Square*, PLS) [47,48], est une méthode linéaire qui aligne deux descripteurs $\mathbf{X}^{(1)}$ et $\mathbf{X}^{(2)}$ en maximisant leur covariance. Comme dans le cadre de *Yan et al.* [21], cette méthode consiste à calculer les vecteurs propres de la matrice de covariance.

L'analyse de corrélations canoniques (*Canonical Correlation Analysis*, CCA) [49] est également une méthode linéaire alignant deux descripteurs en maximisant leur corrélation. Pour se faire, la CCA calcule cette fois les vecteurs propres associés à un produit des matrices de corrélation de $\mathbf{X}^{(1)}$, $\mathbf{X}^{(2)}$ et des deux descripteurs entre eux.

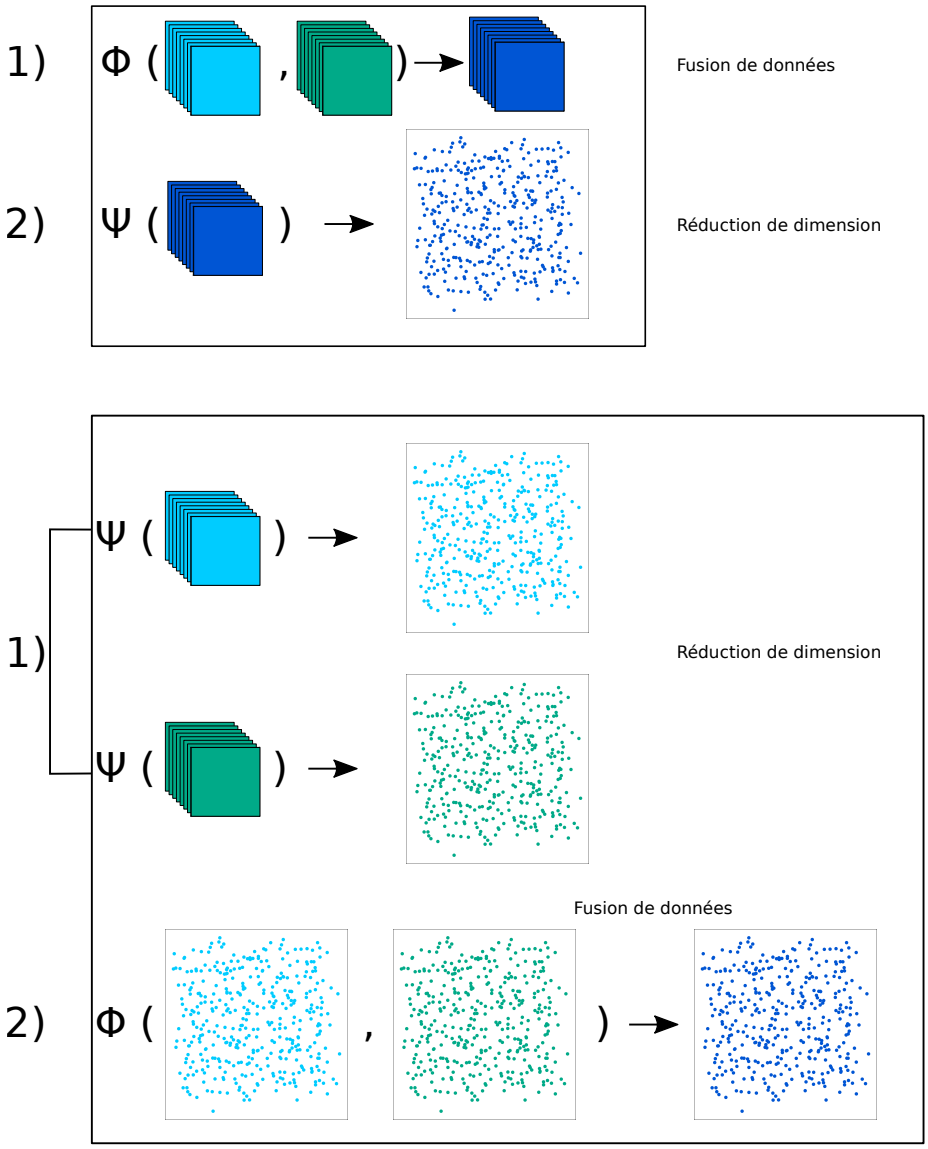


FIGURE 2.12 – Illustration des stratégies de fusion de données précoces (en haut) et tardives (en bas) décrites en Section 2.3.1. La fusion précoce mélange les données avant l'algorithme de réduction de dimension, la fusion tardive réduit les dimensions séparément, puis fusionne les résultats.

2.3.2.2 Méthodes non linéaires

Alignement paire-à-paire : *Ham et al.* introduisent en 2005 [50] l'alignement de variétés non linéaire en forçant les correspondances paire-à-paire. Cette

méthode dérive du formalisme des *Laplacian eigenmaps* pour le calcul des espaces latents correspondant à chaque modalité, et rajoute dans la fonction de coût un terme forçant la proximité entre les espaces latents. Notons respectivement $\mathbf{L}^{(1)}$ et $\mathbf{L}^{(2)}$ les graphes Laplaciens de $\mathbf{Y}^{(1)}$ et de $\mathbf{Y}^{(2)}$. De façon similaire à l'apprentissage de variété monomodal décrit en sous-Section 2.2.2 (voir Équation 2.2), la méthode consiste à minimiser :

$$C(\mathbf{X}^{(1)}, \mathbf{X}^{(2)}) = \sum_{i,j} \|\mathbf{x}_i^{(1)} - \mathbf{x}_j^{(1)}\|^2 W_{ij}^{(1)} + \sum_{i,j} \|\mathbf{x}_i^{(2)} - \mathbf{x}_j^{(2)}\|^2 W_{ij}^{(2)} + \mu \sum_i \|\mathbf{x}_i^{(1)} - \mathbf{x}_i^{(2)}\|^2 \tag{2.10}$$

Le premier terme force les variables latentes $\mathbf{X}^{(1)}$ et $\mathbf{X}^{(2)}$ à être proches paire-à-paire, et les deux termes suivants correspondent à l'apprentissage de variétés unimodal via *Laplacian eigenmaps* sur $\mathbf{Y}^{(1)}$ et $\mathbf{Y}^{(2)}$. Pour pallier les éventuels problèmes d'échelle, on minimise en pratique le quotient de Rayleigh suivant :

$$\tilde{C}(\mathbf{X}^{(1)}, \mathbf{X}^{(2)}) := \frac{C(\mathbf{X}^{(1)}, \mathbf{X}^{(2)})}{\mathbf{X}^{(1)T}\mathbf{X}^{(1)} + \mathbf{X}^{(2)T}\mathbf{X}^{(2)}} \tag{2.11}$$

La minimisation de Eq. (2.11) admet une résolution matricielle en calculant les vecteurs propres de $\mathbf{L}^z$:

$$\mathbf{L}^z = \begin{bmatrix} \mathbf{L}^{(1)} + \mathbf{U} & -\mathbf{U} \\ -\mathbf{U} & \mathbf{L}^{(2)} + \mathbf{U} \end{bmatrix} \tag{2.12}$$

avec $\mathbf{U}$ la matrice diagonale telle que $U_{ii} = \mu \quad \forall i \in [1, N]$, avec μ un paramètre pondérant la force de l'alignement.

Généralisation : L'apprentissage de variétés multiples (*Multiple Manifold Learning*, MML) généralise la formulation de *Ham et al.* Cette variante propose un alignement par voisinages plutôt qu'un alignement paire-à-paire et donne ainsi plus de souplesse dans l'alignement. Cette méthode permet également d'être plus robuste aux valeurs extrêmes ou *outliers* (données en dehors de la distribution, celles-ci pouvant fortement influencer les premières dimensions estimées par les méthodes dérivant des *Laplacian eigenmaps*).

L'alignement relaxé peut s'écrire sous forme matricielle de la façon suivante :

$$\arg\min_{\mathbf{X}} \operatorname{Tr} \mathbf{X}^T \mathbf{L}^z \mathbf{X} \tag{2.13}$$

avec :

$$\operatorname{Tr} \mathbf{X}^T \mathbf{L}^z \mathbf{X} = \sum_{i,j} \|\mathbf{x}_i^{(1)} - \mathbf{x}_j^{(1)}\|^2 W_{ij}^{(1)} + \sum_{i,j} \|\mathbf{x}_i^{(2)} - \mathbf{x}_j^{(2)}\|^2 W_{ij}^{(2)} + \mu \sum_{i,j} \|\mathbf{x}_i^{(1)} - \mathbf{x}_j^{(2)}\|^2 M_{ij} \tag{2.14}$$

avec $\mathbf{M} = [M_{ij}]$ une mesure de similarité entre l'individu i et l'individu j.

Valencia-Aguirre et al. en 2011 [51] proposent la mesure de similarité suivante pour l'alignement de variétés multiples :

$$M_{ij} = \frac{\langle \mathbf{w}_i^{(1)}, \mathbf{w}_j^{(2)} \rangle}{\|\mathbf{w}_i^{(1)}\| \|\mathbf{w}_j^{(2)}\|} \tag{2.15}$$

avec $\mathbf{w}_i^{(m)}$ le vecteur de la matrice d'affinité $\mathbf{W}^{(m)}$ du descripteur m, associé à l'individu i. Cette mesure de similarité est intéressante, elle permet de comparer facilement les voisinages inter-descripteurs. *Clough et al.* proposent en 2020 [52] une autre formulation plus complète de la mesure de similarité inter-voisinages, inspirée de l'appariement de surfaces [53].

L'alignement par apprentissage de variétés multiples a notamment été utilisé par *Valencia-Aguirre et al.* [54] pour l'analyse d'images d'objets pris en photo selon différentes orientations. *Clough et al.* [52] l'ont exploité pour la reconstruction d'images d'Imagerie par résonance magnétique (IRM) de haute qualité. Cette méthode a également été utilisée dans notre équipe pour la quantification du lien entre forme et déformation cardiaque du ventricule droit sur des individus sains et pathologiques [55].

2.3.2.3 Méthodes d'apprentissage profond

CLIP : CLIP (*Contrastive Language-Image Pre-training*) [56] utilise le principe d'attention dans le cadre de réseaux de neurones pour aligner les descripteurs en entrée. Dans le cas de l'article original décrivant CLIP, les descripteurs sont d'un côté une image, de l'autre un texte décrivant le contenu de cette image. L'objectif de ce réseau est de prédire la similarité entre le texte et l'image. Le modèle est composé d'un encodeur d'images et d'un encodeur de textes, produisant chacun un vecteur de représentation. Ces vecteurs sont comparés au sein d'un module d'attention croisé. Il s'agit ici d'un alignement paire à paire où une image correspond à un texte, c'est donc un alignement fort (comme pour la méthode de *Ham et al.* [50]). CLIP est ensuite utilisé pour prédire un texte descriptif des images.

Cette méthode a par exemple été adaptée pour l'estimation d'incertitudes dans le cadre de la segmentation d'images cardiaques, CRISP (*ContRastive Image Segmentation for uncertainty Prediction*) [57]. Dans ce contexte, les auteurs ont fait correspondre le vecteur latent de l'image à segmenter à celui obtenu en apprenant un auto-encodeur variationnel sur les segmentations. Pour une paire image-segmentation donnée, si la représentation de la segmentation obtenue via l'auto-encodeur est proche du vecteur encodé pour l'image, on peut dire que la prédiction est sûre ; sinon, l'incertitude sur la segmentation est élevée.

Réseaux de neurones siamois : Les réseaux de neurones siamois [58] sont une autre méthode qui peut être utilisée pour l'alignement de données. Le principe clé des réseaux de neurones siamois est de quantifier la similarité entre les deux membres d'une paire d'échantillons. Pour cela, les deux échantillons passent dans des encodeurs jumeaux qui permettent de calculer une représentation vectorielle pour chaque échantillon. Le réseau calcule ensuite un score de similarité à partir des représentations produites.

Ces réseaux peuvent par exemple être utilisés pour faire de la recherche d'images similaires au sein d'une base de données non labellisée [59]. Par extension, ils peuvent être également utilisés pour quantifier la similarité entre deux échantillons provenant de modalités différentes. L'idée principale est de créer des branches jumelles pour chaque modalité. Chaque branche est responsable de traiter les données spécifiques à sa modalité, tout en partageant les mêmes architectures mais pas forcément les mêmes poids [60]. Cela permet d'apprendre une représentation vectorielle par modalité, tout en forçant les vecteurs des paires d'échantillons à être proches. Comme pour CLIP, cet alignement est un alignement fort, car basé sur une similarité paire-à-paire.

Dans le domaine médical, les réseaux de neurones siamois ont par exemple été utilisé pour l'établissement d'un spectre continu de sévérité pour différentes maladies (arthrose du genou, rétinopathie du prématuré) [61].

Auto-encodeur variationnel à canaux multiples (MCVAE) : L'auto-encodeur variationnel à canaux multiples (*Multi-Channel Variational AutoEncodeur*, MCVAE) [62] est une extension du VAE pour la prise en charge de plusieurs modalités. Même si elle est principalement exploitée dans un cadre de fusion de descripteurs, elle peut être assimilée à une stratégie d'alignement. Elle associe à chaque canal (ou descripteur) un encodeur et un décodeur spécifiques comme montré sur la Figure 2.13. Les données provenant de chaque canal sont projetées dans l'espace latent, puis les coordonnées sont moyennées pour ne former qu'un seul point à partir duquel on reconstruit les différents canaux.

La reconstruction étant apprise à partir des coordonnées d'un point dans l'espace latent, cette méthode permet la représentation d'échantillons ayant des canaux manquants. Cette propriété est intéressante car elle permet d'estimer une observation probable pour une modalité dont l'observation est inexistante sur un échantillon.

Par rapport au VAE unimodal, la présence de multiples canaux d'entrée et de sortie nécessite de minimiser une nouvelle fonction de coût :

$$\mathbb{E}_c[D_{KL}(q(\mathbf{X}|\mathbf{Y}^{(c)})\|p(\mathbf{X}|\mathbf{Y}^{(1)}\cdots\mathbf{Y}^{(m)}))] \tag{2.16}$$

Ici, $\mathbb{E}_c[Q]$ est la moyenne de la quantité Q sur les canaux c. La fonction de coût du MCVAE revient à sommer les divergences de Kullback-Leibler (D_{KL}) entre une distribution a priori p et la distribution obtenue q associée à chaque canal. Cette méthode est également équipée d'un schéma appelé *variational dropout* [63], permettant d'avoir une représentation latente éparse et de choisir ainsi le nombre de dimensions dans l'espace latent. Le MCVAE a été adapté à des données longitudinales [64] à l'aide d'un réseau de neurone récurent variationnel pour le calcul d'un espace latent par instant de la séquence temporelle d'entrée.

2.3.3 Fusion de données

La fusion de données, dans le cadre de l'apprentissage de représentations, consiste à estimer une unique représentation latente à partir de plusieurs des-

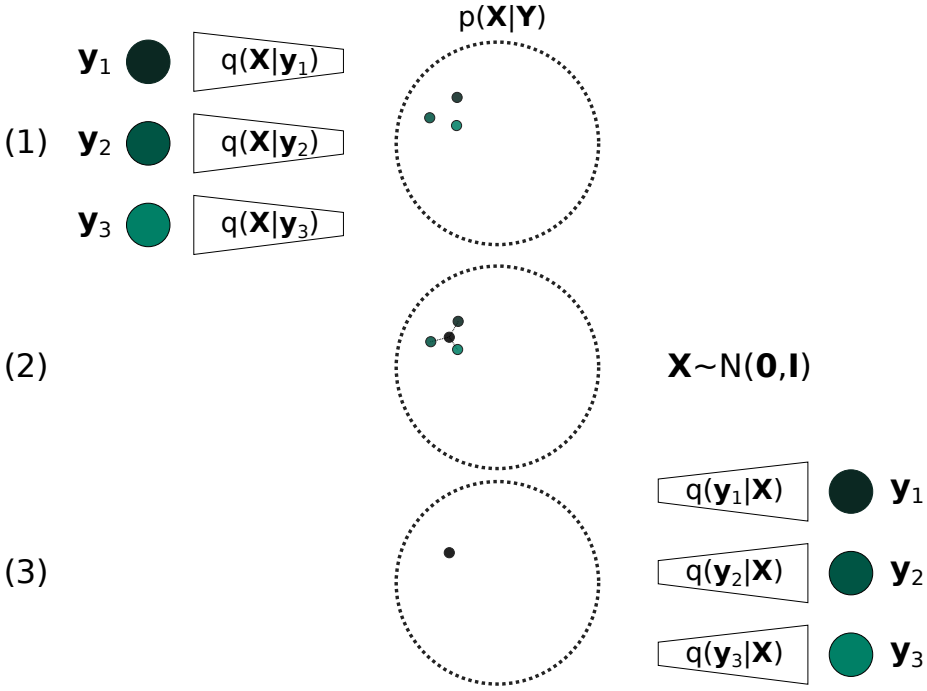


FIGURE 2.13 – Représentation schématique de l'auto encodeur variationnel à canaux multiples. (1) Les observations $\mathbf{Y}$ sont encodées dans l'espace latent $\mathbf{X} \sim \mathcal{N}(0, \mathbf{I})$. Comme pour le VAE (cf. Section 2.2.6), on apprend en pratique pour chaque canal un vecteur moyen μ et un écart-type $\mathbf{\Sigma}$, et on projette dans l'espace latent en échantillonnant à partir de $\mathcal{N}(\mu, \mathbf{\Sigma})$. (2) On moyenne les projections obtenues indépendamment par chaque canal. (3) On reconstruit les différents descripteurs à partir de cette projection.

cripteurs. Cette sous-partie s'attache à décrire deux des principales méthodes de fusion : l'apprentissage à noyaux multiples (*Multiple Kernel Learning*, MKL) et la fusion de réseaux de similarités (*Similarity Network fusion*, SNF), toutes deux assez proches du formalisme unifié de l'apprentissage de variétés [21] décrit en sous-Section 2.2.2. En effet, pour ces deux méthodes, le but est de trouver une unique matrice (ou un graphe) d'affinité à partir de la fusion des descripteurs, qu'on diagonalisera ensuite pour obtenir les directions principales du jeu de données.

Apprentissage à noyaux multiples (MKL) : L'algorithme de *Multiple Kernel Learning* (MKL) [65] [66] consiste à créer une unique matrice d'affinité $\mathbf{W}$ à partir d'une combinaison linéaire des noyaux Gaussiens associés à chacun des M descripteurs. Dans ce modèle, on optimise le changement de base (i.e. le calcul de l'espace latent) et les poids associés à chaque descripteur.

Soit $\{\mathbf{K}^{(m)}\}_{m=1}^{M}$ l'ensemble des M noyaux associées aux M descripteurs

d'entrée, on définit ainsi $\mathbf{W}$ par :

$$\mathbf{W} = \sum_{m=1}^{M} \beta^{(m)} \mathbf{K}^{(m)}, \qquad \text{avec } \beta^{(m)} \geq 0, \tag{2.17}$$

Typiquement, $K_{i,j}^{(m)} = \exp\left(\frac{-\|\mathbf{x}_i^{(m)} - \mathbf{x}_j^{(m)}\|^2)}{\sigma_m^2}\right)$, avec σ_m^2 une constante représentant la taille du noyau.

Notons $\mathbb{K}_i$ le noyau multiple associé à l'individu i :

$$\mathbb{K}_i = \begin{bmatrix} K_{1,i}^{(1)} & \cdots & K_{1,i}^{(M)} \\ \vdots & \ddots & \vdots \\ K_{N,i}^{(1)} & \cdots & K_{N,i}^{(M)} \end{bmatrix} \in \mathbb{R}^{N \times M}, \tag{2.18}$$

L'algorithme *MKL* consiste à trouver la matrice de projection $\mathbf{A} \in \mathbb{R}^{N \times d}$ et le vecteur de poids $\beta \in \mathbb{R}^M$ tels que :

$$\min_{\mathbf{A},\beta} \sum_{i,j=1}^{N} \|\mathbf{A}^T \mathbb{K}_i \beta - \mathbf{A}^T \mathbb{K}_j \beta\|^2 w_{ij} \tag{2.19}$$

avec $\beta = [\beta^{(1)}, \cdots, \beta^{(M)}]^T$. Le terme $\mathbf{A}^T \mathbb{K}_i \beta \in \mathbb{R}^d$ est la projection de l'individu i dans l'espace latent, amenant aux coordonnées $\mathbf{x}_i$.

Ce problème se résout en apprenant de façon alternée les paramètres $\mathbf{A}$ et β, et en itérant ce processus jusqu'à convergence.

Cette méthode a été largement utilisée pour l'analyse de données médicales. *Sanchez-Martinez et al.* [67,68] l'ont notamment utilisée pour la caractérisation du continuum allant de sujets sains à malades dans le cadre de l'insuffisance cardiaque à fraction d'éjection préservée. *Cikes et al.* [69] ont également utilisé *MKL* pour le phénotypage chez les patients suivants une thérapie de resynchronisation cardiaque. *Nogueira et al.* [70] l'ont quand à eux utilisé pour l'analyse de séquences temporelles issues d'un protocole de stress en échocardiographie. *Loncaric et al.* [71] ont aussi utilisé *MKL* en échocardiographie pour le phénotypage de patients avec hypertension artérielle. En bioinformatique, cette stratégie a également servi pour l'analyse de données de génomique [72].

Fusion de réseaux de similarité (SNF) : L'algorithme de fusion de réseaux de similarité (SNF) [73], [74] propose une fusion plus avancée des matrices d'affinité liées aux différents descripteurs, en les mélangeant itérativement selon un processus de diffusion dans les graphes de données.

Comme pour *Diffusion maps*, on définit la matrice d'affinité $\mathbf{W}$ par :

$$W_{ij} = \exp\left(-\frac{d(\mathbf{y}_i, \mathbf{y}_j)}{\mu}\right). \tag{2.20}$$

Pour pallier les problèmes d'échelle, on définit $\mathbf{P}$, version normalisée de $\mathbf{W}$, par :

$$P_{ij} = \begin{cases} \frac{W_{ij}}{2\sum_{j \neq i} W_{ij}} & , \text{si } j \in \mathcal{N}_i, \\ 1/2 & , \text{si } j = i. \end{cases} \tag{2.21}$$

On définit également la matrice $\mathbf{S}$, affinité locale telle que :

$$S_{ij} = \begin{cases} \frac{W_{ij}}{\sum_{k \in \mathcal{N}_i} W_{ik}} & , \text{si j} \in \mathcal{N}_i, \\ 0 & , \text{sinon.} \end{cases} \tag{2.22}$$

On effectue ensuite un processus de fusion itératif entre la matrice $\mathbf{S}$ liée à un descripteur et les matrices $\mathbf{P}$ liées aux autres descripteurs. Ainsi, étant donné une modalité m, l'objectif de cette opération est d'incorporer peu à peu les informations d'affinités locales des autres modalités (encodées dans les matrices $\mathbf{S}$) aux informations de la matrice d'affinité normalisée $\mathbf{P}^{(m)}$.

Cette opération, pour le descripteur m (avec au total M descripteurs) est la suivante :

$$\mathbf{P}^{(m)}_{t+1} = \mathbf{S}^{(m)} \frac{\sum_{l \neq m} \mathbf{P}^{(l)}_t}{M-1} \mathbf{S}^{(m)T} \tag{2.23}$$

Comme les matrices $\mathbf{S}^{(m)}$ sont parcimonieuses, l'information qui est modifiée entre chaque itération est uniquement liée aux voisinages des points sur la modalité concernée.

Après T itérations, on définit le graphe de similarité fusionné des M descripteurs par la moyenne linéaire des graphes provenant des différentes modalités :

$$\mathbf{P}^{(c)} = \frac{\sum_1^M \mathbf{P}^{(m)}_T}{M} \tag{2.24}$$

Cette méthode a été utilisée par les auteurs pour combiner des données de génomiques pour 5 types de cancers [73]. Ils ont également utilisé *SNF* sur plusieurs bases de données d'images (MNIST, Caltech 101). Mes expériences avec *SNF* sur les données d'imagerie médicale ont été moins concluantes, je n'ai pas observé un mélange de données pertinent, principalement car le processus de diffusion tel que formulé dans les articles correspondants diluait trop les informations.

2.3.4 Conclusions sur l'alignement de variétés et les algorithmes de fusion

Nous avons décrit dans ce chapitre deux paradigmes intéressants pour le mélange de données : l'alignement de variétés (Section 2.3.2) et la fusion de données (Section 2.3.3).

L'alignement de variétés cherche à estimer une représentation par descripteur, alors que la fusion fait l'hypothèse forte qu'une seule représentation permet de représenter les différentes données, ce qui n'est pas toujours possible. Une performance moindre de reconstruction entre différents descripteurs a d'ailleurs déjà été rapportée [64].

Cependant, dans le cadre de l'alignement de variétés, la visualisation de plusieurs espaces latents pour un même patient (qui plus est en plus de trois dimensions) est difficile pour l'humain. La multiplication des espaces latents rend ainsi moins intuitive la compréhension des données. Les approches de fusion de

données peuvent ainsi s'avérer plus faciles à appréhender pour les interprétations liées à l'application.

Néanmoins, ces deux types de méthodologies souffrent de la même limitation : elles considèrent les différents descripteurs simultanément, ce qui peut s'avérer critique en termes de capacités de calcul, ou plus simplement de fusions d'information. En effet, les ensembles de données sont généralement régis par une structure hiérarchique liant les différentes modalités, et la connaissance a priori de cette structure est nécessaire à une exploitation cohérente des données.

Les contributions de cette thèse, tout en construisant sur un formalisme de l'apprentissage de représentations proche de ces méthodes, visent à dépasser cette limitation en intégrant progressivement les données à la représentation latente, de manière hierarchique. La première contribution, décrite au Chapitre 5, repose sur une extension de l'apprentissage de variétés décrit dans ce chapitre. Le Chapitre 6 développe quant à lui une généralisation, basée sur les processus Gaussiens à variables latentes, donnant lieu à un modèle hiérarchique plus flexible.

Chapitre 3

État de l'art : Processus Gaussiens à variables latentes

Dans le chapitre précédent, nous avons donné une vue d'ensemble des méthodes statistiques de réduction de dimensions dans le cadre de l'apprentissage de représentations. Nous avons présenté les versions à une unique modalité, puis les éventuelles extensions de ces méthodes à plusieurs modalités. Nous allons à présent discuter plus en détail un ensemble de méthodes probabilistes, les GP-LVMs (*Gaussian Process Latent Variable Models* [42]), sur lequel est construite la deuxième contribution de cette thèse, qui sera présentée dans le Chapitre 6 de ce document.

Les modèles à variables latentes font l'hypothèse que les observations (i.e. les données) sont les expressions réelles de phénomènes sous-jacents, ou latents. Ce chapitre décrit la façon dont ce concept peut être modélisé dans des cadres de données plus ou moins complexes, en commençant par le cas à une seule modalité, puis les modèles permettant d'exploiter conjointement plusieurs modalités.

3.1 Modèles à variables latentes

Étant donné un ensemble d'observations $\mathbf{Y} \in \mathbb{R}^{N \times D}$, régies par des variables latentes $\mathbf{X} \in \mathbb{R}^{N \times d}$, on modélise le lien entre les observations et les variables latentes par une fonction f suivant un processus Gaussien et du bruit :

$$\mathbf{y}_i = f(\mathbf{x}_i, \mathbf{A}) + \eta_i, \tag{3.1}$$

avec $\eta_i \sim \mathcal{N}(\mathbf{0}, \Sigma^{-1}\mathbf{I})$ une loi normale de moyenne $\mathbf{0}$ et de covariance $\Sigma^{-1}\mathbf{I}$, $\mathbf{A}$ la matrice des paramètres de la fonction f, $\mathbf{x}_i \in \mathbb{R}^{d \times 1}$ la variable latente associée à l'individu i et $\mathbf{y}_i \in \mathbb{R}^{D \times 1}$ l'observation associée au même individu i.

L'optimisation des processus Gaussiens consiste à estimer les paramètres de la fonction f permettant de représenter au mieux les observations $\mathbf{Y}$ en estimant les variables latentes $\mathbf{X}$ de dimensionalité la plus petite possible.

3.1.1 De l'ACP aux Processus Gaussiens non linéaires

L'analyse en composantes principale (ACP) a été présentée plus haut, en Section 2.2.1. Il existe une version probabiliste de l'ACP, décrite en 1999 par *Bishop et Tipping* [75], qui se rapproche des GP-LVMs et qui nous permet d'introduire les modèles à variables latentes (LVM) linéaires [76].

On fait l'hypothèse que la relation entre les données $\mathbf{Y}$ et les variables latentes $\mathbf{X}$ est linéaire, c'est-à-dire :

$$f(\mathbf{x}_i, \mathbf{A}) = \mathbf{A}\mathbf{x}_i, \tag{3.2}$$

avec $\mathbf{A} \in \mathbb{R}^{D \times d}$. On cherche à estimer les paramètres qui expliquent les observations $\mathbf{Y}$, c'est-à-dire qui maximisent la vraisemblance de $\mathbf{Y}$ sachant les variables latentes $\mathbf{X}$ et la matrice de paramètres $\mathbf{A}$.

Pour un individu i de l'espace des observations, cette vraisemblance s'écrit de la façon suivante :

$$p(\mathbf{y}_i|\mathbf{x}_i, \mathbf{A}, \Sigma) = \mathcal{N}(\mathbf{y}_i|\mathbf{A}\mathbf{x}_i, \Sigma^{-1}\mathbf{I}). \tag{3.3}$$

Pour obtenir la vraisemblance marginale, on intègre relativement aux variables latentes $\mathbf{x}_i$:

$$p(\mathbf{y}_i|\mathbf{A}, \Sigma) = \int p(\mathbf{y}_i|\mathbf{A}\mathbf{x}_i, \Sigma^{-1}\mathbf{I})p(\mathbf{x}_i)d\mathbf{x}_i. \tag{3.4}$$

Il faut ensuite définir une distribution a priori sur les variables latentes $\mathbf{X}$. Dans le cas de l'ACP probabiliste, la distribution choisie est une Normale centrée réduite :

$$p(\mathbf{x}_i) = \mathcal{N}(\mathbf{x}_i|\mathbf{0}, \mathbf{I}). \tag{3.5}$$

Cette hypothèse permet de trouver une formulation analytique de la vraisemblance marginale pour chaque observation :

$$p(\mathbf{y}_i|\mathbf{A}, \Sigma) = \mathcal{N}(\mathbf{y}_i|\mathbf{0}, \mathbf{A}\mathbf{A}^T + \Sigma^{-1}\mathbf{I}). \tag{3.6}$$

En supposant l'indépendance entre les échantillons $\mathbf{y}_i$, cette vraisemblance s'écrit alors pour l'ensemble des observations de la façon suivante :

$$P(\mathbf{Y} \mid \mathbf{A}, \Sigma) = \prod_{i=1}^{N} \mathcal{N}(\mathbf{y}_i \mid \mathbf{0}, \mathbf{A}\mathbf{A}^T + \Sigma^{-1}\mathbf{I}). \tag{3.7}$$

Pour résoudre ce problème en pratique, on maximise la log-vraisemblance :

$$\log P(\mathbf{Y} \mid \mathbf{A}, \Sigma) = \sum_{i=1}^{N} \log \mathcal{N}(\mathbf{y}_i \mid \mathbf{0}, \mathbf{A}\mathbf{A}^T + \Sigma^{-1}\mathbf{I}). \tag{3.8}$$

Bishop et Tipping [75] ont montré que les estimateurs de maximum de vraisemblance de $\mathbf{A}$ et Σ (c'est à dire les valeurs de $\mathbf{A}$ et Σ maximisant l'équation

(3.8)) avaient des solutions analytiques. Le maximum de vraisemblance est atteint quand $\mathbf{A}$ est la matrice de projection entre l'espace principal des données et l'espace latent, et l'estimateur de Σ peut alors être interprété comme un estimateur de la variance perdue lors de la projection des données ; le modèle est donc une version probabiliste de l'ACP.

Cette approche s'appelle l'analyse en composantes principales probabiliste (PPCA), et est formulée pour des problèmes linéaires (cf. (3.2)). Cependant, cette approche souffre de deux principales limitations : si on considère des fonctions f non linéaires, ce problème est insolvable ; de plus, l'expression de la vraisemblance en fonction des paramètres du modèles n'est pas intéressante, on préférerait avoir accès à la probabilité d'observer la distribution $\mathbf{Y}$ en sachant les variables latentes $\mathbf{X}$.

L'analyse en composantes principales probabiliste duale (*Dual Probabilistic PCA*, DPPCA) [42] propose de pallier cette deuxième limite. Dans ce modèle, on adopte une approche duale à la PPCA : au lieu de marginaliser relativement aux variables latentes, on marginalise sur les paramètres. On utilise donc cette fois un a priori Gaussien, centré réduit, sur les paramètres $\mathbf{A}$:

$$P(\mathbf{A}) = \prod_{j=1}^{D} \mathcal{N}(\mathbf{a}_i \mid \mathbf{0}, \mathbf{I}), \tag{3.9}$$

avec $\mathbf{a}_i$ le vecteur correspondant à la ligne i de la matrice de paramètres $\mathbf{A}$.

Les calculs sont similaires à ceux développés pour la PPCA, et on obtient donc comme formulation de la vraisemblance sur l'ensemble des données, toujours en supposant l'indépendance entre échantillons :

$$P(\mathbf{Y} \mid \mathbf{X}) = \prod_{j=1}^{D} \mathcal{N}(\mathbf{y}_{:,j} \mid \mathbf{0}, \mathbf{X}\mathbf{X}^T + \Sigma^{-1}\mathbf{I}), \tag{3.10}$$

Dans cette approche, $\mathbf{X}\mathbf{X}^T + \Sigma^{-1}\mathbf{I}$ peut être assimilé à un noyau linéaire. On note alors, avec $\mathbf{K} = \mathbf{X}\mathbf{X}^T + \Sigma^{-1}\mathbf{I}$:

$$P(\mathbf{Y} \mid \mathbf{X}) = \prod_{j=1}^{D} \mathcal{N}(\mathbf{y}_j \mid \mathbf{0}, \mathbf{K}). \tag{3.11}$$

Dans le cas où le noyau $\mathbf{K}$ est linéaire, cette approche est nommée analyse en composantes principales probabiliste duale (DPPCA). Cependant, afin de mieux représenter des phénomènes réels, on peut introduire de la non linéarité dans le modèle précédent : on remplace alors le noyau linéaire $\mathbf{K}$ par un noyau non linéaire $\mathbf{K}_x = [k(\mathbf{x}_i, \mathbf{x}_j)] \in \mathbb{R}^{N \times N}$. On nomme ce modèle GP-LVM, et nous en décrivons les spécificités dans la sous-Section suivante 3.1.2.

3.1.2 Introduction de non linéarité : le modèle GP-LVM

En pratique, on utilise généralement pour $\mathbf{K}$ un noyau Gaussien (*Radial Basis Function*, RBF) :

$$k(\mathbf{x}_i, \mathbf{x}_j) = \omega^2 \exp - \frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2l^2} + \sigma^2 \delta_{ij}, \tag{3.12}$$

avec ω et σ les paramètres d'échelle de la distribution, l l'échelle du noyau et δ_{ij} le delta de Dirichlet.

De la même façon que pour la DPPCA, l'espace latent est estimé en recherchant une solution maximale a posteriori, c'est-à-dire en maximisant la log-vraisemblance par rapport à $\mathbf{X}$:

$$\underset{\mathbf{X}}{\text{argmax}} \log P(\mathbf{X} \mid \mathbf{Y}). \tag{3.13}$$

En utilisant la règle de Bayes, cela peut être réécrit de la façon suivante :

$$\underset{\mathbf{X}}{\text{argmax}} \log \frac{P(\mathbf{Y} \mid \mathbf{X})P(\mathbf{X})}{P(\mathbf{Y})}. \tag{3.14}$$

L'équation Eq.(3.14) correspond à la maximisation de $\log P(\mathbf{Y} \mid \mathbf{X})P(\mathbf{X})$, puisque $P(\mathbf{Y})$ est constant par rapport à $\mathbf{X}$. La fonction de coût du GP-LVM est ainsi la fonction suivante :

$$-\sum_{j=1}^{D} \log \mathcal{N}(\mathbf{y}_j | \mathbf{0}, \mathbf{K}). \tag{3.15}$$

Du fait de la non linéarité du noyau, il n'existe plus de solution analytique au maximum de vraisemblance. En pratique, on minimise la log-vraisemblance negative $-\log(\cdot)$ de l'équation (3.15) par un algorithme de descente de gradient (nous avons utilisé dans les applications de cette thèse l'algorithme L-BFGS [77]).

3.1.3 Aparté sur l'optimisation des GP-LVMs

Nous avons observé qu'en pratique, sur des exemples d'analyse d'images, la fonction de coût (voir Equation (3.15)) du processus Gaussien était souvent négative. Cela est dû au fait que la vraisemblance, contrairement à une mesure de probabilité, n'est pas bornée entre 0 et 1. Ainsi, dans les zones où il n'y a pas de variabilité (par exemple les pixels en dehors de la zone d'intérêt), la fonction de coût va fortement décroître et va beaucoup influencer la prédiction, qui va se contenter de très bien prédire les zones à faible variabilité (qui nous intéressent peu ou pas) au lieu de se concentrer à faire le moins d'erreurs possibles sur les zones à forte variabilité. Nous avons considéré que les pixels pour lesquels la fonction de coût était négative correspondaient à des zones de trop faible variabilité, et nous avons donc cherché à n'avoir que des pixels à fonction de coût positive.

On montre ce phénomène sur les données MNIST en Figure 3.1. Pour pallier ce problème et offrir une meilleure évaluation des données (c'est à dire un meilleur espace latent), nous avons envisagé deux solutions :

1. La première consiste à dégrader artificiellement les données en ajoutant un bruit Gaussien d'une certaine amplitude. En pratique, nous avons testé cette solution sur des images de segmentations (données discrètes comprises entre 0 et 1). Ainsi, en ajoutant un bruit d'amplitude strictement inférieure à 0.5, la transformation est réversible. Cette approche fonctionne mais n'est pas très satisfaisante. En effet, l'optimisation des GP-LVMs n'étant pas très stable, l'ajout d'une perturbation aléatoire n'est pas souhaitable. En outre, cette solution implique également une modification des textures de l'image qui peut perturber la représentation apprise.
2. La deuxième solution, qui a été utilisée dans le Chapitre 6 de cette thèse, consiste à ne traiter que la partie à variabilité suffisante de l'image. Cette solution résout le problème de la fonctionnelle négative, et permet de se focaliser sur la zone d'intérêt, qui contient les motifs à forte variabilité. Elle n'induit pas de grandes perturbations dans le traitement des données car on utilise une distance Euclidienne par pixel.

Illustration sur la base de données MNIST : Nous illustrons le phénomène de fonction de coût négative sur le jeu de donnée MNIST [78], qui comporte 60000 images de dimensions 28×28 pixels représentant des chiffres écrits à la main allant de 0 à 9. Dans ce jeu de données, les chiffres sont dessinés au centre de l'image, et les bords sont toujours des pixels noirs. Il y a donc très peu variabilité au sein de la population sur ces pixels. On affiche sur la Figure 3.1 la fonction de coût d'un GP-LVM calculé sur 400 données, et à côté l'écart-type par pixels.

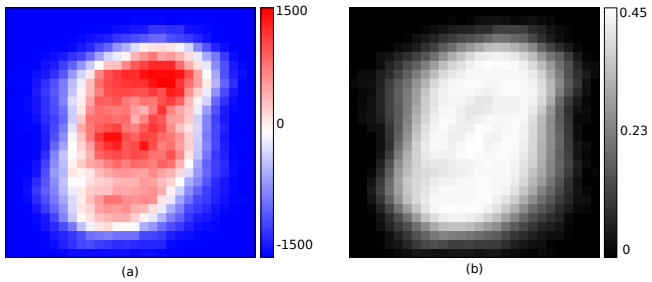


FIGURE 3.1 – Images représentant (a) la fonction de coût par pixel du GP-LVM calculé sur les données MNIST (b) l'écart type des données. Les pixels de l'image (b) qui correspondent aux fortes variabilités (en blanc) coïncident avec les pixels correspondant à une fonction de coût positive dans l'image (a) La fonction de coût est celle décrite par l'équation Eq. (3.15).

Ces deux quantités sont fortement corrélées : plus la variabilité est faible, plus le GP-LVM est sûr de la valeur du pixel, et plus la fonction de coût est négative.

L'optimisation relative à la zone d'intérêt, à forte variabilité, est perturbée par ce phénomène.

Pour s'affranchir de cette fonction de coût négative, on tronque l'image pour n'utiliser que la zone qui varie le plus, afin de mieux représenter la variabilité des motifs du jeu de données. La Figure 3.2 affiche les fonctions de coût avec et sans troncature, et montre que pour les données tronquées, la fonction est de nouveau positive. La Figure 3.3 montre les deux espaces latents calculés dans l'un et l'autre cas.

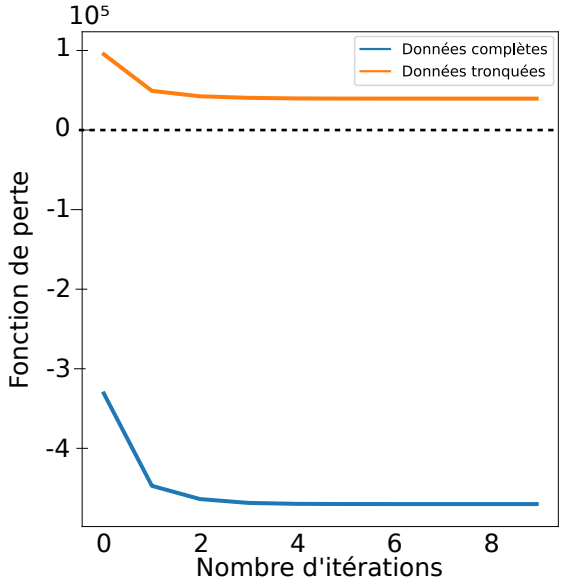


FIGURE 3.2 – Comparaison des fonctions de coût sur les données tronquées (en orange) et non-tronquées (en bleu). Les fonctions ont des allures similaires, mais la fonctionnelle correspondant aux données tronquées est cette fois positive.

L'espace calculé sur les données complètes n'était pas capable de distinguer les chiffres 4, les 7 et les 9 (zone entourée en rouge sur les espaces de la Figure 3.3, les points correspondant à ces chiffres sont superposés). L'espace calculé sur les données tronquées règle en partie ce problème, l'espace obtenu étant ainsi plus représentatif des données.

3.2 GP-LVM à plusieurs descripteurs

Représentation graphique : *Lawrence et Moore* [79] ont proposé en 2007 une version des GP-LVMs prenant en compte d'éventuelles dépendances et permettant de les modéliser sous forme d'un graphe de dépendance orienté. Chaque nœud représente soit un ensemble d'observations, soit une variable latente. Les arrêtes du graphe représentent les interactions entre les différentes entités. Par

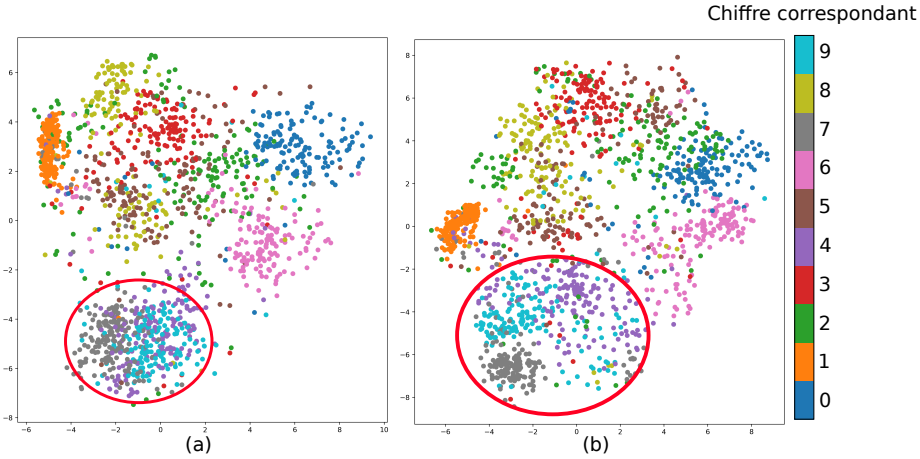


FIGURE 3.3 – Comparaison des espaces obtenus avec un GP-LVM sur un sous-échantillon des données de la base MNIST : (a) fonctionnelle négative (b) fonctionnelle positive. Dans cet exemple-ci, nous avons imposé que les espaces possèdent seulement 2 dimensions latentes. La zone entourée en rouge est une zone qui souligne une différence notable entre ces deux espaces, à savoir la meilleure séparation des chiffres 4, 7 et 9, montrant l'intérêt d'avoir une fonction de coût positive.

exemple, le GP-LVM unimodal présenté en Section 3.1.2 peut être représenté comme en Figure 3.4.

FIGURE 3.4 – Représentation d'un GP-LVM sous forme de graphe de dépendance. Les variables latentes $\mathbf{X}$ régissent les observations $\mathbf{Y}$.

Modélisation de dépendances temporelles : Dans cette même logique, *Lawrence et Moore* ont proposé de représenter de cette façon un phénomène dynamique dont l'espace latent est lui-même régi par le temps $\mathbf{t}$ (voir schéma sur la Figure 3.5) :

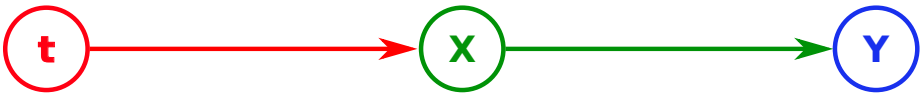


FIGURE 3.5 – Représentation de la dépendance dynamique d'un GP-LVM sous forme de graphe. Le temps $\mathbf{t}$ régit les variables latentes $\mathbf{X}$, qui elles-mêmes régissent les observations $\mathbf{Y}$.

Pour ce nouveau modèle, l'objectif est de maximiser $P(\mathbf{X} \mid \mathbf{Y}, \mathbf{t})$. En utilisant la loi de Bayes :

$$P(\mathbf{X} \mid \mathbf{Y}, \mathbf{t}) = \frac{P(\mathbf{Y}, \mathbf{t} \mid \mathbf{X})P(\mathbf{X})}{P(\mathbf{Y}, \mathbf{t})}. \tag{3.16}$$

$\mathbf{Y}$ et $\mathbf{t}$ sont indépendants sachant $\mathbf{X}$ donc :

$$P(\mathbf{Y}, \mathbf{t} \mid \mathbf{X})P(\mathbf{X}) = P(\mathbf{Y} \mid \mathbf{X})P(\mathbf{t} \mid \mathbf{X}), \tag{3.17}$$

et :

$$P(\mathbf{Y}, \mathbf{t}) = P(\mathbf{Y})P(\mathbf{t}). \tag{3.18}$$

En appliquant de nouveau la loi de Bayes :

$$P(\mathbf{t} \mid \mathbf{X}) = \frac{P(\mathbf{X} \mid \mathbf{t})P(\mathbf{t})}{P(\mathbf{X})}. \tag{3.19}$$

En combinant les équations 3.16 3.17 3.18 et 3.19, on obtient :

$$P(\mathbf{X} \mid \mathbf{Y}, \mathbf{t}) = \frac{P(\mathbf{Y} \mid \mathbf{X})P(\mathbf{X} \mid \mathbf{t})}{P(\mathbf{Y})} \tag{3.20}$$

$P(\mathbf{Y})$ étant constante, l'équation 3.20 revient finalement à maximiser la log-vraisemblance :

$$\log(P(\mathbf{X} \mid \mathbf{Y}, \mathbf{t})) = \log(P(\mathbf{Y} \mid \mathbf{X})) + \log(P(\mathbf{X} \mid \mathbf{t})). \tag{3.21}$$

Le premier terme est exactement le GP-LVM de $\mathbf{X}$ sur $\mathbf{Y}$, et le gradient du second terme vaut :

$$\frac{\partial \log P(\mathbf{X} \mid \mathbf{t})}{\partial \mathbf{X}} = \mathbf{K}_t^{-1}\mathbf{X}, \tag{3.22}$$

où $\mathbf{K}_t$ est le noyau temporel.

Ce terme peut être combiné avec le gradient du premier terme afin de trouver l'espace latent par maximum a posteriori (MAP) [79].

Application à l'apprentissage multi-descripteurs : On peut généraliser ce concept à des graphes plus complexes, qui vont représenter des structures de dépendances dans les données. Pour les résoudre, il suffit ensuite de raisonner comme précédemment et de définir des GP-LVMs entre les différentes entités constituant l'espace de données. *Lawrence et al.* en donnent un exemple [79] que nous allons décrire ici.

Supposons que nous ayons un espace de données défini de la façon suivante : on connaît les dynamiques sous-jacentes du système, et on sait que chacune des deux modalités ($\mathbf{Y}_1$ et $\mathbf{Y}_2$) découle indépendamment d'une variable latente ($\mathbf{X}_1$ et $\mathbf{X}_2$). Ces deux variables latentes sont elles-mêmes gouvernées par le même phénomène latent $\mathbf{X}_3$ (voir Figure 3.6).

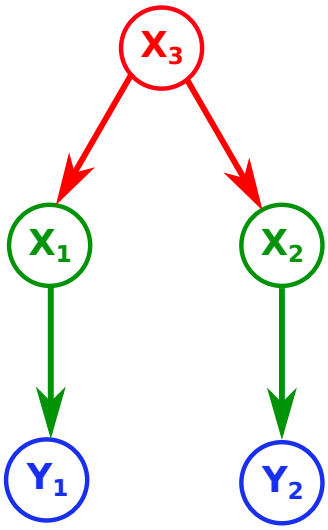


FIGURE 3.6 – Représentation d'un GP-LVM hiérarchique sous forme de graphe. (Exemple tiré de *Lawrence et al.* [79].)

Dans l'exemple de l'article de *Lawrence et al.*, ce système modélise une interaction entre deux *stickmans* (dessins d'homme-bâtons) se tapant dans la main au sein d'une suite temporelle de modèles articulés. $\mathbf{Y}_1$ et $\mathbf{Y}_2$ représentent respectivement les coordonnées spatiales du premier et du second *stickman*, et chaque instant de la séquence temporelle est considéré comme un échantillon. $\mathbf{X}_1$ et $\mathbf{X}_2$ représentent les phénomènes latents déterminant les positions respectives des deux individus, tandis que $\mathbf{X}_3$ représente leur interaction, et la façon dont elle conditionne leurs mouvements respectifs au cours de la séquence d'images. La modélisation de ce phénomène par GP-LVM revient à optimiser :

$$\arg \max_{\mathbf{X}_1,\mathbf{X}_2,\mathbf{X}_3} \log(P(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3 \mid \mathbf{Y}_1, \mathbf{Y}_2)). \tag{3.23}$$

En utilisant les mêmes principes que pour l'exemple temporel de la section précédente, c'est-à-dire en exploitant la formule de Bayes et l'indépendance de $\mathbf{Y}_1$ et $\mathbf{Y}_2$, on obtient :

$$P(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3 \mid \mathbf{Y}_1, \mathbf{Y}_2) = \frac{P(\mathbf{X}_1 \mid \mathbf{Y}_1)P(\mathbf{X}_2 \mid \mathbf{Y}_2)P(\mathbf{X}_3 \mid \mathbf{Y}_3)P(\mathbf{X}_3)}{P(\mathbf{Y}_1, \mathbf{Y}_2)}. \tag{3.24}$$

$P(\mathbf{Y}_1, \mathbf{Y}_2)$ étant constante, et en utilisant l'a priori Gaussien sur $P(\mathbf{X}_3)$, l'équation 3.24 revient à l'optimisation jointe des GP-LVMs $\mathbf{X}_1 \rightarrow \mathbf{Y}_1$, $\mathbf{X}_2 \rightarrow \mathbf{Y}_2$, $\mathbf{X}_3 \rightarrow \mathbf{X}_1$ et $\mathbf{X}_3 \rightarrow \mathbf{X}_2$.

On peut ainsi exploiter des structures de dépendance dans les données connues a priori. Les GP-LVMs multimodaux sont flexibles grâce à cette représentation sous forme de graphe, et seront adaptés pour la représentation des lésions du

myocarde (infarctus et MVO) au sein d'une population de patients souffrant d'infarctus aigü du myocarde, que nous développerons dans le Chapitre 6 de cette thèse.

3.2.1 Exemple d'application : CelebA

Nous illustrons dans cette sous-section le concept de hiérarchie sur le jeu de données CelebA, décrit en sous-Section 2.2.3. Dans le Chapitre 2, nous avions testé entre autres les GP-LVMs sur les images, et nous avions trouvé les limitations suivantes :

1. La pauvreté de la métrique utilisée ne permet pas d'exploiter correctement la texture des images : les fonds noirs peuvent par exemple être assimilés à des chevelures,
2. Certaines informations cachées dans les données (l'expression faciale, l'orientation du visage) sont mal représentées.

Pour dépasser en partie ces limitations, on se propose ici d'étudier l'apport des points clés du visage (considérés comme une nouvelle donnée, le premier niveau dans la hiérarchie) en exploitant la structure hiérarchique du jeu de données. L'algorithme utilisé ici est basé sur les concepts décrits dans ce chapitre et sera développé au Chapitre 6 de ce document.

Tout d'abord, nous lançons un GP-LVM sur les données des marqueurs de position, de dimensionalité 10 (2 positions pour chacun des 5 marqueurs). Nous autorisons 3 dimensions dans l'espace latent. Les projections sur les différentes dimensions de cet espace latent sont affichées sur les Figures 3.7 et 3.8, et les points de l'espace latent sont colorés suivant les indices présentés en Section 2.2.3. Comme attendu, le sourire et l'orientation du visage sont bien représentés, mais pas la valeur d'intensité moyenne du centre de l'image. Cet espace a donc appris correctement les caractéristiques des marqueurs de position du visage, et peut servir de soutien et d'apport pour l'apprentissage du contenu image.

Nous avons ensuite appliqué l'algorithme du GP-LVM hiérarchique (détails au Chapitre 6) pour exploiter la donnée image. Nous avons cette fois utilisé 10 dimensions latentes. L'espace correspondant aux marqueurs de position du visage ne comportant que 3 dimensions latentes, nous avons complété les 7 dimensions manquantes avec des zéros. Les projections coloriées relativement aux indices sont présentées sur la Figure 3.9. On observe cette fois-ci que les 3 indices de référence sont présents dans l'espace latent. Comparativement à l'espace sans l'apport des marqueurs de positions (voir Chapitre 2), ce nouvel espace hiérarchique restitue l'information du sourire, qui était absente, et conserve un bon étalement des indices relatifs à la texture de l'image et à l'orientation des visages. L'espace hiérarchique encode donc des informations plus riches sur 10 dimensions, tout en étant guidé par le premier niveau de la hiérarchie, et cette méthodologie permet de contrebalancer en partie la faiblesse de la métrique utilisée. Afin d'apporter une comparaison plus quantitative, nous avons également regardé les corrélations entre les espaces et ces indices (orientation, sourire

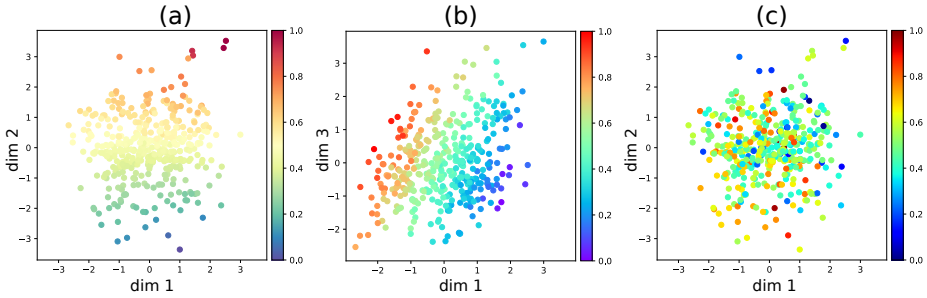


FIGURE 3.7 – Différentes projections de l'espace latents des marqueurs de positions du visage, coloriées selon (a) l'orientation du visage (dimension 1 contre dimension 2), (b) le sourire (dimension 1 contre dimension 3), (c) la valeur d'intensité moyenne du centre du visage encadré par les yeux et la bouche (dimension 1 contre dimension 2). Cette figure montre que l'espace latent est très ordonné selon les indices relatifs aux marqueurs de position(a) et (b), ce qui est logique car le GP-LVM a pris en entrée les marqueurs de position. En revanche, l'intensité du centre de l'image (c) n'est pas corrélée avec les marqueurs de position (ce qui était également attendu).

et intensité du centre de l'image), comme synthétisé en Figure 3.10. Nous observons que les espaces distribuent tous deux les données de façon similaire, comme le montrent les colonnes associées à l'intensité du centre du visage et à son orientation. Cependant, l'espace hiérarchique encode également le sourire sur les dimensions 6 et 8, et plus faiblement sur les dimensions 4 et 5.

Cette expérience sert de preuve de concept pour illustrer l'apport d'une intégration hiérarchique de données, et donc son potentiel pour l'application visée. Le Chapitre 6 présentera une évaluation bien plus complète de cette méthodologie, en comparant notamment différentes stratégies pour construire la hiérarchie, et en examinant les différences par rapport à d'autres schémas tels que la fusion ou la concaténation de données (si celles-ci sont du même type).

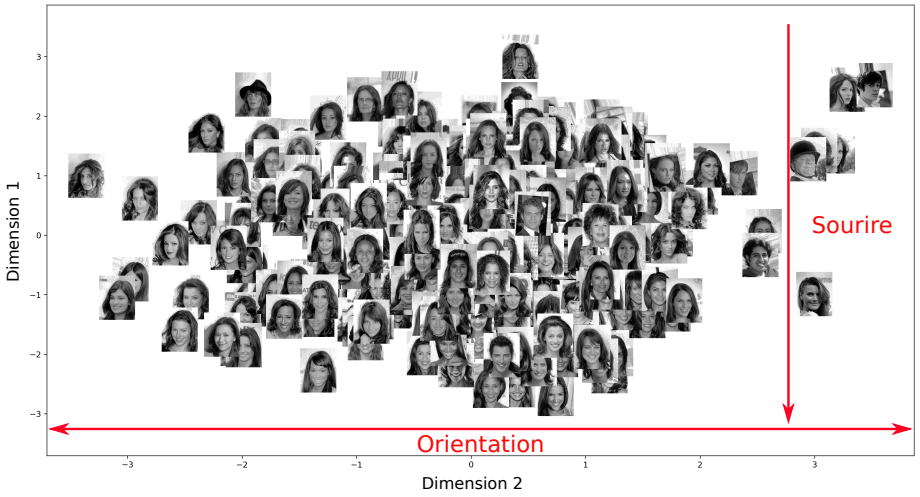


FIGURE 3.8 – Vue d'ensemble des images associées à chaque point de l'espace latent estimé à partir des marqueurs de positions. On observe la variation d'orientation sur la dimension 2 (en abscisse) et la variation de sourire sur la dimension 1 (en ordonnée). L'espace des marqueurs de position encode bien les informations qui nous intéressent (sourire et orientation), il paraît donc être une bonne base pour l'apprentissage hiérarchique. On remarque néanmoins des sujets mal positionnés par rapport au contenu image, soit relativement au contenu du centre de l'image (non corrélé avec les marqueurs de position, voir Figure 3.7), mais aussi par rapport au contenu périphérique (par exemple, femme avec un chapeau ou homme avec un casque), en raison de la métrique très simple utilisée.

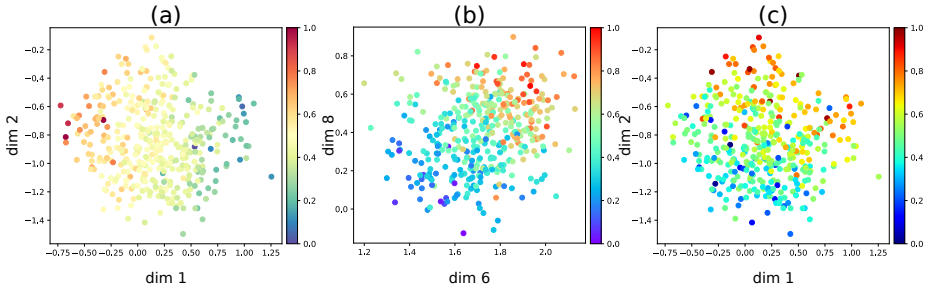


FIGURE 3.9 – Différentes projections de l'espace latent hiérarchique, coloriées selon (a) l'orientation du visage (dimension 1 contre dimension 2), (b) le sourire (dimension 6 contre dimension 8), (c) la valeur d'intensité moyenne du centre du visage encadré par les yeux et la bouche (dimension 1 contre dimension 2). L'espace hiérarchique est capable de représenter à la fois les informations liées aux marqueurs de position et celles liées à la donnée image.

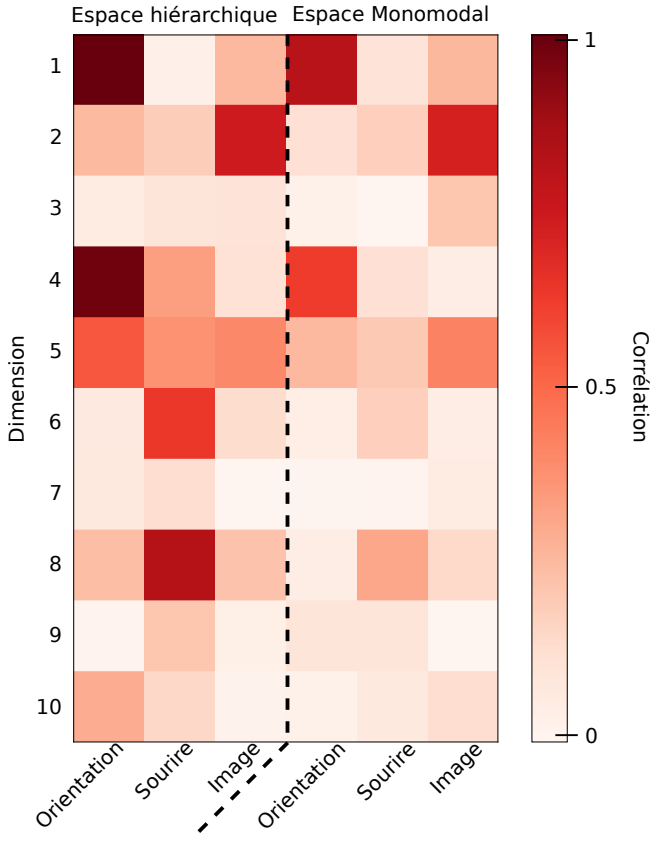


FIGURE 3.10 – Corrélations entre les 10 dimensions des deux espaces latents (espace hiérarchique et espace monomodal estimé uniquement à partir des images) et les indices physiologiques (orientation du visage, sourire et intensité du centre de l'image). Les espaces encodent des informations similaires, mais les corrélations dans l'espace hiérarchique sont renforcées, et le sourire (qui était quasiment absent pour l'espace monomodal) est également encodé.

Chapitre 4

Données et pré-traitement

Ce chapitre présente les données que nous avons utilisées pour les deux applications principales de cette thèse, correspondant aux Chapitres 5 et 6 qui suivent. Elles ont été récoltées dans le cadre de l'étude clinique MIMI [14] (ClinicalTrials ID : NCT01360242). Cette cohorte multi-centrique contient les données de 160 patients atteints d'infarctus aigü du myocarde (123 ayant des images à un format pouvant être analysé). Nous avions ainsi accès aux données IRM de réhaussement, à la fois tardif (LGE, pour 123 patients) et précoce (EGE, pour 117 patients). Nous avions également à disposition les segmentations de l'infarctus du myocarde (sur les données LGE) et du MVO (sur les données LGE et EGE), qui nous ont été fournies par nos collaborateurs Pierre Croisille (radiologue) et Magalie Viallon (physicienne médicale) du Centre Hospitalier Universitaire de Saint-Étienne.

4.1 Prétraitement des données images

4.1.1 Caractéristiques de l'étude MIMI

Les patients inclus dans l'étude MIMI sont des adultes présentant, moins de 12h après l'apparition des premiers symptômes, un infarctus du myocarde avec élevation du segment ST (STEMI, voir Section 1.1.1) supérieure à 1mm dans au moins 2 membres contigüs, ou supérieure à 2mm dans au moins 2 dérivations précordiale de l'ECG, patients pour lesquels une intervention coronarienne percutanée primaire était programmée. Certains patients présentant des contre-indications ont été exclus de l'étude (voir *Belle et al.* pour plus de détails sur les critères d'exclusion [14]).

Les patients ont été traités par une angioplastie (permettant de restaurer l'afflux sanguin), complétée par la pose d'un stent soit immédiatement, soit retardée de 24 à 48h, les patients ayant été répartis aléatoirement dans ces deux groupes. Les cliniciens faisaient l'hypothèse que la pose retardée d'un stent pouvait mettre en jeu des mécanismes de protection amenant à une restauration

de la perfusion du myocarde, et donc des lésions (infarctus et obstruction microvasculaire (MVO)) de moindre importance. Le critère d'évaluation principal était l'étendue du MVO, quantifié par une IRM cardiaque 5 jours après l'inclusion du patient (sur des machines 1.5 T). Sur la base de ces mesures scalaires, l'étude n'avait pas observé de différences significatives entre les deux groupes de patients, et avait même constaté une étendue du MVO légèrement plus grande dans le groupe où la pose du stent avait été retardée.

Dans une étude postérieure [80], ces données ont été ré-analysées avec des outils plus élaborés permettant d'aller jusqu'à des comparaisons à l'échelle des pixels, après une standardisation des données sur une même référence géométrique (voir détails ci-dessous en Section 4.1.3). L'hypothèse était que l'analyse des mesures scalaires réalisée dans l'étude de *Belle et al.* [14] tronquait fortement le contenu image et pouvait ainsi manquer des subtilités des motifs de lésions, pouvant potentiellement conduire à des conclusions différentes sur l'intérêt ou non d'un stenting retardé. Néanmoins, la nouvelle analyse a conduit à des conclusions similaires.

Dans le cadre de cette thèse et de notre équipe de recherche, l'étude MIMI reste néanmoins très intéressante sur plusieurs points :

- Ses données étaient disponibles (segmentées et standardisées sur une même référence géométrique) peu après le début de ma thèse, ce qui m'a permis de travailler rapidement sur des données réelles (contribution détaillée au Chapitre 5) et ainsi dépasser les premiers tests faits sur des données synthétiques et des bases de données publiques comme CelebA (voir chapitres précédents),
- Elles permettent de formaliser une hiérarchie simple et pertinente pour l'application cardiaque concernée (voir Figure 4.5) : d'un contenu plus simple vers un contenu plus élaboré qu'il est difficile de caractériser directement par de l'apprentissage de représentation (segmentations puis apparence des images comme au Chapitre 5, ou motif d'infarctus puis motif de MVO comme au Chapitre 6),
- Au niveau de notre équipe de recherche, elle a permis de mettre en place les outils d'analyse de base qui sont désormais généralisés à d'autres cohortes : HIBISCUS-STEMI (350 patients / ClinicalTrials ID : NCT03070496) et CARIM (2000 patients / ClinicalTrials ID : NCT02967965). Les segmentations ont également été revues précisément pour servir de base d'entrainement dans le récent challenge MYOSAIQ [1], qui visait la segmentation automatique d'images de LGE, dans le cadre de la conférence FIMH qui a eu lieu à Lyon en Juin 2023.

4.1.2 Segmentations

Dans cette thèse, nous avons exploité deux modalités d'images disponibles et segmentées dans le cadre de l'étude MIMI : l'image en réhaussement précoce (EGE, où est visible le MVO) et l'image en réhaussement tardif (LGE, où

1. `https://www.creatis.insa-lyon.fr/Challenge/myosaiq/index.html`

sont visibles l'infarctus et le MVO, même si le EGE est plus recommandé pour évaluer le MVO). Ces images ont été segmentées manuellement par un expert, et contrôlées par deux autres, à l'aide d'un logiciel commercial (CVI42 v.5.1.0 Circle Cardiovascular Imaging, Calgary, Canada). Les segmentations résultantes sont l'endocarde, l'épicarde, la zone infarcie et le MVO pour les images LGE ; et l'endocarde, l'épicarde et le MVO pour les images LGE. L'infarctus correspond à une zone en hyper-intensité à l'intérieur du myocarde, débutant au niveau de l'endocarde, et a été segmenté de façon semi-automatique. Le MVO correspond à une zone en hypo-intensité à l'intérieur de la zone infarcie, et a été segmenté manuellement. Les piles d'images disponibles comportent 17 ± 2 coupes 2D couvrant le ventricule gauche, pour une résolution médiane de $1.57 \times 1.57.00$ mm. La Figure 4.1 illustre cela sur les données EGE et LGE d'un même patient, avec une coupe 2D au niveau du milieu de la cavité cardiaque.

4.1.3 Alignement et normalisation

Les patients diffèrent néanmoins en termes d'anatomie, de nombre de coupes dans une acquisition, et éventuellement leur orientation. Cela peut être également le cas (dans une moindre mesure) entre deux types d'acquisition (EGE et LGE). Avant toute analyse statistique allant jusqu'à l'échelle du pixel, et à plus forte raison avant un apprentissage de représentation sur ces données, il est donc nécessaire d'aligner ces données sur une même géométrie de référence. Cet alignement a été effectué par mon co-encadrant Nicolas Duchateau comme décrit dans [80], et repose sur des principes assez classiques inspirés d'outils d'atlas statistiques et géométrie computationnelle. Il peut être résumé par les étapes suivantes :

1. Le myocarde de chaque individu est paramétrisé par des coordonnées radiales, circonférentielles et grand-axe, comprises entre 0 et 1 et disponibles en chaque pixel à l'intérieur du myocarde sur chaque coupe de la base à l'apex,
2. Une géométrie de référence est estimée, également paramétrisée de la même manière, comportant 21 coupes 2D composées de 80×80 pixels chacune,
3. Les données pixel de chaque acquisition sont enfin transportées sur cette géométrie de référence, à l'aide d'une interpolation (méthode linéaire basée sur une grille éparse associée aux coordonnées du myocarde [80]).

Cela permet de transporter non seulement les informations de segmentation des lésion, mais aussi le contenu des images pour des analyses plus fines, en particulier des zones "grises". Cette statégie de paramétrisation a été préférée par rapport à un recalage et une analyse de maillages, pour garder un meilleur contrôle du contenu des images au niveau de chaque pixel et de la correspondance des coupes de la base à l'apex (certaines régions comme l'anneau mitral posant par ailleurs certains défis pour ce type d'analyse, car le myocarde est "coupé" par le début de l'aorte). La Figure 4.2 récapitule la stratégie d'alignement utilisée.

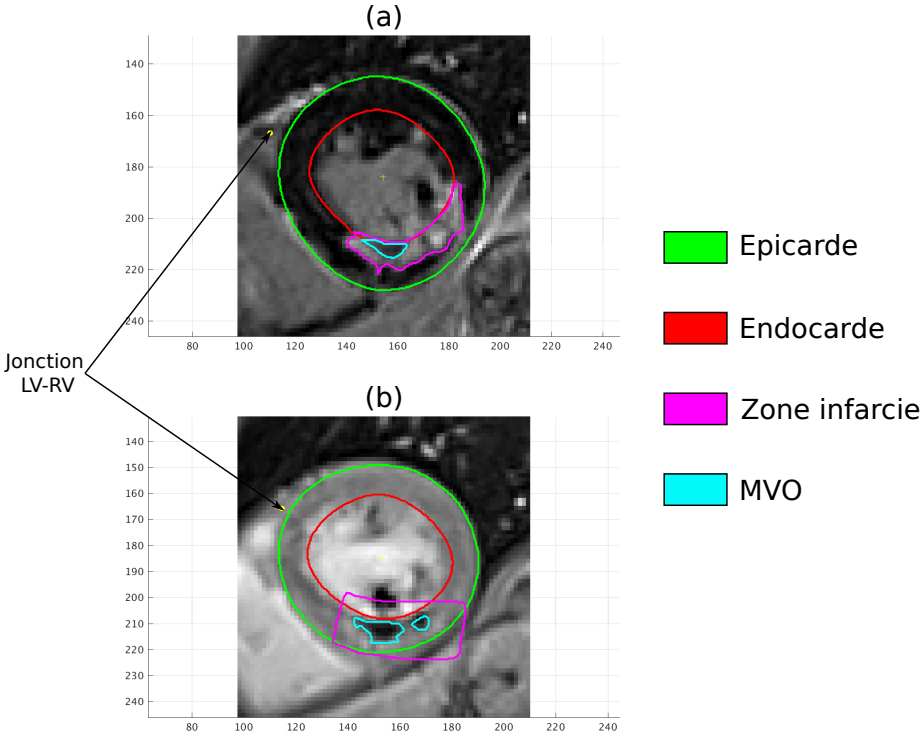


FIGURE 4.1 – Exemple des segmentations d'une coupe à mi-cavité (a) LGE et (b) EGE chez un patient issu de l'étude MIMI. Le point jaune représente la jonction entre le ventricule gauche (VG) et le ventricule droit (VD) au niveau de la paroi antérieure. L'épicarde est représenté en vert, l'endocarde en rouge. La zone infarcie est indiquée précisément sur la modalité LGE mais n'est pas disponible sur la modalité EGE (on représente en violet la zone d'intérêt correspondante). Le MVO est représenté en cyan sur les deux modalités. On peut noter sur ces segmentations que la zone du MVO est légèrement plus importante sur l'image EGE.

4.1.4 Caractéristiques de la population

Les infarctus des individus de la base de données MIMI sont liés aux trois principales artères coronaires. La répartition des individus selon leur traitement et l'artère d'origine de leur infarctus est donnée dans le Tableau 4.1.

Nous utilisons également une représentation "aplatie" du ventricule gauche similaire au Bull's eye généralement découpé en 17 segments AHA pour une analyse régionale (voir Figure 1.6). Néanmoins dans notre cas, nous sommes capables de proposer un Bull's eye bien plus détaillé jusqu'à l'échelle des pixels (seule la direction radiale est condensée). La Figure 4.3 utilise cette représentation pour résumer la répartition des infarctus au sein de la population MIMI, en séparant les sujets selon le territoire coronaire concerné.

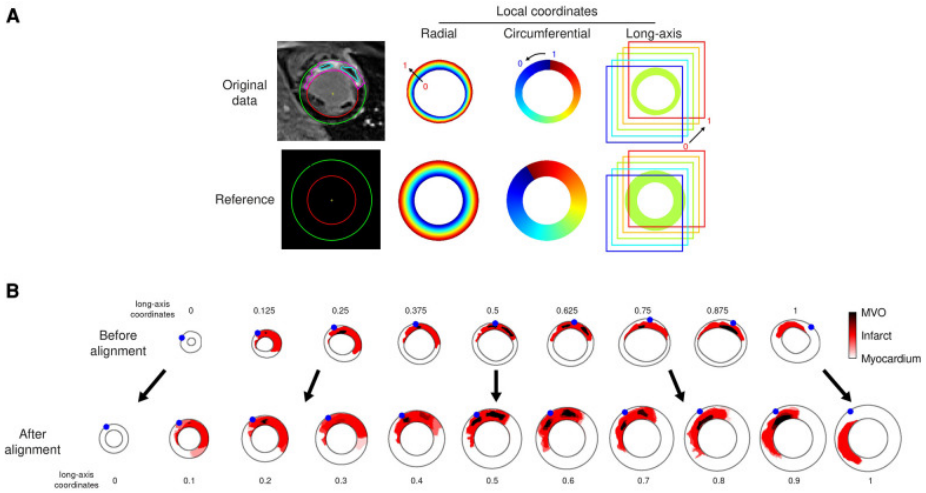


FIGURE 4.2 – Illustration de la stratégie d'alignement utilisée sur les données de l'étude MIMI. Image fournie par N. Duchateau et tirée de [80]. (A) Coordonnées radiales, circonférentielles et grand-axe d'une coupe réelle et celles de la géométrie de référence. (B) Alignement de toutes les coupes d'un individu sur la géométrie de référence.

La zone infarcie est en général large, étendue, et de forme relativement régulière. À l'inverse, les motifs de MVO sont petits, irréguliers, et se recouvrent rarement. Comme décrit dans les chapitres précédents (Chapitres 2 et 3), la métrique des méthodes d'apprentissage de représentation utilisées dans cette thèse est la distance Euclidienne entre paires de pixels, assez limitée par rapport à la façon dont opèrent les réseaux de neurones convolutionnels. Cette limite est encore plus critique si nous analysons directement les images de segmentations, car si 2 pixels de MVO ne sont pas parfaitement superposés, ils sont considérés comme éloignés.

Dans le cadre de cette thèse, pour dépasser ces limites mais également dans le contexte de techniques pouvant être gourmandes en quantité de données pour bien représenter une population (comme l'apprentissage de représentation), nous avons notamment :

- Diminué les facteurs non pertinents de variabilité dans les données, au-delà de l'alignement spatial précédemment décrit en Section 4.1.3. Pour cela, nous avons notamment réorienté les images pour enlever la variabilité liée à la position du territoire coronaire. Pour ce faire, nous avons calculé la position moyenne de l'infarctus pour chaque sujet, puis sa moyenne par sous-groupe (donc par territoire coronaire), et tourné l'ensemble des données d'un sous-groupe le long de la circonférence pour que les moyennes des sous-groupes soient toutes en correspondance sur celle du territoire LAD. Du point de vue du MVO, cela permet d'augmenter artificiellement

Artère responsable	Pose de stent immédiate	Pose de stent retardée	Total
LAD mid	11	14	25
LAD proximal	12	8	20
LCX	8	9	17
RCA	34	27	61
TOTAL	65	58	**123**

TABLE 4.1 – Répartition de la population étudiée selon le type de traitement et l'artère coronaire responsable.

le recouvrement des motifs de MVO, et dans le même temps d'augmenter le nombre de pixels présentant une variabilité suffisante pour les analyses par GP-LVM (voir expérience sur ce point en Section 3.1.3).

- Augmenté la quantité d'échantillons, par exemple en exploitant de façon indépendante toutes les coupes 2D d'un même sujet, et non plus des piles 3D de coupes. Nous pouvons ainsi atteindre 2500 coupes 2D, donnant lieu à 1726 coupes exploitables pour l'analyse si l'on omet les coupes comportant des valeurs manquantes (par exemple, localement au niveau du départ de l'aorte sur les coupes basales), l'utilisation de la valeur *NaN / Not A Number* pouvant être difficile à mettre en place en pratique dans les méthodes d'apprentissage que nous avons exploitées.

Ce jeu de données (coupes 2D avec alignement des barycentres) a été utilisé pour les applications des Chapitres 5 et 6.

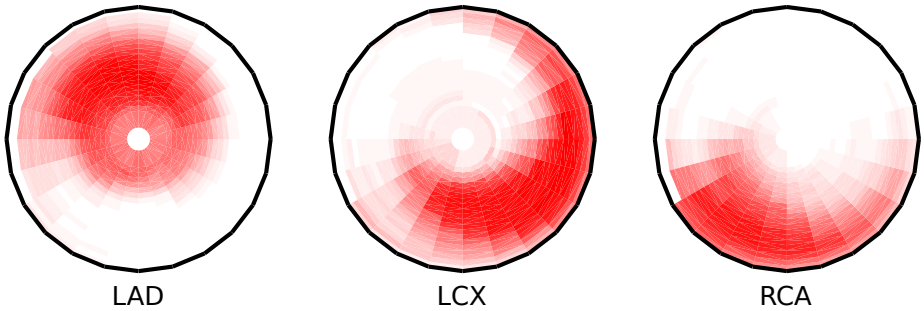


FIGURE 4.3 – Moyennes linéaires de la quantité d'infarctus provenant des différents territoires au sein de la base de données MIMI [14] utilisée dans cette thèse. Les motifs de lésions sont cohérents avec les segments AHA concernés (Figure 1.6) mais présentent une distribution spatiale plus complexe et allant au-delà du découpage en 17 segments, encourageant les analyses plus détaillées que nous effectuons. Image adaptée de [80].

La Figure 4.4 récapitule les différentes opérations réalisées pour le prétraitement des données issues de la cohorte MIMI, et les jeux de données ainsi à disposition.

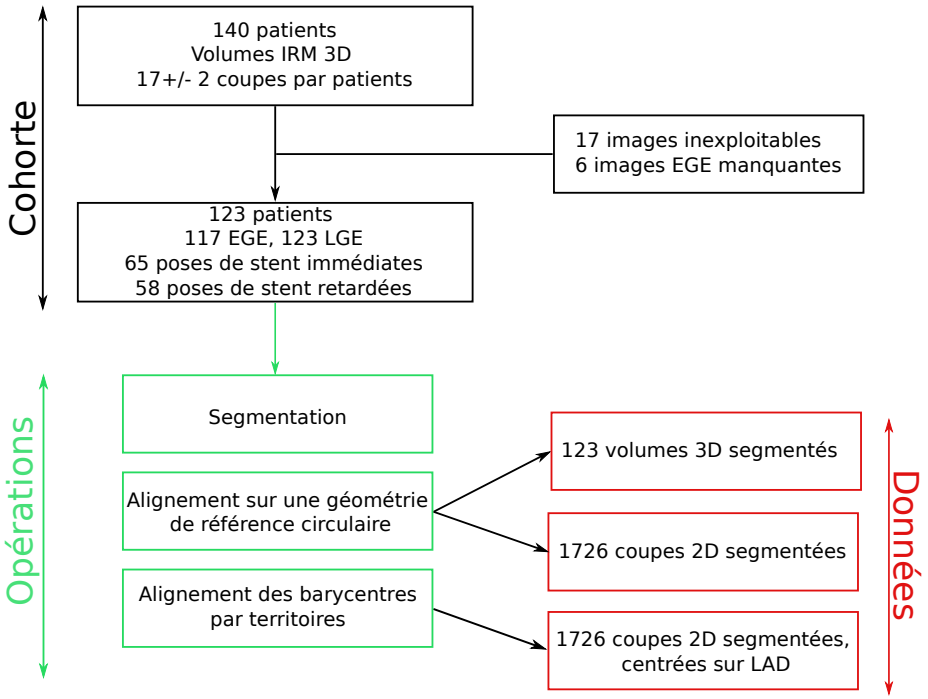


FIGURE 4.4 – Récapitulatif du pré-traitement des données de la cohorte MIMI. Les volumes sont segmentés, puis alignés sur une géométrie de référence. Les coupes sont ensuite tournées afin d'aligner les barycentres des infarctus provenant d'artères différentes.

La Figure 4.5 résume les 5 types de descripteurs de haute dimension que nous avons exploités une fois ces opérations effectuées. Les images présentées sur cette figure sont issues de la même coupe, du même patient.

Les chapitres suivants approfondissent l'analyse de ces données avec des techniques d'apprentissage de représentation non linéaires (apprentissage de variétés et GP-LVMs), plus adaptés pour ce type de données de haute dimension, comme synthétisé en Figure 1.9.

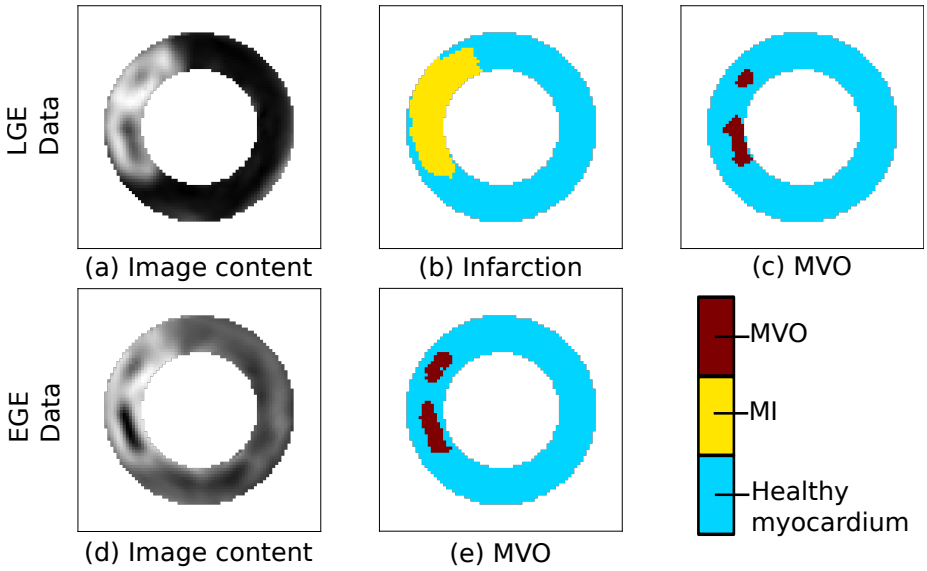


FIGURE 4.5 – Les différents types d'images disponibles pour l'analyse à partir des acquisitions d'IRM cardiaque à réhaussement. Seul le contenu du myocarde est affiché, après alignement sur une géométrie de référence. Première ligne : Données réhaussement au Gadolinium tardif (LGE) : (a) image LGE, (b) segmentation de l'infarctus en jaune, (c) segmentation du MVO en rouge foncé. Deuxième ligne : Données réhaussement au Gadolinium précoce (EGE) : (d) image EGE, (e) segmentation du MVO en rouge foncé.

Chapitre 5

Apprentissage de variété hiérarchique

Adapté de :
Freiche B, Clarysse P, Viallon M, Croisille P, Duchateau N. Characterizing myocardial ischemia and reperfusion patterns with hierarchical manifold learning. Proc. Statistical Atlases and Computational Models of the Heart (STACOM), MICCAI'21 Workshop, LNCS 2022 ;13131 :66-74.

5.1 Contexte

Les mécanismes ischémiques qui suivent l'obstruction d'une artère coronaire peuvent entraîner des lésions myocardiques structurelles et fonctionnelles. Dans le cas de l'infarctus aigu du myocarde, les bénéfices des traitements qui rétablissent la circulation coronaire sont contrebalancés par les lésions de reperfusion potentielles (obstruction microvasculaire, MVO) dues à un reflux sanguin soudain dans des zones qui en étaient privées [81]. Dans ce contexte, l'imagerie, et en particulier l'imagerie par résonance magnétique cardiaque, joue un rôle crucial pour comprendre les mécanismes d'ischémie-reperfusion [13]. Cependant, en raison d'outils d'analyse limités, la richesse des images acquises est sous-exploitée dans l'analyse clinique. Les caractéristiques des lésions sont limitées à de simples descripteurs scalaires (étendue et transmuralité, principalement) [82], et le contenu des images n'est pas exploité (par exemple, l'hétérogénéité des valeurs des pixels au sein des lésions segmentées).

L'apprentissage de représentations offre des outils efficaces pour une meilleure caractérisation des motifs des lésions à l'intérieur d'une population. Il permet de cartographier les données de haute dimension (par exemple, les images) en une représentation latente simplifiée qui facilite l'analyse des tendances individuelles ou de sous-groupes. Dans ce domaine, l'apprentissage de variétés offre un cadre solide dans lequel les distances statistiques dans l'espace latent peuvent être exploitées pour une telle mise en correspondance. Il suppose que les échantillons

d'entrée se situent sur une variété mathématique (non linéaire) qui est inconnue mais qui peut être apprise à partir des données.

Cependant, l'analyse du contenu des images à l'intérieur du myocarde n'est pas simple. Par exemple, dans les images de réhaussement tardif (LGE), le MVO est représentée par des zones sombres à l'intérieur de régions claires et plus grandes correspondant à l'infarctus. Les niveaux de gris du MVO et des tissus sains peuvent être proches (voir Figure 5.1), ce qui peut engendrer des erreurs dans la comparaison d'images à partir d'une métrique simple, en particulier pour les MVO de grande taille. D'autres problèmes critiques peuvent survenir en cas d'artefacts présents sur les images. Cette analyse pourrait être considérablement plus robuste si l'on utilisait des informations supplémentaires issues des images, jusqu'aux images mêmes (et non leurs segmentations ou des indices scalaires extraits de ces images).

Pour fusionner les informations provenant de différents descripteurs d'imagerie, plusieurs stratégies de fusion dans le domaine de l'apprentissage de variétés ont été décrites au Chapitre 2 de ce document. On peut citer par exemple l'apprentissage à noyaux multiples (MKL) [66] et la fusion de réseaux de similarité (SNF) [74]. Néanmoins, ces méthodes effectuent une fusion simultanée de tous les descripteurs, ce qui peut être sous-optimal dans notre contexte. Un meilleur schéma d'intégration possible consiste en un processus d'apprentissage hiérarchique, visant à guider l'analyse d'une population basée sur un descripteur par une analyse préalablement faite sur un autre descripteur, représentant un niveau inférieur dans la hiérarchie. Cette approche fait déjà partie du raisonnement clinique standard comme dans le cas des arbres de décision cliniques [83]. Elle est aussi intégrée dans les outils d'aide à la décision, par exemple via l'utilisation de forêts aléatoires [84]. Cependant, ce type d'apprentissage est difficilement extensible à de multiples descripteurs à haute dimension à partir d'images. *Bhatia et al.* [85] ont proposé un schéma d'apprentissage de variétés hiérarchique intéressant pour réaliser cette intégration de données multiples, mais celui-ci n'a été exploité que pour étudier différentes résolutions d'une unique modalité d'images médicales.

Dans ce travail, nous souhaitons intégrer les données de plusieurs descripteurs extraits d'images médicales de manière hiérarchique, et estimer ainsi une représentation unique pour une population de patients. Du point de vue applicatif, nous visons à améliorer l'analyse de l'hétérogénéité des tissus dans les images LGE grâce à la connaissance préalable des lésions segmentées, via une stratégie pouvant être considérée comme une manière hiérarchique d'estimer un espace latent. Nous proposons d'utiliser le cadre de l'apprentissage de variétés décrit au Chapitre 2 de façon hiérarchique, de telle sorte que l'intégration d'un descripteur de niveau supérieur (ici les images LGE) soit guidée par celle d'un descripteur de niveau inférieur (ici, les zones d'infarctus et de MVO segmentées sur ces mêmes images). La stratégie que nous employons étant non-supervisée, nous concevons également deux méthodes de sélection des hyperparamètres pertinents a posteriori. Nous démontrons la pertinence de cette approche sur la population MIMI décrite au Chapitre 4 afin d'améliorer l'analyse des patrons de lésions du myocarde au-delà des segmentations actuellement utilisées.

5.2 Méthodes

5.2.1 Définition du problème d'apprentissage de représentations hiérarchique

Pour cette application, nous construisons un modèle hiérarchique à deux niveaux, où $\mathbf{y}_i^{(0)}$ représente le i-ème échantillon du niveau inférieur (niveau "parent" ; il correspond à l'image des lésions segmentées, où les valeurs des pixels se situent dans l'intervalle [0,2], avec 0, 1 et 2 représentant respectivement le myocarde sain, l'infarctus et le MVO), et $\mathbf{y}_i^{(1)}$ correspond au même échantillon du niveau supérieur (niveau "enfant" ; l'image LGE en nuances de gris). Nous cherchons à estimer l'espace latent du niveau supérieur/enfant $\mathbf{X}^{(1)} = [\mathbf{x}_i^{(1)}]_{i\in[1,N]}$ guidé par l'espace latent du niveau inférieur $\mathbf{X}^{(0)} = [\mathbf{x}_i^{(0)}]_{i\in[1,N]}$, avec N le nombre d'échantillons.

5.2.2 Méthode spectrale

Dans ce travail, l'apprentissage de variétés est inspiré des *Diffusion maps* [28] (voir Section 2.2.2), méthode qui a servi de base aux stratégies de fusion [66], [74] et hiérarchiques [85] existant dans la littérature).

Pour rappel, pour chaque niveau $m = \{0, 1\}$ de la hiérarchie, les affinités par paire entre les individus sont encodées dans la matrice $\mathbf{W}^{(m)} = [W_{ij}^{(m)}] \in \mathbb{R}^{N\times N}$ définie comme suit :

$$W_{ij}^{(m)} = \begin{cases} \exp(-\frac{\|\mathbf{x}_i^{(m)} - \mathbf{x}_j^{(m)}\|^2}{2\sigma^2}) & \text{si } j \in \mathcal{N}_k(i), \\ 0 & \text{sinon,} \end{cases} \tag{5.1}$$

où $\mathcal{N}_k(i)$ représente le voisinage du i-ième échantillon, sur la base des k échantillons les plus proches. Le Laplacien du graphe est défini à partir de cette matrice comme $\mathbf{L}^{(m)} = \mathbf{D}^{(m)} - \mathbf{W}^{(m)}$, où $\mathbf{D}^{(m)}$ est une matrice diagonale telle que $D_{ii}^{(m)} = \sum_j W_{ij}^{(m)}$.

Comme décrit au Chapitre 2 en sous-Section 2.2.2, l'algorithme des *Diffusion maps* consiste à effectuer la décomposition spectrale de ce Laplacien pour estimer l'espace latent $\mathbf{X}^{(m)}$. En pratique, on y parvient en diagonalisant la matrice $\tilde{\mathbf{W}}^{(m)} = (\mathbf{D}^{(m)})^{-\frac{1}{2}}\mathbf{W}^{(m)}(\mathbf{D}^{(m)})^{-\frac{1}{2}}$, ce qui correspond à travailler avec le Laplacien normalisé du graphe. La matrice $\tilde{\mathbf{W}}^{(m)}$ peut être considérée comme une matrice de chaîne de Markov, où $\tilde{W}_{ij}^{(m)}$ représente la probabilité de passer de l'échantillon i à j en une étape d'une marche aléatoire sur le graphe [28]. L'espace latent correspond aux vecteurs propres associés aux plus grandes valeurs propres de $\tilde{\mathbf{W}}^{(m)}$, après suppression du cas trivial associé à la valeur propre 1. Il s'agit des principales directions de diffusion à travers la variété, approximée par le graphe constitué des échantillons disponibles.

5.2.3 Apprentissage de variétés hiérarchique

L'intégration hiérarchique proposée dans *Bhatia et al.* [85] s'appuie sur ce cadre et minimise la fonction de coût suivante afin d'estimer $\mathbf{X}^{(1)}$ à partir de $\mathbf{X}^{(0)}$:

$$\underset{\mathbf{X}^{(1)}}{\arg\min}\ (1-\mu)\sum_i\sum_j \|\mathbf{x}_i^{(0)}-\mathbf{x}_j^{(1)}\|^2 W_{ij}^{(1)} + \mu\sum_i \|\mathbf{x}_i^{(1)}-\mathbf{x}_i^{(0)}\|^2, \qquad (5.2)$$

où $\mathbf{X}^{(0)} = [\mathbf{x}_i^{(0)}]$ a été précédemment estimé en appliquant la méthode des *Diffusion maps* aux données du niveau inférieur, et $\mu \in [0,1]$ équilibre les contributions des niveaux supérieur et inférieur. Si $\mu = 1$, la solution optimale est $\mathbf{X}^{(1)} = \mathbf{X}^{(0)}$, de sorte que l'espace latent hiérarchique est dans ce cas l'espace latent du niveau inférieur. Si $\mu = 0$, cela correspond à l'application des *Diffusion maps* au niveau supérieur uniquement. Dans leur article, *Bhatia et al.* ont montré qu'il existe une solution analytique à la minimisation de cette fonction de coût pour $\mu \neq 0$:

$$\mathbf{X}^{(1)} = (\mu\mathbf{I} + 2(1-\mu)\mathbf{L}^{(1)})^{-1}\mu\mathbf{X}^{(0)}, \qquad (5.3)$$

où $\mathbf{I}$ est la matrice identité.

5.2.4 Optimisation des hyperparamètres

Bhatia et al. fixent arbitrairement le paramètre de pondération μ. De notre côté, nous proposons deux stratégies complémentaires pour trouver le meilleur espace latent pour notre application.

- Tout d'abord, nous avons calculé a posteriori chaque terme de la fonction d'énergie (Eq. 5.2) et défini la valeur optimale μ comme la valeur pour laquelle les deux termes sont équilibrés (Fig. 5.2b).
- En outre, nous avons quantifié les distances point à point entre l'espace latent hiérarchique $\mathbf{X}^{(1)}$ et les espaces estimés pour les niveaux supérieurs et inférieurs, considérés indépendamment. Pour réduire le biais dans les distances, nous avons remis à l'échelle les espaces latents de manière globale afin que les écarts types le long de leur première dimension se correspondent, et nous avons déterminé le signe des vecteurs propres qui produisaient la meilleure correspondance. Le μ optimal utilisant cette seconde stratégie correspond à un espace latent à égale distance des niveaux supérieur et inférieur (voir Figure 5.2c).

Nous avons implémenté la méthode en Python 3.7.6. L'ensemble de l'algorithme a été calculé sur un ordinateur portable standard en quelques secondes. La partie la plus contraignante est le calcul des matrices d'affinité pour l'ensemble des images car il nécessite de calculer les distances paires à paires entre les individus de la base de données. La partie hiérarchique de l'algorithme (Équation (5.2)) est très rapide car elle ne nécessite aucune étape d'optimisation.

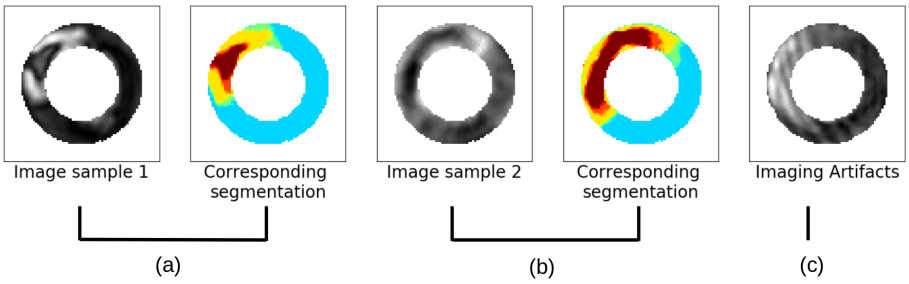


FIGURE 5.1 – Paires représentatives d'images LGE et de la segmentation correspondante. La MVO et l'infarctus correspondent respectivement aux régions sombres/rouges et claires/jaunes dans les images en niveaux de gris/segmentées. Quelques exemples difficiles sont présentés : un très grand MVO qui couvre la majeure partie de la lésion et pourrait être confondu avec du myocarde sain (b), et une coupe avec des artéfacts typiques de l'IRM (c). À noter que pendant l'étape d'alignement des données d'imagerie vers une référence commune, les contenus d'imagerie peuvent avoir été interpolés à l'intérieur d'une coupe et entre les coupes, ce qui conduit à des valeurs non catégoriques pour les images segmentées (pas strictement bleu/jaune/rouge).

5.3 Expériences et résultats

5.3.1 Données

Pour ces expériences, nous avons utilisé les données de la cohorte MIMI [14] décrites au Chapitre 4. Nous avons ici utilisé les segmentations du MVO et de l'infarctus (obtenues à partir des images LGE), afin d'exploiter les images brutes LGE, couramment sous-utilisées en clinique. Nous avons effectué nos expériences sur les données 2D alignés sur le barycentre de l'artère LAD, et alignées sur une même géométrie pour permettre la comparaison pixel à pixel (voir Chapitre 4). Nous avions au total 1711 coupes pour chaque modalité.

5.3.2 Organisation de l'espace latent

Nous avons d'abord appliqué l'algorithme *Diffusion Maps* aux images segmentées dans lesquelles l'infarctus et le MVO sont étiquetés, ce qui a permis d'obtenir l'espace latent de niveau parent $\mathbf{X}^{(0)}$ (Fig. 5.2a-b-c, colonne de gauche). Nous avons ensuite calculé la matrice d'affinité associée au contenu de l'image LGE pour définir le Laplacien du graphe $\mathbf{L}^{(1)}$, et estimé l'espace latent de niveau enfant $\mathbf{X}^{(1)}$ à partir de l'équation 5.3 pour plusieurs valeurs de μ couvrant l'intervalle $[0, 1]$. À des fins de comparaison, la méthode des *Diffusion maps* a également été appliquée directement aux images LGE (Fig. 5.2a-b-c, colonne de droite).

La largeur σ du noyau impliqué dans les matrices d'affinité $\mathbf{W}^{(m)}$ a été

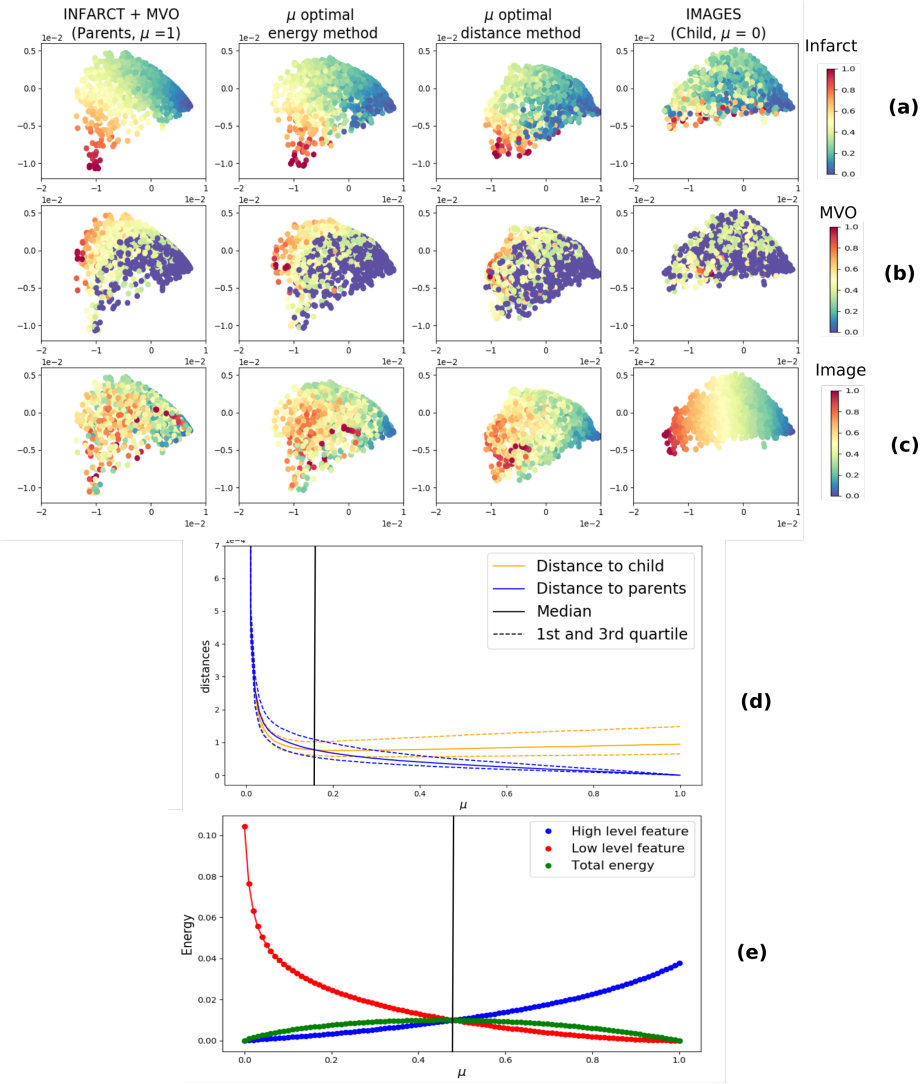


FIGURE 5.2 – Trois premières lignes (a-b-c) : Espaces latents obtenus pour les données du parent (colonne de gauche) et de l'enfant (colonne de droite) indépendamment, et pour l'espace latent hiérarchique optimal (colonnes du milieu). Les deux premières dimensions sont affichées et colorées en fonction de l'infarctus (a), de la taille de la MVO (les points violets correspondent aux coupes sans MVO) (b) ou de la valeur moyenne des pixels dans les images (c). (d-e) : Les deux μ optimaux (ligne noire verticale) ont été obtenus à partir du croisement des courbes de distance (d) ou des courbes d'énergie (e), comme expliqué dans la Section 5.2.4

fixée expérimentalement pour chaque espace latent. Dans la littérature, elle est généralement définie comme la distance moyenne entre un échantillon et son k-ième plus proche voisin. Cependant, ce choix n'était pas pertinent dans notre cas, en particulier pour les données MVO. Par exemple, si σ est trop petit, les directions principales de diffusion peuvent être fortement affectées par un échantillon spécifique. Inversement, si σ est trop grand, il peut conduire à une compression des espaces latents vers zéro. Nous avons donc fixé de manière heuristique σ à la valeur la plus basse permettant une diffusion significative des données dans l'espace latent (dans notre cas, $\sigma = 35$ pour les données de segmentation, et $\sigma = 25$ pour l'intégration des données image).

Comme le montrent les deux colonnes centrales de la Fig. 5.2a-b, les valeurs optimales de μ conduisent à des espaces latents intermédiaires auxquels contribuent à la fois les niveaux parent et enfant. Les échantillons ne sont plus entièrement organisés en fonction des caractéristiques de segmentation (quantité d'infarctus et de MVO encodée dans l'échelle de couleurs de la Fig. 5.2a-b). Elles ne sont pas non plus fortement désorganisées comme celles obtenues avec les données de l'enfant uniquement, mais contiennent une partie de l'information sur l'apparence de l'image (la valeur moyenne du pixel est encodée dans l'échelle de couleurs de la Fig. 5.2c).

Les courbes d'énergie et de distance de la Figure 5.2d-e confirment que $\mu = 1$ conduit à une correspondance avec l'espace latent des données parent. En revanche, un saut est observé à l'approche de $\mu = 0$, car l'Equation 5.3 reviendrait à $\mathbf{X}^{(1)} = 0$. Cela peut s'expliquer par la taille des espaces latents obtenus lorsque l'on s'approche de $\mu = 0$ (typiquement pour $\mu \leq 0.1$). Comme les espaces sont vraiment petits (en raison de limites numériques), leur remise à l'échelle peut être moins précise, ce qui entraîne un saut dans les courbes de distance. Sur cette population, une valeur de $\mu = 0.47$ conduit à des énergies équilibrées dans l'Eq. 5.3, bien qu'une valeur plus petite de 0.16 soit nécessaire pour obtenir un espace latent hiérarchique à égale distance des espaces latents enfant et parent. Dans notre cas, la solution basée sur l'énergie est plus proche de la structure de l'espace latent parental, tandis que celle basée sur la distance est plus proche de l'espace latent enfant.

5.3.3 Consistance des voisinages

La Figure 5.3 complète ces observations en montrant des cas représentatifs choisis dans les espaces latents. La ligne (a) montre une coupe avec un modèle d'infarctus standard contenant un petit MVO. Les quatre images et segmentations affichées à gauche de la figure représentent les quatre voisins les plus proches dans l'espace latent parent, tandis que celles de droite représentent les quatre voisins de l'espace latent hiérarchique (solution à énergie équilibrée). Nous observons que les voisins des espaces latents parent et hiérarchique ont des segmentations proches de la coupe analysée. Toutefois, les images issues de la méthode hiérarchique sont plus proches (par rapport à l'espace latent parent) du sujet original. Cela signifie que l'espace latent est plus fidèle à l'apparence de l'image LGE. La ligne (b) affiche les voisins d'une coupe présentant un MVO

important qui couvre la majeure partie de l'infarctus. Dans ce cas, nous affichons à gauche les quatre voisins les plus proches de l'espace latent basée sur l'image, tandis que les images de droite correspondent aux voisins de l'espace latent hiérarchique. Nous observons ici que les voisins basés sur l'image présentent des motifs de MVO très différents, malgré l'apparence proche de l'image. En revanche, les voisins hiérarchiques sont beaucoup plus cohérents avec le patron original. Cela souligne l'intérêt de guider la représentation hiérarchique par les données segmentées, ce qui permet d'obtenir des représentations plus robustes dans le cas de contenus d'images difficiles pour des échantillons inhabituels.

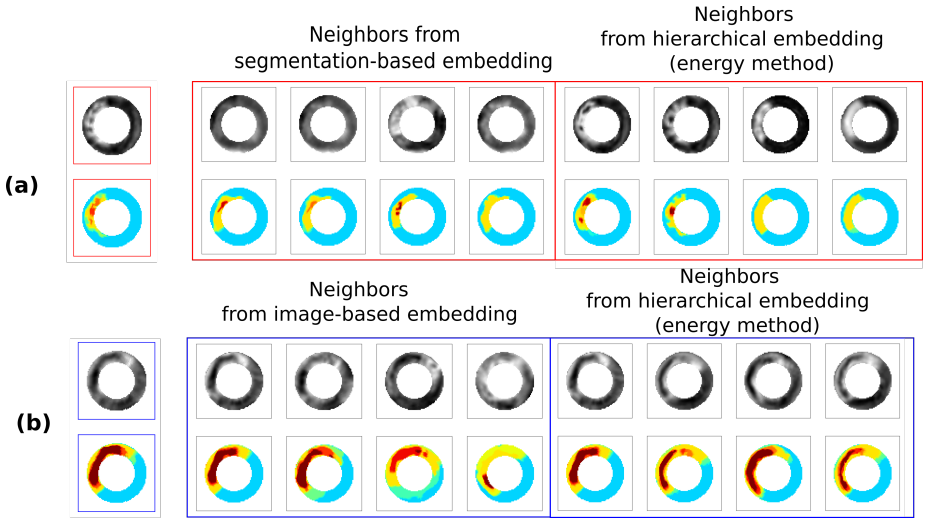


FIGURE 5.3 – Illustration de la robustesse face à des contenus d'image difficiles. La première colonne représente le contenu de l'image et la segmentation pour deux échantillons représentatifs : un infarctus standard avec un petit MVO (a) et un grand MVO couvrant la majeure partie de l'infarctus (b). Les autres colonnes montrent les quatre plus proches voisins pour ces deux cas, choisis à partir de l'espace latent basé sur la segmentation ou sur l'image, comparés à ceux de l'espace latent hiérarchique.

5.4 Conclusions et perspectives

Nous avons démontré dans cette section la pertinence d'une approche hiérarchique pour l'analyse du contenu des images LGE. Notre approche hiérarchique consiste à guider la représentation d'un contenu de niveau enfant (difficile à représenter/ images) par un contenu de niveau parent (plus facile à représenter/segmentations) correspondant aux images segmentées. Nous avons également introduit deux façons de sélectionner un paramètre de pondération pertinent μ pour équilibrer la contribution de chaque niveau dans la hiérarchie.

Notre approche s'inspire de la manière hiérarchique dont les médecins intègrent plusieurs données provenant de différentes sources, pour plus de robustesse et de confiance dans leur diagnostic. Dans notre cas, la hiérarchie permet d'intégrer les connaissances préalables correspondant à la localisation de la lésion. Elle peut être considérée comme un moyen de contourner la simplicité d'une métrique (par exemple pixel à pixel) entre les échantillons, qui peut être corrompue par des motifs de lésion spécifiques ou des artéfacts d'image. Nous n'avons utilisé que des métriques simples et un cadre d'apprentissage de variétés bien connu (*Diffusion maps*), qui peut s'avérer limité pour des ensembles de données complexes. Nos expériences ont montré qu'il est possible d'obtenir des représentations significatives même pour les cas difficiles inclus dans notre base de données.

Cet algorithme d'apprentissage hiérarchique pourrait être étendu à d'autres protocoles d'imagerie présentant plusieurs niveaux de complexité des données (par exemple, les examens typiques d'échocardiographie ou d'IRM). L'espace latent peut servir de représentation intermédiaire pour alimenter un algorithme de classification, ou comme moyen simplifié d'examiner des données complexes, comme dans notre application. Il permet d'extraire les principales dimensions de diffusion dans ces espaces de données.

De meilleures métriques d'images et des algorithmes plus puissants peuvent améliorer ces résultats, comme nous le montrons dans le Chapitre 6, mais notre but ici était de démontrer l'intérêt d'une telle hiérarchie sur une application pertinente. Les travaux présentés au Chapitre 6 exploitent d'autres modalités issues du même jeu de données, et dépassent certaines limitations de ce chapitre en utilisant les processus Gaussiens pour les modèles à variables latentes (GP-LVM) introduits au Chapitre 3.

Chapitre 6

Intégration hiérarchique de données avec des Processus Gaussiens : application à la caractérisation des motifs d'ischémie-reperfusion cardiaque

Adapté de :
Freiche B, Bernardino G, Clarysse P, Duchateau N. Hierarchical data integration with Gaussian processes : application to the characterization of cardiac ischemia-reperfusion patterns. En cours de révision à IEEE Transactions on Medical Imaging, 2023.

6.1 Contexte

Les maladies complexes peuvent rarement être expliquées par une seule source d'information (un seul type de mesure, d'acquisition d'imagerie ou même de modalité), qui n'offre que des informations partielles sur l'état du patient. Les médecins sont capables de fusionner mentalement plusieurs types d'informations pour caractériser l'état des patients, à la fois à des fins de compréhension (approche non supervisée) et de diagnostic ou de pronostic (approche supervisée). Pour plus de robustesse et d'efficacité, ils incorporent généralement ces éléments d'information de manière incrémentale, dans un ordre basé sur leur expérience (par exemple, des mesures plus simples aux mesures plus complexes ou plus coûteuses) ou sur des lignes directrices editées par des comités d'experts.

Cependant, cette fusion d'informations reste très qualitative et subjective, d'où le besoin de méthodes automatisées et reproductibles.

La fusion de ces informations demeure un défi majeur pour les approches automatisées, étant donné l'hétérogénéité des types et des dimensions des descripteurs de données, en particulier lors du traitement de données d'imagerie de haute dimension. L'intégration progressive des données (c'est-à-dire en suivant une hiérarchie dans les descripteurs de données) pourrait être très bénéfique pour limiter la charge de calcul et les instabilités potentielles par rapport aux stratégies qui fusionnent toutes les données simultanément.

Dans ce contexte, nous proposons une nouvelle approche basée sur les GP-LVMs, décrits au Chapitre 3 de ce document, pour modéliser explicitement les dépendances hiérarchiques entre les descripteurs de données et apprendre une représentation latente d'une population de manière non supervisée. Nous démontrons sa pertinence sur un problème médical à forte prévalence dans la population générale, à savoir la caractérisation des schémas d'ischémie-reperfusion cardiaque à partir d'images par RM à rehaussement différé.

6.1.1 Fusion de données pour l'apprentissage de représentations

6.1.1.1 Réduction de dimensions non supervisée

Les techniques de réduction de dimensions permettent de représenter l'information utile cachée dans les données provenant de descripteurs d'imagerie de haute dimension dans un espace de données plus simple, appelé espace latent.

Parmi les approches non supervisées, l'apprentissage de variétés réduit la dimension en exploitant la structure de l'espace de données. Il a été démontré que la plupart des méthodes linéaires et non linéaires d'apprentissage de variétés peuvent être regroupées dans un cadre commun basé sur l'utilisation des graphes d'affinité [21]. L'idée principale est de construire une matrice de similarité dont les éléments encodent une mesure significative entre les échantillons de haute dimension, puis d'effectuer la décomposition spectrale de cette matrice. Les différentes distances utilisées conduisent à des représentations distinctes : par exemple, Isomap [23] fonctionne globalement avec la distance géodésique, tandis que les *Diffusion Maps* [28] fonctionnent localement avec la distance de diffusion.

Les auto-encodeurs (AEs) sont des réseaux de neurones profonds qui estiment une représentation latente de faible dimension d'une population tout en étant capable de reconstruire de manière optimale chaque échantillon de haute dimension à partir de cet espace latent. La partie encodage est composée de couches de neurones en nombre décroissant, conduisant à un goulot d'étranglement, qui est une représentation de faible dimension d'un échantillon dans l'espace latent. A partir de l'espace latent, le décodeur reconstruit les données d'entrée de haute dimension. Les auto-encodeur variationnels (VAEs) [35] sont une extension probabiliste de AEs. Ils associent à chaque échantillon une distribution latente normale au lieu d'un vecteur de valeurs. A partir de la distribution apprise, ils construisent un espace latent et reconstruisent les données par échan-

tillonnage. Les AEs sont très populaires en analyse d'images médicales, mais ils souffrent encore de difficultés d'apprentissage et d'instabilité, ainsi que d'une interprétabilité et de propriétés statistiques limitées de l'espace latent estimé, sujets qui font l'objet de recherches actives ces dernières années [39, 86].

Les GP-LVMs [42] constituent un cadre complémentaire et prometteur pour la réduction de dimension. Ils estiment un lien régi par une distribution Gaussienne non linéaire entre les observations et l'espace latent en modélisant explicitement leur relation. Leur flexibilité permet d'inclure des connaissances préalables dans le modèle (par exemple, la dynamique [79]) et ouvre donc la voie à des extensions pertinentes aux données multimodales ou aux descripteurs multiples. Les GP-LVMs nécessitent moins de données que les réseaux neuronaux, mais ils ont du mal à s'adapter à de grands ensembles de données. Une version basée sur des réseaux de neurones, les processus Gaussiens profonds [87], a été développée pour les ensembles de données plus importants.

6.1.1.2 Fusion de données

L'utilisation de données comportant de multiples descripteurs (ou multi-descripteurs) permet d'obtenir des informations plus riches, mais présente plusieurs difficultés, principalement liées à la haute dimensionnalité et à l'hétérogénéité des descripteurs, une concaténation directe conduisant à des résultats sous-optimaux. Pour estimer une représentation de faible dimension à partir de données multi-descripteurs, deux types de méthodes ont émergé : la fusion et l'alignement [88] (cf. sous-Sections 2.3.2 et 2.3.3).

Les méthodes de fusion regroupent les différents descripteurs de données en un seul espace latent. L'apprentissage de noyaux multiples (MKL) [66] élargit le cadre général de l'apprentissage de variétés [21] en codant les affinités par paire d'échantillons dans une matrice de noyau, différente pour chaque descripteur de données, et en optimisant simultanément l'espace de faible dimension et la combinaison de ces matrices d'affinité. De manière comparable, l'algorithme des réseaux à fusion de similarité (*SNF*) [73] fusionne ces matrices d'affinité par un processus de diffusion itératif qui revient à estimer un seul graphe qui représente l'ensemble des données.

Les méthodes d'alignement, quant à elles, apprennent plusieurs espaces latents (un pour chaque modalité) tout en optimisant leur mise en correspondance. Par exemple, l'apprentissage de variétés multiples (*MML*) [51, 52] code dans une matrice de blocs à la fois les affinités spécifiques à la modalité entre les échantillons (blocs diagonaux) et les affinités inter-modalités entre les échantillons (blocs extra-diagonaux, qui forcent l'alignement des espaces latents). Les auto-encodeurs variationnels multi-canaux (*MCVAEs*) [62] réalisent également l'alignement de données en généralisant le cadre de VAEs à plusieurs canaux, chacun associé à une modalité ou à un descripteur de données différent, en contrôlant le codage de chaque canal et sa reconstruction. À l'instar des MCVAEs, les VAEs partiels [89] permettent de sélectionner une variable pertinente parmi plusieurs variables d'entrées. Cette méthode vise à prédire une quantité avec le moins d'entrées possible, à l'aide d'un réseau de neurones profond ca-

pable de gérer les variables manquantes et d'apprendre à partir d'observations partielles.

Dans des travaux antérieurs, nous avons démontré la pertinence des approches de fusion [67] et d'alignement [55] pour analyser conjointement plusieurs types de descripteurs de haute dimension extraits d'images cardiaques, et exploiter la représentation latente pour caractériser des maladies cardiaques spécifiques. Toutefois, ces approches, de fusion et d'alignement, combinent simultanément tous les descripteurs de données, ce qui peut être critique en termes d'efficacité et de robustesse, en particulier pour les descripteurs de données de dimensions élevées et/ou très différentes. En revanche, une intégration hiérarchique ou incrémentale de ces données présente un fort potentiel pour surmonter ce problème.

6.1.1.3 Fusion de données hiérarchique

Dans les approches basées sur les noyaux, des processus itératifs ont été proposés pour évoluer progressivement vers une matrice de noyau consensuelle [90], [91], mais ces méthodes n'ont été utilisées que pour l'incorporation incrémentale de nouveaux échantillons (un concept proche de l'apprentissage de curriculum [92]) et non pour l'analyse de données multimodales. Dans le cadre de l'apprentissage de variétés, des contraintes proches de celles utilisées pour l'alignement de variétés [52,54] ont été proposées pour relier des niveaux consécutifs dans une hiérarchie de données [85]. Cette méthode a notamment été expérimentée pour prendre en compte des échelles multiples pour l'analyse régionale d'images médicales.

En tant qu'approche alternative et plus générique, les GP-LVMs peuvent être adaptés pour encoder explicitement les relations hiérarchiques entre les différents niveaux de données. Dans [79], les auteurs ont d'abord illustré cette approche sur des séries temporelles de mouvements de *stickman* (dessins d'hommes-bâtons) avec une hiérarchie en deux étapes où le temps est utilisé comme premier niveau pour guider la représentation de la trajectoire du *stickman*. Ils ont également développé cette idée pour guider la représentation d'un corps articulé en mouvement en modélisant explicitement le mouvement d'organes distincts en tant que niveaux inférieurs dans la hiérarchie.

Dans des travaux antérieurs, nous avons démontré la pertinence d'une hiérarchie en deux étapes basée sur l'apprentissage de variétés [85] pour guider l'analyse de l'apparence des images dans des données IRM [46]. Dans ce chapitre, nous proposons une généralisation originale basée sur les processus Gaussiens, qui est à la fois intuitive et offre une flexibilité pour l'intégration de données multiples suivant une hiérarchie prescrite.

6.1.2 Caractérisation des motifs d'infarctus aigü du myocarde

L'infarctus aigu du myocarde survient lorsque le sang provenant d'une artère coronaire est insuffisant pour répondre aux besoins en oxygène du myocarde, gé-

néralement en raison d'une obstruction des artères coronaires. Si l'ischémie se prolonge, elle entraîne la mort du myocarde, ce qui déclenche un remodelage affectant la forme et la fonction cardiaques. Cela peut finalement entraîner des complications graves, voire la mort du patient. Lorsqu'il est détecté suffisamment tôt, le cœur du patient est reperfusé (revascularisé, par exemple par angioplastie). Le rétablissement de l'approvisionnement en sang limitera idéalement l'expansion de l'infarctus du myocarde, mais cela peut entrainer une inflammation grave en raison de ce brusque apport d'oxygène. Ces dommages causés par la reperfusion sont appelés micro-obstruction vasculaires (*Microvascular obstruction*, MVO). La compréhension des mécanismes d'ischémie-reperfusion est un sujet de recherche clinique majeur [16] pour décider en fin de compte d'une thérapie de revascularisation adaptée et surveiller l'évolution de l'état du patient.

L'imagerie médicale est très utile pour étudier les lésions ischémiques et le phénomène de non reflux (*no-reflow*), en utilisant généralement l'IRM [13, 93]. Le protocole d'IRM commence par l'acquisition de cartes T1 et T2 natives. Ensuite, un agent de contraste (Gadolinium) est injecté au patient. Environ 3-4 minutes après l'injection de l'agent de contraste, des images EGE (*Early Gadolinium Enhancement*) sont acquises pour évaluer les MVOs précoces. Dix minutes après l'injection de l'agent de contraste, les images LGE (*Late Gadolinium Enhancement*) sont acquises pour évaluer l'infarctus du myocarde infarctus du myocarde (MI) et les MVO tardifs (Fig. 6.1). Malheureusement, les médecins manquent d'outils pour exploiter la richesse des informations disponibles dans les images. Ils simplifient les images en segmentant le myocarde et ses lésions (sans tenir compte de la richesse potentielle de l'apparence de l'image), et se contentent généralement de simples mesures scalaires extraites des segmentations (taille de l'infarctus et transmuralité, principalement). Cela limite considérablement la compréhension des mécanismes complexes en jeu. En outre, il convient d'être particulièrement prudent lors de l'analyse d'images médicales appartenant à une variété non linéaire, comme c'est le cas pour les données d'infarctus (MI) et d'obstruction micro-vasculaire (MVO) (Fig. 6.2).

Pour améliorer sensiblement la compréhension des mécanismes d'ischémie-reperfusion, il est nécessaire (i) d'intégrer des données d'imagerie complexes dans l'analyse de populations de patients atteints d'infarctus du myocarde, ce qui peut être réalisé par l'apprentissage de représentations pertinentes, et (ii) d'intégrer plusieurs types d'acquisitions ou de descripteurs de données complexes extraits des images, potentiellement à différents niveaux de complexité. Nous abordons ces deux aspects dans ce chapitre au moyen d'une méthode originale d'apprentissage de représentations explicitement conçue pour prendre en compte les hiérarchies entre différents descripteurs.

6.1.3 Approche proposée et contribution

Dans ce travail, nous proposons un cadre original pour l'intégration hiérarchique de plusieurs types de données extraites d'images médicales, appliqué à la caractérisation des lésions d'infarctus et de reperfusion à partir de données IRM.

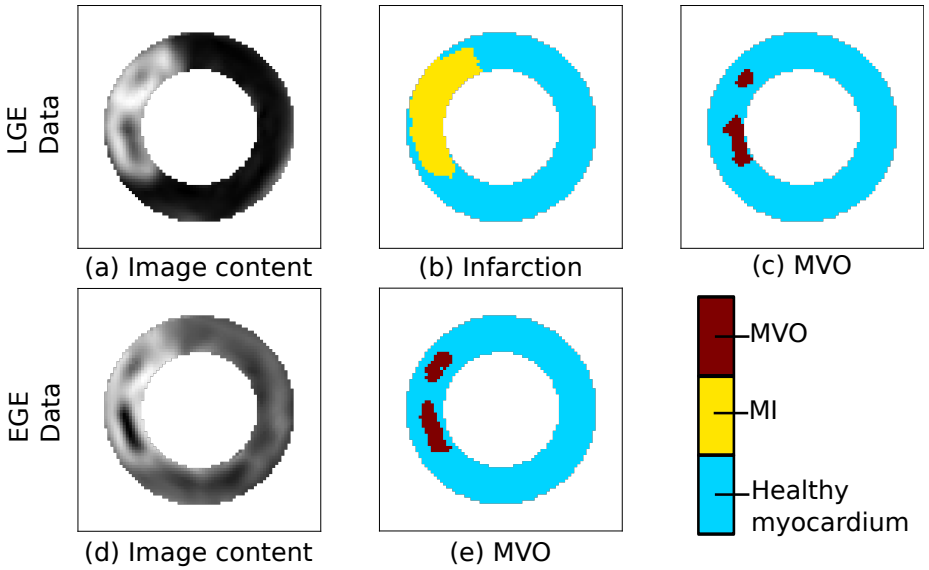
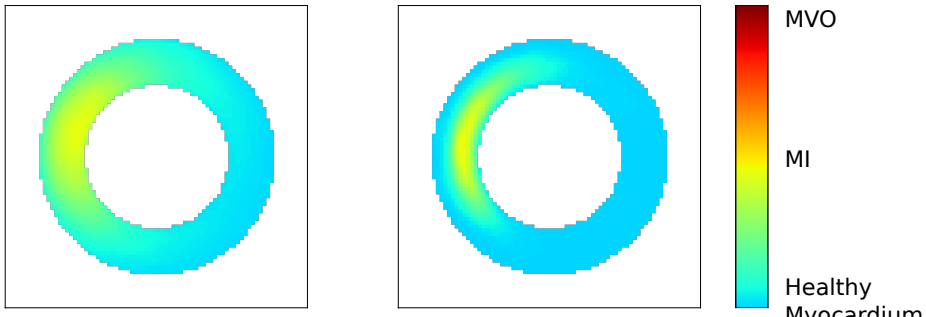


FIGURE 6.1 – Les différents types d'images que nous sommes amenés à analyser avec l'IRM cardiaque à réhaussement différé. Seuls le contenu du myocarde est représenté, sur la même géométrie, décrite au Chapitre 4. Première ligne : donnée LGE : (a) image LGE, (b) segmentation de l'infarctus MI en jaune, (c) segmentation du MVO en rouge foncé. Deuxième ligne : donnée EGE : (d) image EGE, (e) segmentation du MVO en rouge foncé. Le myocarde sain est représenté sur les segmentations en bleu clair.

Ce travail peut être considéré comme une généralisation substantielle de notre travail précédent [46] (cf. Chapitre 5), qui a démontré que l'intégration hiérarchique des données (avec de l'apprentissage de variétés classique) est pertinente pour une telle application.

Ici, nous nous appuyons sur des processus Gaussiens pour modéliser explicitement les dépendances entre les descripteurs de données, tout en considérant également le lien entre la représentation latente et les observations, ainsi que les distributions a priori sur les espaces latents. Plus précisément, nous étendons le cadre des processus hiérarchiques GP-LVMs [79] présenté au Chapitre 3 pour construire de telles hiérarchies de données. Notre approche comprend notamment un processus Gaussien additionnel qui sert de terme de régularisation pour les hiérarchies sur des données réelles difficiles. Nous démontrons que cette méthodologie permet des représentations plus fiables et plus robustes, par rapport à des variantes plus simples de la hiérarchie et à des représentations à descripteur unique.



(a) Average MI pattern (b) Average MVO pattern

FIGURE 6.2 – Moyenne linéaire des patrons de lésion ((a) MI et (b) MVO) pour tous les patients de la base de données analysée (toutes les lésions ont été réorientées vers le même territoire, celui de l'artère coronaire droite (LAD)). Le motif moyen est flou car les images MI et MVO proviennent d'une variété non linéaire. Ce n'est peut-être pas critique pour l'infarctus MI, mais le motif de MVO ne ressemble pas du tout à une lésion réelle de MVO. Cette Figure souligne la nécessité d'une approche non linéaire pour la réduction de la dimensionnalité.

6.2 Méthodes

6.2.1 Données de l'étude

6.2.1.1 Données images

Nous nous concentrons sur les lésions myocardiques dues aux lésions d'ischémie et de reperfusion dans le contexte de l'infarctus aigü du myocarde. Plus précisément, nous considérons les données d'un essai clinique anonymisé pour lequel des images par RM de réhaussement précoce et tardif étaient disponibles : l'étude Minimalist Immediate Mechanical Intervention (MIMI) [14] (ClinicalTrials ID : NCT01360242). Le protocole de l'étude a été approuvé par le comité d'éthique local (IRB 2010-048), et est conforme à la Déclaration d'Helsinki et aux lois françaises. Tous les sujets ont donné leur consentement éclairé par écrit. Cette étude a été conçue à l'origine pour comparer deux stratégies de reperfusion : le stenting (pose d'un stent) différé et le stenting immédiat, le "délai" étant de 24-48 heures. L'acquisition des images a eu lieu 4 à 6 jours après l'intervention médicale, de sorte que le MVO est encore visible sur les images EGE et LGE, et que la forme de la cavité cardiaque ne s'est pas encore remodelée, ce qui limite les facteurs externes susceptibles d'affecter l'analyse.

Notre ensemble de données comprenait les images IRM de 123 patients, avec une résolution médiane de $1,5625 \times 1,5625 \times 5$ mm, sur lesquelles le ventricule gauche (VG) s'étendait sur 17 ± 2 coupes. Ces images ont été recadrées autour du VG et redimensionnées à 80×80 pixels. Les segmentations des zones MI et MVO ont été obtenues de manière semi-automatique à partir d'un logiciel commercial (CVI42 v.5.1.0 Circle Cardiovascular Imaging, Calgary, Canada) par un

observateur expérimenté et contrôlées par deux autres observateurs expérimentés.

6.2.1.2 Alignement spatial

Nous avons identifié la jonction antérieure VG-VD sur chaque coupe, et les coupes apicale et basale sur chaque pile d'images, ce qui a permis de paramétrer le myocarde en définissant les coordonnées radiales, circonférentielles dans le plan image et le niveau de coupe en longitudinal, pour chaque individu. Nous avons ensuite utilisé cette paramétrisation pour transporter les données d'image et les segmentations vers une géométrie commune (interpolation linéaire pour les données sur une grille dispersée), afin de normaliser les données d'entrée avant l'apprentissage automatique.

Nous avons procédé à une étape supplémentaire d'alignement spatial afin de bénéficier d'une population plus importante et de nous concentrer sur la forme de l'infarctus tout en minimisant d'autres facteurs non pertinents tels que son emplacement autour du myocarde. Tous les motifs d'infarctus en 3D ont été tournés selon la circonférence de manière à ce que les centres de masse de tous les infarctus associés à un territoire coronaire donné soient alignés sur le centre de masse de l'artère descendante gauche (LAD).

6.2.1.3 Coupes analysées

Dans la suite de ce chapitre, l'analyse est effectuée sur les coupes 2D afin d'accélérer les calculs, en particulier pour les expériences visant à évaluer le rôle des hyperparamètres et à choisir une stratégie optimale. En outre, la visualisation est plus facile, un atout évident pour comparer les motifs de lésion parfois visibles sur quelques coupes (comme pour MVO). Nous avons écarté les coupes sans lésions et concentré l'analyse sur les coupes restantes, considérées comme des échantillons indépendants (pour bénéficier d'une population plus large). En conséquence, notre population se compose de 1726 échantillons associés à 123 patients.

Les motifs de lésions MI sont larges, ont des formes relativement régulières et se chevauchent partiellement sur différentes coupes. Au contraire, les motifs de MVO sont petits, irréguliers et ne se chevauchent pas. Leur analyse statistique, même avec des techniques d'apprentissage de représentation non linéaire, est donc beaucoup plus difficile. Cependant, les lésions MVO sont par principe contenues dans la région de l'infarctus MI, ce qui signifie que la représentation des motifs MVO pourrait être guidée par une représentation précédemment obtenue à partir de MI. C'est ce que nous proposons d'étudier en formulant cette relation par le biais d'un processus hiérarchique, dans le cadre des GP-LVMs.

6.2.2 GP-LVM hiérarchique

6.2.2.1 Modèle hiérarchique simple

Le schéma hiérarchique présenté au Chapitre 3 peut être adapté à différentes structures de données. Une première façon de considérer la hiérarchie entre deux descripteurs est présentée sur la Figure 6.3a. Considérons deux variables latentes $\mathbf{X}^{(0)}$ et $\mathbf{X}^{(1)}$, avec les observations correspondantes $\mathbf{Y}^{(0)}$ et $\mathbf{Y}^{(1)}$. $\mathbf{X}^{(0)}$ est ici le premier niveau de la hiérarchie, $\mathbf{X}^{(1)}$ le second. L'obtention des variables latentes revient à résoudre :

$$\underset{\mathbf{X}^{(0)},\mathbf{X}^{(1)}}{\operatorname{argmax}} \log P(\mathbf{X}^{(0)}, \mathbf{X}^{(1)} \mid \mathbf{Y}^{(0)}, \mathbf{Y}^{(1)}) \tag{6.1}$$

En utilisant de nouveau la règle de Bayes :

$$P(\mathbf{X}^{(0)}, \mathbf{X}^{(1)} \mid \mathbf{Y}^{(0)}, \mathbf{Y}^{(1)}) = \frac{P(\mathbf{Y}^{(0)}, \mathbf{Y}^{(1)} \mid \mathbf{X}^{(0)}, \mathbf{X}^{(1)}) P(\mathbf{X}^{(0)}, \mathbf{X}^{(1)})}{P(\mathbf{Y}^{(0)}, \mathbf{Y}^{(1)})}. \tag{6.2}$$

Comme $\mathbf{Y}^{(0)}$ et $\mathbf{Y}^{(1)}$ sont indépendants, nous avons également :

$$P(\mathbf{Y}^{(0)}, \mathbf{Y}^{(1)} \mid \mathbf{X}^{(0)}, \mathbf{X}^{(1)}) = P(\mathbf{Y}^{(0)} \mid \mathbf{X}^{(0)}, \mathbf{X}^{(1)}) P(\mathbf{Y}^{(1)} \mid \mathbf{X}^{(0)}, \mathbf{X}^{(1)}). \tag{6.3}$$

Comme $\mathbf{Y}^{(0)}$ est indépendant de $\mathbf{X}^{(1)}$, et $\mathbf{Y}^{(1)}$ est indépendant de $\mathbf{X}^{(0)}$ sachant $\mathbf{X}^{(1)}$:

$$P(\mathbf{Y}^{(0)} \mid \mathbf{X}^{(0)}, \mathbf{X}^{(1)}) = P(\mathbf{Y}^{(0)} \mid \mathbf{X}^{(0)}), \tag{6.4}$$

$$P(\mathbf{Y}^{(1)} \mid \mathbf{X}^{(0)}, \mathbf{X}^{(1)}) = P(\mathbf{Y}^{(1)} \mid \mathbf{X}^{(1)}). \tag{6.5}$$

De plus, nous avons :

$$P(\mathbf{X}^{(0)}, \mathbf{X}^{(1)}) = P(\mathbf{X}^{(1)} \mid \mathbf{X}^{(0)}) P(\mathbf{X}^{(0)}). \tag{6.6}$$

Finalement, la combinaison des équations (6.2), (6.3), (6.4), (6.5), (6.6) mène à :

$$P(\mathbf{X}^{(0)}, \mathbf{X}^{(1)} \mid \mathbf{Y}^{(0)}, \mathbf{Y}^{(1)}) = \frac{P(\mathbf{Y}^{(0)} \mid \mathbf{X}^{(0)}) P(\mathbf{Y}^{(1)} \mid \mathbf{X}^{(1)}) P(\mathbf{X}^{(1)} \mid \mathbf{X}^{(0)}) P(\mathbf{X}^{(0)})}{P(\mathbf{Y}^{(0)}, \mathbf{Y}^{(1)})}. \tag{6.7}$$

Comme pour le GP-LVM classique, nous définissons un a priori Gaussien sur $\mathbf{X}^{(0)}$. De plus, $P(\mathbf{Y}^{(0)}, \mathbf{Y}^{(1)})$ est constant par rapport à $\mathbf{X}^{(0)}$ et $\mathbf{X}^{(1)}$, de sorte que la résolution du problème d'optimisation (6.1) revient à résoudre conjointement chaque processus Gaussien.

6.2.2.2 Modèle hiérarchique amélioré

En pratique, le modèle présenté dans la section précédente peut être difficile à optimiser sur des données réelles. Nous avons donc apporté deux adaptations substantielles à ce modèle, ce qui a conduit au nouveau modèle illustré à la Fig. 6.3b :

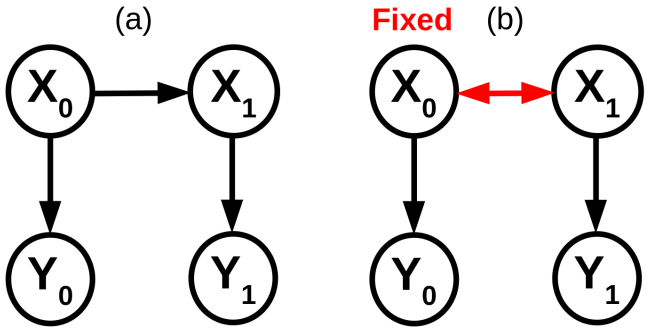


FIGURE 6.3 – Modèles GP-LVM hiérarchiques (a) simple (Sec. 6.2.2.1) et (b) amélioré (Sec. 6.2.2.2).

- Nous avons ajouté un processus Gaussien retour entre $\mathbf{X}^{(0)}$ et $\mathbf{X}^{(1)}$,
- Nous avons effectué le calcul de l'espace $\mathbf{X}^{(0)}$ en premier, puis l'espace $\mathbf{X}^{(1)}$ en fixant les valeurs de $\mathbf{X}^{(0)}$, ce qui revient à forcer la relation hiérarchique entre nos descripteurs de données.

L'entraînement revient donc à un processus itératif, partant du premier niveau de la hiérarchie ($\mathbf{X}^{(0)}$) vers le second ($\mathbf{X}^{(1)}$). On contrôle les contributions respectives des niveaux grâce à la valeur du lien inter-descripteurs, qui pondère la fonction de perte.

6.2.2.3 Réglage des hyperparamètres et procédure d'entraînement

Nous avons entraîné nos modèles avec Python 3.7.6 en utilisant la bibliothèque gpflow [94] (version 2.3.0). Le code et les données de démonstration seront accessibles après acceptation.

Le modèle hiérarchique que nous proposons est constitué de quatre GP-LVMs à optimiser : deux basés sur les observations ($\mathbf{X}^{(0)} \rightarrow \mathbf{Y}^{(0)}$ et $\mathbf{X}^{(1)} \rightarrow \mathbf{Y}^{(1)}$), et deux sur les variables latentes ($\mathbf{X}^{(0)} \rightarrow \mathbf{X}^{(1)}$ et $\mathbf{X}^{(1)} \rightarrow \mathbf{X}^{(0)}$). Pour chacun d'entre eux, il y a trois hyperparamètres à optimiser (ω, l, σ), soit un total de 12 hyperparamètres. Chaque espace latent possède un hyperparamètre supplémentaire pour le calcul de la vraisemblance, nous avons donc 14 hyperparamètres à optimiser. Enfin, en plus de ces hyperparamètres liés au GP-LVM, il existe 4 hyperparamètres supplémentaires contrôlant les poids relatifs de chaque GP-LVMs. En pratique, nous fixons le poids de $\mathbf{X}^{(0)} \rightarrow \mathbf{Y}^{(0)}$ et de $\mathbf{X}^{(1)} \rightarrow \mathbf{Y}^{(1)}$ à 1, et on utilise une valeur identique (à déterminer) pour $\mathbf{X}^{(0)} \rightarrow \mathbf{X}^{(1)}$ et $\mathbf{X}^{(1)} \rightarrow \mathbf{X}^{(0)}$, ce qui revient à ne déterminer qu'un seul hyperparamètre supplémentaire.

En ce qui concerne la procédure d'apprentissage, nous initialisons les espaces latents $\mathbf{X}^{(0)}$ et $\mathbf{X}^{(1)}$ avec les premiers vecteurs de l'ACP de leurs observations correspondantes ($\mathbf{Y}^{(0)}$ et $\mathbf{Y}^{(1)}$, respectivement). Si les hyperparamètres sont

fixés, nous n'entraînons que l'espace latent, sinon nous optimisons alternativement les hyperparamètres et l'espace latent par un processus itératif. En pratique, nous entraînons les espaces latents hiérarchiques en utilisant la procédure suivante :

1. Entraîner les GP-LVMs basés sur les observations (par descente de gradient) pour obtenir les hyperparamètres de $\mathbf{X}^{(0)} \rightarrow \mathbf{Y}^{(0)}$ et $\mathbf{X}^{(1)} \rightarrow \mathbf{Y}^{(1)}$.
2. Réutiliser les valeurs des hyperparamètres obtenus pour les GP-LVMs restants. Les GP-LVMs ayant $\mathbf{X}^{(0)}$ (respectivement $\mathbf{X}^{(1)}$) comme espace latent associé obtiennent les hyperparamètres appris de $\mathbf{X}^{(0)} \rightarrow \mathbf{Y}^{(0)}$ (respectivement $\mathbf{X}^{(1)} \rightarrow \mathbf{Y}^{(1)}$).

Nous avons observé que les résultats n'étaient pas robustes lorsque nous utilisions plusieurs valeurs d'échelles de longueur du noyau pour la hiérarchie, nous n'avons utilisé pour la hiérarchie que la moyenne des valeurs apprises par les GP-LVM de la première étape d'entraînement.

6.2.3 Validation

Notre stratégie à base de GP-LVMs effectue un apprentissage non supervisé de la représentation, qui est difficile à valider. C'est pourquoi nous nous sommes concentrés sur les mesures de la sous-section suivante pour quantifier la qualité de la représentation apprise.

6.2.3.1 Distribution des échantillons dans l'espace latent

Nous avons défini cinq variables physiologiques à partir des segmentations de MI et du MVO, qui correspondent aux mesures globales généralement effectuées par les médecins pour quantifier les lésions myocardiques (cf. Tab.6.1). Ces étiquettes ont été utilisées pour quantifier la qualité des espaces latents estimés, soit en examinant qualitativement la distribution des échantillons colorés par une étiquette donnée (Fig. 6.5, 6.6 et 6.8), soit en quantifiant la corrélation des dimensions latentes avec ces étiquettes (Fig. 6.8).

6.2.3.2 Reconstruction d'images de haute dimension

Nous avons également examiné la qualité des échantillons reconstruits à partir de l'espace latent, qui reflète la quantité d'informations pertinentes provenant des données d'entrée qui sont effectivement encodées dans l'espace latent. Contrairement aux VAEs, la reconstruction ne fait pas intrinsèquement partie de la méthode. Nous avons donc utilisé la régression Ridge à noyaux multi-échelles proposée par *Bermanis et al.* [32, 33], présentée au Chapitre 2, pour estimer les données d'origine à partir de l'espace latent. Cette stratégie multi-échelle est plus robuste aux changements locaux de la densité des échantillons dans l'espace latent, comme cela a été observé sur nos données.

Les expériences ont été réalisées en utilisant une stratégie d'exclusion d'un patient pour la reconstruction (c'est-à-dire en retirant toutes les coupes d'un

Étiquette	Valeur
INF	Étendue de l'infarctus (% du myocarde)
MVO	Étendue du MVO (% du myocarde)
Transmuralité (TM)	Étendue transmurale moyenne de l'infarctus (%)
Angle	Localisation du centre de l'infarctus (°)
ESA	Étendue de l'infarctus à la surface endocardique (% de l'endocarde)

TABLE 6.1 – Étiquettes physiologiques utilisées pour quantifier la pertinence de l'espace latent

patient donné de l'entraînement). Les coupes ne contenant pas de MVO n'ont pas été évaluées.

Pour notre application, nous avons examiné spécifiquement :

— L'erreur de reconstruction en tant que distance L^2 entre les échantillons reconstruits et les échantillons originaux (Fig. 6.4),

— Le coefficient de Dice et la distance de Hausdorff (non représentée ici) entre les échantillons reconstruits et les échantillons originaux (Fig. 6.7), après seuillage à 0.5 et 1.5 (les régions du myocarde sain, MI, et MVO étant codées dans les images originales comme 0, 1, et 2, respectivement),

— Les motifs des lésions encodées le long des directions les plus corrélées aux quantitées définies ci-dessus (Fig. 6.9), ces directions étant obtenues par méthode des moindres carrés orthogonale (O-PLS) [95].

6.3 Expérience et résultats

6.3.1 Choix des hyperparmètres

Nous avons entraîné tous les modèles décrits dans cette section avec les mêmes hyperparamètres. Une fois les hyperparamètres fixés (i.e. après la première étape d'entraînement, basée sur les observations, décrite dans la sous-Section 6.2.2.3), nous avons entraîné un modèle GP-LVM sur les images MI ou MVO pour un nombre de dimensions latentes allant de 1 à 30, avec pour but de trouver un nombre approprié de dimensions pour la représentation hiérarchique. La Figure 6.4 montre l'erreur de reconstruction en fonction du nombre de dimensions latentes.

En examinant les courbes médianes, nous avons observé que l'erreur de reconstruction a déjà diminué de respectivement 60% (pour MI) et 74% (pour MVO) lorsque l'on atteint la sixième dimension (comparativement à la réduction de l'erreur de reconstruction totale). C'est également le début d'un plateau pour la courbe MVO : nous avons donc choisi 6 dimensions pour chacun des espaces latents dans la suite de ce chapitre. Comme notre but est de fusionner les informations, nous avons autorisé seulement 6 dimensions également pour les espaces latents hiérarchiques, qui sont supposés représenter à la fois MI et

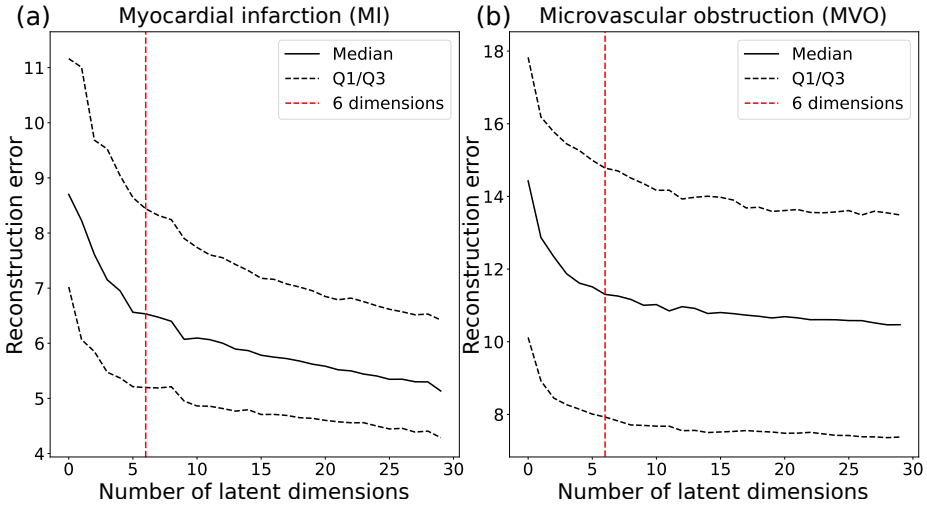


FIGURE 6.4 – Erreur de reconstruction (médiane et premier/troisième quartiles) en fonction du nombre de dimensions latentes pour GP-LVMs sur (a) MI uniquement et (b) MVO uniquement. La ligne verticale rouge indique le nombre de dimensions retenues.

Hyperparamètres	**GP-LVM (MI)**	**GP-LVM (MVO)**
ω	0.10	0.23
σ	$9e-3$	$1e-6$
facteur d'échelle	6 valeurs variant de 0.55 à 1.24	6 valeurs variant de 0.11 à 1.73

TABLE 6.2 – Valeurs des hyperparamètres pour les GP-LVMs sur seulement MI et seulement MVO, respectivement.

MVO. Les hyperparamètres numériques retenus pour toutes les expériences des sections suivantes sont résumés dans la Table 6.2.

6.3.2 Distribution des échantillons dans l'espace latent

Nous avons appliqué le GP-LVM hiérarchique décrit dans le schéma de la Figure 6.3b avec $\mathbf{Y}^{(0)}$ = MI et $\mathbf{Y}^{(1)}$ = MVO, et une valeur de lien latent de 3. Les deux premières dimensions de l'espace latent $\mathbf{X}^{(1)}$ sont représentées sur la Figure 6.5 et colorées selon la quantité de MI ou MVO dans chaque coupe. L'espace est assez bien structuré en ce qui concerne MI et MVO, ce qui reflète l'intégration conjointe de ces deux types d'informations. L'étendue de MI et de MVO se reflète principalement le long de l'axe horizontal (correspondant à la première dimension de l'espace latent), les infarctus les plus petits étant proches de l'origine. Cette partie centrale correspond également aux coupes sans MVO (ou avec un MVO extrêmement petit). Cette figure est instructive mais

Modèle	Nombre de dimensions	Description
Coupe standard	-	Prédiction à partir d'une coupe standard
MI	6	GP-LVM sur les données MI
MVO	6	GP-LVM sur les données MVO
Concaténation	3+3 = 6	Concaténation des GP-LVMs 3D sur MI et MVO
MI+MVO	6	GP-LVM sur la somme des segmentation de MI et MVO
Fusion Naïve	6	GP-LVM Multi-modal avec un unique espace latent
Hiérarchie (couplée)	6	GP-LVM hiérarchique en laissant $\mathbf{X}^{(0)}$ libre
Hiérarchie ($\mathbf{X_0}$ fixé)	6	GP-LVM hiérarchique en fixant $\mathbf{X}^{(0)}$

TABLE 6.3 – Les différents GP-LVMs testés pour la prédiction de MI et MVO (Tab. 6.1) à partir de leurs espaces latents, évalués en Fig. 6.7.

incomplète, car nous n'affichons que les deux premières dimensions d'un espace à 6 dimensions. Par exemple, les coupes presque entièrement infarcies (en rouge foncé) semblent (par erreur) réparties dans la zone centrale de l'espace. Des motifs similaires sont regroupés dans l'espace latent sur d'autres dimensions, l'examen de la sixième dimension de l'espace confirmant par exemple la pertinence de la position de ces échantillons par rapport aux autres. Il convient également de noter que le code couleur ne reflète que la quantité totale de MI ou de MVO dans chaque coupe, alors que les reconstructions montrent en réalité des formes plus complexes, comme on peut le voir sur la Figure 6.9.

Nous avons comparé ces résultats à ceux obtenus à partir d'un GP-LVM monomodal entraîné sur les segmentations MI ou MVO indépendamment, dont les deux premières dimensions de chaque espace latent sont représentées sur la Figure 6.6. À première vue, ces deux espaces représentent bien le type de données qu'ils encodent respectivement, avec MI et MVO répartis de façon homogène à travers ces espaces. Cependant, ils présentent plusieurs défauts quand à leur capacité à représenter la population de patients :

- Les échantillons MVO peuvent être côte à côte bien qu'ils puissent également correspondre à des motifs MI assez différents,
- En outre, comme il n'y a pas une quantité infinie d'échantillons, plusieurs paires de motifs MVO qui ne se chevauchent pas peuvent être associés à la même distance et (faussement) mises en correspondance avec des emplacements proches,
- Enfin, les coupes avec peu ou pas de MVO sont toutes affectées au même point au centre de l'espace latent (point bleu clair dans la partie droite de la Figure 6.6b).

Se concentrer uniquement sur MVO sans tenir compte de MI conduit donc à des malentendus sur la disposition réelle des échantillons, alors que l'établissement d'une hiérarchie entre MI (premier niveau) et MVO (deuxième niveau)

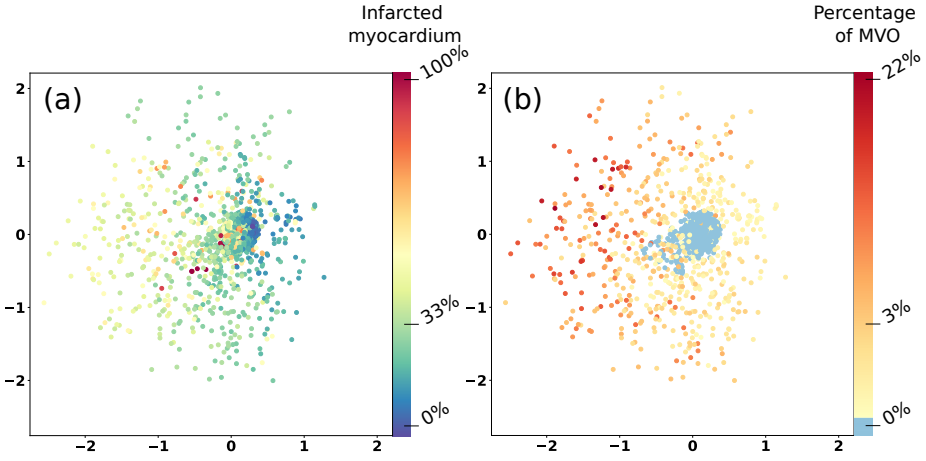


FIGURE 6.5 – Les deux premières dimensions latentes pour le modèle hiérarchique GP-LVM. Le code couleur représente la quantité de (a) MI et (b) MVO dans chaque coupe.

permet de surmonter ces problèmes.

6.3.3 Pertinence des échantillons reconstruits

Nous avons complété les observations qualitatives des espaces latents en quantifiant la fidélité des échantillons reconstruits, à l'aide du coefficient de Dice et de la distance de Hausdorff, comme expliqué dans la Section 6.2.3.2. En particulier, nous avons comparé notre modèle à plusieurs variantes énumérées dans la Table 6.3, les résultats pour le coefficient de Dice étant affichés sous forme de diagrammes en boîte dans la Fig. 6.7. La figure quantifie également la précision de l'utilisation d'une seule coupe représentative pour les segmentations MI et MVO de tous les sujets (notée '*single slice*' dans la figure). Pour MI, cette coupe unique correspondait à la médiane de tous les motifs MI. Pour MVO, ce calcul n'est pas possible car ces lésions sont très petites et ne se chevauchent pas souvent. Nous avons donc calculé les occurences des pixels de MVO et avons fixé le seuil de cette carte à 20%, ce qui correspond à un motif représentatif de MVO. Ces calculs permettent de se rendre compte de l'écart de la population à une valeur médiane relativement à ces métriques.

Les approches hiérarchiques, avec ou sans $\mathbf{X}^{(0)}$, ont donné des résultats satisfaisants pour prédire les segmentations MI et MVO pour chaque coupe. Les différentes valeurs de liens offrent différents équilibres entre MI et MVO. L'espace latent représenté sur la Figure 6.5 correspond au deuxième diagramme en boîte rouge de la Figure 6.7 (a) et (b). Trois observations principales peuvent être faites à partir de cette figure :

— Les résultats des espaces hiérarchiques sont proches de ceux des espaces à 6 dimensions appris sur une seule modalité, tout en étant capables de

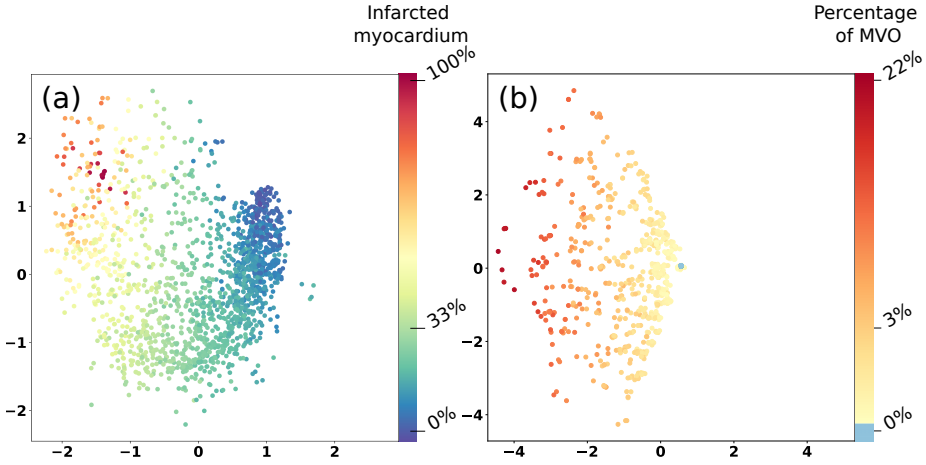


FIGURE 6.6 – Les deux premières dimensions latentes des GP-LVMs sur (a) MI seulement et (b) MVO seulement. Le code couleur est similaire à celui de la Fig. 6.5) et corresponds à la quantité de MI ou de MVO.

prédire les deux modalités,

— Les autres modèles testés ici n'ont pas réussi à produire une représentation conjointe de MI et MVO, et encodent plutôt soit MI, soit MVO,

— La stratégie où $\mathbf{X}^{(0)}$ n'est pas fixé est légèrement plus performante que celle où $\mathbf{X}^{(0)}$ est fixé, car elle est plus flexible.

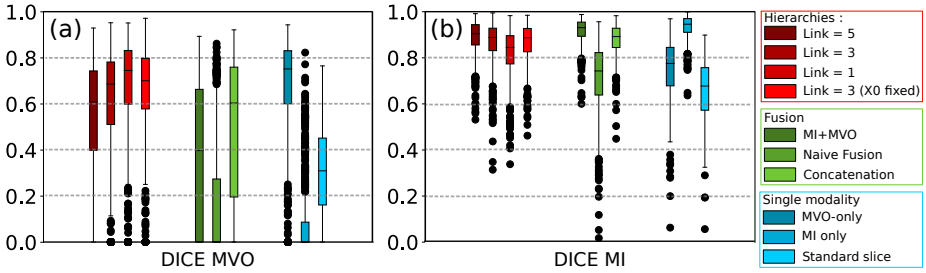


FIGURE 6.7 – Coefficient de Dice pour la reconstruction de MVO (a) et de MI (b) à partir d'un espace latent donné (voir la liste des GP-LVMs dans la Table 6.3).

6.3.4 Mélange du premier niveau de données

Nous avons mené une expérience supplémentaire pour confirmer que l'amélioration de l'encodage obtenue avec le cadre hiérarchique était due au lien existant entre les deux modalités, et non à un a priori trop fort. Pour ce faire,

$\mathbf{X}^{(0)}$ a été mélangé, rompant ainsi tout lien entre les deux niveaux, ce qui invalide l'hypothèse initiale. Nous avons estimé trois espaces latents avec différents réarrangements aléatoires de la première modalité, en utilisant exactement les mêmes paramètres que notre espace hiérarchique de référence (lien = 3). Les Figures 6.8a et b montrent la projection bidimensionnelle de l'un des espaces latents réorganisés, relativement à la quantité de MI et de MVO. Une disposition similaire est observée pour les deux autres espaces mélangés (non affichés ici). Cela signifie que la quantité de MVO a été apprise et que le motif MI n'a été respecté que lorsqu'il n'y avait pas de MVO dans la coupe, ce qui a produit un espace désordonné lorsque l'on s'éloigne du centre. La Figure 6.8c complète ce tableau en montrant les corrélations entre chacune des 6 premières dimensions latentes et les quantités (mélangés) de MI et de MVO.

Les corrélations sont globalement plus fortes pour les deux modalités dans l'espace hiérarchique, et l'impact de la fusion est significatif. C'est particulièrement vrai pour le MVO, ce qui montre que cette information a été mieux apprise avec le soutien d'un lien hiérarchique structuré. Même lorsque les valeurs sont faibles, la dimension en question est corrélée à la fois à MI et MVO, ce qui indique que les modalités sont apprises conjointement et que le lien hiérarchique a un impact réel sur la représentation finale des deux quantités. La corrélation confirme également l'observation qualitative selon laquelle les espaces mélangés présentent un comportement similaire dans l'ensemble. Dans les trois espaces mélangés, les dernières dimensions sont fortement corrélées avec la quantité MI, tandis que le MVO est corrélé avec la première dimension.

6.3.5 Consistence physiologique de l'espace latent

Enfin, nous avons utilisé l'algorithme O-PLS pour parcourir l'espace latent le long de la direction qui encode le mieux une étiquette physiologique donnée (voir la liste des étiquettes dans le Tab. 6.1). Nous avons reconstruit des images de haute dimension correspondant à des points synthétiques échantillonnés dans un intervalle de $+/-2\sigma$ le long de ces directions, σ représentant l'écart-type de la projection de l'espace latent sur cette direction. À des fins de comparaison, nous avons réalisé la même expérience pour la concaténation des espaces latents tridimensionnels MVO et MI. Nous avons choisi cet espace car c'est l'espace multi-modal non hiérarchique le plus performant au regard des Dice (cf. Figure 6.7).

La reconstruction de l'espace hiérarchique (Figure 6.9) reflète correctement l'étiquette que chaque direction est sensée encoder. Ceci est particulièrement bien rendu pour la mesure 'Surface de l'endocarde infarcie (ESA)' : la coupe à $+2\sigma$ présente une grande zone de MI, très large près de l'endocarde. En revanche, les reconstructions de l'espace latent de concaténation pour MI, la transmuralité et ESA présentent toutes un comportement similaire ne correspondant pas spécifiquement à l'étiquette souhaitée.

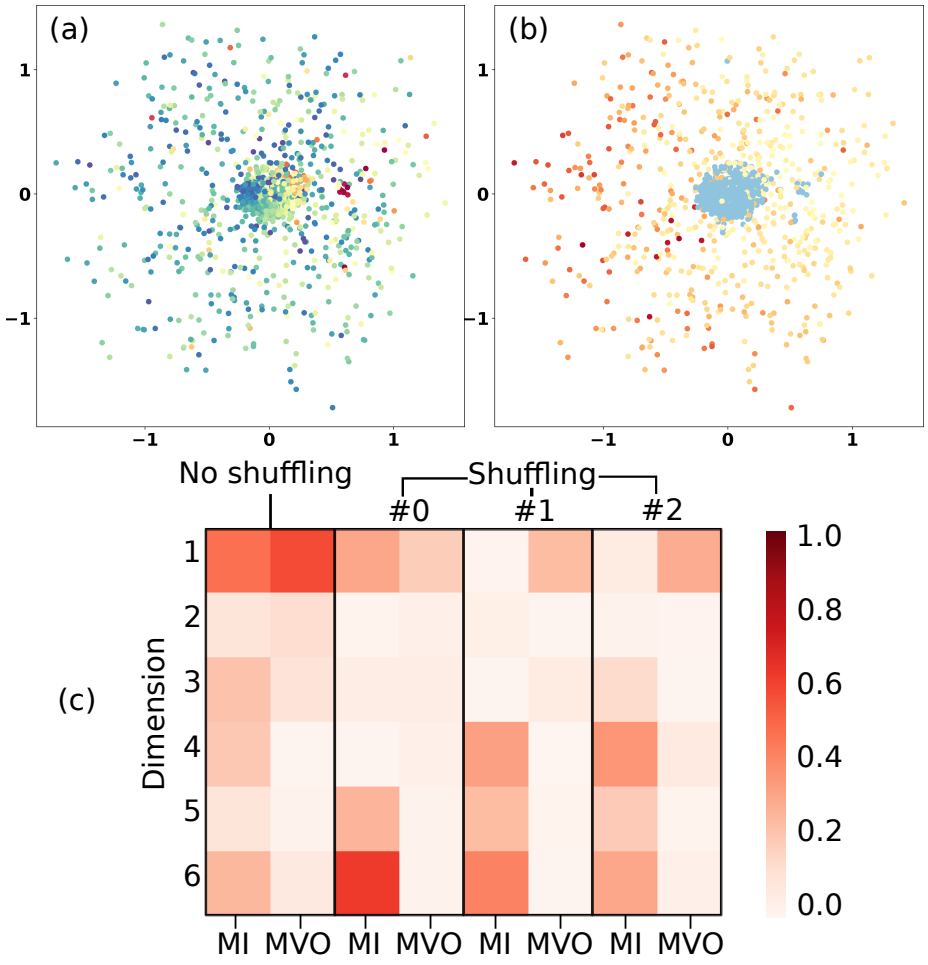


FIGURE 6.8 – Espace latent hiérarchique obtenu lorsque la première modalité a été mélangée, ce qui rompt l'a priori hiérarchique. (a) et (b) Espace latent estimé coloré respectivement par MI et par MVO. (c) Corrélation entre chaque dimension de l'espace latent et les deux étiquettes physiologiques, sans (à gauche) / avec brassage (trois colonnes à l'extrême droite). La rupture de l'a priori hiérarchique entraîne une décorrélation entre les deux étiquettes et une certaine redondance entre les dimensions, ce qui souligne la pertinence du lien hiérarchique dans les données.

6.4 Conclusions et perspectives

Dans ce chapitre, nous avons présenté une stratégie d'apprentissage non supervisée d'une représentation d'une population à partir de multiples descripteurs de haute dimension. La méthode est hiérarchique, ce qui signifie que les

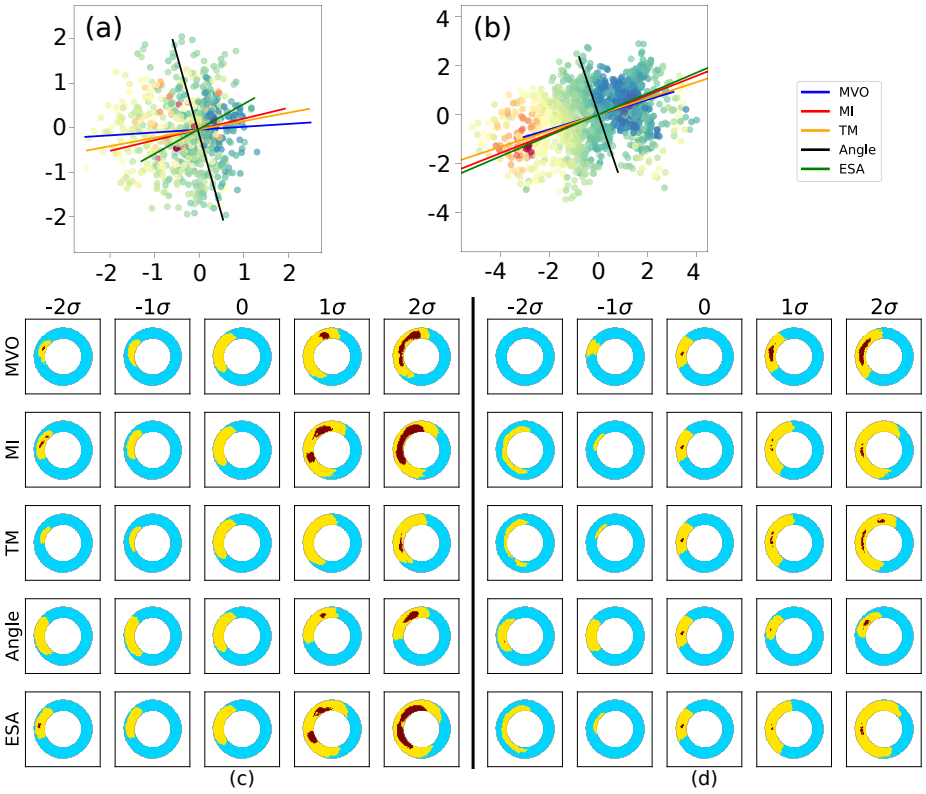


FIGURE 6.9 – Cohérence de l'espace latent en fonction des étiquettes physiologiques énumérées dans le Tab. 6.1. Projections bidimensionnelles des espaces latents hiérarchiques (a) et de concaténation (b), colorées par l'étendue de MVO. Les lignes droites colorées représentent les directions principales des cinq étiquettes physiologiques, estimées avec O-PLS. Reconstruction d'images de haute dimension obtenues par échantillonnage le long de ces axes, pour les espaces hiérarchique (c) et de concaténation (d), respectivement. L'espace hiérarchique est capable de reconstruire des motifs spécifiques à partir des étiquettes, contrairement à l'espace de concaténation qui ne fait pas la distinction entre MI (2ème ligne), la transmuralité (3ème ligne) et l'ESA (5ème ligne).

niveaux précédents de la hiérarchie (imposée) des descripteurs sont utilisés pour guider la représentation des niveaux suivants (potentiellement associés à des descripteurs plus complexes). Plus précisément, notre méthode est basée sur les modèles GP-LVMs, et étend ce cadre pour la construction d'une hiérarchie à deux niveaux avec des considérations spécifiques sur le lien mutuel entre les deux niveaux. Nous avons démontré sa pertinence sur des données d'imagerie cardiaque réelles dans une étude sur l'impact d'une revascularisation retardée, afin de mieux représenter la distribution des motifs de lésion liés à l'infarctus

aigu et à la reperfusion au sein d'une population. Notre modèle a été capable d'apprendre une représentation latente pertinente de faible dimension de ces données, guidée par un premier niveau de hiérarchie, permettant une analyse statistique des données de niveau supérieur.

Les deux variantes (avec et sans niveau supérieur fixe) du modèle hiérarchique proposé ont surpassé les modèles à modalité unique et les modèles de fusion sur tous les aspects évalués : la distribution de l'espace latent, son pouvoir de représentation et la cohérence avec les étiquettes physiologiques utilisées dans la routine clinique. Le modèle d'apprentissage conjoint permet une meilleure circulation de l'information entre les deux espaces latents $\mathbf{X}_0$ et $\mathbf{X}_1$, ce qui explique un pouvoir de représentation légèrement supérieur. Pour ce modèle, les deux espaces latents ($\mathbf{X}_0$ et $\mathbf{X}_1$) sont des compromis différents entre la représentation des observations $\mathbf{Y}_0$ et $\mathbf{Y}_1$. Le modèle hiérarchique avec $\mathbf{X}_0$ fixe est intéressant pour imiter la manière dont les cliniciens intègrent progressivement différents niveaux d'information : dans notre application, nous aimerions estimer une représentation de MVO qui bénéficie des informations du niveau précédent (MI), mais sans remettre en question la représentation originale de MI. Il s'agit également d'un atout considérable lorsque l'on traite de grandes quantités de données et que des niveaux supplémentaires peuvent être ajoutés à la hiérarchie sans avoir à réentraîner l'ensemble du modèle.

Notre modèle est adaptable, mais cette flexibilité a un prix : sur des données réelles, il peut être difficile de trouver l'équilibre entre chaque modalité. En pratique, si nous laissons les paramètres du modèle complétement libres lors de l'optimisation, l'espace latent est dominé par les représentations MI ou MVO et ne parvient pas à apprendre un compromis équilibré entre elles. En outre, les GP-LVMs fournissent une manière intuitive de représenter la hiérarchie souhaitée, mais sont difficiles à optimiser. Pour mieux contrôler ces aspects, nous avons fixé les hyperparamètres avant l'apprentissage des différents espaces latents. En théorie, ces hyperparamètres peuvent être appris par descente de gradient, mais cela nécessiterait l'ajout d'un terme de régularisation supplémentaire pour contraindre l'équilibre entre les modalités. Néanmoins, la flexibilité du cadre des GP-LVMs permet d'apprendre à partir d'espaces de données et de hiérarchies beaucoup plus complexes. Le principal obstacle consiste alors à établir les relations hiérarchiques entre les nombreuses modalités. Dans ce travail, nous avons choisi une hiérarchie simple au vu de la complexité des données pour nous concentrer sur sa mise en œuvre avec les GP-LVMs. Une alternative serait d'apprendre l'ordre hiérarchique en même temps que l'intégration des modalités, ce qui, bien que séduisant, peut être très difficile dans ce cadre, mais a été étudié avec l'apprentissage par renforcement pour la sélection active de modalités sur des données plus simples et pour des problèmes supervisés [96].

L'apprentissage non supervisé est intéressant pour caractériser la distribution d'une population à partir de descripteurs complexes, comme c'est le cas dans cette étude (caractérisation de motifs de lésions spécifiques à partir de données d'imagerie que les cliniciens simplifient à l'excès). Les méthodes non supervisées sont difficiles à valider, étant donné l'absence de références. Dans notre application, nous avons réalisé une série d'expériences approfondies pour examiner à

la fois l'espace latent, les données reconstruites et les étiquettes physiologiques, ce qui a permis de comprendre le comportement du modèle hiérarchique et de montrer ses meilleures performances pour notre tâche. Les GP-LVMs peuvent également être utilisés pour des problèmes supervisés, ce qui signifie que notre méthodologie pourrait être directement appliquée à de tels problèmes.

Les mécanismes d'ischémie-reperfusion sont complexes et évoluent dans le temps. Dans ce chapitre, nous nous sommes concentrés sur une partie des données d'imagerie utilisées par les cliniciens pour caractériser ces modèles, à savoir les segmentations d'IRM LGE, qui reflètent la forme des lésions. Une caractérisation plus complète pourrait intégrer également des données d'imagerie liées au contenu du tissu myocardique (images T1 et T2) et à la déformation (séquences CINE ou DENSE), ce qui est envisagé dans des travaux futurs sur des cohortes plus importantes disposant de données longitudinales.

Chapitre 7

Conclusions et Perspectives

7.1 Contributions

Dans cette thèse, nous avons exploré l'intégration hiérarchique de données pour l'analyse non supervisée de multiples descripteurs de haute dimension extraits de l'imagerie cardiaque. Nous avons proposé deux stratégies d'intégration basé sur deux formalismes différents pour représenter de façon pertinente la charge lésionnelle dans une population de patients avec infarctus aigu du myocarde.

Dans une première contribution, nous avons représenté le contenu brut des images IRM de réhaussement tardif (LGE), avec le support des segmentations de la zone infarcie et du MVO. Ce contenu image est riche mais sous-exploité en pratique car son analyse directe est difficile. En particulier, l'aspect du MVO sur ces images (zone en hypo-intensité) est similaire au myocarde sain. Cette analyse a exploité l'apprentissage de variétés non supervisé en étendant l'algorithme *Diffusion Maps* à une approche hiérarchique. Nous sommes partis de la formulation de *Bhatia et al.* [85] mais cette fois pour l'exploitation d'une hiérarchie de descripteurs. Nous avons également proposé deux méthodes de détermination des hyperparamètres du modèle. Ce travail a permis une représentation plus fidèle de la population d'images LGE en résolvant les principaux défauts de l'espace latent calculé sur les images seules. Il a donné lieu à une communication au Workshop STACOM associé à la conférence MICCAI en 2021 [46].

Dans un deuxième temps, nous avons proposé une méthodologie inspirée des travaux de *Lawrence et al.* [79], basée sur les modèles GP-LVMs. Nous avons exploité une version hiérarchique de GP-LVM pour la représentation des motifs de MVO grâce aux segmentations de la zone infarcie. Ces motifs ont un aspect particulier (formes petites et irrégulières) qui rend une analyse directe difficile. Une représentation cohérente de ces motifs a été obtenue en utilisant de façon hiérarchique la donnée liée aux motifs d'infarctus. Nous avons proposé deux approches hiérarchiques : l'une apprend conjointement l'espace associé à l'infarctus et celui associé au MVO, tandis que l'autre apprend d'abord l'espace

de l'infarctus, le fixe et s'en sert pour représenter les motifs de MVO. Cette deuxième méthodologie est particulièrement intéressante car elle permet de ne pas avoir à recalculer entièrement la hiérarchie lors de l'intégration d'un nouveau descripteur. Nous avons montré la pertinence de cette stratégie en examinant les reconstructions obtenues le long d'axes correspondants à des indices cliniques pertinents. Cette partie a été soumise à *IEEE Transactions on Medical Imaging*.

7.2 Perspectives

7.2.1 Perspectives méthodologiques

Une des limites principales des analyses décrites dans ce document est la métrique utilisée pour comparer les sujets. Dans cette thèse, nous nous sommes principalement intéressés à l'aspect hiérarchique et donc aux façons d'exploiter le lien entre descripteurs. La distance Euclidienne utilisée réduit les capacités de représentation des algorithmes. En effet, cette distance ne permet pas d'exploiter les voisinages des pixels (comme le ferait un réseau de neurones convolutif par exemple). Cette limite a été particulièrement critique pour l'étude des motifs de MVO dans cette thèse. Une façon de dépasser cette limite serait par exemple d'utiliser des GP-LVMs profonds, comme ceux proposés par *Damianou et al.* [87], et d'y adapter la méthode des GP-LVMs hiérarchiques. Les couches de convolution des réseaux de neurones permettraient de mieux exploiter la structure spatiale et les textures des images.

Ce travail s'est concentré sur la construction de hiérarchies entre deux descripteurs, mais il serait pertinent et intéressant de l'étendre à des hiérarchies à plus de niveaux. La flexibilité de la hiérarchie basée sur les GP-LVMs permet en théorie cette extension, sous réserve de la perte d'information au fil de la hiérarchie et des difficultés d'optimisation des GP-LVMs.

Une dernière piste d'amélioration de la méthodologie développée dans cette thèse serait d'apprendre directement les hiérarchies, et non de les imposer selon un a priori par exemple sur la complexité des descripteurs. Un travail a déjà été publié dans ce sens au sein de notre équipe en utilisant de l'apprentissage par renforcement [96]. Cette approche a cependant été utilisée pour de l'apprentissage supervisé et ne permet pas en l'état l'apprentissage d'une cartographie des sujets. De plus, le processus d'apprentissage utilisé n'est pas un processus hiérarchique : les auteurs réapprennent complètement la hiérarchie à chaque fois qu'une modalité y est ajoutée.

7.2.2 Perspectives cliniques

Pour obtenir un nombre plus conséquent de données et augmenter la variabilité des motifs observés, nous avons considéré les coupes en 2D des patients comme des échantillons à part entière de la base de données. Cela n'est pas entièrement satisfaisant car nous ne pouvons pas remonter à la donnée patient (en 3D), et nos espaces représentent ainsi une population de motifs pathologiques

plutôt qu'une population de patients. Il pourrait être pertinent de développer une métrique adaptée aux volumes pour fournir aux médecins une représentation basée sur les données 3D.

Sur les données que nous utilisons, il serait également intéressant d'incorporer d'autres descripteurs d'images à la hiérarchie. Une première piste pertinente serait d'ajouter la donnée de réhaussement précoce (EGE), présente et exploitable sur 117 patients de la base de donnée MIMI. Cela permettrait une meilleure compréhension des phénomènes d'ischémie-reperfusion liés à un infarctus aigü du myocarde, car la donnée EGE est la modalité la plus adéquate pour évaluer la zone d'obstruction microvasculaire. Une autre option intéressante serait d'évaluer le potentiel de notre approche sur la base de données HIBISCUS-STEMI [15], qui contient également les images T1 et T2 des patients. Cette base de données, en cours d'exploration dans l'équipe, contient notamment les images acquises 1 mois et 12 mois après l'infarctus, dans une perspective de suivi longitudinal des patients.

7.3 Conclusion

Nous avons développé dans ce manuscrit des méthodes d'analyse statistique de populations pour aller plus loin dans l'analyse des mécanismes d'ischémie-reperfusion, dans le cadre de l'infarctus du myocarde. Nous avons exploité des descripteurs d'image de haute dimension, les segmentations du myocarde et les images de réhaussement tardif (LGE) natives, sous-exploitées en routine clinique. La démarche générale de cette thèse a été de s'inspirer du raisonnement des médecins, en incorporant dans les algorithmes d'apprentissage non supervisé une intégration hiérarchique des données. Nous avons montré la pertinence d'une telle stratégie pour l'analyse des motifs de MVO et d'infarctus sur les images LGE, pour une population de patients atteints d'infarctus aigü du myocarde. Ces approches sont relativement intuitives et posent une première base solide pour leur extension à des hiérarchies plus complexes et d'autres types d'images.

Bibliographie

[1] T. Ryan, J. Anderson, E. Antman *et al.*, "ACC/AHA guidelines for the management of patients with acute myocardial infarction : executive summary. A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee on Management of Acute Myocardial Infarction)," *Circulation*, vol. 94, pp. 2341–50, 1996.

[2] J. C. Riera, "Risk stratification after acute myocardial infarction," *Revista espanola de cardiologia*, vol. 56, no. 3, pp. 303–313, 2003.

[3] P. A. Cowper, J. D. Knight, L. Davidson-Ray, E. D. Peterson, T. Y. Wang, D. B. Mark, and T.-A. Investigators, "Acute and 1-year hospitalization costs for acute myocardial infarction treated with percutaneous coronary intervention : Results from the translate-acs registry," *Journal of the American Heart Association*, vol. 8, no. 8, p. e011322, 2019.

[4] W. RK *et al.*, "Cardiomyopathy : an overview," *Am Fam Physician*, 2009.

[5] B. Ibanez, S. James, S. Agewall, M. J. Antunes, C. Bucciarelli-Ducci, H. Bueno, A. L. P. Caforio, F. Crea, J. A. Goudevenos, S. Halvorsen, G. Hindricks, A. Kastrati, M. J. Lenzen, E. Prescott, M. Roffi, M. Valgimigli, C. Varenhorst, P. Vranckx, P. Widimský, and E. S. D. Group, "2017 ESC Guidelines for the management of acute myocardial infarction in patients presenting with ST-segment elevation : The Task Force for the management of acute myocardial infarction in patients presenting with ST-segment elevation of the European Society of Cardiology (ESC)," *European Heart Journal*, vol. 39, no. 2, pp. 119–177, 08 2017. [Online]. Available : https://doi.org/10.1093/eurheartj/ehx393

[6] B. Asatryan, L. Vaisnora, and N. Manavifar, "Electrocardiographic diagnosis of life-threatening stemi equivalents," *JACC : Case Reports*, vol. 1, no. 4, pp. 666–668, 2019. [Online]. Available : https://www.jacc.org/doi/abs/10.1016/j.jaccas.2019.10.030

[7] F. Liu, C. Liu, L. Zhao, X. Zhang, X. Wu, X. Xu, Y. Liu, C. Ma, S. Wei, Z. He *et al.*, "An open access database for evaluating the algorithms of electrocardiogram rhythm and morphology abnormality detection," *Journal of Medical Imaging and Health Informatics*, vol. 8, no. 7, pp. 1368–1373, 2018.

[8] E. A. P. Alday, A. Gu, A. J. Shah, C. Robichaux, A.-K. I. Wong, C. Liu, F. Liu, A. B. Rad, A. Elola, S. Seyedi *et al.*, "Classification of 12-lead ecgs : the physionet/computing in cardiology challenge 2020," *Physiological measurement*, vol. 41, no. 12, p. 124003, 2020.

[9] F. Flachskampf, N. Schmid, C. Rost *et al.*, "Cardiac imaging after myocardial infarction," *Eur Heart J*, vol. 32, pp. 272–83, 2011.

[10] V. Mor-Avi, R. Lang, L. Badano *et al.*, "Current and evolving echocardiographic techniques for the quantitative evaluation of cardiac mechanics : ASE/EAE consensus statement on methodology and indications endorsed by the Japanese Society of Echocardiography," *J Am Soc Echocardiogr*, vol. 24, pp. 277–313, 2011.

[11] M. D. Cerqueira, N. J. Weissman, V. Dilsizian, A. K. Jacobs, S. Kaul, W. K. Laskey, D. J. Pennell, J. A. Rumberger, T. Ryan *et al.*, "Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart : a statement for healthcare professionals from the cardiac imaging committee of the council on clinical cardiology of the american heart association," *Circulation*, vol. 105, no. 4, pp. 539–542, 2002.

[12] R. M. Lang, L. P. Badano, V. Mor-Avi, J. Afilalo, A. Armstrong, L. Ernande, F. A. Flachskampf, E. Foster, S. A. Goldstein, T. Kuznetsova *et al.*, "Recommendations for cardiac chamber quantification by echocardiography in adults : an update from the american society of echocardiography and the european association of cardiovascular imaging," *European Heart Journal-Cardiovascular Imaging*, vol. 16, no. 3, pp. 233–271, 2015.

[13] H. Bulluck *et al.*, "Cardiovascular magnetic resonance in acute st-segment–elevation myocardial infarction : recent advances, controversies, and future directions," *Circulation*, vol. 137, no. 18, pp. 1949–1964, 2018.

[14] L. Belle *et al.*, "Comparison of immediate with delayed stenting using the Minimalist Immediate Mechanical Intervention approach in acute ST-segment-elevation myocardial infarction : the MIMI study," *Circ Cardiovasc Interv*, vol. 9, p. e003388, 2016.

[15] N. Dufay, G. Cavillon, C. Jossan, G. Blanchard, C. Amaz, G. Tabdjoun, L. Ma, T. Bochaton, C. C. Da Silva, and M. Ovize, "Hibiscus-stemi biobank : High quality samples for powerful research on stemi biomarkers," *Archives of Cardiovascular Diseases Supplements*, vol. 12, no. 1, p. 11, 2020.

[16] G. Heusch, "Coronary microvascular obstruction : the new frontier in cardioprotection," *Basic Research in Cardiology*, vol. 114, no. 6, p. 45, 2019.

[17] M. A. Konstam, D. G. Kramer, A. R. Patel, M. S. Maron, and J. E. Udelson, "Left ventricular remodeling in heart failure : current concepts in clinical significance and assessment," *JACC : Cardiovascular imaging*, vol. 4, no. 1, pp. 98–108, 2011.

[18] B. Ibanez, A. Aletras, A. Arai *et al.*, "Cardiac MRI endpoints in myocardial infarction experimental and clinical trials : JACC Scientific Expert Panel," *J Am Coll Cardiol*, vol. 16, pp. 238–56, 2019.

[19] M. S. Amzulescu, M. De Craene, H. Langet, A. Pasquet, D. Vancraeynest, A.-C. Pouleur, J.-L. Vanoverschelde, and B. Gerber, "Myocardial strain imaging : review of general principles, validation, and sources of discrepancies," *European Heart Journal-Cardiovascular Imaging*, vol. 20, no. 6, pp. 605–619, 2019.

[20] S. Sanchez-Martinez, N. Duchateau, T. Erdei *et al.*, "Machine learning analysis of left ventricular function to characterize heart failure with preserved ejection fraction," *Circulation Cardiovascular Imaging*, vol. 11, p. e007138, 2018.

[21] S. Yan *et al.*, "Graph embedding and extensions : A general framework for dimensionality reduction," *IEEE Trans Pattern Anal Mach Intell*, vol. 29, pp. 40–51, 2007.

[22] L. Van der Maaten, E. Postma, and H. Van Den Herik, "Dimensionality reduction : A comparative review," *Technical Report TiCC TR 2009-005*, 2009.

[23] J. B. Tenenbaum *et al.*, "A global geometric framework for nonlinear dimensionality reduction," *science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[24] Y. Bengio, A. Courville, and P. Vincent, "Representation learning : a review and new perspectives," *IEEE Trans Pattern Anal Mach Intell*, vol. 35, pp. 1798–828, 2013.

[25] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *Proceedings of International Conference on Computer Vision (ICCV)*, December 2015.

[26] M. Turk and A. Pentland, "Eigenfaces for recognition," *J Cogn Neurosci*, vol. 3, pp. 71–86, 1991.

[27] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Comput*, vol. 15, pp. 1373–96, 2003.

[28] R. R. Coifman and S. Lafon, "Diffusion maps," *Applied and Computational Harmonic Analysis*, vol. 21, no. 1, pp. 5–30, 2006, special Issue : Diffusion Maps and Wavelets. [Online]. Available : https://www.sciencedirect.com/science/article/pii/S1063520306000546

[29] S. Lafon, Y. Keller, and R. Coifman, "Data fusion and multi-cue data matching by diffusion maps," *IEEE Trans Partten Anal Mach Intell*, vol. 28, pp. 1784–1797, 2006.

[30] L. Van der Maaten and G. Hinton, "Visualizing data using t-sne." *Journal of machine learning research*, vol. 9, no. 11, 2008.

[31] L. McInnes, J. Healy, and J. Melville, "Umap : Uniform manifold approximation and projection for dimension reduction," *arXiv preprint arXiv :1802.03426*, 2018.

[32] A. Bermanis *et al.*, "Multiscale data sampling and function extension," *Applied and Computational Harmonic Analysis*, vol. 34, no. 1, pp. 15–29, 2013.

[33] N. Duchateau, M. De Craene, M. Sitges *et al.*, "Adaptation of multiscale function extension to inexact matching. Application to the mapping of individuals to a learnt manifold," *Proc. SEE-GSI'13, LNCS*, vol. 8085, pp. 578–86, 2013.

[34] H. Bourlard and Y. Kamp, "Auto-association by multilayer perceptrons and singular value decomposition," *Biological cybernetics*, vol. 59, no. 4-5, pp. 291–294, 1988.

[35] D. Kingma and M. Welling, "Auto-encoding variational Bayes," *arXiv preprint arXiv :1312.6114*, 2013.

[36] R. Wei and A. Mahmood, "Recent advances in variational autoencoders with representation learning for biomedical informatics : A survey," *IEEE Access*, vol. 9, pp. 4939–4956, 2021.

[37] J. Ehrhardt and M. Wilms, "Chapter 8 - autoencoders and variational autoencoders in medical image analysis," *Biomedical Image Synthesis and Simulation, Burgos & Svoboda, eds.*, pp. 129–162, 2022.

[38] I. Higgins, L. Matthey, A. Pal, C. Burgess, X. Glorot, M. Botvinick, S. Mohamed, and A. Lerchner, "beta-vae : Learning basic visual concepts with a constrained variational framework," in *International conference on learning representations*, 2017.

[39] X. Liu, P. Sanchez, S. Thermos *et al.*, "Learning disentangled representations in the imaging domain," *Med Image Anal*, vol. 80, p. 102516, 2022.

[40] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[41] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words : Transformers for image recognition at scale," 2021.

[42] N. Lawrence, "Gaussian process latent variable models for visualisation of high dimensional data," *Advances in neural information processing systems*, vol. 16, 2003.

[43] A.-G. A. Bulluck H, Rosmini S *et al.*, "Impact of microvascular obstruction on semiautomated techniques for quantifying acute and chronic myocardial infarction by cardiovascular magnetic resonance." *Open Heart*, 2016.

[44] W. A. Romero R., M. Viallon, J. Spaltenstein, L. Petrusca, O. Bernard, L. Belle, P. Clarysse, and P. Croisille, "Cmrsegtools : An open-source software enabling reproducible research in segmentation of acute myocardial infarct in cmr images," *PLOS ONE*, vol. 17, no. 9, pp. 1–17, 09 2022. [Online]. Available : https://doi.org/10.1371/journal.pone.0274491

[45] A. T. Yan, A. J. Shayne, K. A. Brown, S. N. Gupta, C. W. Chan, T. M. Luu, M. F. D. Carli, H. G. Reynolds, W. G. Stevenson, and R. Y. Kwong, "Characterization of the peri-infarct zone by contrast-enhanced cardiac magnetic resonance imaging is a powerful predictor of

post–myocardial infarction mortality," *Circulation*, vol. 114, no. 1, pp. 32–39, 2006. [Online]. Available : https://www.ahajournals.org/doi/abs/10.1161/CIRCULATIONAHA.106.613414

[46] B. Freiche *et al.*, "Characterizing myocardial ischemia and reperfusion patterns with hierarchical manifold learning," pp. 66–74, 2021.

[47] e. a. Wegelin, J.A., "A survey of partial least squares (pls) methods, with emphasis on the two-block case." *Tech. Rep. University of Washington, Department of Statistics*.

[48] H. Wold, "Partial least squares," *Encyclopedia of Statistical Sciences, Vol. 6, John Wiley, New York*.

[49] H. Hotelling, "Relations between two sets of variates," pp. 162–190, 1992.

[50] J. Ham, D. D. Lee, and L. K. Saul, "Semisupervised alignment of manifolds," in *International Conference on Artificial Intelligence and Statistics*, 2005.

[51] J. Valencia-Aguirre, A. Álvarez-Meza, G. Daza-Santacoloma, C. Acosta-Medina, and C. G. Castellanos-Domínguez, "Multiple manifold learning by nonlinear dimensionality reduction," pp. 206–213, 2011.

[52] J. R. Clough, D. R. Balfour, G. Cruz, P. K. Marsden, C. Prieto, A. J. Reader, and A. P. King, "Weighted manifold alignment using wave kernel signatures for aligning medical image datasets," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 4, pp. 988–997, 2019.

[53] M. Aubry, U. Schlickewei, and D. Cremers, "The wave kernel signature : a quantum mechanical approach to shape analysis," *IEEE ICCV Workshops*, p. 1626–1633, 2011.

[54] J. Valencia-Aguirre *et al.*, "Multiple manifold learning by nonlinear dimensionality reduction," pp. 206–213, 2011.

[55] M. Di Folco, P. Moceri, P. Clarysse, and N. Duchateau, "Characterizing interactions between cardiac shape and deformation by non-linear manifold learning," *Medical Image Analysis*, vol. 75, p. 102278, 2022. [Online]. Available : https://www.sciencedirect.com/science/article/pii/S1361841521003236

[56] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark *et al.*, "Learning transferable visual models from natural language supervision," pp. 8748–8763, 2021.

[57] T. Judge, O. Bernard, M. Porumb, A. Chartsias, A. Beqiri, and P.-M. Jodoin, "Crisp-reliable uncertainty estimation for medical image segmentation," pp. 492–502, 2022.

[58] J. Bromley, I. Guyon, Y. LeCun, E. Säckinger, and R. Shah, "Signature verification using a" siamese" time delay neural network," *Advances in neural information processing systems*, vol. 6, 1993.

[59] G. Koch, R. Zemel, R. Salakhutdinov *et al.*, "Siamese neural networks for one-shot image recognition," vol. 2, no. 1, 2015.

[60] A. Rossi, M. Hosseinzadeh, M. Bianchini, F. Scarselli, and H. Huisman, "Multi-modal siamese network for diagnostically similar lesion retrieval in prostate mri," *IEEE Transactions on Medical Imaging*, vol. 40, no. 3, pp. 986–995, 2020.

[61] M. D. Li, K. Chang, B. Bearce, C. Y. Chang, A. J. Huang, J. P. Campbell, J. M. Brown, P. Singh, K. V. Hoebel, D. Erdoğmuş *et al.*, "Siamese neural networks for continuous disease severity evaluation and change detection in medical imaging," *NPJ digital medicine*, vol. 3, no. 1, p. 48, 2020.

[62] L. Antelmi *et al.*, "Sparse multi-channel variational autoencoder for the joint analysis of heterogeneous data," pp. 302–311, 2019.

[63] D. P. Kingma, T. Salimans, and M. Welling, "Variational dropout and the local reparameterization trick," *Advances in neural information processing systems*, vol. 28, 2015.

[64] G. Martí-Juan, M. Lorenzi, G. Piella, A. D. N. Initiative *et al.*, "Mc-rvae : Multi-channel recurrent variational autoencoder for multimodal alzheimer's disease progression modelling," *NeuroImage*, vol. 268, p. 119892, 2023.

[65] Y.-Y. Lin *et al.*, "Dimensionality reduction for data in multiple feature representations," *Advances in Neural Information Processing Systems*, vol. 21, 2008.

[66] ——, "Multiple kernel learning for dimensionality reduction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 6, pp. 1147–1160, 2010.

[67] S. Sanchez-Martinez, N. Duchateau, T. Erdei, A. G. Fraser, B. H. Bijnens, and G. Piella, "Characterization of myocardial motion patterns by unsupervised multiple kernel learning," *Medical image analysis*, vol. 35, pp. 70–82, 2017.

[68] S. Sanchez-Martinez, N. Duchateau, T. Erdei, G. Kunszt, S. Aakhus, A. Degiovanni, P. Marino, E. Carluccio, G. Piella, A. G. Fraser *et al.*, "Machine learning analysis of left ventricular function to characterize heart failure with preserved ejection fraction," *Circulation : cardiovascular imaging*, vol. 11, no. 4, p. e007138, 2018.

[69] M. Cikes, S. Sanchez-Martinez, B. Claggett, N. Duchateau, G. Piella, C. Butakoff, A. C. Pouleur, D. Knappe, T. Biering-Sørensen, V. Kutyifa *et al.*, "Machine learning-based phenogrouping in heart failure to identify responders to cardiac resynchronization therapy," *European journal of heart failure*, vol. 21, no. 1, pp. 74–85, 2019.

[70] M. Nogueira, M. De Craene, S. Sanchez-Martinez, D. Chowdhury, B. Bijnens, and G. Piella, "Analysis of nonstandardized stress echocardiography sequences using multiview dimensionality reduction," *Medical Image Analysis*, vol. 60, p. 101594, 2020.

[71] F. Loncaric, P.-M. M. Castellote, S. Sanchez-Martinez, D. Fabijanovic, L. Nunno, M. Mimbrero, L. Sanchis, A. Doltra, S. Montserrat, M. Cikes

et al., "Automated pattern recognition in whole-cardiac cycle echocardiographic data : capturing functional phenotypes with machine learning," *Journal of the American Society of Echocardiography*, vol. 34, no. 11, pp. 1170–1183, 2021.

[72] J. Mariette and N. Villa-Vialaneix, "Unsupervised multiple kernel learning for heterogeneous data integration," *Bioinformatics*, vol. 34, no. 6, pp. 1009–1015, 2018.

[73] B. Wang, A. M. Mezlini, F. Demir, M. Fiume, Z. Tu, M. Brudno, B. Haibe-Kains, and A. Goldenberg, "Similarity network fusion for aggregating data types on a genomic scale," *Nature methods*, vol. 11, no. 3, pp. 333–337, 2014.

[74] B. Wang, J. Jiang, W. Wang, Z.-H. Zhou, and Z. Tu, "Unsupervised metric fusion by cross diffusion," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 2997–3004.

[75] M. E. Tipping and C. M. Bishop, "Probabilistic principal component analysis," *Journal of the Royal Statistical Society Series B : Statistical Methodology*, vol. 61, no. 3, pp. 611–622, 1999.

[76] N. Lawrence and A. Hyvärinen, "Probabilistic non-linear principal component analysis with gaussian process latent variable models." *Journal of machine learning research*, vol. 6, no. 11, 2005.

[77] D. C. Liu and J. Nocedal, "On the limited memory bfgs method for large scale optimization," *Mathematical programming*, vol. 45, no. 1-3, pp. 503–528, 1989.

[78] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, pp. 2278–2324, 1998.

[79] N. D. Lawrence and A. J. Moore, "Hierarchical Gaussian process latent variable models," pp. 481–488, 2007.

[80] N. Duchateau, M. Viallon, L. Petrusca, P. Clarysse, N. Mewton, L. Belle, and P. Croisille, "Pixel-wise statistical analysis of myocardial injury in STEMI patients with delayed enhancement MRI," *Front Cardiovasc Med*, vol. 10, 2023.

[81] S. Bekkers *et al.*, "Microvascular obstruction : underlying pathophysiology and clinical diagnosis," *J Am Coll Cardiol*, vol. 55, pp. 1649–60, 2010.

[82] J. Alexandre *et al.*, "Scar extent evaluated by late gadolinium enhancement CMR : a powerful predictor of long term appropriate ICD therapy in patients with coronary artery disease," *J Cardiovasc Magn Reson*, vol. 15, p. 12, 2013.

[83] P. Shekelle, "Clinical practice guidelines : what's next ?" *JAMA*, vol. 320, pp. 757–8, 2018.

[84] A. Criminisi and J. Shotton, *Decision forests for computer vision and medical image analysis*. Springer Publishing Company, 2013.

[85] K. K. Bhatia, A. Rao, A. N. Price, R. Wolz, J. V. Hajnal, and D. Rueckert, "Hierarchical manifold learning for regional image analysis," *IEEE Transactions on Medical Imaging*, vol. 33, no. 2, pp. 444–461, 2014.

[86] F. Leeb, S. Bauer, M. Besserve, and B. Schölkopf, "Exploring the latent space of autoencoders with interventional assays," *Advances in Neural Information Processing Systems*, vol. 35, pp. 21 562–21 574, 2022.

[87] A. Damianou and N. D. Lawrence, "Deep gaussian processes," pp. 207–215, 2013.

[88] Y. Li, M. Yang, and Z. Zhang, "A survey on multi-view representation learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, no. 10, pp. 1863–1883, 2019.

[89] C. Ma, S. Tschiatschek, K. Palla *et al.*, "EDDI : efficient dynamic discovery of high-value information with partial VAE," *Proc. ICLR*, 2019.

[90] P. Zhou, Y.-D. Shen, L. Du, F. Ye, and X. Li, "Incremental multi-view spectral clustering," *Knowledge-Based Systems*, vol. 174, pp. 73–86, 2019.

[91] H. Yin, W. Hu, Z. Zhang, J. Lou, and M. Miao, "Incremental multi-view spectral clustering with sparse and connected graph learning," *Neural Networks*, vol. 144, pp. 260–270, 2021.

[92] Y. Bengio, J. Louradour, R. Collobert *et al.*, "Curriculum learning," *Proc. ICML*, pp. 41–8, 2009.

[93] C. W. H. Beijnink, N. W. van der Hoeven, L. S. F. Konijnenberg, R. J. Kim, S. C. A. M. Bekkers, R. A. Kloner, H. Everaars, S. El Messaoudi, A. C. van Rossum, N. van Royen, and R. Nijveldt, "Cardiac MRI to visualize myocardial damage after ST-segment elevation myocardial infarction : A review of its histologic validation," *Radiology*, vol. 301, no. 1, pp. 4–18, 2021.

[94] A. Matthews, M. van der Wilk, T. Nickson *et al.*, "GPflow : A Gaussian Process Library using TensorFlow ," *J Mach Learn Res*, vol. 18, pp. 1–6, 2017.

[95] J. Trygg and S. Wold, "Orthogonal projection to latent structures," *J Chemometr*, vol. 16, pp. 119–28, 2002.

[96] G. Bernardino, A. Jonsson, F. Loncaric *et al.*, "Reinforcement learning for active modality selection during diagnosis," *Proc. MICCAI, LNCS*, vol. 13431, pp. 592–601, 2022.