Methods for automation of vascular lesions detection in computed tomography images

Maria Alejandra Zuluaga Valencia

No. 10-2011

Université Claude Bernard Lyon 1

Universidad de los Andes

ÉCOLE DOCTORALE EEA
ÉLECTRONIQUE, ÉLECTROTECHNIQUE, AUTOMATIQUE

THÈSE

pour obtenir le grade de docteur de

l'Université Claude Bernard Lyon 1

Specialité : Images et Systèmes

et

Universidad de los Andes, Bogotá, Colombie

présentée et soutenue publiquement par

Maria Alejandra Zuluaga Valencia

Methods for automation of vascular lesions detection in computed tomography images

Soutenance prévue le 12 janvier 2011 devant le jury composé de

Rapporteurs	Isabelle Bloch	Telecom ParisTech
	Olivier Salvado	University of Queensland
Examinateurs	Isabelle Magnin	CREATIS, INSERM U1044
	José Tiberio Hernández	Universidad de los Andes
	Edgar Eduardo Romero	Universidad Nacional de Colombia
Directeurs	Maciej Orkisz	Université Claude Bernard Lyon 1
	Marcela Hernández Hoyos	Universidad de los Andes

No. 10-2011

Université Claude Bernard Lyon 1

Universidad de los Andes

ÉCOLE DOCTORALE EEA
ÉLECTRONIQUE, ÉLECTROTECHNIQUE, AUTOMATIQUE

THÈSE

pour obtenir le grade de docteur de

l'Université Claude Bernard Lyon 1

Specialité : Images et Systèmes

et

Universidad de los Andes, Bogotá, Colombie

présentée et soutenue publiquement par

Maria Alejandra Zuluaga Valencia

Methods for automation of vascular lesions detection in computed tomography images

Soutenance prévue le 12 janvier 2011 devant le jury composé de

Rapporteurs	Isabelle Bloch	Telecom ParisTech
	Olivier Salvado	University of Queensland
Examinateurs	Isabelle Magnin	CREATIS, INSERM U1044
	José Tiberio Hernández	Universidad de los Andes
	Edgar Eduardo Romero	Universidad Nacional de Colombia
Directeurs	Maciej Orkisz	Université Claude Bernard Lyon 1
	Marcela Hernández Hoyos	Universidad de los Andes

Methods for automation of vascular lesion detection in computed tomography images

This thesis presents a framework for the detection and diagnosis of vascular lesions with a special emphasis on coronary heart disease. Coronary heart disease remains to be the first cause of mortality worldwide.

Typically, the problem of vascular lesion identification has been solved by trying to model the abnormalities (lesions). The main drawback of this approach is that lesions are highly heterogeneous, which makes the detection of previously unseen abnormalities difficult. We have selected not to model lesions directly, but to treat them as anomalies which are seen as low probability density points. We propose the use of two classification frameworks based on support vector machines (SVM) for the density level detection problem. The main advantage of these two methods is that the learning stage does not require labeled data representing lesions, which is always difficult to obtain. The first method is completely unsupervised, whereas the second one only requires a limited number of labels for normality.

The use of these anomaly detection algorithms requires the use of features such that anomalies are represented as points with low probability density. For this purpose, we developed an intensity-based metric, denoted concentric rings, designed to capture the nearly symmetric intensity profiles of healthy vessels, as well as discrepancies with respect to the normal behavior. Moreover, we have selected a large set of alternative candidate features to use as input for the classifiers. Experiments on synthetic data and cardiac CT data demonstrated that our metric has a good performance in the detection of anomalies, when used with the selected classifiers.

Combination of other features with the concentric rings metric has potential to improve the classification performance. We defined an unsupervised feature selection scheme that allows the definition of an optimal subset of features. We compared it with existent supervised feature selection methods. These experiments showed that, in general, the combination of features improves the classifiers performance, and that the best results are achieved with the combination selected by our scheme, associated with the proposed anomaly detection algorithms.

Finally, we propose to use image registration in order to compare the classification results at different cardiac phases. The objective here is to match the regions detected as anomalous in different time-frames. In this way, more than attract the physician's attention to the anomaly detected as potential lesion, we want to aid in validating the diagnosis by automatically displaying the same suspected region reconstructed in different time-frames.

Keywords: Medical Imaging; Computer Assisted Diagnosis; Machine Learning; Support Vector Machines; Feature Selection; Computed Tomography; Vascular disease; Atherosclerosis

Méthodes d'automatisation de la détection des lésions vasculaires dans des images de tomodensitométrie

Résume: Les travaux de cette thèse sont consacrés à la détection et le diagnostic des lésions vasculaires, particulièrement dans le cas la maladie coronaire. La maladie coronaire continue à ètre la première cause de mortalité dans les pays industrialisés.

En général, l'identification des lésions vasculaires est abordée en essayant de modéliser les anormalités (lésions). Le principal inconvénient de cette approche est que les lésions sont très hétérogènes, ce qui rend difficile la détection de nouvelles lésions qui n'ont pas été prises en compte par le modèle. Dans cette thèse, nous proposons de ne pas modeliser directement les lésions, mais de supposer que les lésions sont des événements anormaux qui se manifestent comme points avec une faible densité de probabilité. Nous proposons l'utilisation de deux méthodes de classification basées sur les Machines à Vecteurs de Support (SVM) pour résoudre le problème de détection du niveau de densité. Le principal avantage de ces deux méthodes est que la phase d'apprentissage ne requiert pas de données étiquetées reprsentant les lésions. La première méthode est complètement non supervisée, alors que la seconde exige des étiquettes seulement pour les cas qu'on appelle sains ou normaux.

L'utilisation des algorithmes de classification sélectionnés nécessite des descripteurs tels que les anomalies soient représentées comme des points avec une densité de probabilité faible. À cette fin, nous avons développé une métrique basée sur l'intensité de l'image, que nous avons appelée concentric rings. Cette métrique est sensible à la quasi-symétrie des profils d'intensité des vaisseaux sains, mais aussi aux écarts par rapport à cette symétrie, observés dans des cas pathologiques. De plus, nous avons sélectionné plusieurs autres descripteurs candidats à utiliser comme entrée pour les classifieurs. Des expériences sur des données synthétiques et des données de CT cardiaques démontrent que notre métrique a une bonne performance dans la détection d'anomalies, lorsqu'elle est utilisée avec les classifieurs retenus.

Une combinaison de plusieurs descripteurs candidats avec la métrique concentric rings peut améliorer la performance de la détection. Nous avons défini un schéma non supervisé de sélection de descripteurs qui permet de déterminer un sous-ensemble optimal de descripteurs. Nous avons confronté les résultats de détection réalisée en utilisant le sous-ensemble de

déscripteurs sélectionné par notre méthode avec les performances obtenues avec des sous-ensembles sélectionnés par des méthodes supervisés existantes. Ces expriences montrent qu'une combinaison de descripteurs bien choisis améliore effectivement les performance des classifieurs et que les meilleurs résultats s'obtiennent avec le sous-ensemble sélectionnépar notre méthode, en association avec les algorithmes de détection retenus.

Finalement, nous proposons de réaliser un récalage local entre deux images représentant différentes phases du cycle cardiaque, afin de confronter les résultats de détection dans ces images (phases). L'objectif ici est non seulement d'attirer l'attention du praticien sur les anomalies détectées comme lésions potentielles, mais aussi de l'aider à conforter son diagnostic en visualisant automatiquement la même région reconstruite à différents instants du cycle cardiaque.

Mots clés: Imagerie médicale; Diagnose assisté par ordinateur; Machine Learning; Support Vector Machines; Selection de déscripteurs; Tomographie; Maladie vasculaire; Atherosclerose

Thèse préparée à
Creatis
Bât. Blaise Pascal
7 Av. Jean Capelle
69621 Villeurbanne Cedex
France

Synthèse

La maladie coronaire continue à être la principale cause de mortalité dans le monde, provoquant 7,2 millions de décès par an. Chez les patients symptomatiques, le diagnostic de la présence et sévérité de la maladie vasculaire est essentiel pour déterminer un traitement médical approprié. La référence pour le diagnostic de la maladie coronaire est encore la coronarographie. Le plus grand avantage de cette angiographie conventionnelle est sa haute résolution spatiale et la possibilité d'exercer directement des interventions telles que la dilatation par ballonnet ou pose de stent coronaire. Cependant, en raison de son caractère invasif, la coronarographie a un risque, faible mais non négligeable, de complications.

Un développement rapide de la tomodensitomtrie (CT) au cours des 6-8 dernières années a rendu possible l'angiographie CT des artères coronaires, dans un contexte clinique. Cette technique non invasive permet la visualisation non seulement de la lumière artérielle coronaire, mais aussi de la présence de plaque d'athérosclérose. Cependant, son utilisation soulève encore plusieurs préoccupations telles que les doses de rayonnement plus élevés qu'en angiographie conventionnelle, ainsi que des difficultés affectant l'évaluation de la maladie coronaire, notamment de l'étendue et du type de plaque:

- Dans la littérature, il n'y a pas de consensus quant aux nombres CT (unités Hounsfield) qui différencient les composants de la plaque.
- Les artères coronaires suivent les mouvements du cœur provoquant des artefacts dans les images reconstruites. Les distorsions résultantes obligent le radiologue à visualiser l'artère d'intérêt dans des différentes phases, généralement diastole et systole, afin de créer un modèle mental de l'image qui leur permettra de poser un diagnostic.
- La qualité des images varie selon le protocole d'acquisition et les caractéristiques individuelles du patient.
- L'analyse, l'interprétation et la documentation des examens CT coronaires sont complexes et pas suffisamment standardisées. Bien qu'il existe des pratiques communes chez les médecins il n'y a pas de protocole d'évaluation bien établi et universel.

Ce contexte médical est décrit dans le chapitre 1.

Pour surmonter les difficultés mentionnées ci-dessus, dans ce travail, nous proposons l'élaboration d'un cadre méthodologique visant à l'identification des lesions vasculaires. L'objectif de notre proposition n'est pas de remplacer le travail des cliniciens, mais de

les aider dans le diagnostic en soulignant des lésions potentielles. Ce cadre méthologique est décrit dans le chapitre 2.

Le cœur de notre proposition repose sur une stratégie de classification. Notre approche est basée sur l'apprentissage automatique, cherchant à classer comme normales ou anormales les coupes orthogonales à un axe précédemment extrait. Une préoccupation majeure d'un tel choix tient au fait que la plupart des approches existantes basées sur l'apprentissage automatique et visant à la détection des lésions vasculaires, suivent un schéma de classification supervisée. Ces méthodes reposent sur la disponibilité d'étiquettes précises et représentatives, aussi bien pour la normalité que pour toutes les formes de lésions. Un tel étiquetage étant très laborieux et demandant une grande expertise et beaucoup de temps, il n'est pas souvent possible de l'obtenir. Deuxièmement, les méthodes supervisées ont des difficultés pour détecter de nouvelles formes de lesions qui n'étaient pas représentées dans la base d'apprentissage. Comme les lésions sont par nature hétérogènes, des tendances anormales inconnues peuvent souvent apparaître. Pour surmonter ces problèmes, nous avons proposé l'utilisation de méthodes qui réduisent au minimum la dépendance par rapport aux ensembles de données étiquetées. En outre, nous avons formulé le problème d'une façon qui, à notre connaissance, n'a jamais été utilisée dans ce domaine. Nous formulons notre problème comme un problème de détection des anomalies. Nous considérons qu'une anomalie peut être définie comme une observation qui semble être incompatible avec le reste de l'ensemble des données considérées. Dans cette perspective, nous nous concentrons sur une seule classe, la normalité, en supposant que tout ce qui s'en écarte peut être considéré comme anormal. Par conséquent, notre approche consiste à modéliser la normalité par sa haute densité de probabilité et à considérer que les données non-conformes au comportement attendu ont une densité de probabilité faible. Notre formulation permet l'utilisation de méthodes d'apprentissage semi-supervisées et non-supervisées.

L'hypothèse que les anomalies sont des points de densité de probabilité faible nous amène à formuler le problème comme un problème de détection du niveau de densité (DLD, pour density level detection en anglais). A cet effet, deux mesures sont définies. La première permet de mesurer la concentration absolue d'une distribution de probabilité, tandis que la seconda mesure la concentration d'une distribution, par rapport à une autre distribution qu'on appele *mélange*. Ces mesures permettent la détermination des niveaux de densité qui differencient les points de densité probabilité faible de ceux qui sont très concentrés. Sur la base de ces deux mesures, nous formulons deux algorithmes qui permettent de détecter les niveaux de densité. Les deux méthodes sont basées sur des séparateurs à vaste marge (SVM).

Le premier algorithme, nommé DLD-SVM, fait usage de la mesure de la concentration absolue. Sa caractéristique principale est qu'il est complètement non-supervisé. Cela signifie qu'il ne nécessite pas d'étiquettes à l'étape de l'apprentissage, car il cherche tout simplement les régions de forte densité de probabilité, puis il identifie les anomalies avec le complément des régions normales. Le deuxième algorithme, nommé LPU (learning from positive and unlabeled data), est adapté aux situations où un petit ensemble de données étiquetées (normales) et une grande quantité de données non-étiquetées sont

disponibles. Le but de cet algorithme est de trouver la densité relative des échantillons provenant de la première série de manière à être en mesure de détecter des échantillons de la même distribution probabiliste.

Une fois les classifieurs décrits, nous introduisons une nouvelle métrique utilisée pour calculer leurs entrées. La métrique est appelée anneaux concentriques. Elle a été conçue dans le but de capturer les profils d'intensité et la symétrie axiale des vaisseaux normaux, tout en étant sensible aux écarts par rapport à ce comportement normal.

L'objectif de la définition d'une nouvelle métrique n'a pas été de trouver quelque chose qui pourrait résoudre complètement le problème de classification. Tout d'abord, nous croyons que cela n'est pas possible et le deuxièmement, il a été montré que la combinaison des descripteurs de nature différente améliore la classification. C'est pourquoi, outre les anneaux concentriques, nous avons également sélectionné un ensemble d'autres déscripteurs candidats à tester avec nos classifieurs. Vu que la détection d'anomalies vasculaires avec des techniques d'apprentisage automatique est assez nouvelle, il n'est pas possible d'établir un ensemble de déscripteurs performants à partir de l'état de l'art dans le domaine. Par conséquent, nous utilisons des déscripteurs globaux, couramment utilisés dans le rehaussement ou la segmentation vasculaires, pour définir un ensemble initial de candidats. Notre critère pour sélectionner ce type de descripteurs repose sur le fait qu'ils sont censés donner de fortes réponses à des endroits situés dans les régions de lumière normale, tandis que leurs réponses sont susceptibles de s'écarter de la normalité en présence d'une lésion. Le réglage des paramètres de tous les descripteurs utilisés dans notre étude, ainsi que l'évaluation des performances individuelles des anneaux concentriques en association avec nos classifieurs, sont décrits dans le chapitre 3. Le réglage des paramètres de notre métrique a été réalisé par une série d'évaluations sur des fantômes artificiels et sa validation a posteriori sur des données CT de patients. Les résultats ont montré une bonne spécificité et une bonne sensibilité, aussi bien en association avec DLD-SVM qu'avec LPU. En outre, les résultats de classification par les deux algorithmes démontrent un niveau substantiel de la coïncidence avec les annotations des observateurs, mesuré par le coefficient Kappa (0,72% et 0,71%, respectivement). Les résultats ont permis de conclure que la métrique des anneaux concentriques peut être appliquée à la détection des lésions vasculaires. Comparés individuellement à chacune des autres métriques sélectionées, les anneaux concentriques se montrèrent les meilleurs avec DLD-SVM, et quatrièmes avec LPU.

Dans ce travail, nous explorons également différentes approches de sélection des descripteurs pour améliorer les performances de la détection des anomalies vasculaires. Notre objectif est de déterminer les combinaisons de descripteurs qui améliorent la performance de classification et de définir des lignes directrices pour leur sélection. Pour cette tâche, nous avons d'abord défini une stratégie non-supervisée utilisant des informations fournies par les classifieurs basés sur la notion de densité de probabilité, à savoir le risque empirique. L'avantage de cette stratégie, outre son caractère non-supervisé, est d'être intégrée avec les classifieurs au sens de l'utilisation des mémes mécanismes. En outre, nous avons évalué trois stratégies de sélection existantes, toutes supervisés, connues sous les noms F-score, forêts aléatoires combinées avec SVM et

SVM-RFE. L'évaluation des performances de ces différentes stratégies fait l'objet du chapitre 4. Cette évaluation a montré que les propositions non-supervisée à l'aide DLD-SVM et semi-supervisée avec LPU, ont été les plus performantes. Ici, la spécificité a eu des valeurs de 96,37 et 86,71%, et les sensibilités de 76,52 et 83,84%, respectivement. En règle générale, DLD-SVM a tendance à avoir une meilleure spécificité que LPU, tandis que LPU tend à avoir une plus grande sensibilité. Les anneaux concentriques figurent toujours parmi les descripteurs sélectionnées en utilisant DLD-SVM et LPU. Cela confirme le potentiel de notre métrique dans le problème de la détection des lésions vasculaires.

Dans le cadre de notre évaluation, nous avons également comparé la performance de DLD-SVM et LPU avec d'autres méthodes de l'état de l'art pour la classification. Les méthodes évaluées se composait de deux classifieurs classiques, forêts aléatoires et SVM à marge douce, et une méthode de détection d'anomalies très populaire, one-class SVM. Non seulement DLD-SVM et LPU ont largement dépassé les autres méthodes, mais cette évaluation nous a permis de valider la pertinence de la perspective de la détection des anomalies. Dans tous les cas, les trois méthodes de détection des anomalies ont dépassé les classifieurs classiques.

Dans le chapitre 5, nous présentons une stratégie visant à comparer les anomalies détectées à différents instants du cycle cardiaque. Notre proposition adapte un algorithme existant de récalage non-rigide, de manière à être utilisé à des régions particulières de l'image. Sur la base des anomalies détectées, nous construisons une région d'intérêt (ROI), dans l'image correspondant à une phase, et nous identifions, dans la deuxième phase, une région avec une forte probabilité de contenir les informations de la première ROI. Les deux ROI sont ensuite reécalés spatialement. La méthode a été évaluée sur 10 différents ensembles de données, 3 ROI par ensemble de données, en montrant une bonne correspondance. Le principal avantage de cette implémentation est que, en raison de son caractère local, sa charge de calcul est faible, ce qui la rend envisageable pour la pratique clinique de routine. Ainsi, le clinicien pourra prononcer le diagnostic final, en analysant, dans deux phases différentes, la même région suspecte à laquelle le système attirera automatiquement son attention.

Acknowledgements

First of all, I would like to thank my advisors Marcela Hernández and Maciej Orkisz for giving me the opportunity to work and learn by their side during my PhD. It has been a pleasure to work with you, despite the long, tiring and sometimes difficult journeys.

I am very grateful with my reviewers Isabelle Bloch and Olivier Salvado for accepting to read and correct this manuscript, and for being part of my jury. I am also thankful with Isabelle Magnin for her interest in my work and her wise advices about my research. All my gratitude to the jury members for accepting to be part of the committee.

I am indebted to all the people I had the chance to work with. I should first of all mention Edgar Delgado, for his precious help in coding and in the visualization of my results. I am also grateful to Alfredo Morales, Vincent Doré, Juan C. Prieto and, of course, Mónica Lozano, for all the time devoted to the challenge adventures. I also thank the physicians that were involved in this project at one point or another. I am very thankful to Philippe Douek, for his wise comments on the clinical motivation of this work, and to Julio César Dávila and Luis Felipe Uriza for the enormous time they spent doing annotations.

I owe my deepest gratitude to all those who kindly shared some of their knowledge with no other interest than making things better. Thanks a lot to Marc Robini and Fernando Lozano for the valuable discussions, Patrick and Olivier for their suggestions on the final presentation. A special word goes to Don Hush for his invaluable help, his disposition to answer all my questions and his great enthusiasm. You represent what I consider science should be.

I would like to thank all those who went along with me during these years. First, I mention the uniandinos: Carolina, Torrado, Lina, Andrew, Soledad, Javier, Camila, Alejandro, Gabriel and M. Alejandra. Thanks for making Uniandes such a warm place to work. I shall also mention my collegues and friends from the French-side: Pierre F., Marion, Xavier, Emily, Elena, Jean-François, Rodica, Eduardo, Vivi and Feng. I really enjoyed France.

A big thank-you goes to my family for their unconditional support. I thank Fernando for his encouragement and patience. Finally, I shall thank my grandmother. The fact of being here today, I owe it largely to her.

A todos, gracias.

Contents

List of Figures

LIST OF FIGURES

List of Tables

LIST OF TABLES

1

Introduction

Coronary heart disease (CHD) remains to be the main cause of mortality worldwide, World Health Organization (2008) reporting 7.2 millions of deaths per year. In symptomatic patients, diagnosis of the presence and severity of the vascular disease is critical for determining appropriate clinical management. Advances in the resolution of multidetector computed tomographic (CT) angiography have encouraged its use in the assessment of vascular diseases. Nevertheless, the detection and quantification of vascular lesions continue to be a challenging and tedious work for physicians who have to explore a vast amount of data using different visualization schemes based on advanced post-processing techniques. Therefore, our main objective is to develop automatic or semi-automatic tools that ease the diagnosis stage.

This introductory chapter addresses coronary heart disease and CT as an alternative tool for its diagnosis and the development of tools that facilitate this task. Initially, basic anatomical aspects of the cardiovascular system (Section 1.1.1) and, more particularly, of coronary arteries (Section 1.1.2) and the CHD (Section 1.1.3) are briefly reviewed. Then, advantages and drawbacks of CT technology in the diagnosis of CHD are discussed (Section 1.1.4). In Section 1.2, we introduce computer aided diagnosis systems as an alternative in CHD assessment. Our specific proposal in order to reduce the associated problems of CHD diagnosis and the contributions of this thesis are presented in section 1.3 to finally conclude with an outline of the organization of this manuscript in section 1.4.

1.1 Medical Context

1.1.1 The Cardiovascular System

The cardiovascular system or circulatory system is an organ system that passes nutrients, gases, hormones, etc. to and from cells in the body. The main components of the human cardiovascular system are the heart and the blood vessels (arteries and veins).

Both arteries and veins have a three layer structure of the wall (Figure 1.1). The interior of the vessel enclosed by the wall is the vessel lumen. The *tunica intima* is the

thinnest and innermost layer. It is composed of a single layer of endothelial cells and a small amount of connective tissue. The second and thickest layer is the *tunica media* that provides structural support, vasoreactivity and elasticity. It is composed of smooth muscle cells, elastic fibers and connective tissue, which vary in amount depending on the type of vessel. The *tunica adventitia* is the outermost layer. It is entirely made of fibrous connective tissue and it contains nerves that supply the vessel as well as nutrient capillaries in the larger blood vessels (Gray, 2000).

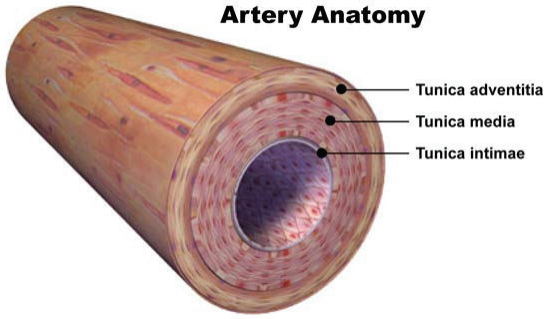


Figure 1.1: Blood vessel anatomy - The vessel wall is made up of three layers tunica adventitia, tunica media and tunica intima. The region enclosed by the vessel wall is called the vessel lumen (source: LiveScience http://www.livescience.com/health/060619_synthetic_arteries.html

1.1.2 The Heart and the Coronary Arteries

The heart is a muscular organ responsible for pumping blood throughout the vessels by repeated, rhythmic contractions. The cardiac muscle (myocardium) receives its own blood supply from the coronary arteries. The coronary arteries are therefore vital for the correct functioning of the whole circulatory system. They branch off from the ascending aorta near the point where the aorta and the left ventricle meet. A branching point is called an *ostium*. Two major branches of the coronary arteries arise from the aorta (Figure 1.2):

The right coronary artery (RCA). It branches into the right marginal artery and the posterior descending artery. In general, the RCA supplies blood to the right atrium, the right ventricle, the bottom part of both ventricles and back of the septum.

The left main coronary artery (LM). The LM divides after a short course into the circumflex artery (LCX) and the left anterior descending artery (LAD). The former supplies blood to the left atrium, side and back of the left ventricle, whereas the latter supplies the front and bottom of the left ventricle and the front of the septum. Both LCX and LAD give rise to other branches. Amongst others, the left marginal artery branches from the LCX while the diagonal artery rises from the LAD.

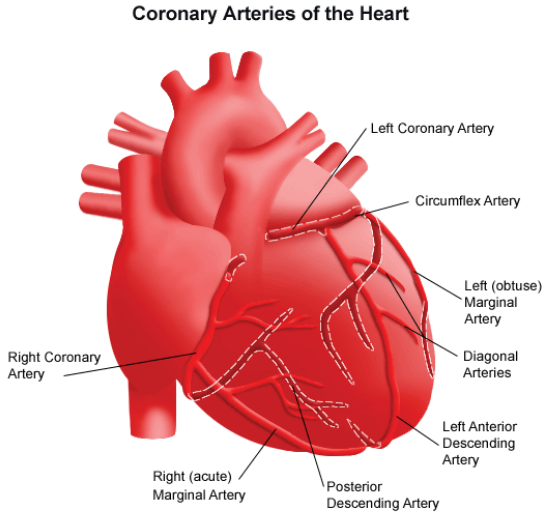


Figure 1.2: The coronary arteries - Main coronary arteries representation (source: Jefferson University Hospitals http://www.jeffersonhospital.org/Tests-and-Treatments/coronary-artery-bypass-grafting.aspx)

1.1.3 Coronary Heart Disease

Coronary heart disease (CHD), also known as coronary artery disease, is a condition characterized by the development of atherosclerotic plaques in the coronary arteries. Atherosclerosis is a chronic immuno-inflammatory disease in which the artery wall thickens as the result of a build-up of different materials (*i.e.* cholesterol, calcium, fibro-fatty deposits). The wall thickening causes progressive narrowing and hardening of the arteries over time.

Stary (2000) proposed a classification of atherosclerotic plaques into eight different types (Table 1.1). Types range from minimal intimal change to changes associated with severe clinical manifestations. This classification scheme has been accepted by the American Heart Association.

CHD can remain asymptomatic. This occurs in the presence of lesions up to type III. CHD associated with the presence of type IV or higher lesions can lead to serious complications such as angina pectoris, acute myocardial infarction, silent ischemia, arrhythmias, left ventricular dysfunction, ischemic cardiomyopathy or sudden death.

1.1.4 CHD Assessment through Computed Tomography Angiography

The standard of reference for diagnosis of CHD still is conventional coronary angiography (CCA) (Figure 1.3). The greatest advantage of conventional angiography is high spatial resolution and the option of direct performance of interventions such as balloon dilatation or coronary stent placement. However, only one-third of all conventional

Table 1.1: Stary's classification of atherosclerotic lesions. Atherosclerotic lesions classification as proposed by Stary (2000) and accepted by the American Heart Association.

Classification	Description
Type I	Microscopically and chemically detectable lipid deposits in the intima and the cell reactions associated with such deposits. No tissue injury.
Type II	Fatty streaks. Streaks may be visible as yellow-colored streaks, patches, or spots on the intimal surface of the artery.
Type III	Microscopic evidence of tissue injury. Also called preatheroma.
Type IV	Extensive lipid core, massive structural injury. Also called atheroma.
Type V	Increased amount of smooth muscle and collagen. It can be multilayered (several lipid cores, separated by layers of fibrous connective tissue that are stacked irregularly). Denoted as fibroatheroma.
Type VI	Development of disruptions of the lesion surface, hematoma or hemorrhage, and thrombotic deposits
Type VII	The dominant feature of the lesion is the presence of calcium.
Type VIII	The normal intima is replaced and thickened with fibrous connective tissue. The lipid core is minimal or absent

coronary angiographic examinations in the United States are performed in conjunction with an interventional procedure, while the rest are performed only for diagnostic purposes (Schoepf et al., 2004). Furthermore, because of its invasive nature, CCA has a low, but non-negligible, risk of procedure related complications (Zanzonico et al., 2006). Given the high incidence of coronary artery disease, performing coronary angiography for diagnostic purposes seriously affects limited health-care resources. On the other hand, it is preferable for a patient not to undergo an unnecessary and potentially risky invasive test, if a reliable noninvasive imaging modality to visualize the coronary arteries is available.

A surprisingly rapid development of CT technology over the past 6-8 years, especially the introduction of 64-slice CT with gantry rotation times well below 500 ms, has made imaging of the heart and coronary arteries be CT possible in a clinical setting (Achenbach, 2007) (Figure 1.4). CT is a noninvasive technique that allows, next to the assessment of the coronary lumen, the evaluation of the presence, extent, and type of coronary plaque (Leber et al., 2004; Schroeder et al., 2001). Such non-invasive, comprehensive plaque assessment may be relevant for improving risk stratification when combined with current risk measures: the severity of stenosis and the amount of calcium (Schroeder et al., 2001). The latter represents an advantage w.r.t. CCA that only provides information on the coronary lumen. Nevertheless, CT cannot distinguish among all types of plaque. It is capable of differentiating between calcified and soft plaque (Figure 1.5), but it can hardly distinguish the different types of soft plaque.

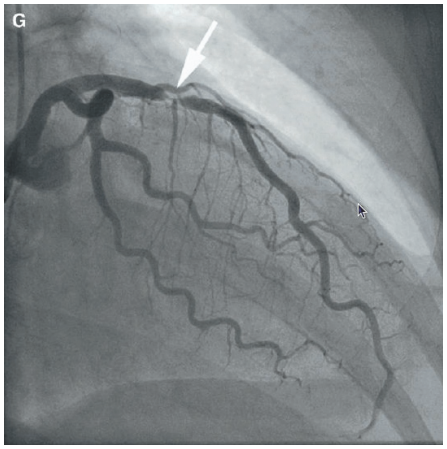
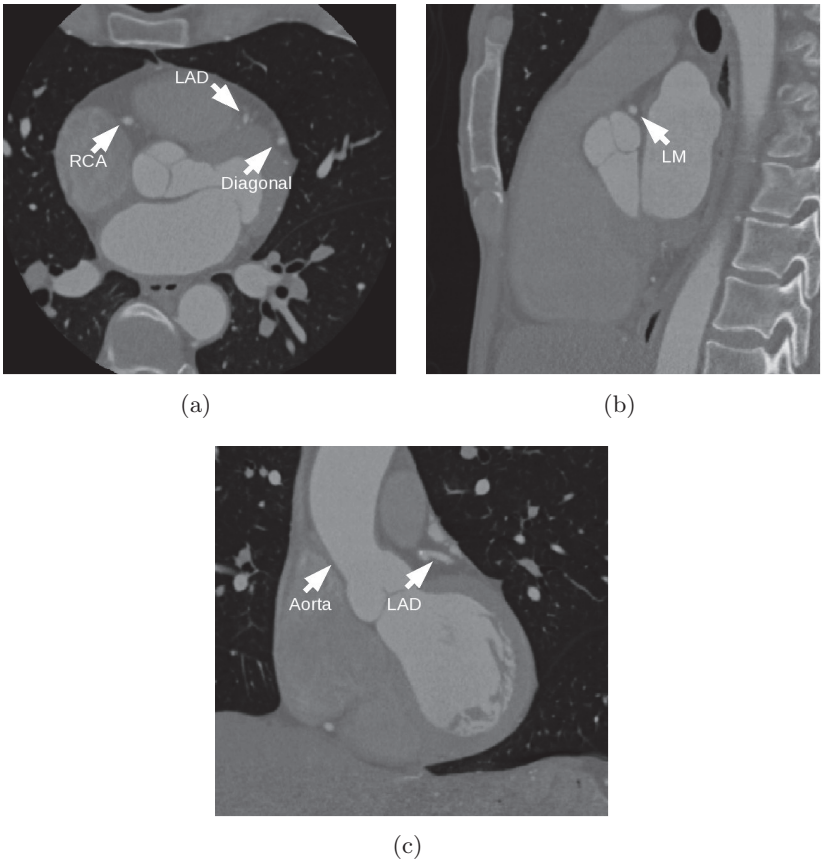


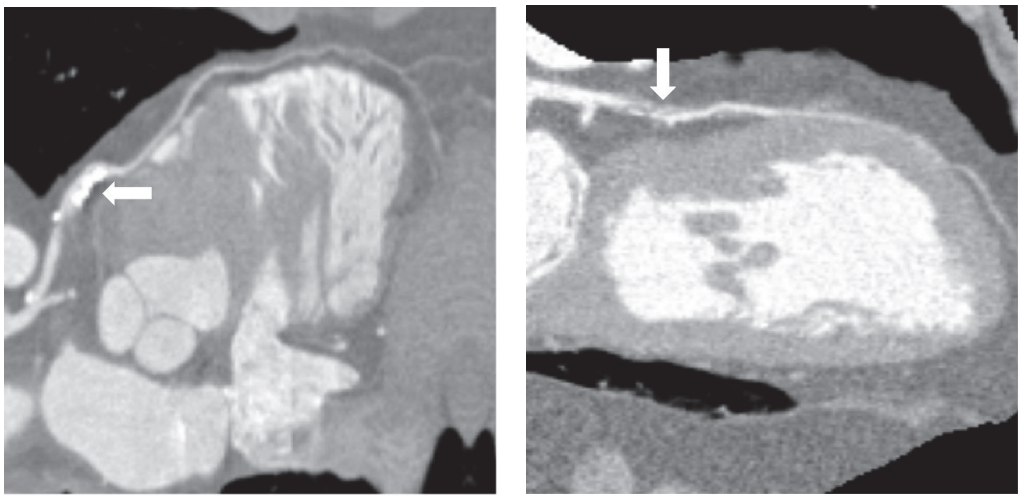
Figure 1.3: Conventional Coronary Angiography - The arrow shows a detected stenosis. Source: (Achenbach, 2007))



(a) (b)

(c)

Figure 1.4: Cardiac CT - Multi-planar reformation (MPR) view of a 3D cardiac data set. (a) axial view, (b) coronal view and (c) sagittal view.



(a) (b)

Figure 1.5: Visualization of atherosclerotic plaques in CT - Visualization of atherosclerotic plaques using a curved planar reformation (CPR) view. (a) The arrow points to a calcified plaques along the LAD (b) The arrow points to a soft plaque in the LAD. Images obtained using MeVisLab.

Although accuracy for the detection of hemodynamically relevant coronary artery stenoses is high for state-of-the-art multidetector CT (MDCT) systems (Achenbach, 2007) and despite its potential as a diagnosis tool for CHD, the use of CT still raises several concerns. One of the principal concerns is that CT yields higher radiation doses than traditional CCA (Lesage, 2009). In the following, some of the major difficulties of CT technology that affect CHD diagnosis are described.

Characteristic Hounsfield Units of the Plaque

In a CT image, each pixel is assigned a numerical value (CT number), which is the average of all the attenuation values contained within the corresponding voxel. This number can be represented in terms of Hounsfield Units (HU).

The HU scale is a linear transformation of the original linear attenuation coefficient measurement into one in which the radiodensity of distilled water at standard pressure and temperature (STP) is defined as zero Hounsfield units (HU), while the radiodensity of air at STP is defined as -1000 HU. For a material X with linear attenuation coefficient μ_X, the corresponding HU value is therefore given by

$$HU = \frac{\mu_X - \mu_{water}}{\mu_{water} - \mu_{air}} \times 1000 \tag{1.1}$$

where μ_{water} and μ_{air} are the linear attenuation coefficients of water and air, respectively. These standards were chosen as they are universally available references and suited for the purpose for which CT was developed: imaging the internal anatomy.

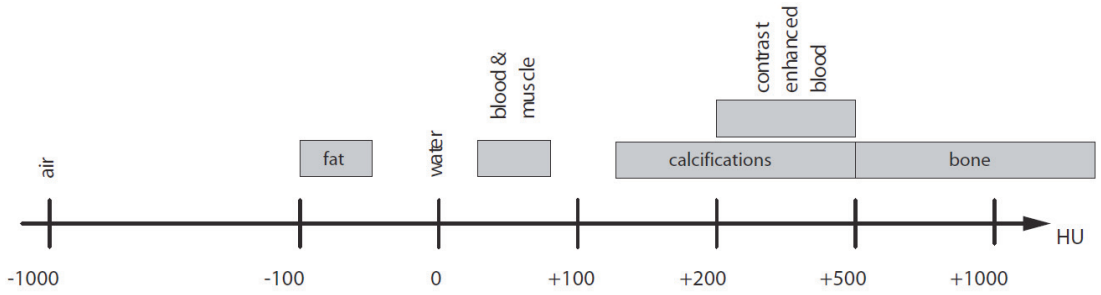


Figure 1.6: Example of HU ranges - Example of HU ranges for selected tissues. Blood is meant with contrast-enhancing agent injected. Source: (Saur, 2009).

Although one can find in literature HU ranges for certain tissues (Figure 1.6), there is no consensus on their actual bounds. More particularly, neighboring ranges partially overlap. As an example, in literature it is not possible to find clearly defined thresholds that differentiate plaque components as a function of the Hounsfield Units. Saur (2009) summarized the ranges proposed by different studies on CT images aiming at the detection of coronary plaques. Results demonstrate that there is not a general consensus (Table 1.2). Moreover, additional difficulties arise from the fact that some of these values overlap with the intensity ranges that are proposed for vessel lumen. This mainly comes from the use of the contrast material that modifies the radio-opacity of the blood in order to permit the visualization of the lumen. As an example, let us cite de Weert et al. (2008) who used a threshold of 130 HU for calcifications and had to resort to manual separation between them and contrast-enhanced arterial blood, since the latter reached on average 217.4 ± 36.9 HU, which clearly overlaps with the calcium range. Although the quoted work reports results in carotid arteries, the same can be observed in coronaries.

Temporal and spatial resolution

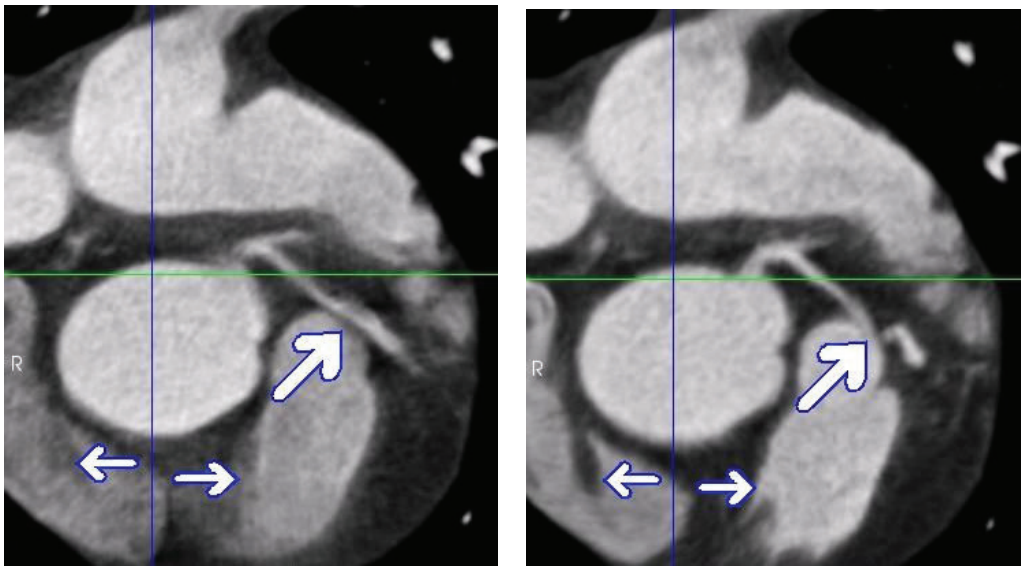
Coronary arteries undergo heterogeneous movement and deformations throughout the cardiac cycle causing motion artifacts. ECG synchronization is performed to reduce cardiac motion and minimize artifacts. Retrospective ECG gating is a method that is based on the simultaneous acquisition of CT data and the ECG signal. The acquired projected raw data are selected for image reconstruction with respect to a predefined cardiac phase (also denoted time-frame or time-point). However, when the velocity exceeds the temporal resolution of the scanner (Husmann et al., 2007) motion artifacts cannot be avoided. Moreover, frequently there are diverging motion patterns between the left and right coronary systems and often also between proximal and distal portions of the same coronary artery (Lu et al., 2001) leading again to motion correction problems.

Coronary arteries are relatively small and thin. Typically, they have radii in a range

Table 1.2: Reported HU value ranges of lipid, fibrous, and calcified tissue. HU value ranges (mean ± std) for lipid, fibrous, and calcified tissue as reported in various CT studies. Namely, I:(Schroeder et al., 2001), II: (Leber et al., 2004), III: (Brodoefel et al., 2008), IV: (Hein et al., 2007), V: (Pohle et al., 2007), VI: (Carrascosa et al., 2003), VII: (Motoyama et al., 2007) and VIII: (Sun et al., 2008). Voltage and current parameters are also reported to ease comparison. Table information obtained and adapted from (Saur, 2009)

Study	Lipid	Fibrous	Calcified	Voltage (kV)	Current (mA)
I	-42 - +47	61 - 112	126 - 736	140	60
	(14 ± 26)	(91 ± 21)	(419 ± 194)		
II	14 - 82	34 to 125	162 to 820	120	450
	(49 ± 22)	(91 ± 22)	(391 ± 156)		
III	-10 - +69	70 to 158	> 437	120	400
IV	-100 - +20	20 to 130	350 to 1000	-	-
V	-39 - +167	60 - 201	-	120	-
	(58 ± 43)	(121 ± 34)	-		
VI	(76 ± 44)	(149 ± 37)	(449 ± 221)	120	360
VII	-15 - +33	32 to 130	221 to 1134	135	400
	(11 ± 12)	(78 ± 21)	(516 ± 198)		
VIII	7 - 149	22 - 154	295 - 1325	120 - 135	350 - 450
	(79 ± 34)	(90 ± 27)	(772 ± 251)		

from 1 to 10 voxels. The temporal and spatial resolution of CT is still somewhat limited relative to the rapid motion and small dimensions of the coronary arteries (Achenbach, 2007). In consequence, due to their reduced size, coronary arteries can be seriously affected by the partial volume effects and reconstruction artifacts caused by motion.



(a) (b)

Figure 1.7: Effects of coronary motion - Same CT slice at different instants of the cardiac cycle (after rigid registration). The coronary artery pointed by big diagonal arrows is better depicted in (a), while some other regions (small arrows) are more contrasted in (b).

Figure 1.7 illustrates the effects of motion in image quality. The same 2D slice of the heart is presented at different heart cycle phases. While Figure 1.7(a) presents a normal separation between the left coronary artery and the neighboring cardiac chamber, Figure 1.7(b) contains an artifact that locally alters the image contrast. Vessel segmentation using the right image would lead to an erroneous result that leaks into the cavity. An analysis, slice-by-slice, of the 3D images demonstrates that this situation is very common: arteries are well defined in some heart phases while in others they are severely altered. This type of situation can lead to what is illustrated in Figure 1.8. The 3D reconstruction shows an interruption of the vessel. This can be a stenosis or it can be produced by a motion artifact.

As a consequence, radiologists are obliged to visualize the vessel of interest in different phases, typically one at diastole and another at systole, in order to create a mental image model that will allow them to give a diagnosis. As a consequence of movement, the spatial location of the structures of interest is not preserved between phases. The radiologist needs to search for the structure of interest at different locations in each evaluated phase. The amount of data in CT acquisitions is large (typically 512×512

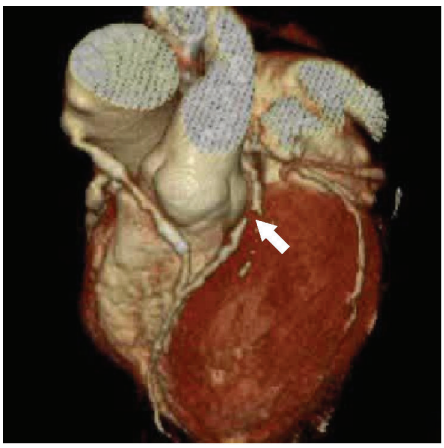


Figure 1.8: 3D reconstruction of the heart and coronary arteries - The reconstruction shows an interruption in the vessel (arrow). This can be caused by a stenosis or an image artifact.

× 512 voxels per phase), which makes this manual labor time consuming.

Acquisition protocol

There is no general consensus on the protocol that should be followed in CT image acquisition for CHD diagnosis. A wide range of studies on the assessment of CHD using CT found in literature use different protocols. Table 1.3 compares four studies for CHD diagnosis using the same CT technology. Although even three of the studies use CT scanners from the same vendor, the protocols differ from each other. Furthermore, the information provided by each study is different which makes difficult the reproducibility of the acquisitions.

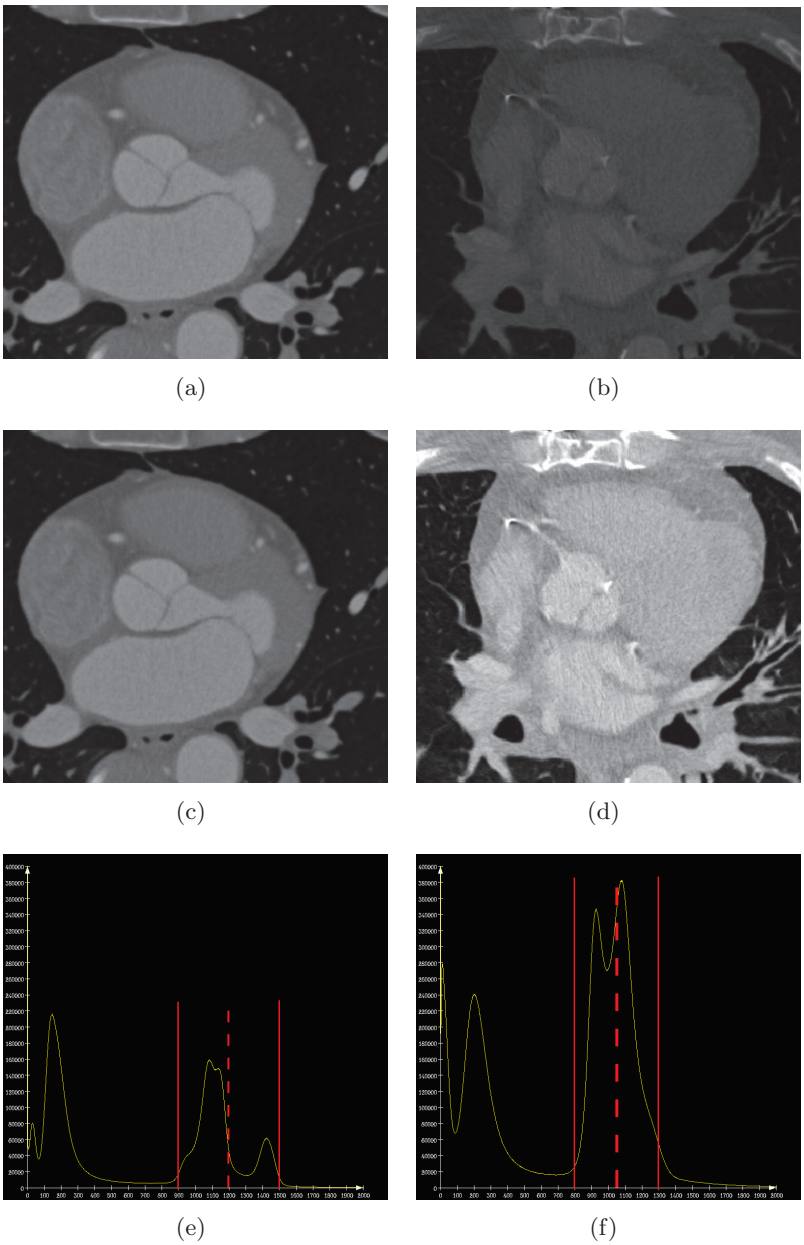
With the lack of a standardized acquisition protocol, image quality and dynamics can significantly vary from one center to another, which makes the evaluation more or less difficult. Even with the same acquisition protocol followed within a given center, significant differences between acquisitions may occur. Figure 1.9 illustrates such an example, which can be explained by different circulation time between the contrast injection and the actual acquisition from one patient to another. As a consequence, Figure 1.9 (c) shows a high quality image whereas the result shown in Figure 1.9 (d) is of low quality. The dynamics of both images differs substantially (Figures 1.9(e) and (f)).

Evaluation protocol

The analysis, interpretation, and documentation of coronary CT examinations are complex and not sufficiently standardized (Schoenhagen et al., 2004). Although there are common practices among physicians there is not a well established universal evaluation protocol.

Table 1.3: Examples of CT acquisition protocols. Comparison of 4 CT acquisition protocols used in the study of CHD detection. For a fair comparison, studies using CT scanners with the same number of slices (16) were kept. Protocol I denotes (Achenbach et al., 2004), II is (Caussin et al., 2004), III is (Leber et al., 2004), IV is (Pohle et al., 2007) and V is (Motoyama et al., 2007).

Parameters	I	II	III	IV	V
Scan					
Collimation (mm)	12 × 0.75	-	12 × 0.75	16 × 0.75	16 × 0.5
Table feed (mm)	2.8	-	-	-	-
Rotation (ms)	420	420	-	420 or 375	400
Voltage (kV)	120	-	120	120	135
Tube current (mA)	400	-	450 (55% of cycle). 80% reduction for remaining time	-	360
Radiation dose (mSv)	4.3	-	4.3	-	-
Contrast					
Agent	-	-	Solutrast 300	-	
Quantity (mL)	80	100	80	60	80
Injection rate (mL/s)	4	-	5	4 or 5	3
Reconstruction					
Algorithm	ECG half scan	-	-	ECG half scan	-
Temporal resolution (ms)	210		210	210 or 188	75
Kernel	B35f	-	-	B35f	-
Slice thickness (mm)	1.0	0.75	-	1.0	-
Increment (mm)	0.5	0.5	-	0.5	-
Pixel size (mm)	-	-	0.39 × 0.39	-	-
Scanner					
Model			Sensation		Aquilion
Vendor			Siemens		Toshiba



(a) (b)

(c) (d)

(e) (f)

Figure 1.9: Quality variation in cardiac CT images - Two CT images acquired at the same medical center. (a) and (b) show the original window-level configuration. (c) A high quality image where the vessels are well contrasted (d) Bad quality image with low contrast and image artifacts. The corresponding image histograms are shown in (e) and (f). The red lines denote the limits of the selected window on each image to visualize arteries and the dashed line represents its center.

Two types of visualization are common in CHD assessment. The first one is multi-planar reformation (MPR), where the radiologist scrolls through the 2D slices of axial, coronal and sagittal views (Figure 1.4). The second one is curved planar reformation (CPR) (Kanitsar et al., 2002). The goal of CPR visualization (*e.g.* Figure 1.5) is to make a tubular structure visible in its entire length within one single image (Kanitsar et al., 2002), so that the radiologists can sweep through the vessel searching for possible lesions. The CPR display mode requires a previous extraction of the centerline of the tubular structure of interest.

1.2 Computer-Aided Diagnosis (CAD)

Computer-aided diagnosis (CAD) is a rapidly expanding area of medical image analysis that aims to help clinicians in making a proper diagnosis by, for example, detecting abnormalities, classifying and diagnosing these, and quantifying the spread of disease (van Ginneken, 2010). In general, CAD systems aid the radiologist in order to enable a faster or more accurate and reproducible diagnosis.

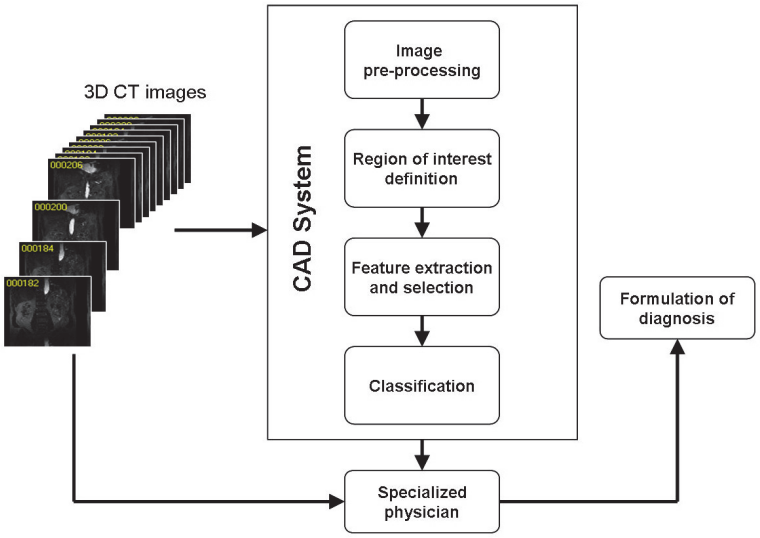


Figure 1.10: CAD system - Image-based CAD system architecture and its incorporation into medical diagnosis. Adapted from (Lu et al., 2009).

Typically, image-based CAD systems consist of four main modules (Lu et al., 2009) (Figure 1.10).

- *Image pre-processing* stage aims at improving image quality by denoising and artifact reduction. It can also handle data standardization that allows comparison of images obtained under different conditions.
- *Region of interest (ROI) definition* seeks to reduce the size of the analyzed data by defining local regions where a lesion can be found. ROI definition can be done

manually, semi- or fully automatically. Segmentation of the structures of interest also applies to this module. In that case, subsequent analysis is performed on the segmented structures instead of a defined ROI.

- *Feature extraction* deals with the obtainment of measurements from the pre-processed images. These measurements are subsequently used to identify the lesion. Since the amount of measurements can be very large, optionally, a subsequent step of *feature selection* can be performed. Here, the most robust features are kept in order to improve classification and speed up the process.

- *Classification* uses the extracted features to detect abnormalities in the datasets. Typically, classification first requires a training stage were training samples are provided so that a classification function is inferred. The obtained function is later used to classify new samples of data. Depending on the type of training data, classification can be either supervised or unsupervised. Supervised learning infers a classification function from labeled training data. Each sample of labeled data is made up of a pair consisting of an input (a feature vector) and a desired output (a label denoting normality or abnormality). Contrarily, unsupervised learning does not require labeled samples. This type of algorithms only requires the input feature vector to infer the classification function.

1.3 Our Approach

Owing to all the difficulties mentioned in section 1.1.4, CHD assessment by means of CT remains a difficult task. It is necessary to acquire sufficient experience to reliably diagnose CHD, and it has been demonstrated that acquiring moderate expertise in coronary CT angiography is slow and may take more than 1 year (Pugliese et al., 2009). In addition, up-to-date, the assessment of CHD is done by manually searching anomalies at a selected time-point of a dataset, which are subsequently confirmed/rejected using one or several additional time-point(s). Although existing platforms provide aid tools, the whole process exclusively relies on the reader's expertise since there are no universally accepted evaluation protocols. The associated difficulties and the large amount of data that needs to be processed makes CHD assessment a potential application for CAD systems.

Section 1.3.1 gives an overview of the working scheme that we propose with the aim to reduce the difficulty in the CHD diagnosis and assessment. Afterward, Section 1.3.2 presents the specific contributions of our work.

1.3.1 Overview

In this work we propose to formulate a methodological framework, inspired from CAD systems, that aims at reducing the time devoted to the evaluation of CT datasets for the diagnosis of CHD. Our final goal is to identify potential locations for vascular lesions so as to call the attention of the physicians. Additionally, we mimic the physician's

procedure, by registering the detected lesions at one time-point (typically systole) with a second time-point (typically at diastole) so as to provide different sources of information of the potential lesion to the physician. We consider that this proposal represents an advantage w.r.t. existent systems in two key aspects:

1. It eases the vascular lesion identification by marking potential lesions on a particular dataset so that the physician does not have to perform the search task but only confirm/reject.

2. It reduces the time devoted to navigation through different cardiac phases since we propose a scheme to register each potential lesion with a second image. Therefore, the physician has all the required information to provide a diagnosis without the necessity to search for it.

To achieve our goal, this thesis investigates the design of classification strategies in unsupervised and semi-supervised environments. We focus on these two types of environments because we want to avoid or minimize the dependence on labeled data. The obtainment of accurate and representative labeled data is an expensive task that we want to avoid.

Our research has also pointed to the development of strategies for feature construction and selection. We have evaluated how these features influence the accuracy of the classification results in unsupervised and semi-supervised environments.

1.3.2 Contributions of this Thesis

Problem formulation. We have formulated the vascular lesion identification as an anomaly detection problem (Section 2.3). The anomaly detection philosophy has been successfully applied to other domains and, to the best of our knowledge, we are the first to propose such an approach in the vascular imaging domain. Under this perspective, we have chosen to formulate our problem as a density level detection (DLD) problem and we have selected to solve it through the use of support vector machines (SVM). One of the motivations of our approach is to avoid or minimize the dependence on labeled data. Depending on the availability of labeled data for the training stage, we introduce two different algorithms. The first one, denoted DLD-SVM, is completely unsupervised, whereas the second one, denoted learning from positive and unlabeled samples (LPU), requires labels from only one of the classes.

Definition of metric. We have developed a metric, denoted `Concentric rings`, that is well adapted to the anomaly detection problem (Section 2.4.1). This metric has been designed so that it can model normality inside a 2-D cross section centered in the vessel axis. `Concentric rings` captures the intensity profiles and the axial symmetry of normal vessel, by computing intensity integrals in sectors subdividing each ring (Figure 1.11). Furthermore, we have proposed to combine various features in order to increase the discriminative power of the

detection. To this purpose, we have proposed the use of known measures that were initially designed as vesselness criteria for vascular enhancement / segmentation (Section 2.4.2). Our selection is motivated by the good response of these metrics in normal regions (lumen), while having responses that deviate from the normality in the presence of a lesion. `Concentric rings` metric showed the best individual performance in DLD-SVM, it was ranked third in LPU and it is always among the group of selected metrics that make up the best feature set.

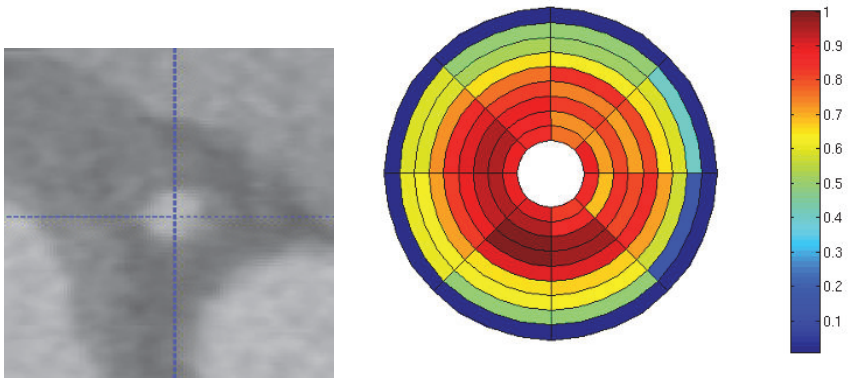
Feature selection scheme. In order to define a strategy to select the set of metrics that best fits our problem, we have introduced two novel feature selection schemes associated to DLD-SVM and LPU methods (Section 2.5.1). The proposed approach exploits the information provided by the two algorithms, more precisely the empirical risk, to be used as a selection criterion. This methodology can therefore be either unsupervised, if used with DLD-SVM or semi-supervised if LPU is employed. We compared the anomaly detection performance when using thus selected subsets of features, with the results based on conventional supervised selection schemes. The feature subsets obtained through our two methods gave out the best detection results.

Performance improvement. We have proposed a scheme to improve the performance of the LPU method by progressively increasing the associated training data sets (Section 2.3.4). This algorithm refinement seeks to reduce undesired effects of classifiers caused by finite samples. We propose to incrementally add samples to the labeled dataset, while also adding new unseen data to the unlabeled data set. The process can take advantage of progressive inclusion of new patients through time, as well as of new detection results validated by experts in routinely use of the system. A minor contribution that also enhances the overall performance resides in a selection of an algorithm, not so well-known, which can be applied as an additional step in the learning phase in any algorithm based on SVM (Section 2.3.5). The selected method seeks the improvement by an additional variation of one of the parameters of the learnt decision function of the SVM.

Inter-phase registration to validate detection. We propose a strategy to validate potential lesions that is inspired in a common diagnostic procedure followed by physicians in cardiac CT. The method consists in confronting information from different cardiac phases to give out a final diagnosis. We propose to register the lesions detected at one phase with a second one in order to compare the information of both volumes the way physicians do (Section 5.3).

1.4 Outline

This document is organized as follows:



(a) (b)

Figure 1.11: An example of the proposed Concentric rings metric - (a) Healthy coronary cross-section (b) Color-coded visual representation of the metric response. The colors encode the normalized values of intensity integrals within each sector of each ring.

Chapter 2 describes the proposed methodological framework and all the necessary steps required to construct it. First, we discuss the hypothesis over which our proposal is based and then we describe each of the elements that compose it. The framework is made up of three main elements that are largely described in the chapter. These are: the classifiers, the metrics and the feature selection strategy.

Chapter 3 is devoted to the experimental design of the metric we have proposed to use in our framework. Once the metric is properly tuned up, the remaining candidate metrics are also parametrized, so that they can be directly compared to our approach. Additionally, the evaluation protocol followed in the different experiments is described.

Chapter 4 is devoted to the selection of an optimal feature set based on the different sets of candidate metrics that were defined in **Chapter 2** and parametrized in **Chapter 3**. The performance of the metrics is first individually evaluated and this information is used to define in which order they will be evaluated by the different feature selection algorithms. The quality of the selected group of metrics is assessed through its use as inputs for the selected classifiers.

Chapter 5 introduces a methodology that eases the task of comparing different phases to formulate a diagnosis of CHD. The method constructs a ROI surrounding potential lesions and maps the ROI to a different time-point. The two ROI are then non-rigidly registered. The two registered ROIs are compared so as to confirm / reject potential lesions.

A final chapter is devoted to give overall conclusions and point out research perspectives.

2

Automatic Detection of Vascular Abnormalities: Framework Definition

Glossary

In the present chapter, several expressions will be used, which may have different meanings in various fields, namely in the medical and statistical domain. To avoid misunderstandings, we provide definitions that will be used along the chapter.

Abnormality. In a wide sense, something deviating from the normal or differing from the typical.

Anomaly. In a restrictive meaning (Barnett and Lewis, 1994), an observation (or subset of observations) that appears to be inconsistent with the remainder of the considered set of data. It is also referred to as **outlier**.

Lesion. In medicine, abnormal alteration of organ morphology and/or tissues due to a disease or trauma. In our case, we consider arterial lesions that include the alterations of vessel wall and lumen, typically due to atherosclerosis.

Anomaly detection. Refers to detecting patterns in a given data set that do not conform to an established normal behavior (Chandola et al., 2007).

Classification. As defined in the fields of machine learning and pattern recognition, classification is a procedure that assigns each piece of input data (*instance*) into one of a given number of categories (*classes*). **Binary classification** refers to a particular case of classification where only two classes are involved.

Learning. Refers to evolving the behavior of the computer, based on empirical data (examples), in order to improve decisions or recognition of complex patterns. Formally, an algorithm is said to learn from an experience E with respect to

some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E (Mitchell, 1997).

2.1 Introduction

It has been demonstrated that acquiring moderate expertise in coronary CT angiography is slow and may take more than 1 year (Pugliese et al., 2009). Discontinuities due to stenoses, contrast variations due to calcifications and nearby structures (*e.g.* heart chambers) and artifacts provoked by heart and respiratory movement during the acquisition, cause undesired effects in the image quality and difficult medical diagnosis. These effects lead to long evaluation times and may cause errors in the diagnosis of the coronary disease (Pugliese et al., 2009). Therefore, a tool that can automatically identify potential vascular lesions and call the attention of the physician is desirable.

Most of the algorithms focused on the identification of vascular lesions share the characteristic that they are directed towards abnormality modeling, *i.e.* identifying the particularities of lesions. However, lesions are heterogeneous by nature, so obtaining a model that copes with all possible abnormalities remains a challenge for this type of approach. In consequence, manual interaction is required to retrieve undetected lesions. Furthermore, since these methods rely on the identification of a particular lesion, (*i.e.* an aneurysm, an atherosclerotic plaque or a particular type of plaque), they are not capable of detecting other types of abnormalities.

Recently, a novel technique (Kang et al., 2009; Wong and Chung, 2006) has been proposed to overcome the weaknesses of the classical approaches. Here, abnormalities are identified by modeling their opposite: the normal vessels. Wong and Chung (2006) use a simple shape of normal vessels to identify different vascular abnormalities (*e.g.* stenotic atherosclerotic plaque, saccular and fusiform aneurysmal lumens) in an indirect fashion, instead of directly manipulating the complex-shaped abnormalities. Although the method is reported to be highly accurate, it does not work over gray-scale images but on topologically and morphologically correct vascular segmentations, which can imply a significant amount of user interaction.

Likewise, Kang et al. (2009) use an active tube model that can be deformed to be registered onto the vessel lumen. In their work, three types of regions are defined: normal, stenotic, and aneurysmal, to classify the vessel segment analyzed by the use of an algorithm based on a cross-sectional distance field.

We propose a method that aims at accurate identification of vascular lesions with minimal user interaction. Our approach relies on machine learning (ML) techniques. Although a few state-of-the-art methods use ML techniques, we differ in two key points.

1. We neither model nor try to directly identify particular abnormalities. Our proposal follows a similar strategy as (Kang et al., 2009; Wong and Chung, 2006), where normal vessels are modeled in order to identify lesions. We assume that lesions can be considered as local outliers compared to the typical appearance of a healthy artery. Therefore, we define image features that allow the description of normality and the indirect detection of the lesions.

2. The training stage can be performed without or with a small amount of labeled data. Actually, one solution requires no labeled data, while the second solution requires labels only for normality.

In the following sections, we first review state-of-the-art methods that tackle the detection of vascular lesions (Section 2.2). Section 2.3 is devoted to the description of the detection approaches we propose. This is followed by the definition of a set of candidate features to be used in the classification task (Section 2.4), where we describe a metric that we have developed, as well as other features that we have adopted from literature. Finally, section 2.5 presents the methodology that we followed to select features from the candidate set. Validation of the methods here described is left for Chapter 4.

2.2 Related Work

Although several algorithms exist for the identification of lesions in arteries such as the carotids, particularly using MR images (Adame et al., 2004; Lekadir and Yang, 2006; Salvado, 2006), and in a smaller measure using CT (Vukadinovic et al., 2010), not so many methods have been proposed to be applied in the coronary arteries. Using the target lesion to be detected, methods can be classified into three categories. These are: calcified plaque-oriented (section 2.2.1), soft plaque-oriented (section 2.2.2) and hybrid methods (section 2.2.3).

2.2.1 Calcified Plaque-Oriented Methods

The calcified plaque-oriented category represents a vast majority of the methods dealing with the identification of lesions in coronary arteries. Toumoulin et al. (2003) use a level-set approach devised to extract the outer and inner boundary of the vessel wall. Calcified plaques are identified within thus delimited region. Komatsu et al. (2005) define a *Plaque Map* based on a color-based isometric line method, which colors the relevant limit of CT numbers, and the Bird's Eye View, which shows 3D-contour images. The map serves as visual aid for manual identification of calcified plaque components. Wesarg et al. (2006) segment vessels while excluding calcifications. Since regions containing calcifications have a lower mean diameter, they combine diameter information with intensity analysis to detect calcified plaques. Dehmeshki et al. (2007) apply a threshold and identify connected regions that exceed the threshold. Manual selection of one (or more) of the connected regions is performed for subsequent quantification through a Modified Expectation-Maximization algorithm. Išgum et al. (2007) propose a learning-based scheme where candidates for coronary calcifications are extracted by thresholding and component labeling. Then, 64 image features are extracted and used as input for a two-stage classification. Their method has been tested exclusively in non-contrast enhanced CT. Using electron beam CT (EBCT), Brunner et al. (2008) investigate unsupervised classification to distinguish arterial calcified plaques from bones, metal stents or any other metallic objects that, due to their density (image intensity value) can be mistaken for calcifications. The method first mines the characteristics

of calcified lesions using an optimization criterion and then identifies a subset of lesion features optimal for classification. A two stage clustering is deployed to discriminate between lesions and other candidate regions. Also using EBCT, Kurkure et al. (2008) propose a supervised hierarchical classification-based approach to distinguish the coronary calcifications from other candidate regions. At each level of the hierarchy the method learns an ensemble of classifiers where each classifier is a cost-sensitive learner trained on a distinct asymmetrically sampled data subset. After aorta detection and coronary tree segmentation in a CT angio (*i.e.* with contrast agent) data set, Saur et al. (2008) extract putative calcification candidates using an intensity-based threshold, which results in many false positive (FP) plaques. In a following step the plaques with the highest rank are rigidly registered to plaques in native CT scans (*i.e.* without contrast agent). Finally, a rule based approach is used to maximize the number of detected plaques while minimizing FP.

2.2.2 Soft Plaque-Oriented Methods

Contrary to calcified plaque detection, to date the methods devoted to soft plaque detection in coronary CT are quite limited. Renard and Yang (2008) proposed a method consisting of three steps: extraction of the arterial lumen centerline, segmentation of the lumen and arterial wall separately with locally adaptive region growing, and detection of soft plaques based on effective cross-section areas of the lumen and of the wall. The soft plaque detection criterion is based only on area difference which makes the method prone to errors. Lankton et al. (2009) propose an algorithm based on a localized active contour framework that employs a scale parameter to restrict the statistical characteristics of the vessel into local regions. Simple probabilistic models based on local means are used to extract the vessel and find areas where soft plaques exist. Makrogiannis et al. (2009) introduce nonparametric intensity priors of foreground and background objects into a local, nonparametric dissimilarity measure as a speed function term in the level set framework. They use a database of vessels with positive wall remodeling to demonstrate the capability of the method to extract hypodense regions associated with soft plaques.

2.2.3 Hybrid Methods

Similarly to soft plaque-oriented methods, only few hybrid methods have been reported for CT images. To the best of our knowledge, the first method aiming the detection of different types of plaques is the one proposed by Rinck et al. (2006). Their approach uses two surface representations, one for the contrast-filled vessel lumen and one for the vascular wall. The deviation between these two surfaces is defined as plaque volume. A histogram of the intensity distribution is calculated and presented to the user. The user adjusts thresholds, which divide the histogram into different ranges that are assigned plaque components. Tessmann et al. (2008, 2009) use a learning-based approach where they train a classifier using the AdaBoost algorithm (Freund and Schapire, 1995) to detect both calcified and soft plaque lesions. As input for the training stage, they

use features based on a sampling pattern that approximates the vessel shape. More recently, Arnoldi et al. (2010) evaluated a CAD algorithm that first extracts calcified lesions. Non-calcified plaque is detected as a hypoattenuating area between the external vessel boundary and lumen without calcium. The latter is done under a learning-based approach. However, the description in the article does not provide sufficient details of the method.

2.2.4 Overview

There are still few methods that try to tackle the lesion identification as a whole (*i.e.* hybrid methods). Most existing methods focus in a particular type of lesion. Among the methods that try to identify all types of lesions, two main drawbacks can be identified. First, methods such as the one proposed by Saur (2009) require either a double CT image acquisition,*i.e.* with- and without contrast enhancement, which is not desirable for a patient owing to increased radiation dose, or Dual CT images that are not available in clinical routine in most medical centers. Second, methods based on learning approaches strongly rely on the availability of labeled data that represent all type of possible pathological behaviors. In practice, this is not often possible. In the following, we propose a framework that tries to overcome all the above mentioned challenges. We propose a methodology that aims at identifying all types of vascular lesions requiring only a contrast enhanced acquisition. Moreover, we propose a learning-based approach that minimizes the necessity of labels and thus reduces the dependency on labels and human-induced errors.

2.3 Definition of a Classification Strategy

Classification methods based on ML techniques have been extensively used in breast and lung cancer diagnosis. However, as previously shown, only a few approaches tackle vascular lesion identification from a ML point of view. Moreover, only one of the presented methods (Tessmann et al., 2008, 2009) aims at detecting several types of lesions. Nevertheless, we have chosen to follow a learning-based approach for several reasons. First, this type of approach has been proven robust in other medical domains. As a consequence, there is an increasing interest in the medical image processing community to investigate this category of methods [1]. Second, and most important, objects such as lesions in medical images are highly heterogeneous, so it is difficult to model them accurately through equations. Therefore, the possibility of learning from examples arises as a promising alternative. With respect to the previous classification approaches that tackle vascular lesions, our proposal differs in two main aspects: very limited use of labels and absence of any attempt to explicitly model the lesions.

A label associated to a data instance denotes if the latter is normal or abnormal. In applications involving diseased organs, it is often very expensive to collect labeled

[1] As an example, the first international Machine Learning in Medical Imaging workshop http://miccai-mlmi.uchicago.edu/

data that are accurate, as well as representative of all type of behaviors. Frequently, only unlabeled samples are available. Hodge and Austin (2004) and Chandola et al. (2007) define three different categories for learning-based algorithms [1] based on the availability of labels. Hodge and Austin (2004) denote these categories as Type 1, 2 and 3 approaches, while Chandola et al. (2007) use more explicit names: Unsupervised (type 1), supervised (type 2) and semi-supervised (type 3). We keep the more explicit names as proposed by Chandola et al. (2007). The definition of each category is as follows:

Unsupervised learning. Detects outliers with no prior knowledge of the data. The approach processes the data as a statistic distribution, pinpoints the most remote points, and flags them as potential outliers.

Supervised learning. Models both normality and abnormality. It requires pre-labeled learning data, where every sample is marked as either normal or abnormal.

Semi-supervised learning. Models only one class. Typically, the modeled class is normality. Few techniques exist that model the anomalous instances from training data, since it is difficult to obtain a training data set that covers every possible abnormal behavior that can occur in the data.

With the exception of Brunner et al. (2008), all of the ML-based approaches aiming at the detection of vascular lesions follow a supervised classification scheme. The main advantage of supervised learning is its high accuracy in detecting many kinds of *known* lesions and the accuracy of the normal learnt behavior. However, there are several drawbacks. First, such methods rely on the availability of accurate and representative labels, which is often not possible and, furthermore, it is expensive and time consuming to obtain them. Second, supervised methods are incapable of detecting *new* emerging abnormalities. Since lesions are by nature heterogeneous, unknown abnormal patterns can often appear.

To overcome the intrinsic problems that arise from the use of supervised learning in the identification of abnormalities, in a general meaning, we propose the use of methods that minimize the dependence on labeled datasets. To do so, we formulate our problem as an anomaly detection problem, in the restrictive meaning of this expression (Chandola et al., 2007; Hodge and Austin, 2004; Markou and Singh, 2003a,b; Patcha and Park, 2007). Following the definition from Barnett and Lewis (1994), we consider that an anomaly can be defined as an observation (or subset of observations) that appears to be inconsistent with the remainder of the considered set of data. With this definition, we consider that lesions can be regarded as anomalies that appear inconsistent w.r.t. healthy vessel sections. While state-of-the-art methods try to find two different classes: normality and abnormality, we focus in only one class, the normality, and assume everything diverging from it is considered abnormal. Therefore, our approach refers to

[1] Their surveys are focused on anomaly detection problems. However, the definitions apply for learning-based algorithms, in general.

the problem of finding patterns in data that do not conform to expected behavior. This formulation permits the use of semi-supervised and unsupervised learning approaches were only labels for normality or no labels at all are required. In literature it is possible to find a vast amount of work on anomaly detection. As an example, let us mention the one-class SVM algorithm that is closely related to our proposal. One-class SVM was formulated by Schölkopf et al. (2001) with the aim of detecting data outliers. We refer the interested reader to more extensive reviews on the subject (Chandola et al., 2007; Hodge and Austin, 2004; Markou and Singh, 2003a,b; Patcha and Park, 2007) and to Appendix A.2 for more details on the one-class SVM algorithm. However, to the best of our knowledge, there is no previous work on the identification of vascular lesions following a semi-supervised/unsupervised anomaly detection approach.

In the following, we first formally describe our problem (sec. 2.3.1) followed by the description of the methods that allow us to solve it through an anomaly detection scheme. The first one (Section 2.3.2) is unsupervised, while the second one (Section 2.3.3) is semi-supervised. The third one (Section 2.3.4) is an extension of the second solution, designed to improve the detection results in time by taking advantage of additional data and of additional labels coming from previous correct classifications. We also describe a post-processing method capable of further improvement of the anomaly detection by any of these methods.

2.3.1 Problem Formulation

Let us denote $Q = (x_1, ..., x_n) \in \mathcal{X}^n$ a collection of samples independent and identically distributed (i.i.d) with respect to a probability distribution P_Q in a space $\mathcal{X}$. These samples represent the *a priori* information we have for our anomaly detection problem. In practice, the sample set can be either labeled or unlabeled. If Q contains no labels at all, our goal will be to decide for each x_i whether or not it belongs to the normal class. Otherwise, if Q is labeled, we assume there is a second subset of unlabeled samples $X = (x'_1, ..., x'_{n'}) \in \mathcal{X}^{n'}$ i.i.d. with respect to a distribution P_X, which we want to classify.

Two density-based functions or similarity measures (Porter et al., 2009) are defined over Q and X. The first measure is the *content density* h, which quantifies the absolute concentration of P_Q and is given by the Radon-Nikodym derivative:

$$h := \frac{dP_Q}{d\mu}, \tag{2.1}$$

where μ is a (known) reference measure. Usually, the reference μ is chosen to be the Lebesgue measure (Hush et al., 2005). The second similarity measure is the *relative content density* ϱ that quantifies the concentration of P_Q w.r.t. P_X and defined as:

$$\varrho := \frac{dP_Q}{dP_X}, \tag{2.2}$$

where it is assumed that ϱ is well defined.

Given a threshold $\rho > 0$, the set $\{x \in \mathcal{X} : h > \rho\}$ is denoted as the ρ-level set of the density h. Let us recall that anomalies are assumed to be points with a low probability density value. Therefore, it can be stated that the set $\{h \leq \rho\}$ of points below a threshold level $\rho > 0$ comprises the anomalous set, while its complement $\{h > \rho\}$ is called the normal set (Hush et al., 2005).

To interpret the density ϱ, let us assume that some samples in X are generated by P_Q while the remaining ones are generated by another distribution P_O. Thus, P_X is a mixture distribution:

$$P_X = \beta P_Q + (1 - \beta) P_O, \qquad \text{where } 0 < \beta < 1. \tag{2.3}$$

With an appropriate choice of t, the set $\{x \in X : \varrho(x) > t\}$ consists of samples that are more likely to be generated by P_Q than by P_O. In this case, the samples generated by P_Q are considered as normality, while the rest are considered as anomalous.

Both similarity measure formulations seek to identify the elements for which the density exceeds a threshold (ρ or t). To accomplish this, it is necessary to solve a density level detection (DLD) problem. The DLD problem can be used to detect anomalies. As already mentioned in the introduction, the anomaly detection is performed indirectly: the normal set is estimated by identifying the elements of $\mathcal{X}$ that exceed the threshold, then the anomalous set is identified with the complement of the normal set.

In the subsequent, we will formulate some of the basics for the DLD problem solution under both conditions. Depending on the type of available data, one of the two measurements, content density or relative content density, should be used. The next sections formulate the mathematical background of each approach, describe the data conditions that lead to use either one of the approaches and present how the DLD problem can be solved as a classification problem. Afterwards, we introduce two methods to improve some classification associated problems when no sufficient data is available. The first approach is a novel proposal that is inspired in the flow of data in the clinical enviroment. The second approach is an enhancement algorithm, not so widely used, but for which it has been demonstrated that it can significantly improve performance.

2.3.2 Solution 1: Density Level Detection - Support Vector Machine (DLD-SVM)

The original DLD - Support Vector Machine (DLD-SVM) problem was formulated by Steinwart et al. (2005a,b). In their seminal work they state that to solve the DLD problem it is necessary to find an estimate of the set $\{h > \rho\}$. For this purpose, a real valued function f is constructed, which approximates the set $\{h > \rho\}$ with the set $\{f > 0\}$. In general, $\{f > 0\}$ does not exactly coincide with $\{h > \rho\}$. Therefore, a *performance measure* is needed to describe how well $\{f > 0\}$ approximates the set $\{h > \rho\}$. For measurable functions $f : \mathcal{X} \to \mathbb{R}$, the best performance measure is:

$$\mathcal{S}(f) := \mu(\{f > 0\} \triangle \{h > \rho\}), \tag{2.4}$$

where $\triangle$ denotes the symmetric difference (Figure 2.1[1]).

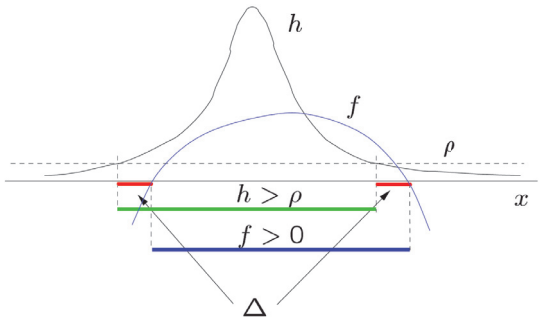


Figure 2.1: Performance of function f evaluated through the symmetric difference - The symmetric difference, defined as $A \triangle B = A \cup B \setminus A \cap B$, represents the number of cases where the function f falsely predicts a sample as belonging to the high density region, plus the number of cases where the function f falsely predicts a sample as not belonging to the high density region.

A DLD algorithm should seek $\hat{f}$ such that $\mathcal{S}(\hat{f})$ is close to zero. Unfortunately, there is no known and computationally feasible method to estimate it from a training set with guaranteed accuracy. However, the problem can be reformulated by use of a risk function.

Let now μ be a known probability measure on $\mathcal{X}$, the risk defined as:

$$\mathcal{R}(f) := \frac{1}{1+\rho} P_Q(f \leq 0) + \frac{\rho}{1+\rho} \mu(f > 0), \tag{2.5}$$

Steinwart et al. (2005a) demonstrated that a function f minimizing $\mathcal{R}$ also minimizes $\mathcal{S}$ and that there is a tight relationship between them. Thus $\mathcal{R}$ is a performance measure for the DLD problem, and choosing $\hat{f}$ that minimizes $\mathcal{S}$ can be done by choosing $\hat{f}$ that minimizes $\mathcal{R}$. Moreover, unlike $\mathcal{S}$, the risk $\mathcal{R}$ can be empirically estimated from sample data.

The definition of $\mathcal{R}$ as a performance measure also permits the interpretation of the DLD problem as a supervised binary classification problem. Given the label set $Y := \{-1, 1\}$, let $x \in \mathcal{X}$ and $y \in Y$ respectively denote the values of random variables $\mathbf{x}$ and $\mathbf{y}$. The binary classification problem is formed by identifying P_Q and μ with the conditional distributions $P_{x|y=1}$ and $P_{x|y=-1}$ respectively and defining the class marginals $P(y = 1) := 1/(1+\rho)$ and $P(y = -1) := \rho/(1+\rho)$. It should be remarked that DLD-SVM, in principle, is unsupervised. The only available data comes from the unlabeled set Q. Therefore, to solve the supervised binary classification problem surrogate labeled sets need to be created. The training set $\mathcal{T}$ is built as follows. First, a set T is formed such that $T = Q$, and all its elements are assigned a label $y = 1$. For the sake of coherence, from now on we denote $|Q| = |T| = n_1$. Then, a set T' of n_{-1}

[1]Illustration resulting from discussions with Dr. Don Hush, researcher in Machine Learning and pattern recognition at Los Alamos National Library (Los Alamos, USA)

synthetic samples is formed. These are artificially generated i.i.d. with respect to μ, and a label $y = -1$ is assigned to each of them. The union of the two sets $\mathcal{T} = \{T, T'\}$ of size $n = n_1 + n_{-1}$ is used as a training set to choose a function $\hat{f}$, which minimizes $\mathcal{R}(\hat{f})$.

Once a training data is defined for the binary classification problem it is possible to construct a Support Vector Machine (SVM) (Cortes and Vapnik, 1995) that solves it. Let $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a positive definite kernel with reproducing kernel Hilbert space (RKHS) $\mathcal{H}$. Let $l : Y \times \mathbb{R} \to \mathbb{R}^*_+$ be the hinge loss function, *i.e.* $l(y,t) := \max\{0, 1 - yt\}$, $y \in Y$, $t \in \mathbb{R}$. According to Steinwart et al. (2005a,b), for a DLD problem with a training set $\mathcal{T}$, a regularization parameter $\lambda > 0$, and $\rho > 0$, the SVM chooses a decision function f that minimizes in $\mathcal{H} \times \mathbb{R}$ the following criterion:

$$\lambda \parallel f \parallel^2_{\mathcal{H}} + \frac{1}{(1+\rho)} \frac{1}{|T|} \sum_{x \in T} l(1, f(x)) + \frac{\rho}{1+\rho} \frac{1}{|T'|} \sum_{x \in T'} (l(-1, f(x))) \,. \tag{2.6}$$

The kernel k is built using Gaussian radial basis functions $e^{-\sigma^2 |x - x_i|^2}$.

Under this formulation, given a set of ρ values, the DLD-SVM approach will select the $\hat{f}$ that minimizes $\mathcal{R}$. The risk is empirically obtained from the sample data through

$$\hat{\mathcal{R}}(f) = \frac{1}{1+\rho} \frac{1}{|T|} \sum_{x \in T} I(f(x) \leq 0) + \frac{\rho}{1+\rho} \frac{1}{|T'|} \sum_{x \in T'} I(f(x) > 0), \tag{2.7}$$

where $I(\cdot)$ returns 1 if the argument statement is true and 0 otherwise. The decision function f minimizing the empirical risk $\hat{\mathcal{R}}(f)$ will allow to identify the ρ value that best reflects the concentration of normality.

2.3.3 Solution 2: Learning from Only Positive and Unlabeled Data (LPU) approach

The second solution deals with situations where data from Q is labeled, while data coming from X is not labeled. In this case, our goal is to detect in X, elements similar to those from Q, denoted targets.

A binary classification problem assumes a data generating distribution as:

$$P_X = p_1 P_{X|y=1} + p_{-1} P_{X|y=-1}, \tag{2.8}$$

where $P_{X|y=1}, P_{X|y=-1}$ denote conditional probabilities in X given $y = 1$, $y = -1$ respectively, and p_1, p_{-1} denote marginal probabilities. We seek a function f that minimizes the binary classification error:

$$e(f) = p_1 P_{X|y=1}(f < 0) + p_{-1} P_{X|y=-1}(f \geq 0). \tag{2.9}$$

Steinwart et al. (2005a) demonstrated that if Eq. 2.3 holds, then with $t = \frac{1}{2\beta}$ any f that minimizes:

$$\mathcal{S}(f) := P_X(\{f > 0\} \triangle \{p > t\}), \tag{2.10}$$

also minimizes the binary classification error $e(f)$, with:

$$p_1 = \beta, \qquad p_{-1} = 1 - \beta, \qquad P_{X|y=1} = P_Q, \qquad P_{X|y=-1} = P_X. \tag{2.11}$$

In order to ensure that Eq. 2.11 is valid, some associations to construct a surrogate training data from the input data are necessary. This is:

- Using unlabeled data coming from X, a new surrogate labeled data set $\bar{X}_{-1}$ is created by assigning the label $y = -1$ to all the samples. As a consequence $\bar{P}_{x|y=-1} = P_X$.
- The labeled data coming from Q builds up a surrogate labeled data $\bar{Q}$ by keeping its original label $y = 1$. Therefore, $\bar{P}_{x|y=1} = P_Q$.
- Surrogate class marginal probabilities are set as follows: $\bar{p}_1 = \beta$ and $\bar{p}_{-1} = 1 - \beta$.

The latter associations convert the semi-supervised problem into a surrogate supervised problem. Thus, this formulation is directly applicable to classification problems, the input data of which consists of labeled samples from one class (Q) and unlabeled samples from a mixture (X). This type of problem is denoted as learning from positive labels and unlabeled data (LPU). It can also be referred to as one-sided problem.

Let us recall that Steinwart et al. (2005a) demonstrated that it is possible to validate the performance of $\hat{f}$ by means of the risk function, without estimating $\mathcal{S}(\hat{f})$. Based on the results obtained in the previous section, we can obtain for the relative content density case: the risk function, the empirical risk function and the formulation of the criterion to be minimized by SVM. For the sake of completeness, we formulate the respective expressions in terms of β:

$$\mathcal{R}(f) := \frac{2\beta}{1+2\beta} P_{\bar{Q}}(f \leq 0) + \frac{1}{1+2\beta} P_{\bar{X}}(f > 0), \tag{2.12}$$

$$\hat{\mathcal{R}}(f) = \frac{2\beta}{1+2\beta} \frac{1}{|\bar{Q}|} \sum_{x \in \bar{Q}} I(f(x) \leq 0) + \frac{1}{1+2\beta} \frac{1}{|\bar{X}|} \sum_{x \in \bar{X}} I(f(x) > 0), \tag{2.13}$$

$$\lambda \parallel f \parallel^2_{\mathcal{H}} + \frac{2\beta}{(1+2\beta)} \frac{1}{|\bar{Q}|} \sum_{x \in \bar{Q}} l(1, f(x)) + \frac{1}{1+2\beta} \frac{1}{|\bar{X}|} \sum_{x \in \bar{X}} \left(l(-1, f(x)) \right). \tag{2.14}$$

Finally, let us remark an important condition that should hold and that is not evident from the LPU formulation. The formulated equations are valid in the infinite sample limit ($|X| = \infty$). When the latter does not hold, finite sample effects represented in classifications errors are introduced. An infinite sample set cannot be obtained. However, it is desired that $|X| \gg |Q|$.

2.3.4 Solution 3: LPU with Progressive Increase of the Training Set

Class imbalance is a common problem in anomaly detection. The problem occurs when most of the data comes from one class. In LPU, this implies having a true t value close to 0 or 1. With imbalanced classes, most learning methods tend to favor a response that assigns all patterns to one class. The latter tends to worsen under limited data conditions.

In DLD-SVM, limited data is not a serious issue since artificial data can be generated. This is not the case for LPU. Let us again remark that the formulations presented in Section 2.3.3 hold in the infinite sample limit. Since this cannot be accomplished in practice, at least it is desired to have $|X| \gg |Q|$.

Under these conditions, we propose an approach to improve the performance of LPU by progressively increasing the two sets that make up the training data set. With the addition of samples to Q our goal is to ameliorate the description of normality, whereas samples are added to X in order to keep the relation $|X| \gg |Q|$ and reduce the finite sample effect. With this, we consider that the trained model M should not be static but should be re-trained every time new data arrives. Our proposal, that is inspired in clinical data flow, is depicted in Figure 2.2. The algorithm workflow is as follows:

1. An initial pair of sets Q_i and X_i is used to train a model M_i (Figure 2.2 (a)).

2. The obtained model M_i is used to label the data from X_i (Figure 2.2 (b)). The clinician validates the samples marked as lesions.

3. A random number of normal samples (samples that the clinician did not consider as lesions) are selected from the labeled set X_i and combined with the set Q_i to build up the enhanced set Q_{i+1} (Figure 2.2 (c)).

4. At the arrival of newly acquired data, this is combined with the set X_i to make up the enhanced set X_{i+1} (Figure 2.2 (d)). It should be remarked that for the purposes of training, it is assumed that no labels are available for X_i, although it has already been labeled.

5. A new refined model M_{i+1} is obtained by using the enhanced sets Q_{i+1} and X_{i+1} for training (Figure 2.2 (e)).

As we already mentioned, the method is quite suitable for a clinical enviroment while reducing the finite sample problem and enhancing the Q set.

2.3.5 Bias Variation as an Alternative to Improve Performance

The tuning of algorithms which involve strong class imbalance often requires a very careful search of the associated parameters (such as λ or σ). Apart from a careful selection of the SVM parameters, it is possible to apply algorithm enhancements that

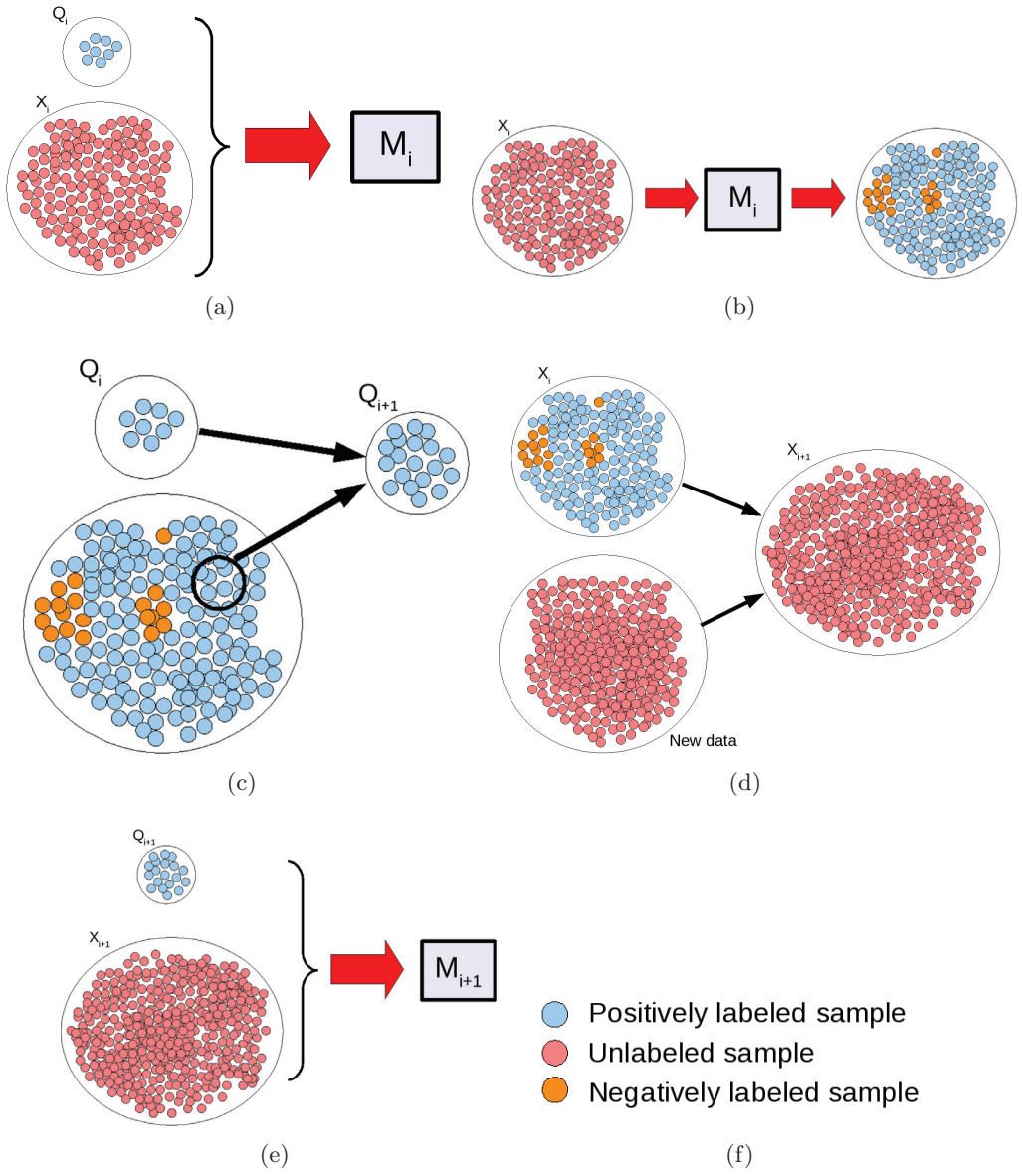


Figure 2.2: The enhanced LPU procedure - (a) A model M_i is obtained from training with the training set $\{Q_i, X_i\}$ (b) The model M_i assigns labels to the set X_i. (c) A random sample is taken from positively labeled samples of X_i and is combined with Q_i to make up the enhanced set Q_{i+1}. (d) Newly acquired data is added to X_i to form the set X_{i+1}. (e) A model M_{i+1} is obtained from training with the training set $\{Q_{i+1}, X_{i+1}\}$. (f) Color map.

help improve the final classification. Enhancement algorithms are not very commonly used in the machine learning domain and are rare to find in literature.

A simple enhancement method consists in following the normal SVM training with a subsequent step that adjusts the bias term in the decision function (parameter b in Eq. A.4) to minimize the risk (or error). This strategy was first used by Fukunaga (1972) to improve the performance of nearest neighbor classifiers. In his work, he proves that this process of adjusting the offset can drastically reduce the error bias of the designed classifier. More recently, Hush et al. (2007) proposed a theorem (Theorem 2.8, in the article) in which it is demonstrated that the procedure is theoretically correct and its performance is *no worse* than using the b value for the standard SVM. To the best of our knowledge, no other works found in literature make use of this enhancement technique to improve classification.

2.4 Definition of Candidate Features for Feature Set Construction

Two key issues need to be addressed to tackle the vascular lesions identification problem within the anomaly detection framework. First, the problem needs to be formulated so as to fit into the DLD problem or the one-sided problem framework. Second, data features differentiating between normal behavior and outliers need to be defined.

If a vessel is modeled as a tube and evaluated as a whole, it is unlikely to fit the problem into the anomaly detection framework. Indeed, diseased vessels cannot be considered as outliers within the set of available radiological images, since these images are mainly acquired in symptomatic patients. However, one characteristic of vascular lesions is that they usually are concentrated in a short segment of a vessel, compared to the vessel entire length. If the problem is tackled on a slice-by-slice basis, due to the concentration of the disease, only a small percentage of the data is abnormal, and it is possible to detect the outlier behavior of the lesion under the DLD philosophy. If at least one slice is detected as anomalous, it is possible to say that the corresponding vessel contains an abnormality (and label it as a diseased vessel). As a consequence, a 2-D approach is selected with cross-sectional planes locally orthogonal to the vessel centerline.

In the following sections, we will describe a number of features that can be used to describe vascular cross-sectional patterns. First, we describe a metric we have developed, which exploits the axial symmetry of normal vessels in order to extract features (section 2.4.1). Then, we present a set of metrics that we have selected as source of candidate features for our anomaly detection problem (section 2.4.2). Some of them are used in literature within the centerline extraction process, in order to seek the best local orientation and location of the centerline. Hence their original formulations include a tentative orientation $\mathbf{d}$ of the cross-sectional plane and a tentative location $\mathbf{p}$ of the centerline point. At the anomaly detection stage, we assume that the vessel centerline is already available. Therefore, we will omit the orientation, and $\mathbf{p}$ (if not omitted) will denote an actual centerline point.

2.4.1 Concentric Rings

For feature extraction, we need to adequately describe a pattern that models normality inside a 2-D cross section. In literature, normal cross-sections are usually expected to be circular or, at least, nearly symmetric *e.g.*: (Krissian et al., 2000; Wörz and Rohr, 2007) assume circular cross-sections with a bar-like intensity profile in big and medium vessels, and a Gaussian-like profile for small ones. Figures 2.3 and 2.4 show several examples of intensity profiles in big, small and diseased vessels.

We propose a metric were the normality is not explicitly modeled by a specific cross-sectional shape or intensity profile. Nevertheless, the metric $\mathcal{M}$ evaluating the local intensity patterns, is designed in a way that it is able to capture the intensity profiles and the axial symmetry of normal vessels. With such a metric, we expect a normal pattern to occur in the majority of cross-sections making it suitable for both DLD-SVM and LPU approaches.

In order to formulate our metric, we first need to define a feature $F(r, \theta)$:

$$F(r,\theta) = \int_{\theta_s}^{\theta_e} \int_{r_s}^{r_e} I(\rho,\vartheta) d\rho d\vartheta, \tag{2.15}$$

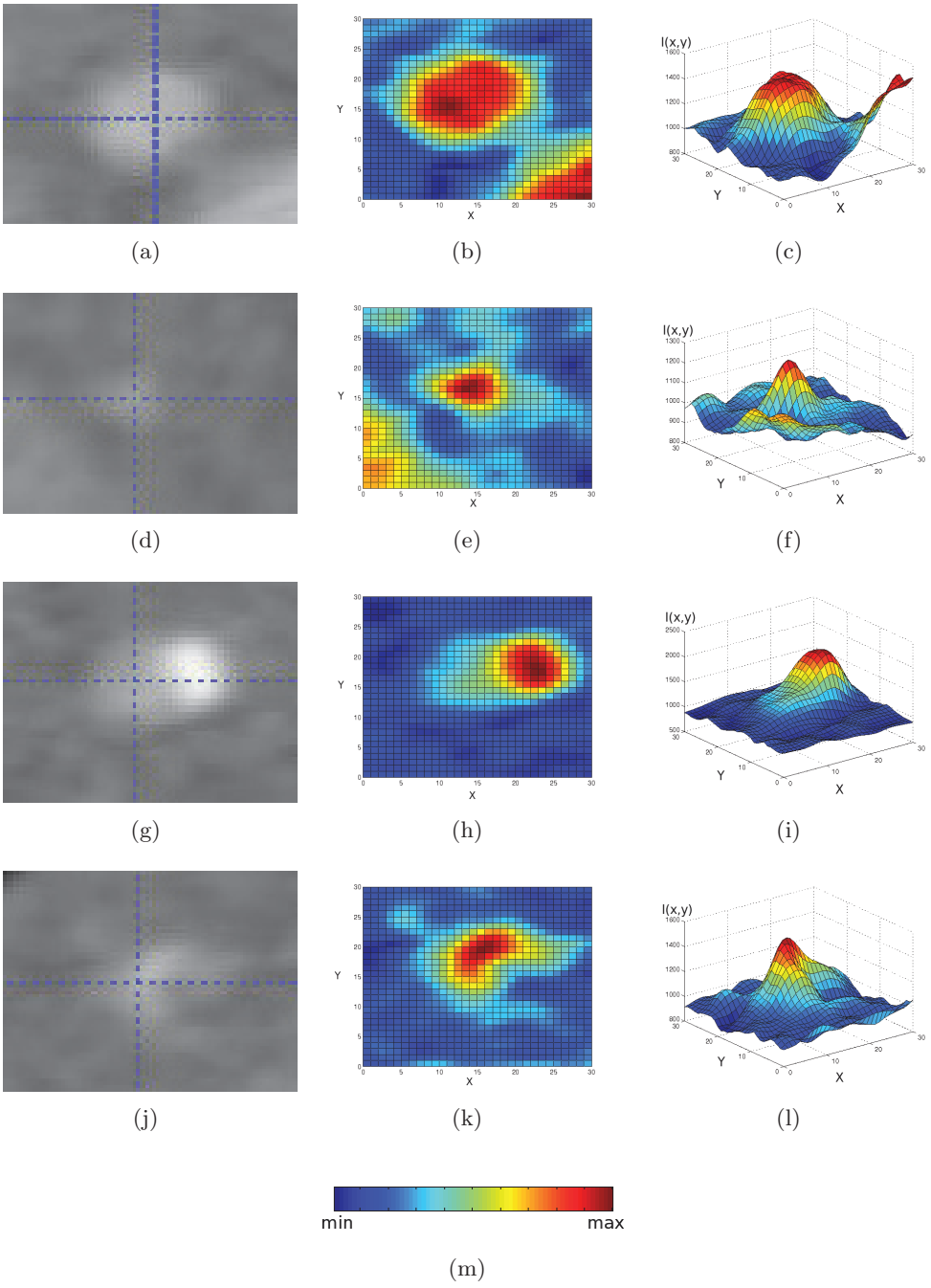
where $[r_s, r_e]$ represents a radial range around the vessel lumen center, $[\theta_s, \theta_e]$ is an angular range, and $I(\rho, \vartheta)$ is the image intensity at the location defined by polar coordinates (ρ, ϑ). Our metric $\mathcal{M}$ is then defined by a set of thus calculated features, using predefined ranges of r and θ. The respective numbers of ranges are denoted by N_r and N_θ, and their sizes are fixed as follows: $\Delta_r = (r_{max} - r_{min})/N_r$ and $\Delta_\theta = 2\pi/N_\theta$. Visually, $\mathcal{M}$ can be seen as a set of concentric rings centered at the vessel centerline and subdivided into sectors (Figure 2.5). However, for the sake of simplicity, we will refer to it as `Concentric Rings`.

The typical behavior of the metric in a normal vessel is expected to have the following characteristics:

1. for a fixed θ value, $F(r, \theta)$ decreases as r increases,

2. for a fixed value of r, $F(r, \theta)$ is nearly constant, while θ varies in the range $[0, 2\pi)$ (rotational invariance).

Our goal, when introducing an integral term in the feature definition (Eq. 2.15), is to reduce sensitivity to noise, compared to a simple sampling scheme at particular positions (following a circular pattern). The latter scheme might include noisy voxels that can cause an erroneous classification. By integrating along a polar neighborhood we overcome this problem. Nevertheless, if the pair $(\Delta_r, \Delta_\theta)$ defines a large region, the integral term can cause a strong homogenization, producing information loss. Therefore, there is an inherent trade-off, when selecting the region size, between noise suppression and the loss of information. Figure 2.5 illustrates the metric response in normal and abnormal cross-sections of a real vessel using $N_r = 9$ and $N_\theta = 8$.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

(j) (k) (l)

(m)

Figure 2.3: Intensity plots of orthogonal 2-D slices in a coronary artery - First column presents normal cross-sections of the proximal left coronary artery (LCA) and of the distal LCA, a cross-section with a calcification and a cross-section with a stenosis. Second column shows 2D views of pseudo-color intensity plots, and the third column shows a 3D view of the same intensity plots. Color map (m) encodes the intensities of the 2D cross-section normalized between its minimum and maximum values.

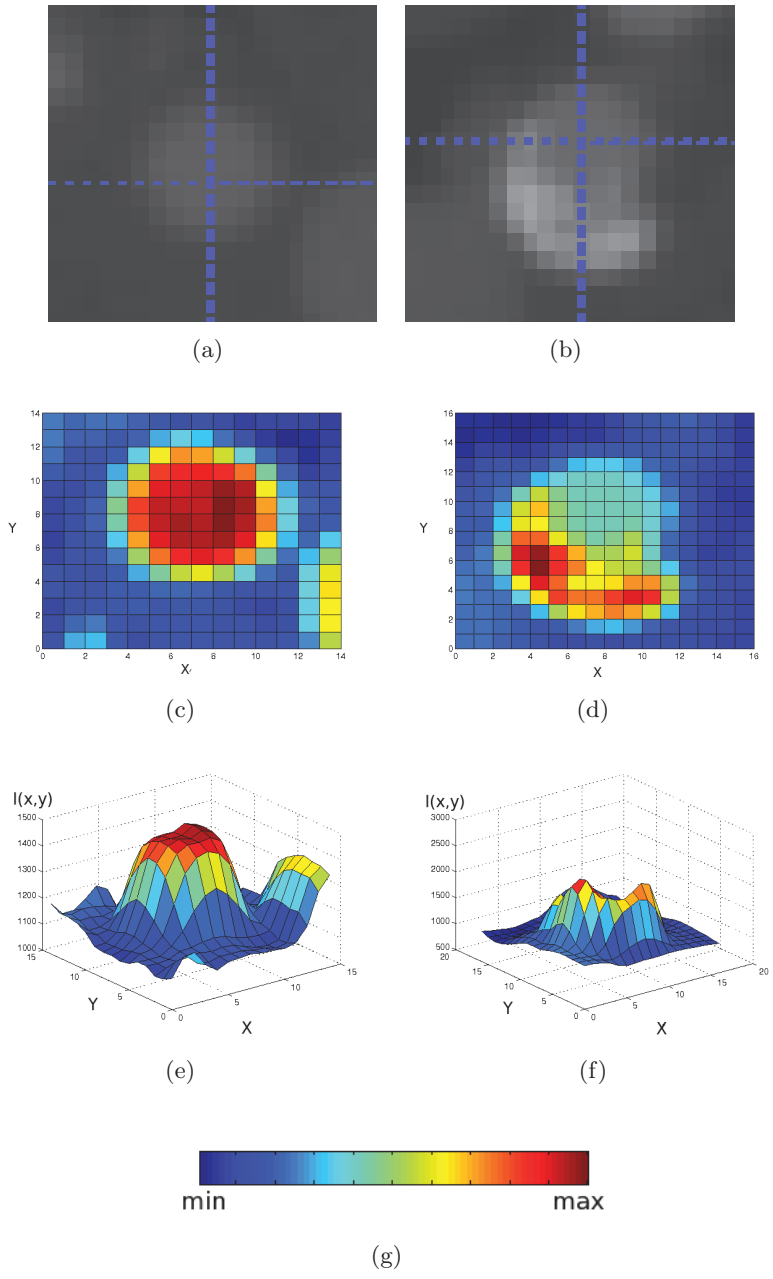
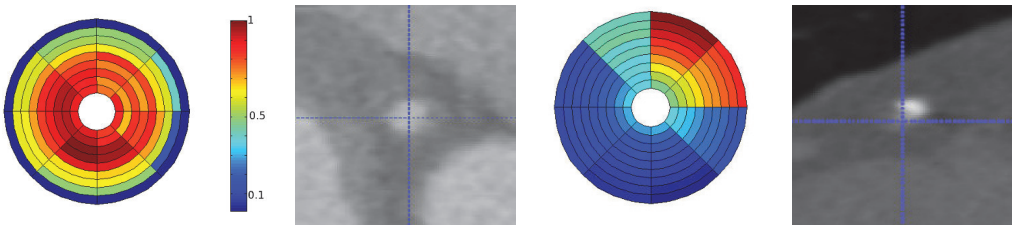


Figure 2.4: Intensity plots of orthogonal 2-D slices in a carotid artery - (a) normal carotid artery, (c) 2D representation of the intensity profile and (e) 3D representation of the intensity profile. In (b), calcified carotid artery and the 2D (d) and 3D (f) intensity plots. Colors encode the intensities of the 2D cross-section normalized between its minimum and maximum values. (g) Color map.



(a) (b) (c) (d)

Figure 2.5: Concentric rings metric representation using the Bull's Eye plot - In a normal case (a), high values are obtained in the inner rings, while low values are obtained in the outer rings. In an anomalous case (c), the pattern varies. Corresponding healthy (b) and a calcified slice (d) are shown (the calcification is the bright spot, top right with regard to the cross-hair indicating the center of the lumen). The colors encode the normalized values of intensity integrals within each sector of each ring.

2.4.2 Other Candidate Features

Apart from our previously defined metric, we are willing to evaluate the performance of other metrics. Since the detection of vascular anomalies through ML techniques is rather new, it is not possible to establish a state-of-the-art set of features for this purpose. Therefore, we make use of global features commonly used in vascular enhancement and/or segmentation to define an initial set of candidates.

Our criterion to select vascular enhancement/segmentation features comes from the fact that these features are supposed to give good responses at lumen positions in normal regions, while their responses are likely to deviate from the normality in the presence of a lesion. As an example, it is well-known that Hessian eigenvalues have poor responses (deviating from the normality) in the vicinity of calcifications. What represents a disadvantage for lumen enhancement or segmentation, can be exploited in our framework to allow anomaly detection.

An additional criterion in the selection of the alternative candidate features is motivated by the possibility of comparing their performance to our metric. Therefore, we keep the slice-by-slice strategy and our features have to be restricted to 2D features.

The candidate features exploit various image properties and different mathematical mechanisms. Their presentation hereafter follows the order from derivative- to integral-based. We thus begin with a metric based on second order derivatives, namely `Hessian eigenvalues`, which seeks normality by evaluating the symmetry of the cross-sectional shape through the eigenvalues of the Hessian matrix. Then we describe measures that exploit the medialness of normal vessels by evaluating first order derivatives at radial positions. These are: `Cores, Flux` and `MFlux` measures. We follow with the `Ribbon` metric that combines integration and differentiation in order to evaluate overall contrast between lumen and background. The `Ball` measure can be considered as integral-based, as it calculates intersection area to evaluate how well the shape of the lumen fits a disc. Another example of integral features are `Inertia moments`

that evaluate the symmetry of the cross-sectional shape by the use of inertia matrix eigenvalues.

Additionally, we have pre-selected a group of local features that have been used in object segmentation through ML techniques and that have recently been applied to detect vascular lesions (Tessmann et al., 2008). These are the so-called `Steerable features` that combine measurements of derivative nature of all orders (from zero to second order).

In the following we provide further details on the selected candidate features.

Hessian eigenvalues

The analysis of second order information of an image provides knowledge about the geometry of vessels. This can be done by evaluating the eigenvalues of the Hessian matrix. For the 2D case, it is defined as:

$$\mathbf{H}(I(x,y)) = \begin{bmatrix} \frac{\partial^2 I}{\partial x^2} & \frac{\partial^2 I}{\partial x \partial y} \\ \frac{\partial^2 I}{\partial y \partial x} & \frac{\partial^2 I}{\partial y^2} \end{bmatrix} \tag{2.16}$$

where the differentiation operation over image $I(x,y) = I(\mathbf{x})$ is defined, according to linear scale space theory (Lindeberg, 1996), as a convolution with Gaussian derivatives:

$$\frac{\partial}{\partial x} I(\mathbf{x},\sigma) = \sigma^{\gamma} I(\mathbf{x}) * \frac{\partial}{\partial x} G(\mathbf{x},\sigma), \tag{2.17}$$

with $G(\mathbf{x},\sigma)$ being a Gaussian function with standard deviation σ, and γ being a parameter introduced by Lindeberg (1996) to define a family of normalized derivatives.

The relationship between the eigenvalues of the Hessian matrix has been exploited by several authors (Bennink et al., 2007; Frangi et al., 1998; Li et al., 2003; Lorenz et al., 1997; Sato et al., 1998) in vessel extraction. In the context of anomaly detection, we calculate the Hessian and its eigenvalues at the vessel centerline point $\mathbf{p}$, within the orthogonal cross-section. As normal cross-sections are nearly symmetric, the two eigenvalues are expected to have similar magnitudes and the same sign. Actually, with the vessel lumen brighter than the surrounding tissues, we should have $\lambda_2 \approx \lambda_1 \ll 0$.

Cores

Core methods extract the medial axes of objects by following a ridge criterion. This criterion can be based on intensity ridge, *e.g.* (Aylward and Bullit, 2002), or on a medialness measure. For the purpose of our anomaly detection problem, we will evaluate the medialness measure used in the core method proposed by Fridman (2004). This measure is obtained by the use of so-called medial atoms. A medial atom is a discrete circular pattern with N equi-angular rays, denoted spokes. Medialness is measured by probing the image by the use of a directional derivative of a Gaussian placed at the

tip of each spoke, where the derivative is taken in the direction of the spoke (Fridman et al., 2003). Mathematically, this is expressed as:

$$\mathcal{C}(\mathbf{p}, \sigma_r) = \sum_{i=1}^{N} D_{i,\sigma_r}(\mathbf{x}_i), \tag{2.18}$$

where $\mathbf{x}_i$ is the tip location of the i-th spoke, $D_{i,\sigma_r}(\mathbf{x})$ is the value, at location $\mathbf{x}$, of the directional image derivative along the i-th spoke, *i.e.* along the vector $\mathbf{x}_i - \mathbf{p}$, and σ_r is the scale of interrogation. The authors recommend $\sigma_r = \|\mathbf{x}_i - \mathbf{p}\|/4$, where $\mathbf{p}$ is the center of the medial atom, in our case a centerline point.

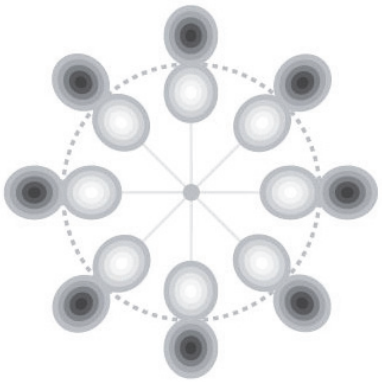


Figure 2.6: The Cores measure - Core evaluation pattern (medial atom) using Gaussian first order derivatives at the tips of the spokes (rays). Illustration obtained from (Lesage, 2009)

Thus defined `Cores` feature is closely related with `Flux` (Lesage et al., 2009a) but differs in the way the derivatives are calculated and in the scales selection.

Flux and MFlux

Flux is a gradient-based measure that exploits the orientation of gradient vectors by computing the gradient flux through the surface of an object. The flux $F(S)$ through a surface S is defined as:

$$F(S) = \int_S \langle \nabla I, \mathbf{n} \rangle \, ds \tag{2.19}$$

where ds is an infinitesimal surface patch, and $\langle \nabla I, \mathbf{n} \rangle$ denotes the scalar product of the image intensity gradient vector ∇I with the inward surface normal $\mathbf{n}$.

We apply a discrete version of the `Flux` metric, based on the implementation by Lesage et al. (2009a). In their proposal, after an equi-angular discretization of the cross-sectional circular contour into N points $\mathbf{x}_i$, `Flux` is defined as:

$$\text{Flux}(\mathbf{p}, r) = \frac{1}{N} \sum_{i=1}^{N} \langle \nabla I(\mathbf{x}_i), \mathbf{n}_i \rangle \tag{2.20}$$

where $\mathbf{p}$ is the center of the circle, r is its radius (*i.e.* $r = \|\mathbf{p} - \mathbf{x}_i\|$), $\nabla I(\mathbf{x}_i)$ is the gradient vector at point $\mathbf{x}_i$, and $\mathbf{n}_i = \frac{\mathbf{p}-\mathbf{x}_i}{r}$ is the inward radial direction.

Based on `Flux`, Lesage et al. (2009a) developed a new measurement denoted Minimal Flux or simply `MFlux` that aims at overcoming the sensitivity of `Flux` to isolated step edges. With N an even number of cross-sectional points, the `MFlux` analyzes $N/2$ pairs of diametrically opposed points $(\mathbf{x}_i, \mathbf{x}_i^\pi)$, *i.e.* $\mathbf{x}_i^\pi = \mathbf{x}_{\frac{N}{2}+i}$, and it retains the minimal flux contribution per pair:

$$\text{MFlux}(\mathbf{p}, r) = \frac{2}{N} \sum_{i=1}^{\frac{N}{2}} \min(\langle \nabla I(\mathbf{x}_i), \mathbf{n}_i \rangle, \langle \nabla I(\mathbf{x}_i^\pi), \mathbf{n}_i^\pi \rangle). \tag{2.21}$$

While Lesage et al. (2009a) use `Flux` and `MFlux` in order to find coronary arteries centerlines, we assume that vessel centerlines have already been extracted, $\mathbf{p}$ is thus a centerline point.

`Flux` and `MFlux` make use of a pre-regularized gradient vector field. As proposed by Lesage et al. (2009a) we use an isotropic Gaussian smoothing to perform the regularization, with a standard deviation σ_s of the same order as the image resolution to preserve the smallest vessels. Another key element of this implementation regards an image pre-processing applied to the original data, which performs the following intensity transformation:

$$\tilde{I}(\mathbf{x}) = \begin{cases} T_{low} & \text{if } I(\mathbf{x}) < T_{low} \\ T_{high} & \text{if } I(\mathbf{x}) > T_{high} \\ I(\mathbf{x}) & \text{otherwise} \end{cases} \tag{2.22}$$

where $I(\mathbf{x})$ is the original CT number of a voxel $\mathbf{x}$, $\tilde{I}(\mathbf{x})$ is the transformed intensity value and T_{low} and T_{high} represent lower and upper thresholds that restrict the intensity dynamics within the range $[T_{low}, T_{high}]$. Definition of the threshold values will be discussed in Chapter 4.

Ribbon metric

`Ribbon` is a region-based metric developed by Florin et al. (2005, 2006), which evaluates the difference between the mean intensity in an inner region A_{int} corresponding to the lumen, and the mean intensity in an outer band B_{ext} surrounding the lumen (Figure 2.7), both having the same area. In angiography images, where the lumen is brighter than the surrounding tissues, this difference is expected to be positive. The metric, denoted R, is therefore defined as:

$$R = \begin{cases} -\infty & , \quad \mu_{int} \leqslant \mu_{ext} \\ \dfrac{\mu_{int} - \mu_{ext}}{\mu_{int} + \mu_{ext}} & , \quad \text{otherwise} \end{cases} \tag{2.23}$$

with

$$\mu_{int} = \frac{1}{N_{int}} \sum_{\mathbf{x} \in A_{int}} I(\mathbf{x}), \qquad \mu_{ext} = \frac{1}{N_{ext}} \sum_{\mathbf{x} \in B_{ext}} I(\mathbf{x})$$

being the respective mean intensities of the inner and outer region, the respective pixel-sizes of which are $N_{in} = |\mathbf{x} \in A_{int}|$ and $N_{ext} = |\mathbf{x} \in B_{ext}|$. Equality of areas implies $N_{in} = N_{ext}$.

The initial ribbon measure defined by Florin et al. (2005) considered ellipsoidal cross-sections, thus making the inner region an ellipse. Since we expect normal cross-sections to be circular or nearly circular, in our implementation the inner region is a disc with a radius r_{int}, and the outer region is a ring delimited by a radius r_{ext}. The equality of areas implies that the radius of the external region has to be set as $r_{ext} = \sqrt{2} r_{int}$. This approach has already been followed by Lesage (2009).

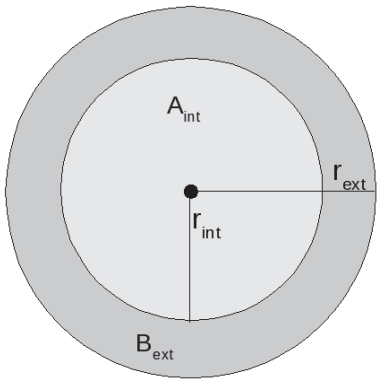


Figure 2.7: The ribbon metric - The metric evaluates the mean intensity difference between a circle A_{int} of radius r_{int} and a surrounding ring B_{ext} of radius r_{ext}.

Ball measure

The `Ball` measure (Nain et al., 2004) characterizes the shape of a region A by comparing it with a local ball-shaped neighborhood $B(\mathbf{x}, r)$ (disk in $\mathcal{R}^2$, solid sphere in $\mathcal{R}^3$). At every point $\mathbf{x}$ inside A, a measure ϵ calculates the percentage of points that fall into the intersection between A and the neighborhood $B(\mathbf{x}, r)$ centered at $\mathbf{x}$ and of radius r. The metric is defined as:

$$\epsilon(\mathbf{x}) = \int_{B(\mathbf{x},r)} \chi(\mathbf{y}) d\mathbf{y}, \tag{2.24}$$

where

$$\chi(\mathbf{y}) = \begin{cases} 1 & \text{if } \mathbf{y} \in A \\ 0 & \text{if } \mathbf{y} \notin A. \end{cases}$$

For the purpose of our problem, the Ball feature is only calculated at the location $\mathbf{p}$ corresponding to a centerline point. As for the function $\chi(\mathbf{y})$, it requires a definition of an inclusion criterion that identifies the pixels belonging to the region A. To be consistent with other metrics, we have used the thresholds defined in Eq. 2.22 as a criterion. Pixels falling into the range (T_{low}, T_{high}) are considered as belonging to A while the rest are not.

Inertia moments

Boldak et al. (2003) and Hernández Hoyos et al. (2006) exploit the analogy between image moments and those from statistics and mechanics, to obtain shape information from local structures. More particularly, they make use of second order moments, also noted inertia moments, of image intensities to extract information about the vessel geometry. Their approach is formulated in a three-dimensional space. In the following, we describe the 2D equivalent formulation that we have applied.

2D geometric moments, m_{pq} of order $n = p+q$, of a continuous function $f(x, y)$ are defined as:

$$m_{pq} = \iint_{\zeta} x^p y^q f(x, y) dx dy \qquad \text{where } p, q = 0, 1, 2, 3... \tag{2.25}$$

Here ζ represents the image region where the function $f(x, y)$ is defined. Adapting this definition, to grayscale images with pixel intensities $I(x, y)$, we have that $f(x, y) = I(x, y)$. Moreover, 2D central moments, μ_{pq} of order $n = p + q$ are expressed as:

$$\mu_{pq} = \iint_{\zeta} (x - \bar{x})^p (y - \bar{y})^q f(x, y) dx dy, \tag{2.26}$$

where $\bar{x}$, $\bar{y}$ represent the coordinates of the gravity center of ζ:

$$\bar{x} = \frac{m_{10}}{m_{00}}, \qquad \bar{y} = \frac{m_{01}}{m_{00}}. \tag{2.27}$$

For the purpose of our problem ζ is a disc of radius r centered in a vessel centerline point $\mathbf{p}$. Furthermore, as the gravity center of a normal cross-section is expected to coincide with the centerline point ((Boldak et al., 2003; Hernández Hoyos et al., 2006)), we take $(\bar{x}, \bar{y}) = \mathbf{p}$.

Using the inertia moments, it is possible to construct the inertia or covariance matrix that allows obtaining orientation information from the image:

$$\mathbf{I} = \begin{bmatrix} \mu'_{20} & \mu'_{11} \\ \mu'_{11} & \mu'_{02} \end{bmatrix}, \tag{2.28}$$

where

$$\mu'_{20} = \frac{\mu_{20}}{\mu_{00}}, \quad \mu'_{02} = \frac{\mu_{02}}{\mu_{00}}, \quad \mu'_{11} = \frac{\mu_{11}}{\mu_{00}} \ .$$

The eigenvectors of the inertia matrix correspond to the major and minor axes of the equivalent ellipse approximating the image intensity pattern. The corresponding eigenvalues are calculated as:

$$\lambda_i = \frac{\mu'_{20} + \mu'_{02}}{2} \pm \frac{\sqrt{4\mu'^2_{11} + (\mu'_{20} - \mu'_{02})^2}}{2}, \tag{2.29}$$

and are proportional to the length of the eigenvector axes. The ratio between the major and minor axes gives an indication of the eccentricity of an object. This is, how much the object approximate shape differs from a circle.

Steerable features

Steerable features were proposed by Zheng et al. (2007, 2008) to capture the orientation and scale of an object to be segmented, which in their case were the heart chambers. Steerable features sample points from the evaluated volume according to a sampling pattern (Figure 2.8). Afterwards, local features are extracted at each sample point, such as: image intensity, its derivatives, etc.

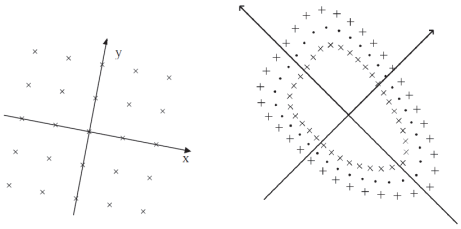


Figure 2.8: Sampling pattern examples - 2D sampling patterns of steerable features. Left, a regular sampling pattern. Right, sampling pattern around an expected shape boundary. Illustration from (Zheng et al., 2007)

In (Zheng et al., 2007), 24 different features are proposed for 3D datasets. Since our features are extracted on a 2D-basis, we eliminate the derivatives along the z direction, which reduces the number of features to 20. Given a point (x,y) with intensity I, the extracted features are denoted: I, $\sqrt{I}$, I^2, I^3, $\log I$, I_x, I_y, $|\nabla I|$, $\sqrt{|\nabla I|}$, ..., I_{xx}, I_{yy}, I_{xy}, $|\mathbf{H}|_F$, $\sqrt{|\mathbf{H}|_F}$...etc. Here, $|\mathbf{H}|_F$ denotes the Frobenius norm of the Hessian. To exploit the symmetrical characteristics of a normal vessel a circular pattern is used.

2.5 Feature Selection: Methodology

An individual set of features, *i.e.* variables extracted from a particular metric, can be used to train a classifier, while obtaining adequate results. Chapter 4 devotes a section to present the individual performance of features extracted from the metrics defined in section 2.4. The performance of individual feature sets rises the idea of combining some of them to compose a larger set that would outperform the classifiers trained with a single feature set.

Our goal is to construct an optimal feature set by combining the metrics defined in section 2.4. In practice, however, this is not a simple task and several issues arise when constructing a feature set. For instance, larger feature sets do not imply higher accuracy of the classifier. Moreover, large feature sets are computationally expensive.A high number of features can induce the so-called curse of dimensionality Bellman (1961) problem during the SVM training, which is related to the exponential growth of the data volume when adding new dimensions (here features). Thus, combining all possible features is not an adequate solution and a subset of features needs to be selected following a predefined criterion. The latter is not obvious. In fact, the best individual

features do not necessarily lead to a good classification or, conversely, poorly performing individual features do not always lead to a bad classification, when combined with others. Apart from these, an incorrect selection of a subset of features can lead to problems such as overfitting. Overfitting refers to the problem caused when a learning algorithm tries to find the best adjustment for the training data. It can occur that it memorizes peculiarities of the training data instead of finding a general predictive rule (Dietterich, 1995).

Feature selection refers to the task of constructing and selecting subsets of features that are useful to construct a good classifier and has been extensively studied (Guyon and Elisseeff, 2003; Liu et al., 2010; Zhao et al., 2010). It differs from the problem of finding all potentially relevant features, which is a suboptimal task.

Different criteria can be used to categorize feature selection algorithms. A widely used categorization proposes to divide the methods into wrappers, filters and embedded. *Wrappers* (Kohavi and John, 1997) use the learning machine of interest as a blackbox to score subsets of variables according to their predictive power. *Filters* select subsets of variables as a pre-processing step, independently of the chosen learning machine. *Embedded* methods perform variable selection within the training process, and feature relevance is obtained analytically from the objective of the learning model. Usually these methods are specific to a given learning machine.

Depending on the type of output that the algorithms provide they can be divided into two main categories: feature ranking and subset selection. *Variable ranking* methods compute a score function S from the input data and sort the features in a decreasing order (assuming a high score indicates a valuable variable) according to S. Then a threshold criterion is used to keep the best ranked features. As an advantage, these approaches are simple, scalable and with good empirical success. However, since they rank features independently, they lack the evaluation of variables interdependence.

On the other hand, *subset selection* methods search for an optimal subset from among all possible features. Wrapper methods usually fall into this category, while filter and embedded methods can either return a subset or a list of ranked features.

One of the main constraints of subset selection comes from the fact that the selection of a subset of variables is a NP-complete problem (Amaldi and Kann, 1998). Therefore, a trade-off is to be found. Typically, greedy search strategies are adopted, which seems to be computationally feasible and robust against overfitting. These are: forward selection and backward elimination. *Forward selection* progressively adds variables to the subset while *backward elimination* starts from the complete set of feature candidates and progressively eliminates the least valuable ones. In both cases, it is necessary to establish a stopping condition (when to stop adding/removing variables). Additionally, forward selection needs to select an initial feature from which to grow the selected subset.

Finally, feature selection algorithms, as classification methods, can be categorized into supervised, semi-supervised and unsupervised, according to the necessity of labeled data. However, most of the efforts have been directed towards feature selection in supervised environments. For instance, the Feature Selection Algorithm Repository (Zhao

et al., 2010) [1] only contains two feature ranking methods that can be used in unsupervised environments. Extensive reviews on feature selection (Guyon and Elisseeff, 2003; Liu and Yu, 2005) focus only on supervised approaches.

On the other hand, most of the existing unsupervised feature selection techniques rely on clustering (Ben-Hur and Guyon, 2003; Brunner et al., 2008; Søndberg-Madsen et al., 2003; Xing and Karp, 2001). Most clustering algorithms generate a cluster even if the data has no inherent cluster structure, so external validation tools are required (Ben-Hur and Guyon, 2003). Although this type of situation is not strictly considered as supervised learning, it does require manual interaction of an expert.

In this section, we propose a novel unsupervised feature selection scheme by making use of the classification algorithms, DLD-SVM and LPU. We propose to exploit the information provided by the the empirical risk $\mathcal{R}$ in these two methods to select features. This proposition falls into the category of wrapper methods. It is further discussed in section 2.5.1. In order to assess the performance of our proposal, we also make use of supervised feature selection schemes and compare the results obtained with each of the approaches. The latter are described in section 2.5.2.

We also add a brief description of an entropy-based unsupervised feature ranking approach (Section 2.5.3). In our work we only use it as a feature parameter tuning instrument (see Sections 3.2.1 and 3.3.1). Entropy is not used for feature selection, as it does not provide an initial estimate of ρ (or t respectively) that we need in our unsupervised (semi-supervised) selection scheme.

2.5.1 DLD-SVM and LPU for Feature Selection

An optimal feature set condition often means the minimal classification error (Peng et al., 2005). In section 2.3.1, we stated that minimizing the empirical error $\mathcal{R}$ also minimizes the classification error (refer to (Steinwart et al., 2005a,b) for further details). Therefore, the empirical risk functions defined in Eqs. 2.5 and 2.12 represent a natural way to perform feature selection in an unsupervised environment.

Intuitively, our proposal seeks to evaluate different subsets of features until a criterion is reached. In our case, the stopping condition is the obtention of a minimum in $\mathcal{R}$. The subset of features reaching this minimum is held. Again, since it is impossible to search across all possible subsets, a search strategy has to be selected. In literature, there is no evidence that a particular search strategy (forward selection or backward elimination) performs better than the others. Selection criteria are empirical. In our case, we have chosen the forward selection for two main reasons. First, the curse of dimensionality might produce undesired results using backward elimination when the size of the available datasets is significantly smaller than the number of variables to evaluate. Second, backward elimination is computationally very expensive.

The selection of a forward strategy imposes an additional challenge: it is necessary to determine which order should be followed for the progressive inclusion of variables. A variable ranking strategy is a reasonable approach to define the inclusion order. Best

[1] http://featureselection.asu.edu/

ranked features are first included whereas variables with a low rank are left for last.

Since we want to keep the unsupervised/semi-supervised nature of our approach, the use of labels should be avoided. Some common criteria used to rank features in an unsupervised environment include saliency, entropy, smoothness, density and reliability (Guyon and Elisseeff, 2003). More complex ranking strategies include the Laplacian score developed by He et al. (2005) and further extensions (Zhao and Liu, 2007). Here, we propose to also use the empirical risk $\mathcal{R}$ to define the inclusion order. In our approach, we first train our classifier with each set of features (features extracted from the same metric) individually. For each set, we keep the empirical risk obtained and use it to sort the features in ascending order. The feature selection algorithm then starts with the feature that performed best by itself (lowest empirical risk) and then it adds features according to the ranking.

Let m_l denote the number of features that compose the l-th evaluated metric, with its associated individual risk $\mathcal{R}_l$, such that $l = 1$ corresponds to the metric with the smallest risk $\mathcal{R}_1$ and $l = L$ corresponds to the metric with the largest risk $\mathcal{R}_L$, L being the total number of metrics. The subscripts denote the individual metrics and their associated risks. Let $\mathbf{X}$ be a matrix representing the entire available training set, which has M columns and N lines. Each line corresponds to one sample, while each column corresponds to one feature, *i.e.* each element of the matrix $\mathbf{X}$ is the value of the m-th feature in the n-th sample. The features are grouped by metrics in such a way that the first m_1 columns belong to the metric corresponding to $l = 1$ and so on. $\mathcal{R}_1$ and $\mathbf{X}$ are the inputs of the algorithm, as well as a vector $\mathbf{F}$ containing the ordered values of m_l, *i.e.* $\mathbf{F}[l] = m_l$.

At each iteration t, the algorithm (Algorithm 2.5.1) evaluates the risk $r_{\text{current}} = \mathcal{R}^t$ associated with a tentative feature set represented by a matrix $\mathbf{X}^t_{\text{train}}$. This evaluation is represented by the function $train(\cdot)$. The superscript $t = l - 1$ denotes the tentative feature set composed of l first metrics, and the risk associated to this set. At the first iteration ($t = 1$), the algorithm constructs $\mathbf{X}^1_{\text{train}}$ by taking the features belonging to the metrics $l = 1$ and $l = 2$, *i.e.* the $m_1 + m_2$ first columns from $\mathbf{X}$. Then it incrementally adds subsequent metrics, *i.e.* m_{t+1} consecutive columns from $\mathbf{X}$, so the number of features included in $\mathbf{X}^t_{\text{train}}$ is $M^t = |\mathbf{X}^t_{\text{train}}| = \sum_{i=1}^{t+1} m_i$. The algorithm stops when the condition $r_{\text{prev}} = \mathcal{R}_{t-1} \leq \mathcal{R}_t$ is accomplished, *i.e.* when the risk associated with $\mathbf{X}^t_{\text{train}}$ stops decreasing. The output of the algorithm is the number of features that should be kept.

It should be remarked that the formulation of the Algorithm 2.5.1 does not include the initial feature ranking using the empirical risk, as an input of the algorithm. With this, we want to make it clear that, even though we have chosen the empirical risk, any other feature ranking approach can be used for this initial step, as long as it provides an initial ordering.

Another important characteristic of this proposal relies in the use of the empirical risk as a criterion to determine the optimal subset. The risk function $\mathcal{R}(f)$ is not strictly attached to DLD-SVM and LPU formulations. Hush et al. (2005) state that $\mathcal{R}$ is a legitimate performance measure for anomaly detection. This means that anomaly

detection problems interpretable as binary classification problem can be formulated in such a way that the empirical risk can be computed (for examples, see (Steinwart et al., 2005a)). Therefore, Algorithm 2.5.1 is not limited to the DLD-SVM and LPU formulation. In fact, any classification scheme that can be formulated by means of the empirical risk can be plugged in into the function $train(\cdot)$ of Algorithm 2.5.1.

Algorithm 2.5.1: UNSUPERVISEDSELECTION$(\mathbf{X}, \mathcal{R}_1, \mathbf{F})$

$r_{\text{prev}} \leftarrow \infty$
$r_{\text{current}} \leftarrow \mathcal{R}_1$
$t \leftarrow 1$
$M^t \leftarrow \mathbf{F}[1]$
while $r_{\text{current}} < r_{prev}$
do
$r_{\text{prev}} \leftarrow r_{current}$
$M^t \leftarrow M^t + \mathbf{F}[t]$
$\mathbf{X}_{\text{train}} \leftarrow \mathbf{X}\left[1 : N, 1 : M^t\right]$
$r_{\text{current}} \leftarrow train(\mathbf{X}_{\text{train}})$
$t \leftarrow t + 1$
return (M^t)

2.5.2 Supervised Feature Selection Strategies

Chen and Lin (2006) empirically showed that SVM-based methods can be successfully combined with various feature selection strategies. This proposal participated in NIPS 2003 Feature Selection Challenge [1] finishing third in the final ranking. The supervised feature selection strategy we present here is inspired by their work. Their seminal idea consisted in combining different feature selection strategies to be used as filters for a posterior classification using SVMs.

From the original proposal by Chen and Lin (2006), we have kept two of the strategies: a variable ranking method denoted F-score (Chen and Lin, 2006; Yang et al., 2008) and random forests (Breiman, 2001), which is a wrapper method. Additionally, we evaluate another popular wrapper method named SVM-Recursive feature elimination (SVM-RFE) described by Guyon et al. (2002). In the following we provide further details on each of these strategies.

F-score

F-score (Chen and Lin, 2006; Yang et al., 2008) is a technique that measures the discrimination of two sets of real numbers. Given M training vectors $\vec{x}_m, m = 1, ..., M$, if the respective mean values and variances of the positively and negatively labeled sets for the i-th feature are $\bar{x}^i_+, \bar{x}^i_-, (\sigma^i_+)^2$ and $(\sigma^i_-)^2$, then the F-score of the ith feature is

[1] http://clopinet.com/isabelle/Projects/NIPS2003/

defined as:

$$F(i) \equiv \frac{\left(\bar{x}^i_+ - \bar{x}^i\right)^2 + \left(\bar{x}^i_- - \bar{x}^i\right)^2}{(\sigma^i_+)^2 + (\sigma^i_-)^2} \tag{2.30}$$

where $\bar{x}^i$ is the mean value of the whole set (union of positive and negative) for the i-th feature. The larger the F-score, the more likely the feature to be discriminative.

A disadvantage of the F-score is that although the score gives an idea of how discriminative a feature can be, since it evaluates features individually, it does not provide any information about mutual information among features.

Using the F-score, our procedure that performs feature selection is as follows:

1. Calculate the F-score of every single feature from each metric.
2. Rank the different metrics based on F-scores obtained from all features composing each individual metric.
3. Build feature sets by combining the first $2, 3, \ldots, 6$ ranked metrics, and evaluate each one of these combinations. To this purpose, for every feature set thus constructed:
   (a) Randomly split the available data into training and validation set.
   (b) Train an SVM using the training set and validate its performance with the validation set.
   (c) Repeat the training-validation process a fixed number K of times and average the obtained error.
4. Select the number of metrics generating the lowest error.
5. Evaluate the selected feature set on a different data set.

The SVM training that is done at step 3b) is not restricted to a particular methodology. In their seminal work, Chen and Lin (2006) use a standard soft margin SVM but, as an example, DLD-SVM or LPU could be used.

The process of randomly splitting the available data into training and validation is commonly referred as cross-validation (CV). The goal of performing several rounds of CV is to reduce the variability by averaging the validation results over these rounds. This version is known as N-cross-validation. Definition of the number of rounds (K), as well as of the partition strategy, is left for Chapter 4.

Random Forests

Random forests (Breiman, 2001) is a classification method that provides feature importance. A random forest (RF) is a classifier consisting of a collection of tree-structured classifiers (decision trees), each of which is constructed by instances with randomly sampled features.

Given X the training set of size N with M variables, each of the trees gives a classification, denoted a vote, for this input set. The forest selects the classification that has received the majority of votes.

Instead of using all the original N samples as training set, each tree grows using N cases randomly drawn from X with replacement (*i.e.* once drawn, a sample returns to the bag and is likely to be drawn again). With this procedure , about one-third of the cases are left out of the sample. The latter are often called out-of-bag data. Then, at each stage (node) of the growth of a tree m_{try} variables, out of M, are randomly drawn ($m_{try} \ll M$) and used to calculate possible splits. The best split based on the m_{try} variables is kept. Each tree is grown to the largest possible extent, *i.e.* as long as the m_{try} variables permit to split the data. The algorithm requires the specification of the parameter m_{try}, as well as the number of trees in the forest n_{tree}.

Breiman (2002) proposes four different measures of variable importance. We have chosen to use permutation importance. Variable importance is estimated by evaluating how much the prediction error increases when out-of-bag data of a variable is permuted while others are left unchanged. The required computations are performed on a tree by tree basis, as the random forest is constructed.

Using random forests as a filter that allows feature ranking, the selection procedure (Svetnik et al., 2004) is as follows:

1. Partition the data to perform N-cross-validation.

2. On thus obtained CV training set, train a model on all variables and use the variable importance measure to rank the metrics.

3. Predict using RF and the CV validation set. Record the error.

4. Use the ranking to remove the least important half of the metrics. Retrain the remainder and predict again, recording the error. Repeat the procedure until only one set of features is left.

5. Repeat from the beginning a number K of times with different CV partitions.

6. Average the results from the K partitions to obtain the error rate.

7. Choose the subset with the lowest error.

8. Evaluate the best features on a different data set.

It should be noted that feature ranking procedure is not recursive. Variable importance is obtained once at step 2 and not anymore for a particular CV iteration. According to Svetnik et al. (2004) a recursive procedure is much greedier and has a worse performance.

As with F-score, the selection of the parameter K of the cross-validation, and the partition strategy are further discussed in Chapter 4.

SVM-RFE

The SVM-RFE algorithm (Guyon et al., 2002) returns the ranking of the features of a classification problem by training a SVM (Schölkopf and Smola, 2001) with a linear kernel. Given the subset of surviving features, the algorithm iterates by removing the feature with smallest ranking criterion at each step until no more features are left over. The ranking criterion consists of a weight vector $\mathbf{w}$ that is constructed from the linear combination of the training patterns, according to Eq. A.9 (see Appendix A for details of the SVM classification formulation). The features ranked list provided as output is used as the criterion for variable selection. We refer the reader to (Guyon et al., 2002) for a more detailed explanation of the algorithm.

The procedure to obtain a subset of relevant features using SVM-RFE is the same as the one used for RF and will therefore not be described again. The only difference relies on the classifier. Instead of using and RF classifier, SVM-RFE is used. Once again, the selection of parameter K of the cross-validation, and the partition strategy are discussed in Chapter 4.

2.5.3 Entropy

Entropy is a commonly used metric for variable ranking in unsupervised environments (Guyon and Elisseeff, 2003). Features with a low entropy have a high level of information for prediction.

The Shannon entropy H of a discrete random variable $\mathbf{x}$ with n possible different values is:

$$H(\mathbf{x}) = E(\mathcal{I}(\mathbf{x})), \tag{2.31}$$

where E is the expected value, and $\mathcal{I}$ is the information content of $\mathbf{x}$. Since $\mathcal{I}(\mathbf{x})$ is a random variable, if p denotes the probability mass function of $\mathbf{x}$, then the entropy can be written as:

$$H(X) = \sum_{i=1}^{n} p(x_i)\mathcal{I}(x_i) = -\sum_{i=1}^{n} p(x_i)\log p(x_i). \tag{2.32}$$

In the case of $p(x_i) = 0$ for some i, the value of the corresponding summand is taken to be zero.

The original entropy definition is valid only for discrete variables, whereas the response of the proposed metrics is of continuous nature. Therefore, it is necessary to make use of the differential entropy, an extension of the original formulation. Given $\mathbf{x}$ a random *continuous* variable with probability density function $f_{\mathbf{x}}(x)$, the differential entropy $h(x)$ is defined as

$$h(x) = -\int f_{\mathbf{x}}(x)\ln f_{\mathbf{x}}(x)dx \tag{2.33}$$

2.6 Discussion

In this chapter we have presented our vascular lesion detection framework. Since our proposal is based on ML techniques, three key elements were included: the definition of a classifier, the selection of metrics and a feature selection scheme.

We have proposed the use of two different classifier schemes: DLD-SVM and LPU. Both methods follow an anomaly detection approach. The reason for our choice is motivated by the fact that vascular lesions are highly heterogeneous so it is difficult to determine a class that can group them for binary or even multiple classification (in the case that each type of lesion is represented by a single class). In fact, Chawla et al. (2003) have shown that the high imbalance in class cardinalities of asymmetric classification causes conventional classification techniques to yield unsatisfactory accuracy. This represents a significant advantage of our proposal w.r.t. state-of-the-art methods aiming at the detection of vascular lesions. While the existent methods use standard binary classification (differentiation of two classes) we are the first to use an anomaly detection scheme. Furthermore, to the best of our knowledge, this is the first time that DLD-SVM and LPU are used in this domain.

Following the seminal idea of modeling anomalies as inconsistent observations w.r.t. the rest of the data, we have designed an intensity-based metric that aims to capture the intensity profiles and the axial symmetry of normal vessels when evaluated on orthogonal cross-sections. Additionally, we propose to include global features typically used for vascular enhancement or segmentation. Our motivation comes from the fact that these are expected to have an homogeneous behavior inside the lumen and to deviate from it when there are surrounding lesions.

In order to define an optimal feature subset to train the selected classifiers, we have proposed a feature selection strategy that uses the empirical risk $\mathcal{R}$ as selection and stopping criterion. Since the empirical risk can be used as performance measure of anomaly detection formulations, the proposed algorithm permits the use of any classification strategy (that can be formulated in terms of the aforementioned risk). This makes the method quite generic up to the point that it can be used for unsupervised, semi-supervised and supervised learning schemes. Nevertheless, we also propose a more classical feature selection scheme that uses different subset selection methods as filters in combination with the DLD-SVM and LPU. Although this scheme has the disadvantage of requiring labels, it serves as an evaluation reference for the risk-based method.

Evaluation of the herein described approach is left for Chapter 4.

3

Metric Parameter Tuning and Experimental Setup

In Chapter 2, we introduced all the elements that build up our anomaly detection framework. Now, an important step is to define an optimal feature subset that maximizes the performance of our two algorithms DLD-SVM and LPU. Nonetheless, before achieving feature selection, it is necessary to tune up the different candidate metrics, which is the purpose of this chapter. In Section 3.2 we elaborate on the experimental design and evaluation of the proposed `Concentric rings` metric. Since the performance of the metric has never been assessed neither in the anomaly detection domain nor in vascular enhancement/segmentation field, we evaluate its use in combination with DLD-SVM and LPU to determine its power in anomaly detection. Once the `Concentric rings` metric is tuned and validated, the remaining candidate feature sets are adjusted such that they can be compared to the `Concentric rings` metric (Section 3.3).

The protocol followed in the different experiments is defined in Section 3.1. Here we describe the available data, the implementation of the classification algorithms and the performance measures used to evaluate the obtained results. Let us remark that the experimental guidelines herein described are also followed in the experiments that will be discussed in Chapter 4.

A final discussion of the obtained results is presented at the end of the chapter.

3.1 Evaluation Protocol

The present section is devoted to describe all the required elements for the evaluation of the methodology that was presented in Chapter 2. First, we specify the characteristics of the input data that was used in our experiments, followed by the configuration of the different classifiers. More particularly, we define the construction of training and testing sets, the tuning parameters of each classifier and the way they are tuned up. Finally, we define the measures that were used to assess the performance of our algorithms.

3.1.1 Experimental Data

Our proposal was evaluated using two different types of data: 3D synthetic images and 3D cardiac CT data sets.

Synthetic Phantoms

This set of data consisted of 42 different volumes containing a variety of cases typically encountered in vascular analysis. The phantom generation procedure was as follows:

1. 42 synthetic tubes with a radius varying from $R = 4$ to $R = 10$ voxels were created. Vessel cross-sections were either circular or elliptical. Additionally, the cross-section intensity profile was bar-like or Gaussian-like (Figure 3.1(a)).

2. From these 42 volumes, 21 were randomly selected to include simulated lesions. Three different types of anomalies were simulated: aneurysms, atherosclerotic calcified plaques and atherosclerotic hypodense plaques (Figure 3.1(b)-(d)).

3. From the remaining data sets, 7 volumes were randomly selected to simulate the presence of tangent structures. Like in Krissian et al. (2000), we say that a structure is tangent to a vessel when their boundaries are near enough to disturb the metric computation (Figure 3.1(e)).

4. 7 further volumes were randomly selected from the remaining data sets to include vessel bifurcations. These were generated at an angle $\alpha \in [\pi/4, \pi/2]$ as suggested by theoretical studies based on hemodynamics (Murray, 1926; Oka and Nakai, 1987) (Figure 3.1(g)).

5. The 7 remaining data sets were left as completely normal.

With the inclusion of adjacent structures in our phantoms, we want to recreate situations that could cause a failure of our method. To make the phantoms more realistic, they were created using the typical Hounsfield Unit values that are found in CT images for blood, background and plaque components, as well as the the typical image dimensions and voxel size. Gaussian noise was also added, resulting in a contrast-to-noise-ratio value of 10.

Cardiac CT Data

Fifteen cardiac CT datasets coming from different patients were obtained from two different sources: the Rotterdam Coronary Artery Algorithm Evaluation which contains data from the Erasmus Medical Center (Rotterdam, The Netherlands) and Hôpital Louis Pradel (Bron, France). Data coming from Erasmus Medical Center (which will be denoted as Set 1) was used for unsupervised feature selection and overall performance evaluation. Data from Hôpital Louis Pradel (the Set 2) was used exclusively for feature selection. The acquisition protocols are next described.

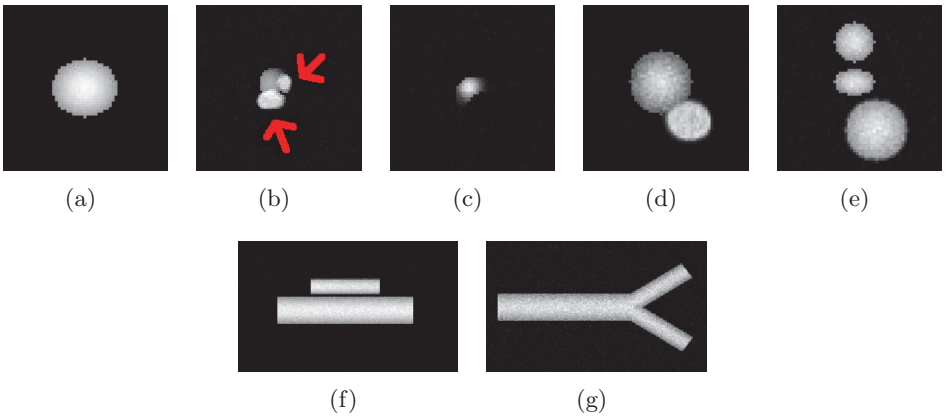


Figure 3.1: Examples of phantoms - (a) normal elliptical cross-section, (b) calcified-plaque (pointed with the arrows) and (c) a stenosis resulting from a soft-plaque. (d) A cross-section with an aneurysm, (e) cross-section with adjacent structures (the vessel of interest is located in the middle of the image), (f) sagittal view of a vessel with an adjacent structure and (g) a sagittal view of a bifurcation.

Set 1. Rotterdam Coronary Artery Algorithm Evaluation Framework. The Rotterdam Coronary Artery Algorithm Evaluation Framework (Schaap et al., 2009) aims at evaluating and comparing different algorithms for coronary artery centerline extraction from CT angiography (CTA) data. For this purpose, thirty-two datasets are provided. The acquisition protocol is as follows:

The CTA data was acquired in the Erasmus MC, University Medical Center Rotterdam, The Netherlands. Thirty-two datasets were randomly selected from a series of patients who underwent a cardiac CTA examination between June 2005 and June 2006. Twenty datasets were acquired with a 64-slice CT scanner and twelve datasets with a dual-source CT scanner (Sensation 64 and Somatom Definition, Siemens Medical Solutions, Forchheim, Germany). A tube voltage of 120 kV was used for both scanners. All datasets were acquired with ECG-pulsing (Weustink et al., 2008). The maximum current (625 mA for the dual-source scanner and 900 mA for the 64-slice scanner) was used in the window from 25% to 70% of the R-R interval and outside this window the tube current was reduced to 20% of the maximum current.
Both scanners operated with a detector width of 0.6 mm. The image data was acquired with a table feed of 3.8 mm per rotation (64-slice datasets) or 3.8 mm to 10 mm, individually adapted to the patient's heart rate (dual-source datasets).
Diastolic reconstructions were used, with reconstruction intervals varying from 250 ms to 400 ms before the R-peak. Three datasets were reconstructed using a sharp (B46f) kernel, all others were reconstructed using a medium-to-smooth (B30f) kernel. The mean voxel size of the datasets is $0.32 \times 0.32 \times 0.4$ mm (Schaap et al., 2009).

From the thirty-two available datasets, we have selected eight cases which correspond to the training set of the Rotterdam Coronary Artery Algorithm Evaluation Framework.

Set 2. Hôpital Louis Pradel (Bron, France). A total of 7 datasets were acquired on a 64-row CT scanner (Brilliance 64 − Philips Healthcare, Cleveland, OH) with a standard scan protocol using the following parameters: 120 kV, 300 mAs, collimation 52×1.5 mm, rotation time 0.35 seconds and scan time 10-14 seconds. Image reconstructions were made with an in-plane pixel size of 0.37×0.37 mm^2, matrix size 512×512, slice thickness 0.9 mm, increment 0.45 mm, with an intermediate reconstruction kernel (B).

Data preparation

Data sets from every source were sorted into different categories. These categories in synthetic data were: normal, aneurysm, soft-plaque, calcification, bifurcation and adjacent structure. For real data, only three categories were defined: normal, calcification and soft-plaque. Bifurcation, adjacent structure and aneurysm were omitted since all vessels had both adjacent structures and bifurcations, while no dataset contained aneurysms. This classification was performed to assure representation of every possible case when configuring the sets required for testing and evaluation (see Section 3.1.2). In set construction, vessels from each category were randomly assigned to a particular set.

The centerlines of 4 different coronary arteries were available (Schaap et al., 2009) for Set 1 (Cardiac CT). The centerlines from three different coronary arteries were manually drawn in Set 2 (Cardiac CT). Similarly, the centerline of one carotid artery (common carotid and internal carotid artery) per dataset was manually drawn in the Carotid CT datasets. The use of reference manually traced centerlines at this stage is justified by the need to evaluate the anomaly detection methods alone, so that the results are not affected by the accuracy of the centerline-extraction algorithm. Using these centerlines, orthogonal cross-sections were computed. Regarding synthetic phantoms, a theoretic centerline was available that was used to obtain the orthogonal cross-sections in every phantom.

Cross-sections from patients' data were manually labeled as normal or pathological (Table 3.1). The annotations were used for the purpose of training in LPU, supervised feature selection and performance evaluation.

It should be remarked that data coming from Cardiac set 2 was labeled by only one observer, while datasets from Cardiac Set 1 were labeled by two observers which allows to measure the agreement in anomaly detection between our methods and each of the observers, as well as the agreement between the two observers and the two methods.

Features extracted using the metrics defined in Section 2.4 were computed on every vessel cross-section. In every case, the features were normalized into the range $[0, 1]$.

3.1.2 Configuration of the Learning Algorithms

Five different learning algorithms are used along this work. The first two are the DLD-SVM and LPU approaches that we have proposed to use in the identification vascular lesions, as well as in the feature selection task. Additionally, we make use of Random

Table 3.1: Available normal and abnormal samples. Discrimination is done per each type of dataset since they are evaluated separately. Cardiac CT (Set 1) has been labeled by two different observers which leads to slightly different numbers of healthy and abnormal slices. Classification according by each observer is separated by a slash

Dataset	Healthy slices	Abnormal slices
Synthetic data	18500	1500
Cardiac CT (Set 1)	10447 / 9095	1010 / 1352
Cardiac CT (Set 2)	2035	195

Forests, SVM-RFE and soft margin SVM. In every case, a set of parameters need to be tuned up.

In the present chapter, we only make use of DLD-SVM and LPU since feature selection is discussed in the next chapter. Nevertheless, for the sake of coherence, we have selected to include the configuration of all the classifiers in the same section so that comparison can be easily achieved. In the following, we detail the necessary parameters of each classifier, its search strategy, implementation details and the data required by each algorithm.

DLD-SVM

The solution of the DLD-SVM formulation requires data division into three subsets: the training set $\mathcal{T} = \{T, T'\}$, the validation set $\mathcal{V} = \{V, V'\}$ and the testing set $\mathcal{W} = \{W, W'\}$, where T, V, W contained samples drawn from the set Q, and T', V', W' contained artificial samples generated according to the distribution defined by μ. The samples from Q were distributed so as to achieve independence and representation of every possible case. To this purpose, cross-sections coming from the same vessel were not included in two different sets. Given m the number of features from each sample in the set Q, artificially generated samples were generated following a uniform distribution on $[0, 1]^m$. For simplicity, we will refer to the artificial samples as the μ-samples. Table 3.2 summarizes the data distribution in function of the datasets previously defined in Section 3.1.1.

The DLD-SVM problem was solved using the `LIBSVM` software (Chang and Lin, 2001). DLD-SVM requires the optimization of three parameters: the σ^2 from the Gaussian kernel function, the density ρ and the regularization parameter λ from Eq. 2.6. These were optimized following the guidelines proposed in (Steinwart et al., 2005a,b).

The learning process was performed for a wide range of values of ρ (step 100 for $\rho \in [100, 2000]$ and a few additional values on each side of this interval: see Table 3.3). For each ρ value, the following steps were performed:

1. For a given pair of parameters λ and σ^2, the SVM was used to find a decision function f, using the training set $\mathcal{T}$.

2. The empirical risk $\mathcal{R}$ associated with f was calculated on the validation set $\mathcal{V}$, according to Eq. 2.7.

Table 3.2: Data distribution for DLD-SVM. Sample sizes for training $\mathcal{T} = \{T, T'\}$, validation $\mathcal{V} = \{V, V'\}$ and testing $\mathcal{W} = \{W, W'\}$ sets using synthetic and patient data. Q denotes the available unlabeled data, while μ denotes the artificially generated samples. Data obtained from Set 2 in cardiac data is exclusively used for the purpose of feature selection.

	Training	**Validation**	**Testing**
Synthetic data			
Q-samples	$\|T\| = 6000$	$\|V\| = 4000$	$\|W\| = 10000$
μ-samples	$\|T'\| = 12000$	$\|V'\| = 100000$	$\|W'\| = 200000$
Cardiac data (Set 1)			
Q-samples	$\|T\| = 3440$	$\|V\| = 2300$	$\|W\| = 5717$
μ-samples	$\|T'\| = 10000$	$\|V'\| = 100000$	$\|W'\| = 100000$
Cardiac data (Set 2)			
Q-samples	$\|T\| = 669$	$\|V\| = 446$	$\|W\| = 1115$
μ-samples	$\|T'\| = 8000$	$\|V'\| = 10000$	$\|W'\| = 50000$

Table 3.3: DLD-SVM parameters. Variation of parameter values for learning (ρ) and grid search: λ and σ^2.

Parameter	**Values**
ρ	0.05, 0.1, 1, 10, 50, 100, 200, 300, ..., 1800, 1900, 2000, 3000, 4000
λ	1.0, 0.5, 0.1, 0.05, 0.01, ..., 10^{-7}
σ^2	$10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}$,..., 10^{2}

The steps 1 and 2 were repeated for various values of the parameters λ and σ^2 by employing a grid search (Chang and Lin, 2001) (Table 3.3 summarizes the grid values for λ and σ^2). For each value of ρ considered, the pair (λ, σ^2) with the smallest empirical risk $\mathcal{R}$ on $\mathcal{V}$ was kept.

Finally, the learned decision function f, corresponding to λ, σ^2 and ρ in a way that the empirical risk was minimum, was applied on the testing set $\mathcal{W}$ to estimate the overall method performance.

LPU

LPU does not require separate sets for training and testing. In fact, all the available data can be used for training. The main difference w.r.t. DLD-SVM relies in the necessity of labeled data for normality. Therefore, two sets are defined. The first set Q containing labeled normal data and the second set X containing unlabeled data. As with DLD-SVM, samples were distributed so as to achieve independence and representation of every possible case. To this purpose, healthy cross-sections coming from the same vessel were not included in two different sets. The cardinality of Q and X when

Table 3.4: Data distribution for LPU. Sample sizes for Q and X sets of the LPU algorithm using synthetic and patient data. The data obtained from Set 2 in cardiac data is exclusively used for the purpose of feature selection.

	Q-samples	X-samples
Synthetic data	1000	19000
Cardiac data (Set 1)	886	10571
Cardiac data (Set 2)	426	1609

Table 3.5: LPU parameters. Variation of parameter values for learning (t) and grid search: K_λ and K_{σ^2}.

Parameter	Values
t	0.1, 0.2, ..., 0.8, 0.9, 0.91, 0.92, 0.93, ..., 0.98, 0.99
K_λ	0.00001, 0.00005, 0.0001, ..., 1, 5, 10, 50, ..., 100, 500, 1000
K_{σ^2}	$10^{-4}, 5\times10^{-4}, 10^{-3}, 5\times10^{-3}$,..., 1

constructed with the different available data is presented in Table 3.4. It is important to remark that the X samples are both used for training and testing.

The `LIBSVM` software (Chang and Lin, 2001) was also used to solve the LPU problem. Again, three parameters need to be optimized: the σ^2 from the Gaussian kernel function, the threshold t and the regularization parameter λ from Eq. 2.14.

The optimization process iterates over a range of values of t (see Table 3.5), performing the following steps:

1. For a given pair of parameters λ and σ^2, the SVM was used to find a decision function f, using the training set $\{Q, X\}$.

2. The empirical risk $\mathcal{R}$ associated with f, was calculated according to Eq. 2.7.

A grid search similar as the one performed for DLD-SVM, is used in LPU to optimize λ and σ^2. Steps 1 and 2 are repeated for a set of values of λ and σ^2. Let n be the total samples available for training and m the number of features in each sample, λ values are defined by:

$$\lambda = \frac{K_\lambda}{\sqrt{n}}, \tag{3.1}$$

where K_λ is a numerical constant that is varied to generate the search grid. The σ^2 parameter is computed as

$$\sigma^2 = \frac{K_{\sigma^2}\sqrt{n}}{m}, \tag{3.2}$$

where again K_{σ^2} is a numerical constant that varies to generate the search grid. Table 3.5 summarizes the constant K values used in the grid search.

The learned decision function f, corresponding to λ, σ^2 and t minimizing the empirical risk, is applied on the set X to evaluate the performance of the method.

Table 3.6: Data distribution for F-score, RF and SVM-RFE. Sample size for patient data using the F-score, RF and SVM-RFE feature selection algorithms. These methods are exclusively used in the task of feature selection. In the three methods, healthy samples receive a positive label (+1), while lesions receive a negative label (−1).

	Healthy slices (positive samples)	**Abnormal slices (negative samples)**
Cardiac data (Set 2)	1840	195

F-score

Data used for feature selection comes exclusively from Set 2, which is entirely labeled. Table 3.6 details the number of healthy and lesioned samples of this set.

Set 2 is made up of 7 different datasets. We have chosen to perform a seven-fold cross-validation using this set. This means that we consider all the samples from one dataset as a subset and, on every iteration, one of the datasets is used for validation, while keeping the six remaining datasets for training. The operation is repeated seven times to collect errors for each case. We could have chosen a different approach where a randomly selected subset S' from Set 2 is used for validation and the remaining data S, where $|S'| < |S|$, is kept for training. This, however, generates interdependence (a particular vessel can make part of both subsets) among sets which we want to avoid.

Since the three supervised feature selection methods we have selected use cross-validation, the previously described data partition scheme was applied not only to F-score, but also to RF and SVM-RFE strategies.

In section 2.5.2, we mentioned that the SVM training step involved in the F-score procedure can be performed with any particular SVM algorithm. For instance, DLD-SVM or LPU approaches can be used. However, we have chosen to use the standard soft margin SVM (Cortes and Vapnik, 1995). Unlike the unsupervised feature selection scheme, where DLD-SVM is used to select a feature subset for a subsequent use by the DLD-SVM classifier, and LPU is used to select a possibly different feature subset for a subsequent use by the LPU classifier, here we propose to select a single feature subset independent of the targeted classifier (*i.e.* DLD-SVM, LPU or another). For this purpose, the most generic version of SVM, the soft margin SVM, is an appropriate choice.

The `LIBSVM` software (Chang and Lin, 2001) was used to solve the soft margin SVM. In this case, two parameters need to be tuned up. The parameter C (see Eq. A.13 in Appendix A) and σ^2 of the Gaussian kernel. In a similar fashion as with DLD-SVM and LPU, a grid search is performed over these two parameters (see Table 3.7). However, experimentally we could determine that variation of C did not induce significant changes. Therefore, $C = 10$ was fixed and the search was performed exclusively on σ^2. The set of values was explored and the one resulting in the minimum training error was kept.

Table 3.7: F-score, RF and SVM-RFE parameters. RF requires tuning of m_{try} and n_{trees}. m_{try} tuning is a function of the number of features m. Both SVM-RFE and soft margin SVM used in the F-score procedure require a grid search over the parameter C (Eq. A.13). Additionally, the soft margin SVM requires a grid search over the σ^2 parameter of the Gaussian kernel.

Parameter	Value
F-score	
σ^2	$10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}$,..., 10^2
C	0.001, 0.005, 0.01, 0.05, ..., 100
SVM-RFE	
C	0.001, 0.005, 0.01, 0.05, ..., 100
Random Forests	
m_{try}	$\frac{m}{2}$, $\frac{m}{4}$, $\sqrt{m}$, m
n_{trees}	10, 50, 100, 500, 1000, 2000, 5000, 10000

Random Forests

The `randomForest` package of the `R` software[1] was used to implement Random Forests. RF requires that two parameters are specified. These are the number of variables in the random subset at each node m_{try} and the number of trees in the forest n_{tree}. Table 3.7 shows the evaluated ranges for both m_{try} and n_{tree}. Experimentations showed that the variation in the number of trees did not affect the results so the value was fixed to $n_{tree} = 1000$. Regarding m_{try}, the classification error was slightly smaller when using $m_{try} = \sqrt{m}$. The result is coherent with previous publications using RF for variable reduction (Svetnik et al., 2004) and it is also coherent with the default value suggested by the `randomForest` package. Therefore, $m_{try} = \sqrt{m}$ was fixed for the feature selection process.

SVM-RFE

An implementation based on the `LIBSVM` Chang and Lin (2001) interface for `R` software was used to solve SVM-RFE. This interface is preferred w.r.t. the `C++ LIBSVM` version used to solve DLD-SVM and LPU, since it contains specific methods associated to the SVM-RFE solution.

SVM-RFE makes use of a linear kernel. Therefore, only the parameter C needs to be tuned up (see Appendix A for the formulation). A coarse grid search was performed through a set of values (Table 3.7) showing no significant variation in the results. C =10 was kept for the experimentations.

[1] `http://www.r-project.org/`

3.1.3 Performance Assessment

The performance of a classification system is typically presented in terms of specificity, sensitivity and accuracy. In order to formulate the expressions of these three measurements, let us first define true positives (TP) as correctly detected anomalies; false negatives (FN) as undetected anomalies; true negatives (TN) as correctly classified negative patterns (*e.g.* normal cases) and false positives (FP) as erroneously classified negative samples. In the following, we first provide our definition for TP, TN, FP and FP to then introduce the different performance measures.

Definition of TP, TN, FP and FN

Both DLD-SVM and LPU suggest that a positive finding refers to the detection of a healthy slice. This is because DLD-SVM seeks for normality, while LPU seeks for elements from the set Q which, in our case, are healthy slices. This perspective does not fit the definitions that are typically used in a clinical context. In the latter, a positive usually refers to a lesion. Therefore, we define a TP as a correctly classified anomaly.

Even though, it is clear that we define TP's as correctly classified anomalies, it is still is necessary to establish a measurement unit. Previous state-of-the art works lack of a clear definition of this aspect, which makes performance comparison difficult.

Table 3.8 summarizes existing works aiming at the evaluation of the coronary heart disease detection methods, and how they define TP, TN, FP and FN. It can be seen that only the clinical review from Reimann et al. (2009) clearly stated how these variables are defined. Although the article of Arnoldi et al. (2010) states that a per-patient and per-vessel analysis is performed, it does not detail how these measurements are performed. In Tessmann et al. (2008), only information of true positives is provided. In a later work (Tessmann et al., 2009), the same authors try to provide more information by performing vessel- and lesion-based evaluations. However, important information is still missing. For instance, in both articles no details are given on what is considered a true negative. True negatives are a key in determining the specificity of a classifier. Moreover, since the above-cited work attempts to discriminate soft plaques and calcifications, no details are given on how mixed plaques are evaluated. We believe that the lack of a clear and standardized definition for TP, TN, FP and FN is an important weakness of the available literature in this field.

The detection framework we have defined, works on a per-slice basis. Therefore, it is straightforward to use slices as the evaluating unit and propose slice-based definitions of TP, TN, FP and FN. In this way, a true positive is any cross-sectional *slice* that is correctly classified as anomalous, a true negative is any cross-sectional *slice* that is correctly classified as normal. False positives and false negatives are cross-sectional *slices* that are misclassified as anomalous and normal, respectively.

However, although the slice-based definitions allow a precise evaluation, they might be not so relevant from a clinical point of view, where the number of detected *lesions* is more pertinent. Moreover, our goal is to detect lesions so as to call the physician's attention. If several contiguous slices are detected as anomalous, the center of such

Table 3.8: TP, TN, FP and FN definition criteria. With the exception of Reimann et al. (2009), previous works lack of a clear definition of the criteria to define TP's, TN's, FP's and FN's.

	Description
TP	
Arnoldi et al. (2010)	Per-patient and per-vessel measurements. No exact information for TP, TN, FP and FN.
Reimann et al. (2009)	Detected segment with stenosis.
Tessmann et al. (2008)	Calcification detected as calcification. Soft plaque detected as soft plaque.
Tessmann et al. (2009)	*Vessel-based:* At least one lesion identified in the vessel. *Lesion-based:* Cylinder with calcification detected as calcification. Cylinder with soft plaque detected as soft plaque.
TN	
Arnoldi et al. (2010)	Not stated.
Reimann et al. (2009)	Detected healthy segment.
Tessmann et al. (2008)	Not stated.
Tessmann et al. (2009)	*Vessel-based:* Not stated. *Lesion-based:* Not stated.
FP	
Arnoldi et al. (2010)	Not stated.
Reimann et al. (2009)	Healthy segment marked as stenosed.
Tessmann et al. (2008)	Not stated.
Tessmann et al. (2009)	*Vessel-based:* Healthy vessel with at least one detected lesion. *Lesion-based:* Not stated.
FN	
Arnoldi et al. (2010)	Not stated.
Reimann et al. (2009)	Stenosed segment marked as healthy.
Tessmann et al. (2008)	Not stated.
Tessmann et al. (2009)	*Vessel-based:* Diseased vessel marked as healthy. *Lesion-based:* Not stated.

a group of slices will be highlighted to call the physician's attention. Under such a condition, a single detected slice suffices to state that a vascular lesion has been detected. Based on these observations, we introduce lesion-based definitions (Fig. 3.2):

True Positive. A number k of contiguously detected anomalous slices that intersect with at least $k-2$ contiguous slices labeled as anomalous are considered as **one** true positive. Here, we consider that one excess slice detected as anomalous on either side of the actual lesion, belong to the true positive group. Indeed, such excess detection(s) do(es) not prevent from calling attention to the actual lesion. Furthermore, interpretation of the transitions between normal slices and lesions is often ambiguous even for the experts.

True Negatives. A number k of contiguously detected normal slices that intersect with k contiguous slices labeled as normal are considered as **one** true negative.

False Positives. A number k of contiguously detected anomalous slices that intersect with k contiguous slices labeled as normal are considered as **one** false positive. As mentioned above, we exclude from false positives the case of a slice detected as anomalous, intersecting with a slice labeled as normal, but adjacent to a correctly detected anomalous group.

False Negatives. A number k of contiguously detected normal slices that intersect with k contiguous slices labeled as anomalous are considered as **one** false negative.

Throughout the remainder of this chapter we will provide the measurements using both definitions: slice- and lesion-based.

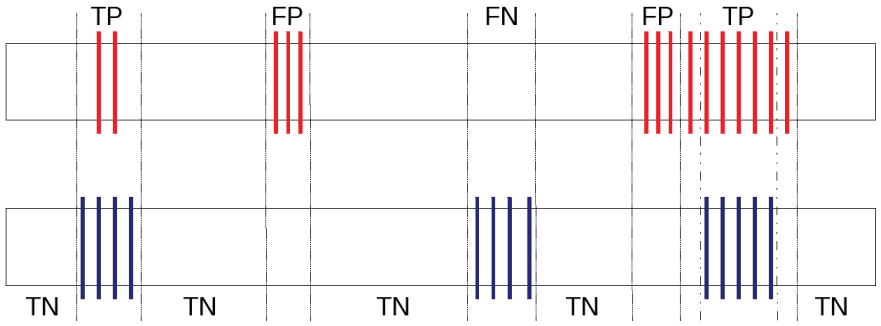


Figure 3.2: Illustration of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) - The reference anomalous slices are depicted as blue lines while the anomalous slices detected by the algorithm are depicted in red. The rectangles represent the vessels and the white region inside them signifies normal slices.

Performance measures

Specificity represents the fraction of negative samples that are correctly classified, *i.e.* the fraction of normal slices that are correctly identified as normal:

$$specificity = \frac{TN}{TN + FP}. \tag{3.3}$$

Sensitivity represents the fractions of positive samples that are correctly classified, *i.e.* slices containing vascular lesions that are correctly identified as anomalous:

$$sensitivity = \frac{TP}{TP + FN}. \tag{3.4}$$

Finally, accuracy evaluates the overall performance of the classifier by measuring the fraction of correctly classified samples. An accuracy of 100% means that the classifier has correctly labeled all the given samples. It is expressed as

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}. \tag{3.5}$$

Anomaly detection problems are highly unbalanced. This means that the sample size corresponding to one class is significantly higher than the size of the second class. Typically, the size of the second class instances is represents 10% or less of the size of the first class instances. In such conditions, accuracy is not an appropriate measure. As an example, assume a test data consisting of 50 positive instances (anomalies) and 1000 negative ones. Suppose also that only one positive sample was correctly labeled ($TP = 1$, $FN = 49$),while all the negative were correctly classified ($TN = 1000$, $FP = 0$). The reported accuracy for such a system is of 95.33% which is high although the classifier is almost incapable of detecting anomalies.

To avoid the problems associated with accuracy in unbalanced data, we propose to use an alternative metric called balanced error rate (BER). BER is the average of the error rates of the positive and negative classes. A BER of 0% means that all the evaluated instances were correctly classified. It is defined as:

$$\begin{aligned} BER &= 1 - \frac{sensitivity + specificity}{2} \\ &= 1 - \frac{1}{2}\left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP}\right). \end{aligned} \tag{3.6}$$

When expressed in terms of the BER, the previous example scores 49% which better reflects the poor performance of the classifier. Since our goal is to highlight possible lesions for a clinician, one can argue that we should focus only in maximizing sensitivity in our system. However, we consider that high sensibilities in combination with low specificties are not very useful for the clinician, since a significant amount of time is then necessary to review all the detected locations and remove a high number of FPs. This is why we prefer to focus on the BER, which searches for the balance between specificity and sensibility.

The BER expresses the performance of a classifier as a function of the error, while accuracy measures the performance in terms of the success rate. The Balanced Classification Rate (BCR) is a measurement closely related to the BER that measures the performance of a classifier in terms of the success rate. Therefore, it can be directly compared with accuracy

$$\begin{aligned} BCR &= \frac{sensitivity + specificity}{2} \\ &= \frac{1}{2}\left(\frac{TP}{TP+FN} + \frac{TN}{TN+FP}\right) \\ &= 1 - BER. \end{aligned} \tag{3.7}$$

Again, the classification example scores 51% using BCR. This is more representative, in terms of success rate, than the score obtained using accuracy.

Since BER and BCR are complementary, we will limit ourselves to use the BER to evaluate performance. Additionally, we will make use of specificity and sensitivity since they provide detailed information of how positive and negative samples are classified. Other well known performance measures, based on TP, FP, TN and FN, might be used, such as Dice similarity index:

$$D = \frac{2TP}{(TP+FP)+(TP+FN)},$$

Jaccard index:

$$J = \frac{TP}{(TP+FP+FN)} = \frac{D}{2-D}$$

or Matthews correlation coefficient:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}},$$

which also provide convenient evaluation in unbalanced cases. However, the former two are rather used in other fields and are rarely mentioned in binary classification literature, while the meaning of the latter is less intuitive. Finally, let us recall that the empirical risk $\mathcal{R}$ (Eq. 2.5) represents a natural way to evaluate the performance of our classifiers. However, we avoid its use since state-of-the-art methods in our particular domain do not use it as a performance measure.

The agreement between the data labeled by two observers and our methods was assessed using two different measures. The first one was the Cohen's kappa coefficient measure Cohen (1960) defined as:

$$\kappa = \frac{Pr(a) - Pr(e)}{1 - Pr(e)}, \tag{3.8}$$

where $Pr(a)$ is the relative observed agreement between raters (in our case, a rater is either an observer or one of the methods), and $Pr(e)$ is the hypothetical probability

of chance agreement, using the observed data to calculate the probabilities of each observer randomly saying each category.

Cohen's kappa coefficient allows to separately measure the agreement in anomaly detection between the methods and each of the observers, as well as the agreement between the two observers and the two methods. To simultaneously assess the agreement between the two observers and each method, Fleiss' kappa (Fleiss, 1971) was used. Fleiss' kappa is an extension of Cohen's kappa, which calculates the agreement degree in classification when more than two raters are present.

3.2 Experimental Design and Evaluation of the Concentric Rings Metric

This section presents a series of experiments on synthetic and real 3D data using `Concentric rings` in combination with DLD-SVM and LPU classification schemes. Synthetic phantoms are used to evaluate the power of the metric in a controlled environment and also to tune up the parameters associated to the metric (Section 3.2.1). The tuned up metric is then evaluated in a real environment using clinical data (Section 3.2.2).

3.2.1 Evaluation on Synthetic Data

Four different parameters need to be tuned-up in the implementation of the `Concentric rings` metric. These are the number of radial and angular ranges, N_r and N_θ, the radius of the evaluated region r_{max} (see Section 2.4.1), as well as the spatial sampling rate for the computation of each feature $F(r, \theta)$.

The r_{max} parameter. r_{max} should be selected so that it covers the vessel cross-section diameter. It is also desirable that it does not fit precisely the vessel diameter, so that it can detect lesions presenting positive remodeling of the vessel wall or aneurysms. Using this information and taking into account the typical ranges of healthy arteries (Funabashi et al., 2003), the maximum radius r_{max} was fixed to 6.5 mm, while r_{min} is always set to zero.

Radial and angular ranges. Experiments to define the number of radial and angular ranges were carried out with a number of features $N = N_\theta \times N_r$ ranging from 24 to 150, which corresponds to 36 different combinations of N_θ and N_r presented in Table 3.9. Both DLD-SVM and LPU approaches were trained using the 36 different combinations and evaluated on the testing datasets (refer to Tables 3.2, 3.4 for details) to determine the best (N_r, N_θ) on each case.

Figure 3.3 shows sensitivity, specificity and BER of anomaly detection using DLD-SVM and LPU for the various combinations of N_θ and N_r used to compute the `Concentric rings` metric. To allow a detailed analysis of the behavior of the methods,

Table 3.9: Parameter values for Concentric rings tuning. Number of radial and angular ranges, N_r and N_θ, subsampling rate, r_{max} and r_{min}.

Parameter	Values
N_r	15, 10, 9, 8, 7, 6
N_θ	10, 9, 8, 6, 5, 4
Subsampling rate	50, 33.33, 25, 20, 10
r_{max}	6.5 mm
r_{min}	0 mm

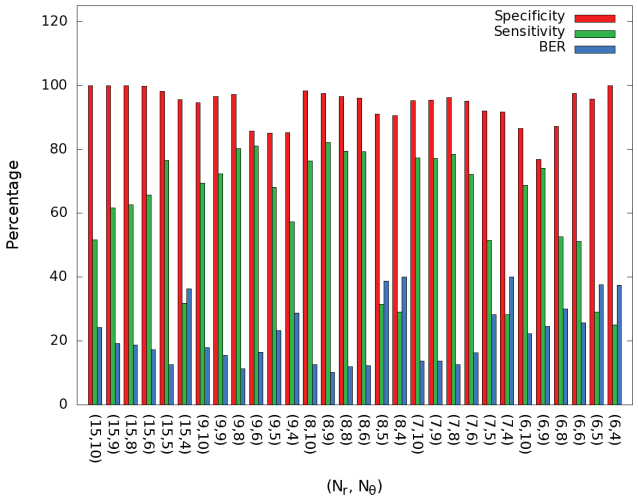
results are shown on a slice-basis. For every pair (N_r, N_θ), the plotted value corresponds to the ρ or t value that gives the lowest empirical risk. The lowest empirical risk also provided the best response in terms of trade-off between specificity, sensitivity and BER.

Performance results for DLD-SVM (Figure 3.3 (a)) illustrate the trade-off between noise removal and information preservation, since when the number of features N is low (integration over large regions) or high (integration over small regions), the results are worse. Additionally, it can be seen that, with a fixed moderate value of N, the configurations with thin rings over wide sectors ($N_r \gg N_\theta$) should be avoided, *e.g.*: with $N = 60$ the configuration $(6, 10)$ performs clearly better than $(15, 4)$.

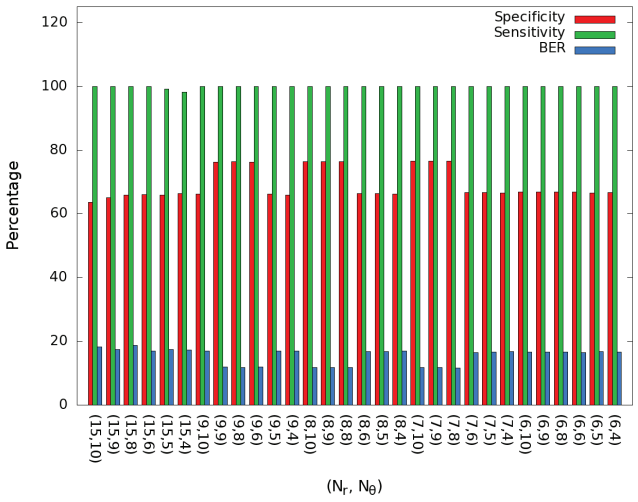
An assessment of the misclassified cross-sections permitted us to identify the principal causes of errors. In general, the DLD-SVM has lower sensitivity than specificity. This is due to the fact that the classification algorithm tends to fail on the outer-most slices of a lesion. At these points, the `Concentric rings` metric not always gives a strong response that would permit the identification of a normal or abnormal pattern. As a consequence, not all the slices of a lesion are detected. Evaluation on a lesion-basis confirms the behavior since the sensitivity increases between 2.1 and 10.3% in the evaluated pairs (N_r, N_θ).

Although specificity presents high scores, from Figure 3.3 (a) it can be seen that even for the (N_r, N_θ) pairs with the best response, it does not score 100%. An inspection of the results allowed us to determine that the DLD-SVM algorithm always classifies bifurcations as anomalies. Figure 3.4 (a) illustrates the problem by plotting two features obtained from the `Concentric rings` metric against each other. The plotted data includes slices from normal vessels containing bifurcations. Lesions are excluded. However, it can be seen that there are some points that strongly deviate from the cloud made up by the majority of the points. The most outlying points were identified as bifurcations. Due to the deviation from a normal symmetric pattern, bifurcations are also detected as anomalies.

Results obtained from LPU for different (N_r, N_θ) pairs are depicted in Figure 3.3 (b). Opposite to what occurs with DLD-SVM, the performance of LPU is more robust to variations in the number of regions $N_r \times N_\theta$. Nevertheless, the range of best performing pairs (N_r, N_θ) coincides with the best performing pairs from DLD-SVM. The pair $(N_r = 8, N_\theta = 9)$ was selected for all the subsequent experiments. Although this is not

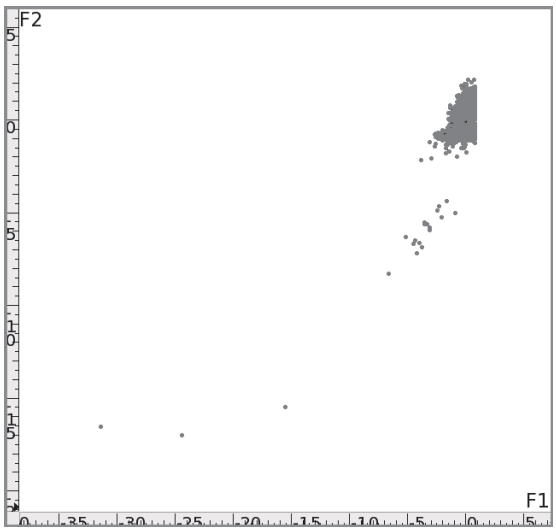


(a)

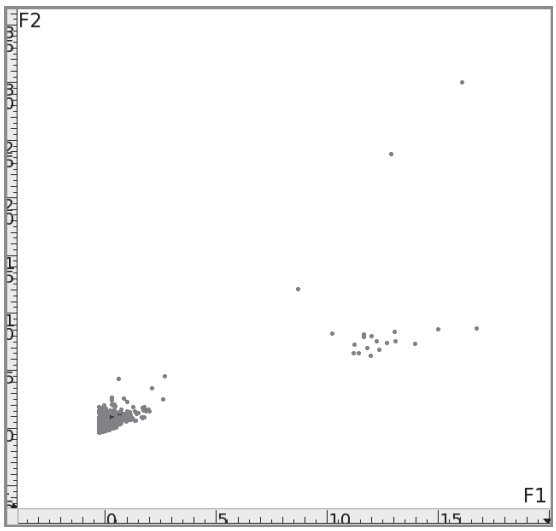


(b)

Figure 3.3: Specificity, sensitivity and BER with varying (N_r, N_θ) - (a) DLD-SVM (b) LPU. Results are shown on a slice basis.



(a)



(b)

Figure 3.4: Illustration of bifurcation misdetection problem in DLD-SVM - Bifurcation samples are seen as outlying points with respect to the vast majority of concentrated points. As an example, the behavior is reflected by (a) plotting two features from the Concentric rings metric and (b) plotting of λ_1 vs. λ_2 parameters from the Hessian metric at a given scale.

Table 3.10: Ring selection summary. Worst and best pairs (N_r, N_θ) according to BER in both DLD-SVM and LPU. Results are presented on a slice basis.

Classifier	Minimum				Maximum			
	Spec.	Sens.	BER	(N_r, N_θ)	Spec.	Sens.	BER	(N_r, N_θ)
DLD-SVM	95.61	31.81	36.29	(15,4)	97.57	82.19	10.12	(8,9)
LPU	65.96	100	18.69	(15,10)	76.33	100	11.83	(8,10)

the best performing configuration for LPU it has the second best performance in terms of the BER. The respective slice-based specificities in DLD-SVM and LPU were 97.57 and 76.34%, sensitivities 82.19 and 100%, and a BER 10.12 and 11.84%. Table 3.10 summarizes the best and worst pairs (N_r, N_θ) in both DLD-SVM and LPU.

The selected pair reflects that the best pairs performances were obtained with intermediate values of $N = N_r \times N_\theta$, such that $N_r \approx N_\theta$: $N = 72$ with (N_r, N_θ)=$(8, 9)$ performed the best. This illustrates the trade-off between noise removal and information preservation, since when the number of features N is low (integration over large regions) or high (integration over small regions), the results are worse. Additionally, it can be seen that, with a fixed moderate value of N, the configurations with thin rings over wide sectors ($N_r \gg N_\theta$) should be avoided, *e.g.*: with $N = 60$ the configuration $(6, 10)$ performs better than $(15, 4)$ in both DLD-SVM and LPU.

It is interesting to compare the behaviors of DLD-SVM and LPU. Contrarily to what occurs in DLD-SVM, under LPU sensitivity outperforms specificity. In the labeled results, this is reflected by overestimated lesions (in terms of the number of detected slices). Although LPU succeeds in detecting all existing lesions (in both slice and lesion-based schemes), it tends to mark slices immediately adjacent to a lesion as anomalous. The latter diminishes the specificity.

A possible explanation for this behavior comes from the semi-supervised nature of the LPU algorithm. Since it tries to identify normality by some provided examples (the labels), slightly diverging normal cases can be excluded. As an example, slices adjacent to lesions typically suffer from variations, *e.g.* intensity, that can cause the misclassification. Our proposal for LPU with progressive increase of the training set, tries to solve some of this issues (see Section 2.3.4).

Subsample rate. Sampling inside the "tiles" formed by the `Concentric rings` metric is performed at subvoxel resolution using tri-linear interpolation. In order to define the sampling rate, we carried out experiments to determine the influence of the selected sampling rate.

Assessing the classification performance as a function of the sampling rate is expensive, while the influence of the parameter is not as significant as the one of the radial and angular ranges. Hence, instead of using the classifiers (DLD-SVM and LPU) to evaluate the sampling rate effect, we have chosen to use two variable ranking measures. The first one is the F-score, which was described in section 2.5.2. The use of synthetic phantoms, where the ground truth is known, allows the usage of a supervised metric

Table 3.11: Classifiers performance as a function of the sampling rate. BER performance of DLD-SVM and LPU for the pair ($N_r = 8$ and $N_\theta = 9$) with varying sampling rate. Only the sampling rates with the best scores were evaluated. Results are presented on a slice basis.

Classification method	25%	20%	10%
DLD-SVM	11.313	11.310	11.310
LPU	11.842	11.842	11.843

such as F-score. However, since our goal is to reduce the dependence on labels in the design and implementation of a vascular classification framework, we have chosen to use also an unsupervised metric. The selected metric is entropy

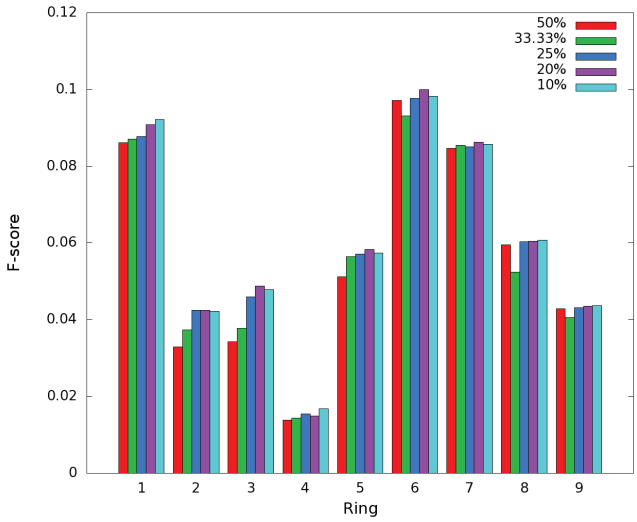
Equations 2.30 (F-score), 2.33 (entropy) were applied to each of the features obtained for a fixed pair ($N_r = 8, N_\theta = 9$) with a varying sampling rate. Table 3.9 presents the evaluated sampling rates. They are all expressed as a percentage of the image intra-slice resolution. Figure 3.5 depicts the obtained scores.

Results show that the scores obtained through entropy and F-score are coherent. In both cases, using a sampling rate between 25 and 10% provides the best scores. To confirm it, we trained DLD-SVM and LPU classifiers using features obtained with the three sample rates and a fixed pair ($N_r = 8, N_\theta = 9$), and assessed their performance. Results, in terms of the BER, demonstrated that classification is robust to the sampling rate parameter. For both classifiers, differences in the BER, appear only at the third decimal digit (Table 3.11). Based on the results, a sampling rate of 20% of the intra-slice resolution was used for all the experiments. It has a good performance and it is computationally less expensive than 10% rate.

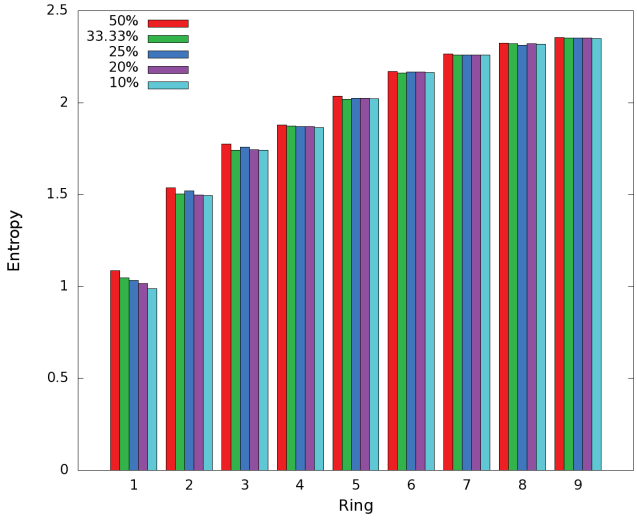
3.2.2 Evaluation on Patients' Data

In the following, we evaluate the performance of DLD-SVM and LPU classifiers, while using the tuned up `COncentric rings` metric. The results obtained when comparing the labels assigned by our method, with respect to each of the observers, are summarized in Tables 3.12 and 3.13. Minimal, maximal and average values of specificity, sensitivity and BER are presented. The values are consistent with the results obtained on synthetic data. Moreover, the behavior of DLD-SVM and LPU remain constant. LPU tends to have a higher sensitivity, at the cost of increasing the number of false positives, while DLD-SVM is more conservative. The main consequence of this behavior is that LPU has lower values of specificity and BER.

Table 3.14 presents the kappa coefficient obtained when comparing the proposed methods with the annotations from two different observers, as well as the comparison between both observers. Although not universally accepted, Landis and Koch (1977) supply a table for comparing the κ coefficients. According to this table, agreement in all cases is considered substantial ($\kappa \in [0.61 - 0.80]$). These results are also consistent with the evaluation of specificity, sensitivity and BER using the observer's annotations as ground truth: our two methods performed better with respect to observer 1 than



(a)



(b)

Figure 3.5: Optimal sampling rate definition - Scores using (a) F-score and (b) entropy are averaged on a ring basis. In both cases, the best responses are obtained between 25 and 10%.

Table 3.12: Performance of Concentric rings and DLD-SVM over real patient data. These values were obtained on the testing subset (Tab. 3.2) of the cardiac set 1, by using observers' annotations as ground truth.

Observer	Specificity			Sensitivity			BER		
	Min.	Max.	Avg.	Min.	Max.	Avg.	Min.	Max.	Avg.
Slice-based									
1	78.22	94.44	87.56	66.10	87.56	83.14	9.00	27.84	14.65
2	69.60	88.43	84.45	49.84	83.67	79.31	13.85	40.28	18.12
Lesion-based									
1	71.45	93.82	72.60	71.33	98.35	91.17	8.15	28.71	18.08
2	65.96	82.12	72.10	60.03	96.72	88.64	11.11	36.15	19.63

Table 3.13: Performance of Concentric rings and LPU over real patient data. Results are obtained by using observer's annotations over a percentage of cardiac set 1 as ground truth.

Observer	Specificity			Sensitivity			BER		
	Min.	Max.	Avg.	Min.	Max.	Avg.	Min.	Max.	Avg.
Slice-based									
1	50.00	79.51	67.21	88.88	100	90.00	15.42	27.71	22.39
2	50.00	75.24	66.56	88.40	100	89.96	16.80	28.09	23.10
Lesion-based									
1	50.00	74.7	55.46	100	100	100	12.65	25.00	22.27
2	50.00	70.12	54.27	100	100	100	14.94	25.00	22.86

with respect to observer 2.

3.3 Other Candidate Features Tuning

This section describes the tuning performed on other candidate features (section 2.4.2). The goal is to parametrize the features such that they are directly comparable to the `Concentric rings`. First, we describe the configuration of the `Steerable features` (sec. 3.3.1) followed by the configuration of the remaining features (sec. 3.3.2). Table 3.16 summarizes all the parameters of every metric.

3.3.1 Steerable Features

`Steerable features` (Zheng et al., 2007) require the definition of two parameters: the sample pattern to be followed and the sample rate at which points will be sampled along the pattern. The `Concentric rings` metric was designed under the assumption that normal vessels are nearly symmetric. For this reason, the metric follows a circular pattern in the extraction of its features. To remain coherent, we use a circular pattern

Table 3.14: Kappa coefficient. Cohen's in the first two columns and Fleiss' in the last one. Cohen's Kappa coefficient between observers 1 and 2 is of 0.79. Cohen's Kappa coefficient between DLD-SVM and LPU was of 0.72.

	Observer 1 vs. Method	Observer 2 vs. Method	All
DLD-SVM	0.72	0.69	0.72
LPU	0.68	0.69	0.69

to sample the steerable features, as well as the same circle size ($r_{min} = 0, r_{max} = 6.5$ mm for coronary arteries).

The sampling rate required a more careful evaluation. Tessmann et al. (2009) propose to sample points following a circular pattern with radius 4 mm. However, their approach differs from ours, since the sampling is done over a cylinder, so it is not directly applicable to our case. Thus, to determine a suitable sampling rate we have used a methodology similar to the one used to tune the `Concentric rings` metric. We have performed an initial sampling using a very fine resolution, *i.e.* 1041 sampling points for the 20 measures. This represented a total of 20820 values. In order to evaluate the discriminative power of each feature, when varying the number of sample points, the F-score was used. The adopted strategy consisted in sub-sampling the spatial locations at which the different measures were evaluated. An incremental sampling rate $\in [2, 3, 4, 10, 15, 20]$ was applied, while evaluating the F-score on each new subset to guarantee that the discriminative power was not lost. Results showed that, even for a high sub-sampling rate, the discriminative power of these features was not significantly altered (Figures 3.6 and 3.7). Setting the sub-sampling rate to 20, the number of sampling points was reduced to 50 per feature, which represents a total of 1041 values.

Although 1041 values per sample can be suitable for evaluating DLD-SVM and LPU approaches, this number can still be too high when the amount of data is limited. Therefore, we evaluated the relevance of each of the features that build up `Steerable features` in order to select the most promising ones. For this purpose, we also made use of the F-score, since it is a metric that allows fast evaluation. Furthermore, we also evaluated the features using entropy by assuming no labels where available. Figures 3.8 and 3.9 summarize the F-score and entropy behavior for each feature.

The plots show that there is some coherence in the results obtained with each of the measures and both F-score and entropy lead to similar feature ranking lists (Table 3.15). The ranking obtained for zero order features is the same with both measures, while in the higher order features both measures agree at least in the worse ranked features.

Using the classification from Table 3.15, twelve measures were selected from steerable features: $\log I$, $\sqrt{I}$, I_x, I_y, $\sqrt{|\nabla I|}$, $\log|\nabla I|$, I_{xx}, I_{yy}, I_{xy}, $|\mathbf{H}|_F$, $\log|\mathbf{H}|_F$ and $\sqrt{|\mathbf{H}|_F}$. We chose to select the best ranked features from each order (zero, first and second) instead of selecting the overall best ranked features, so as to keep information of different nature. If the overall ranking was used, second order features would have been rejected.

Let us recall that F-score and entropy do not measure the same criterion over data,

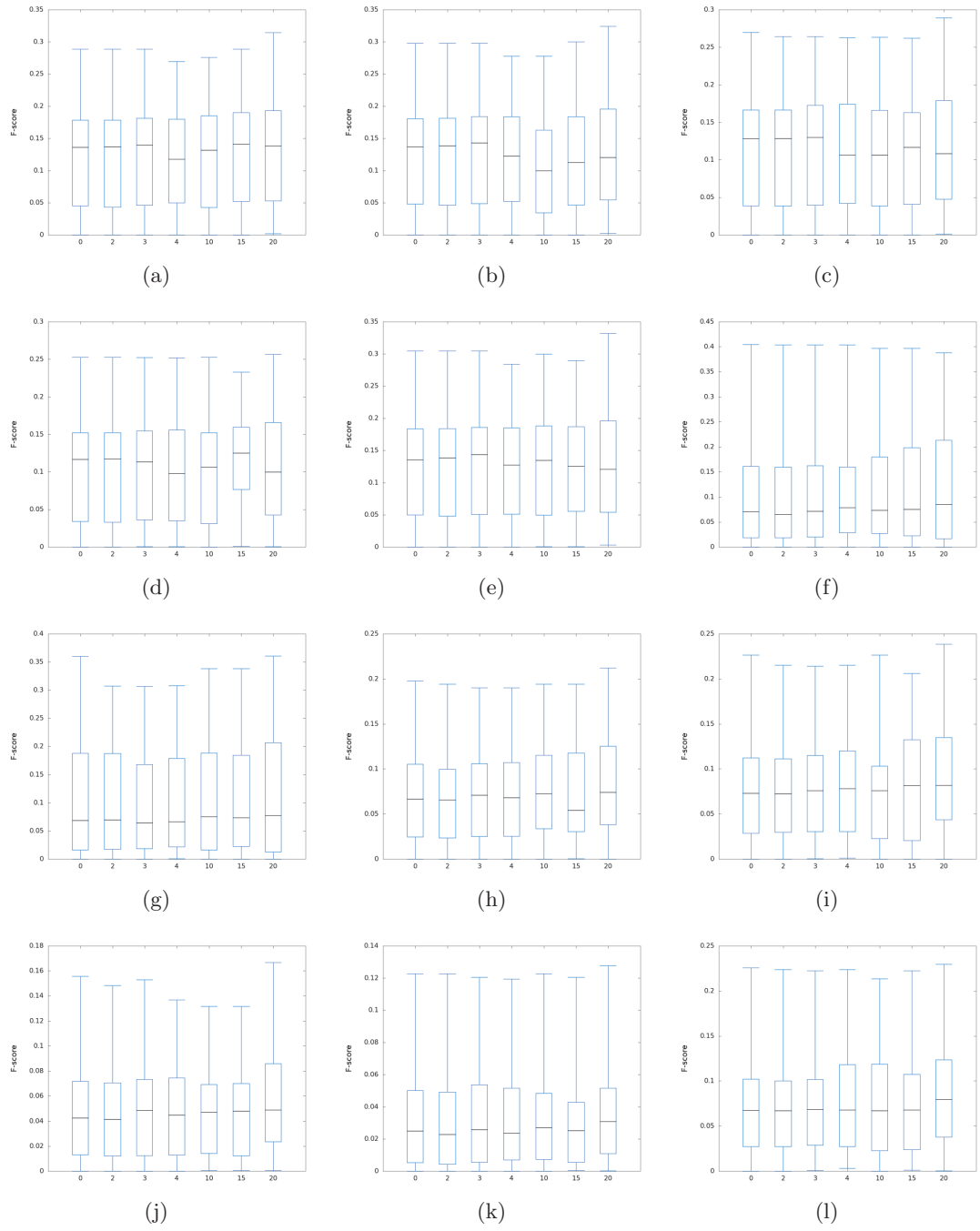


Figure 3.6: F-score as function of spatial locations sub-sampling in zero and first order steerable features. (a) I, (b) $\sqrt{I}$, (c) I^2, (d) I^3, (e) $\log I$, (f) I_x, (g) I_y, (h) $|\nabla I|$, (i) $\sqrt{|\nabla I|}$, (j) $|\nabla I|^2$, (k) $|\nabla I|^3$, (l) $\log |\nabla I|$.

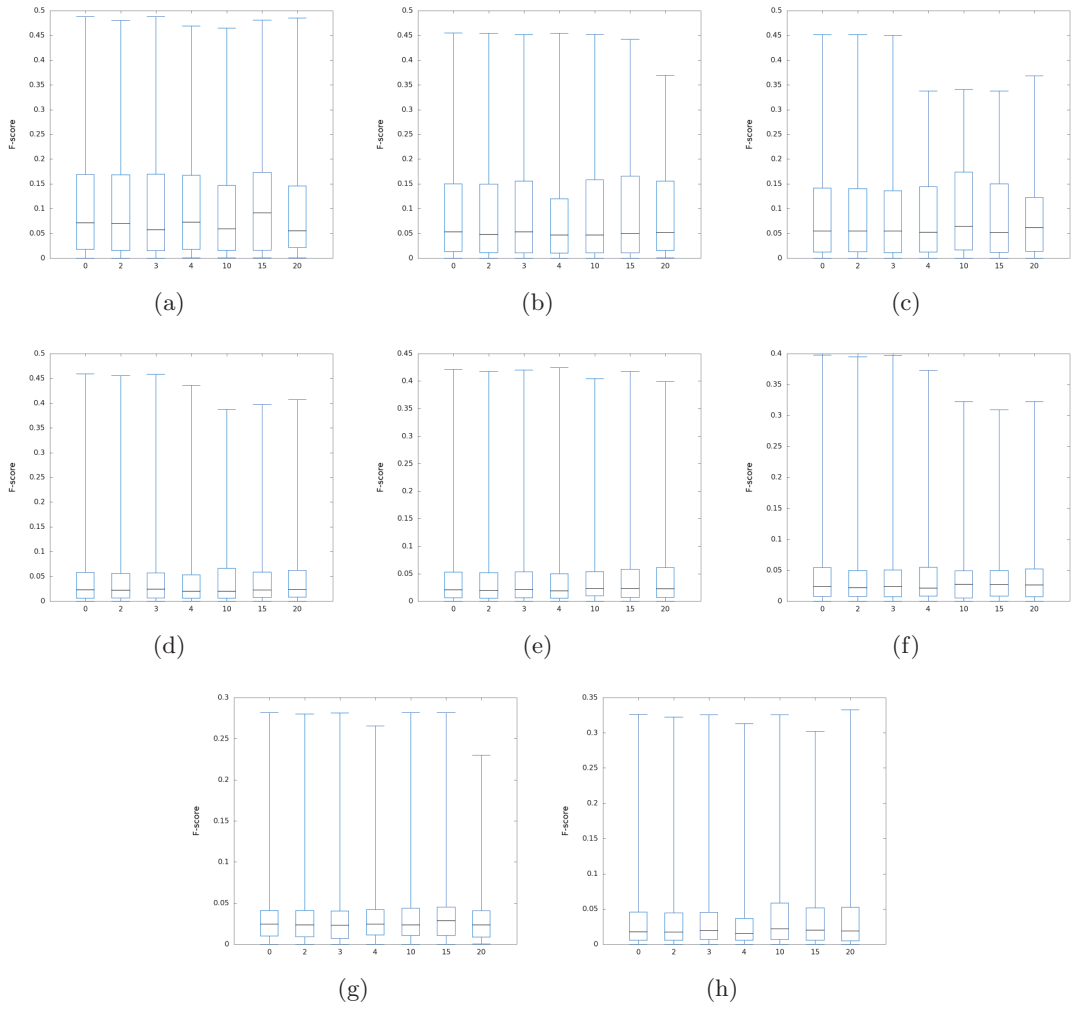


Figure 3.7: F-score as function of spatial locations sub-sampling in second order steerable features. (a) I_{xx}, (b) I_{yy}, (c) I_{xy}, (d) $|\mathbf{H}|_F$, (e) $\sqrt{|\mathbf{H}|_F}$, (f) $|\mathbf{H}|_F^2$, (g) $|\mathbf{H}|_F^3$ and (h) $\log|\mathbf{H}|_F$.

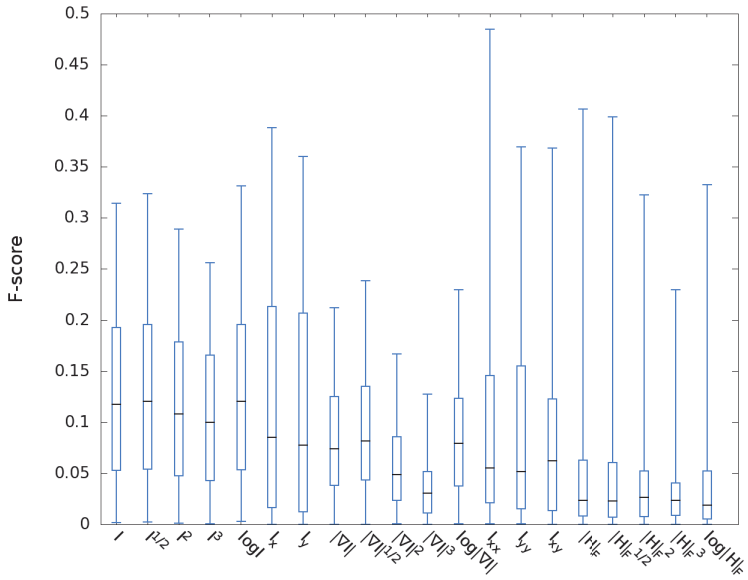


Figure 3.8: F-score summary for steerable features - Features of order zero (I), one (I_x, I_y and $|\nabla I|$) and two (I_{xx}, I_{yy}, I_{xy} and $|\mathbf{H}|_F$) are presented.

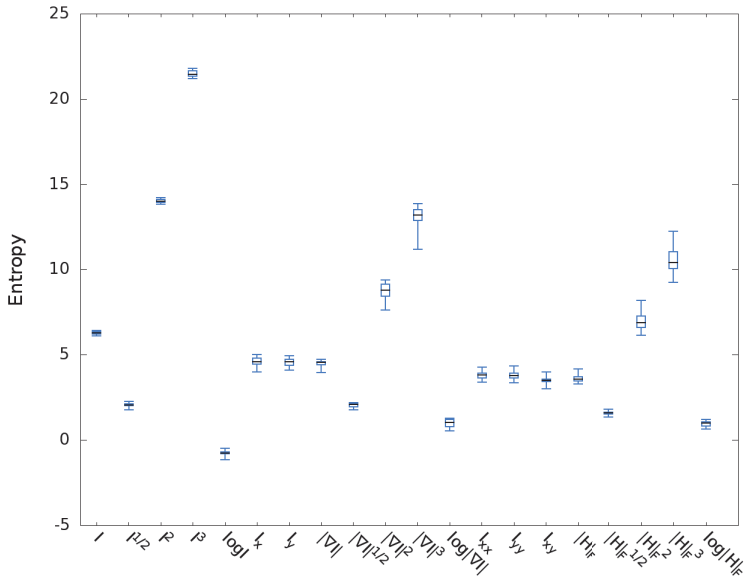


Figure 3.9: Entropy summary for steerable features - Features of order zero (I), one (I_x, I_y and $|\nabla I|$) and two (I_{xx}, I_{yy}, I_{xy} and $||\mathbf{H}||_F$) are presented. Opposite to F-score, best ranked features when using entropy are the ones with the lowest entropy.

Table 3.15: Steerable features ranking using F-score and entropy. The ranking is obtained exclusively by means of the scores obtained from F-score and entropy. No SVM training is performed.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
0 order								
F-score	$\log I$	$\sqrt{I}$	I	I^2	I^3	-	-	-
Entropy	$\log I$	$\sqrt{I}$	I	I^2	I^3	-	-	-
1^{st} order								
F-score	I_x	I_y	$\sqrt{\lvert\nabla I\rvert}$	$\lvert\nabla I\rvert$	$\log\lvert\nabla I\rvert$	$\lvert\nabla I\rvert^2$	$\lvert\nabla I\rvert^3$	-
Entropy	$\log\lvert\nabla I\rvert$	$\sqrt{\lvert\nabla I\rvert}$	I_x	I_y	$\lvert\nabla I\rvert$	$\lvert\nabla I\rvert^2$	$\lvert\nabla I\rvert^3$	-
2^{nd} order								
F-score	I_{yy}	I_{xx}	I_{xy}	$\lvert\mathbf{H}\rvert_F$	$\sqrt{\lvert\mathbf{H}\rvert_F}$	$\log\lvert\mathbf{H}\rvert_F$	$\lvert\mathbf{H}\rvert_F^2$	$\lvert\mathbf{H}\rvert_F^3$
Entropy	$\log\lvert\mathbf{H}\rvert_F$	$\sqrt{\lvert\mathbf{H}\rvert_F}$	$\lvert\mathbf{H}\rvert_F$	I_{xy}	I_{yy}	I_{xx}	$\lvert\mathbf{H}\rvert_F^2$	$\lvert\mathbf{H}\rvert_F^3$

the former being supervised, while the late is not. Therefore, although some coherence is expected, the ranking is not exactly the same. Hence, the selection was done by averaging the ranking position obtained by use of entropy and F-score. Guyon and Elisseeff (2003) suggest the inclusion of noisy variables and reject all the features that are ranked below them. However, in our experiments, the noisy variables were always ranked in the last positions. The selected subset of steerable features was used in cases where the available training data was reduced (such as in feature selection).

3.3.2 Other Features

Hessian eigenvalues. In our experiments, we are interested in observing the behavior of the λ_i values of the Hessian matrix at different scales. Typically, the scale is taken to be the standard deviation of the Gaussian function σ. We associate the evaluated scales to the evaluated radii from `Concentric rings`. Thus, 8 different scales are evaluated and we define the maximum scale to be the $r_{max} = 6.5$ mm parameter from `Concentric rings`. The smallest scale evaluated was 1 mm.

Flux and MFlux. For the sake of coherence, `Flux` and `MFlux` were evaluated at the same number of radii as Hessian eigenvalues. These correspond to the number of rings from the `Concentric rings` metric. The maximum radius is taken to be $r_{max} = 6.5$ mm and the minimal 1 mm. The remaining parameters values were fixed following the seminal proposal from Lesage (2009). These are: number of equi-angular contour points $N = 8$, and Gaussian standard deviation σ of the order of the image resolution used to regularize the gradient field. Concerning exclusively `MFlux`, Lesage (2009) proposes conservative threshold values $T_{low} = -24$ HU and $T_{high} = 576$ HU comprising contrast-enhanced arterial blood, which we have kept.

Core. Similarly to `Flux` and `MFlux`, $N = 8$ points per contour, and 9 different radial extents were evaluated for `Core`, respecting the restriction $r_{max} = 6.5$ mm. The σ_r parameter (see Eq. 2.18) was fixed following the recommendations from Fridman (2004): $\sigma_r = \frac{r}{4}$, where r denotes the evaluated radius.

Ribbon. The radius of the external ring r_{ext} followed the same range of values as all the other metrics (with $r_{max} = 6.5$ mm). The values for r_{int} are obtained from the relationship that holds between the two radii, $r_{ext} = \sqrt{2}r_{int}$. Sampling inside regions A_{int} and B_{ext} was chosen to be the same as the one used for the `Concentric rings` metric (20% voxel intra-slice resolution).

Ball measure. The thresholds required for the Ball Measure were fixed using the same values defined for `MFlux`. However, initial experimentations showed a very weak response of the metric, when using these thresholds. Empirically, the thresholds were re-fixed to $T_{low} = 126$ HU and $T_{high} = 676$ HU. Radii values for the ball were the same as the ones used for all the other metrics.

Inertia moments. The only parameter that needs to be fixed for inertia moments is the size of the region of interest (ROI) where the moments are computed. We used a circular ROI, centered at the vessel axis, with radius varying within the same range as in all the other metrics.

3.4 Discussion

In this chapter we have presented the experimental design of the metric we have formulated, `Concentric rings`, as an alternative input for the anomaly detection problem solved using DLD-SVM and LPU. The experimental design has focused in determining the three main parameters that `Concentric rings` requires. These are the maximum radius, r_{max}, the radial and angular ranges and the sampling rate. Experiments were performed on synthetic phantoms that provided a controlled environment.

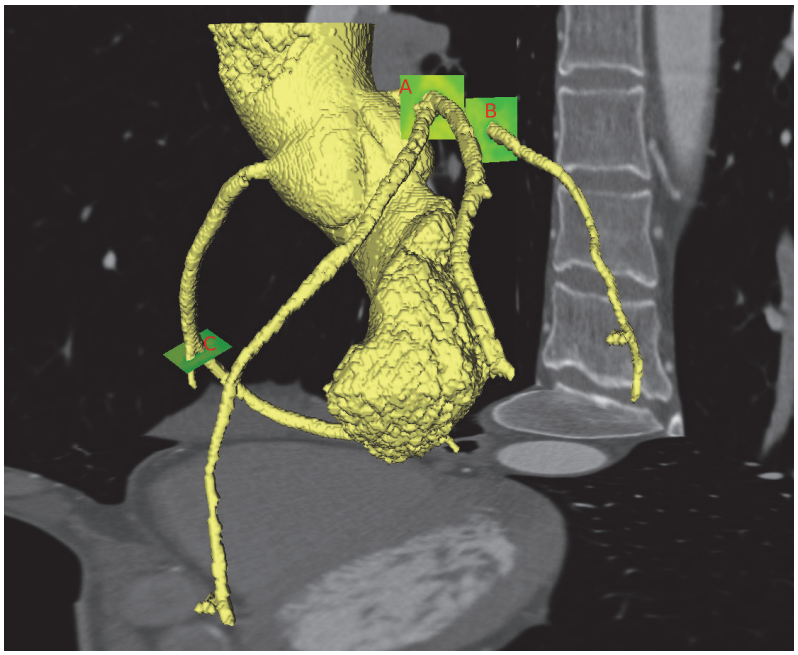
An evaluation of DLD-SVM and LPU, using `Concentric rings`, on patients' data permitted us to demonstrate the feasibility of both our selected classifiers and the designed metric to solve our particular problem. Moreover, the results permitted us to define a behavior pattern for DLD-SVM and LPU. DLD-SVM is a conservative approach where lesions tend to be underestimated. Nevertheless, it has the disadvantage of labeling bifurcations as anomalous. This was however expected, since bifurcations deviate from the normal cross-sectional patterns.

Although LPU does not fail in bifurcations, provided that normal samples of bifurcations are included in the Q set, it tends to overestimate the lesions. We consider that one of the reasons for the overestimation of lesions in LPU comes from a possible lack of sufficiently representative set of normality (the Q set). However, we believe that this effect can be diminished with our LPU with progressive increase of the training set formulation. The latter will be evaluated in the subsequent chapter.

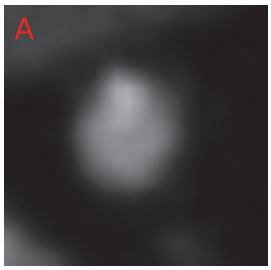
Table 3.16: Summary of other candidate metrics tuning. Parameter values used in the extract features from the different candidate metrics. Parameters common to all metrics are grouped for simplification.

Parameter	Value
Common parameters	
Number of scales/radii	8
Minimum Radius	1 mm
Maximum radius	6.5 mm
Steerable features	
Number of samples per slice	52
Flux, MFlux and Cores	
N	8
MFlux	
T_{low}	-24 HU
T_{high}	576 HU
Cores	
σ_r	$\frac{r}{4}$
Ribbon	
r_{int}	$\frac{r_{ext}}{\sqrt{2}}$
Sampling rate	20% of intra-slice resolution
Ball measure	
T_{low}	126 HU
T_{high}	676 HU

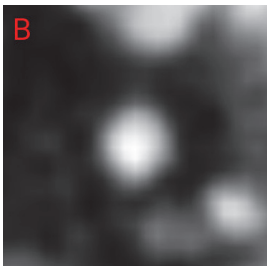
Once again, we should recall that the goal of our proposal is to call the physician's attention by marking potential lesions. Figure 3.10 shows an example of the obtained results in a real dataset, how we foresee it as a clinical application. In Figures 3.11 and 3.12 several examples of labeled vessels using both DLD-SVM and LPU are presented.



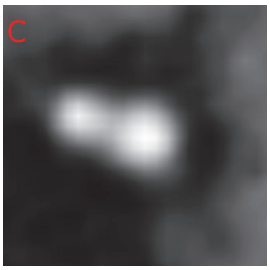
(a)



(b)



(c)



(d)

Figure 3.10: Examples of anomaly detection results using Concentric rings. (a) 3D view of the coronary tree showing the locations of three slices: A containing a calcification, B normal and C containing a bifurcation. Slice A was correctly detected as anomalous by both LPU and DLD-SVM, while slice B was correctly classified as normal. In slice C, LPU correctly detected the bifurcation as normal, and DLD-SVM marks it as anomalous. (b), (c) and (d) respectively represent these three cross-sections.

The final part of this chapter has been devoted to the parametrization of the remaining sets of candidate metrics. We have determined them following the guidelines provided by the parameters defined for the `Concentric rings` metric, so that a fair

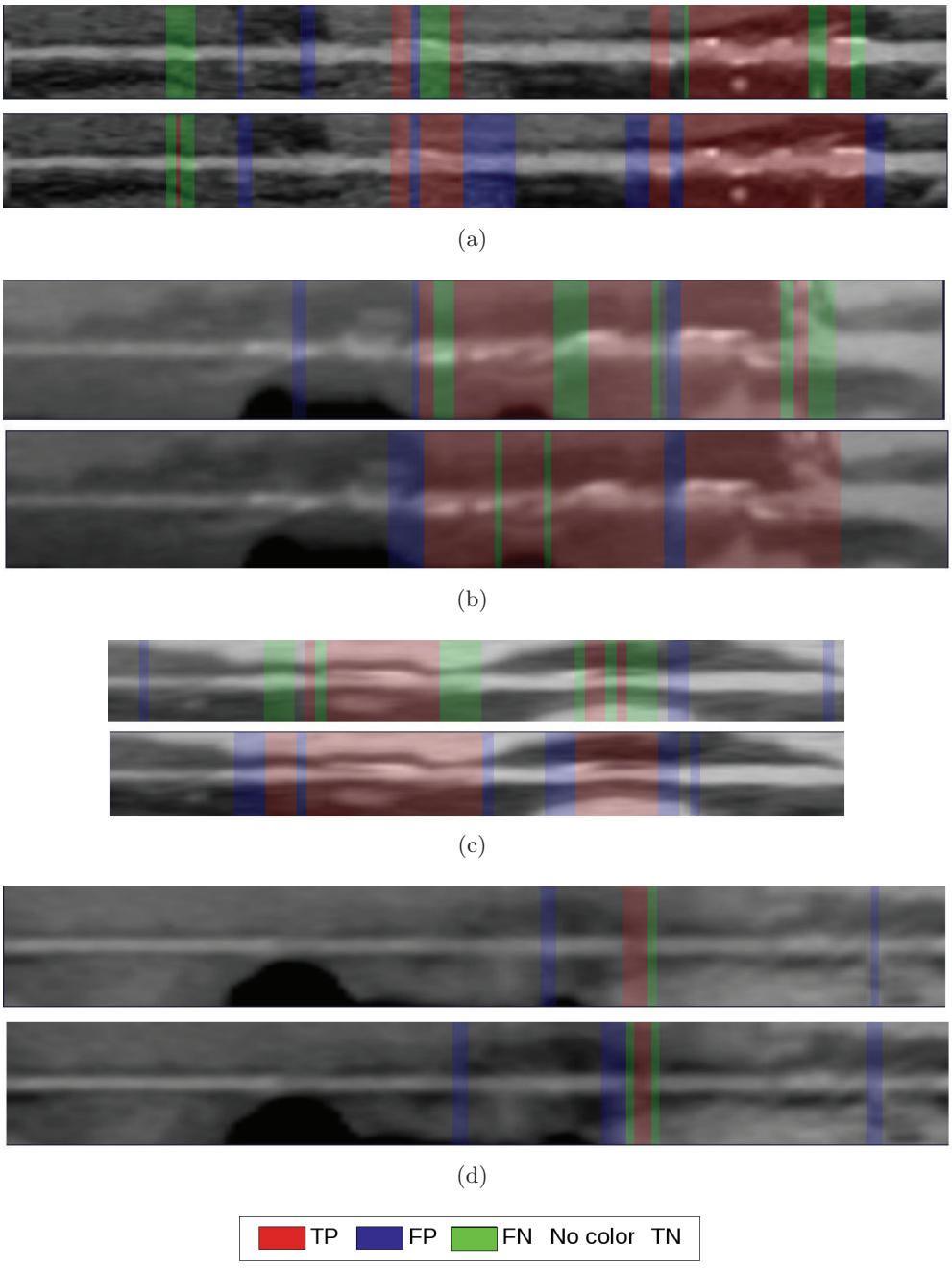


Figure 3.11: Comparison of DLD-SVM and LPU - Each sub image contains on top the result from DLD-SVM and on bottom the result from LPU. The DLD-SVM algorithm tends to under estimate anomalies, whereas LPU overestimates them.

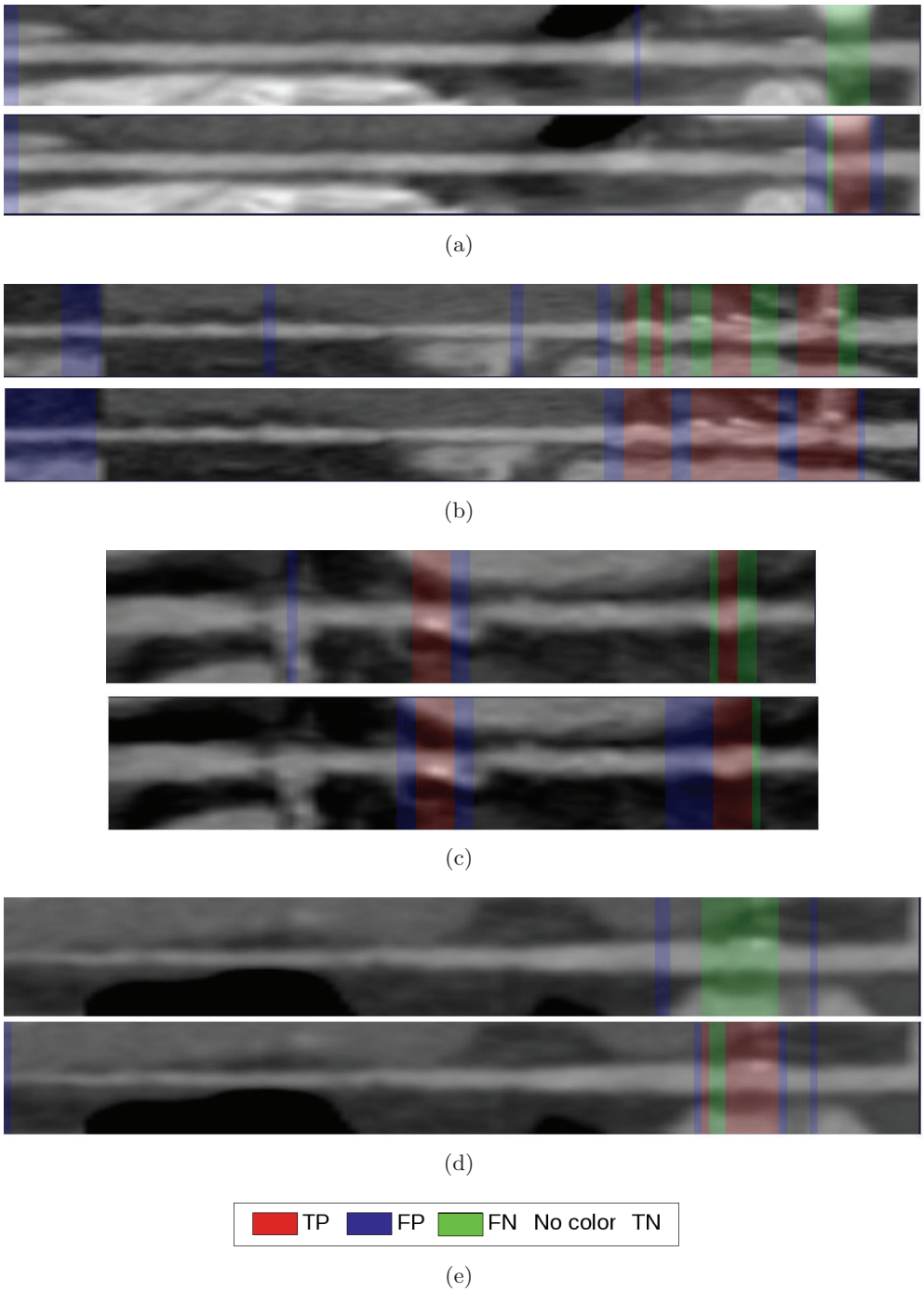


Figure 3.12: Comparison of DLD-SVM and LPU - Each sub image contains on top the result from DLD-SVM and on bottom the result from LPU. The DLD-SVM algorithm tends to under estimate anomalies, whereas LPU overestimates them.

comparison can be performed. The next chapter will make use of all the herein tuned metrics to construct an optimal subset of features.

4

Feature Selection for SVM-based Vascular Anomaly Detection

In the Chapter 2, we introduced all the elements that make up our anomaly detection framework and Chapter 3 was devoted to the experimental design and evaluation of the proposed `Concentric rings` metric and the tuning of the remaining candidate feature sets. Using all the above mentioned elements and the tuned features, this chapter aims at the contruction of an optimal classifier that is obtained by determining the best feature subset among all the proposed candidate metrics. The evaluation of all possible combinations, among the candidate metrics, would be too expensive. Therefore, we make use of the feature selection strategies that were introduced in Section 2.5. Namely, we evaluate our non supervised and semi-supervised feature selection proposal (Section 4.2.1) followed by the more classical supervised schemes (Section 4.2.2). The performance of the feature sets obtained through each method was assessed by using them as inputs for the DLD-SVM an LPU algorithms (Section 4.3). Moreover, in Section 4.4, we evaluate how the selected subsets perform when they are used in conjunction with other state-of-the-art classifiers. Results are presented both raw and after applying the proposed enhancement through bias variation (Section 4.3.2). Finally, we carry out an assessment of our proposed LPU with progressive increase of the training set (Section 4.5) using phantom data.

All the feature selection methods that we have introduced require a criterion that is used to determine which metrics are either removed (if backwards elimination is used) or included (in the case of forward selection). For this matter, the first section of this chapter (Section 4.1) is devoted to assess the individual performance of the different metrics when they are used in conjunction with DLD-SVM and LPU. This is followed by the metric ranking necessary for the supervised strategies (F-score, Random forests and SVM-RFE).

A discussion based on the obtained results is presented on Section 4.6.

4.1 Individual Performance of the Metrics

In this section we assess the different metrics individually in order to obtain a ranking that will be later used for feature selection. First, we assess the performance of the metrics by means of DLD-SVM and LPU, and each method's empirical risk (sec.4.1.1). Second, we make use of the supervised methods to provide metric ranking. Let us remark that by means of the supervised strategies (F-score, RF and SVM-RFE) there is no information of how well a metric performs by itself.

4.1.1 Metric Performance Assessment for Feature Selection Using DLD-SVM and LPU

The set of features defined in Section 2.4.2 [1] , for which tuning parameters were established in the previous section, were individually evaluated in order to assess their performance and to compare it with the results obtained with `Concentric rings`. Figure 4.1 shows the performance of each of the candidate features (including concentric rings) when used to classify anomalies with DLD-SVM on patient's data. Figure 4.2 presents the same results with the LPU algorithm. The results confirm the expected behavior as for the comparison between slice- and lesion-based measures, *i.e.* the sensitivity generally increases, while the specificity decreases, when analyzed on lesion-basis.

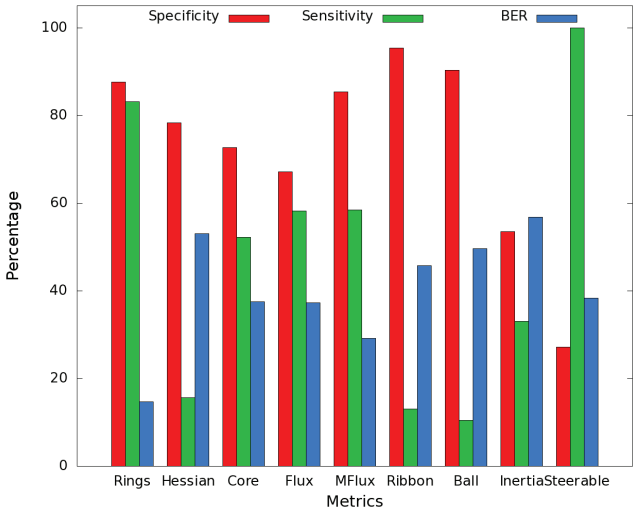
The expected increase in sensitivity results from the fact that one correctly detected anomalous slice within the whole extent of a lesion is enough to consider that the lesion was detected.

The decrease in specificity is explained by the fact that the proposed definition for true negatives and false positives is very strict. Our evaluation scheme allows only one excess slice detected as anomalous on either side of the actual lesion as belonging to the true positive group. However, visual inspection shows that frequently more than one contiguous excess slice is marked as anomalous. In the extreme case, were a vessel has only one lesion, this will result in a specificity of 50%.

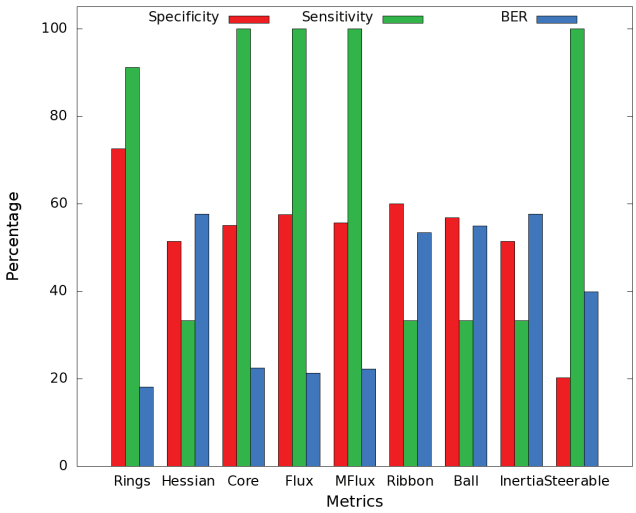
DLD-SVM. The `Concentric Rings` metric has a good performance in terms of specificity and sensitivity which is reflected by the lowest BER on a slice basis. It is the only metric the sensitivity of which does not increase in the lesion-based evaluation. This is interpreted as incapacity of the `Concentric Rings` to detect all the existent anomalies (83.33% of them are detected). At the same time, amongst all the evaluated metrics, the decrease of specificity of the `Concentric Rings` is clearly the smallest. The moderate decrease in specificity suggests that DLD-SVM with `Concentric Rings` tends to mark contiguous false positives slices that form a single false positive lesion, while with other metrics the false positive slices are scattered, thus giving rise to several false positive lesions.

`MFlux`, `Flux` and `Core` are able to detect all the existent lesions. This is reflected by a 100% sensitivity in the lesion-based evaluation, although not all the anomalous

[1] Hessian Eigenvalues, Inertia Moments, Cores, Flux, MFlux, Ball Measure, Ribbon, Steerable features and Concentric rings.



(a)



(b)

Figure 4.1: Individual performance of features using DLD-SVM - Results are presented in (a) slice-basis and (b) lesion-basis.

slices that are part of a lesion are correctly detected in the slice-based evaluation. The specificity of these metrics dramatically diminishes to values close to 50%, when evaluated on a lesion-basis. A visual inspection suggests that these three metrics tend to mark as anomalous several excess slices located at the lesion *border*. False positives between lesions are less frequent. The similar behavior of `MFlux`, `Flux` and `Core` is not surprising since they are closely related.

With the exception of `Steerable features`, all the metrics have better specificity than sensitivity when evaluated on slice basis. In `Steerable features` however, the counterpart of having a perfect sensitivity is that the specificity is very low. This means that very many healthy vessel slices are marked as anomalous. Therefore, the metric is not adequate for lesion detection.

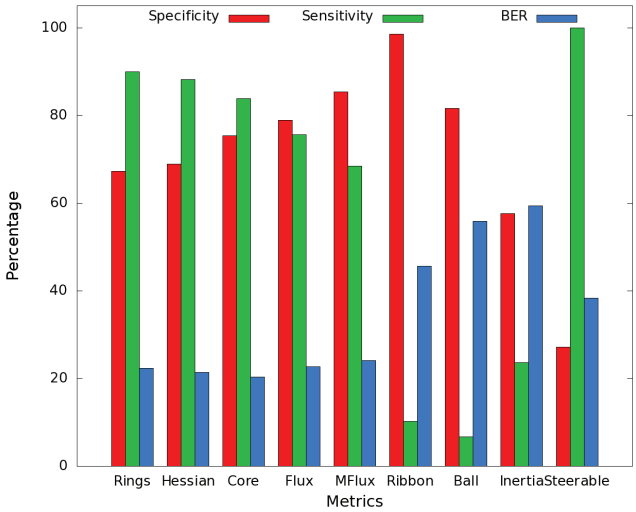
Despite their frequent use in vascular image enhancement and segmentation, `Hessian eigenvalues` and `Inertia moments` show a relatively poor individual performance in anomaly detection. One possible explanation is that both metrics usually aim at describing the shape of the lumen, which does not imply they can detect changes in the vessel wall appearance. Additionally, `Hessian eigenvalues` are sensitive to noise, while `Inertia momentss` integrate voxels intensities regardless of whether or not they belong to lumen, wall or background. Consequently, the inertia moments possibly are not sensitive enough to local changes in lumen and wall, while they detect global changes in the surroundings.

LPU. Contrarily to the behavior of DLD-SVM, the `Concentric rings` metric is not the best performing (in terms of the BER), when using LPU, although the difference w.r.t. the best metrics is very small. Figure 4.2 shows that `Concentric rings` detects all the existent lesions and almost all the slices that are part of the lesions. However, LPU in combination with `Concentric rings` tends to overestimate the lesions. A visual inspection of the labeled slices confirms this behavior. Moreover, this is seriously penalized when evaluating on a lesion basis.

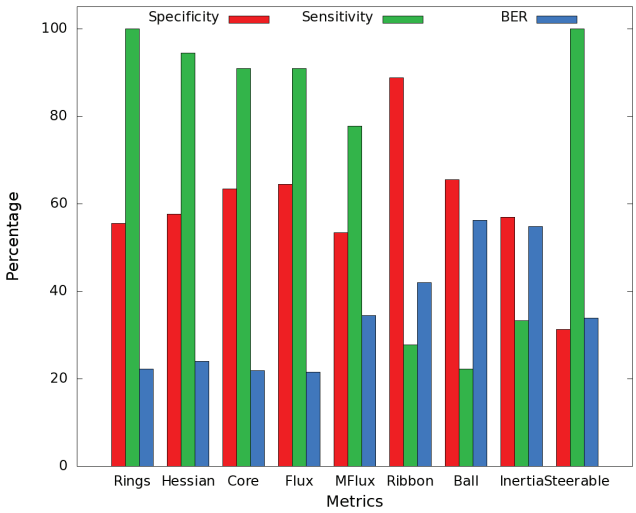
Once again, `MFlux`, `Flux` and `Core` have a closely related behavior among them. Also, they conserve the characteristic of having a good detection capacity. By this, we mean they can detect different type of anomalies instead of focusing on only one type.

While `Steerable features`, `Inertia moments` and `Ball` measure continue to have a rather low performance, the performance of `Hessian eigenvalues` is significantly better than the one observed with DLD-SVM. This phenomenon is explained by the way the classification algorithms work. DLD-SVM tries to detect high probability density points by discarding low probability points. On the other hand, LPU can be seen as a method that tries to find the concentration of a pattern set (the set Q) on a set of data denoted X. In other words, LPU searches for concentration of points following a pattern (given by Q) while DLD-SVM tries to find high concentration vs. low concentration points without specific pattern as an extra condition. The inclusion of an *additional* condition in the classification scheme, via labeled data Q, has a positive effect in the capacity of `Hessian eigenvalues` to detect anomalies.

Table 4.1 summarizes the feature ranking based on the empirical risk using DLD-



(a)



(b)

Figure 4.2: Individual performance of features using LPU - Results are presented in (a) slice-basis and (b) lesion-basis.

Table 4.1: Individual performance results using DLD-SVM and LPU. Feature ranking based on the empirical risk obtained from the individual performance evaluation of each metric.

Ranking	DLD-SVM	LPU
1	Rings	Core
2	Flux	Hessian
3	Core	Rings
4	MFlux	Flux
5	Ball	MFlux
6	Ribbon	Steerable
7	Steerable	Ribbon
8	Hessian	Ball
9	Inertia	Inertia

Table 4.2: Initial feature ranking using F-score, random forests (RF) and SVM-RFE. Initial ranking using F-score, random forests (RF) and SVM-RFE. Steerable features are evaluated over the reduced subset (see Section 3.3.1) and also individually.

Ranking	F-score	RF	SVM-RFE
1	MFlux	Rings	Core
2	Core	Ball	Ribbon
3	Flux	Core	Rings
4	Steerable	Ribbon	MFlux
5	Rings	Flux	Flux
6	Ribbon	MFlux	Hessian
7	Hessian	Hessian	Steerable
8	Inertia	Steerable	Inertia
9	Ball	Inertia	Ball

SVM and LPU for classification. Obtained results are used as a start up for feature selection using DLD-SVM and LPU.

4.1.2 Metric Ranking for the Supervised Approaches

Table 4.2 summarizes the initial feature ranking obtained with each of the supervised approaches. Based on this information, the supervised feature selection strategies described in section 2.5.2 were applied to each of the ranked sets. In the following, results obtained with each strategy are further described.

F-score. Metric individual ranking by means of the F-score ranks `MFlux`, `Core` and `Flux` in the first three positions (Table 4.2). It is not surprising to see these three metrics close to each other, since they are closely related. Therefore, their discriminative power should be very similar.

The `Concentric rings` metric has a rather low ranking (5^{th} position), when using F-score. This can be explained by the fact that `Concentric rings` metric is composed of many *local* features. Its discriminative power does not rely on how well each feature discriminates alone, but on a cooperation among features. Only a small subset of features provides a strong response in the presence of a lesion, and this subset changes depending on the location of the lesion. Therefore, the average individual discriminative power of each feature composing a metric is low. The F-score measures the discriminative power of each **feature**, and then the **metric** ranking provided in Table 4.2 averages the F-score of the features that compose a metric, to give a final ranking.

Random forests. It is surprising that a metric such as `Ball measure` is highly ranked using random forests while it is among the worst metrics in all the other rankings. The only possible explanation for this relies on the fact that random forests work under a principle that is different from the principle of SVM's. The criterion used to split nodes in random forests seem to suit the information by `Ball measure`. This is confirmed with the individual performance of the metric when used with random forests (see section 4.4).

SVM-RFE. Metric ranking using SVM-RFE has several points in common with the other ranking schemes. The worst and best ranked metrics are more or less the same than with other strategies (although a strict order is never preserved). However, it should be remarked that the classifier error obtained with SVM-RFE when performing the training for ranking was considerably higher than with random forests. We compare SVM-RFE only with RF because they follow the same principle of using all variables in order to provide feature importance. A further evaluation on this effect will be provided in section 4.2.2.

4.2 Feature Selection

We evaluated a combination of all the features generated by all the different metrics. In the absence of any selection strategy, the results in terms of BER were as follows: 43.76% with DLD-SVM and 42.17% with LPU. Furthermore, the training times were long: 315 and 192 min, respectively. The high computational times and the relatively low performances justify the use of different feature selection strategies to determine which combination of metrics leads to a better performance of the classifiers. In section 4.2.1, we perform feature selection with our proposed scheme, followed by feature selection with different supervised approaches (section 4.2.2).

4.2.1 Feature Selection Based on DLD-SVM and LPU

A major concern raises when performing unsupervised feature selection based on the empirical risk, which is related to the selection of ρ and t (see Eqs. 2.5 and 2.12). The

problem raises from the same situation so, for simplicity, we will explain it only in terms of ρ (DLD-SVM). The feature selection scheme implies the optimization of each feature set. This is, finding a σ and a ρ value that minimize the empirical risk $\mathcal{R}$. However, it can happen that for two (or more) given subsets the best risk does not appear at the same ρ value. Moreover, let us recall that ρ affects the computation of $\mathcal{R}$, then varying ρ would introduce more complexity to the feature selection problem.

A trivial solution to the problem is to perform feature selection on a dataset where the ρ value is already known. There are major drawbacks for this solution. First, it will turn the method supervised. Second, while in LPU it is rather simple to know t in advance, if labeled data is provided (it is the proportion of normal data w.r.t the overall number of samples), the task is not as simple for DLD-SVM. Although ρ reflects the absolute concentration of points in a set (normality in our case), it is not possible to obtain an analytical relationship between ρ and the actual concentration. Therefore, to have an estimate of ρ, it would be necessary to perform an initial experimentation to roughly estimate ρ and then perform the feature selection with the determined ρ value.

Nevertheless, this trivial solution can be adapted to our particular case, whereas the algorithm remains unsupervised. Our approach uses the empirical risk of single metrics as the inclusion criterion of the forward strategy. Therefore, we select the ρ (or t) value from the best performing metric to be used all along the feature selection process. Roughly, this allows the estimation of ρ and t without requiring to know their values in advance. However, it should be noted that this approach would no more be valid if we used a different criterion to determine the inclusion order, such that ρ (respectively t) was not estimated at the individual performance evaluation stage.

Figure 4.3 shows the evolution of the empirical risk when using a forward strategy to select features with DLD-SVM and LPU respectively. In both cases, results suggest that the best subset of features is obtained using the features form the four individually best-ranked metrics from Table 4.1.

4.2.2 Supervised Feature Selection

Based on the metric ranking results from Table 4.2, the three different supervised strategies, F-score, random forests and SVM-RFE, were performed.

F-score

Figure 4.4 presents the obtained average error of each of the evaluated feature sets. It should be noted that the use of only one metric (the best ranked) presents the highest average error, which suggests that it is not adequate to only use `MFlux`. The curve of classification error from Figure 4.4 suggests that the first 5 ranked metrics should be kept to build up the feature set. `Hessian eigenvalues`, `Inertia moments` and the `Ball` measure are excluded. The fact that `Hessian eigenvalues`, `Inertia moments` are excluded from the optimal set confirms the hypothesis that they are not suited for describing changes in the vessel wall.

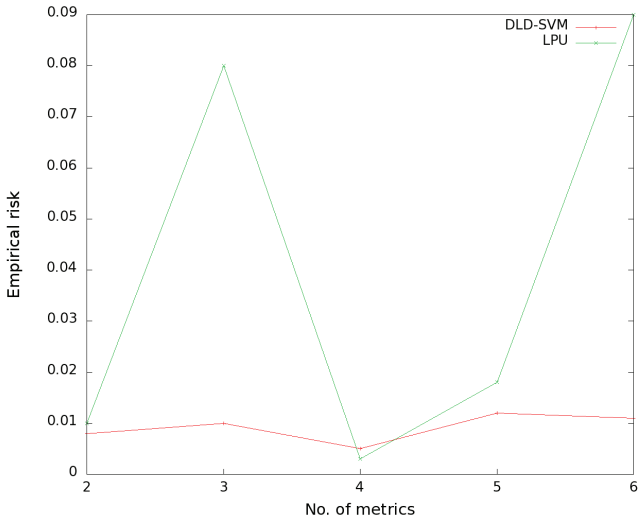


Figure 4.3: Empirical risk evolution in feature selection with DLD-SVM and LPU - The best response is obtained when four metrics are used to generate the feature set using LPU and DLD-SVM.

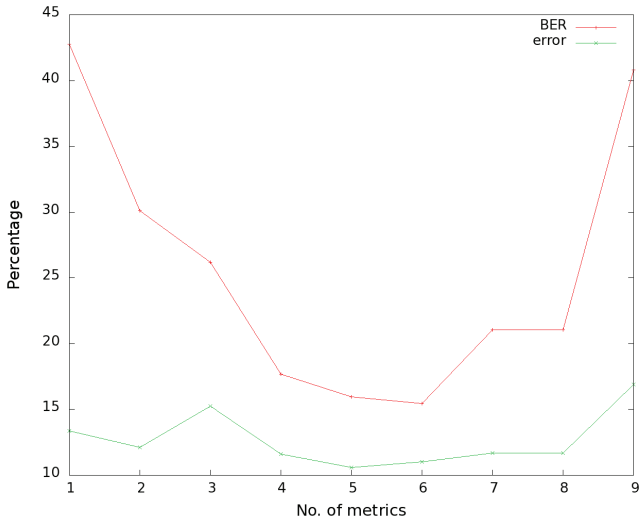


Figure 4.4: Classification error and BER evolution using F-score as a criterion for feature selection - Classification error and BER evolution as a function of the number of included features in the training-validation sets. The abscissa represents the number of included metrics based on the F-score ranking from Table 4.2. The average error obtained with only one metric is given as a reference.

It worths to further elaborate on the results from Figure 4.4, where the classification error is compared with the balanced error rate (BER) that we use to quantitatively evaluate our results. To perform a better analysis, Figure 4.4 shows the evolution of the error for all metrics, instead of only the first six ranked that were evaluated.

The classification error refers to the ratio between the number of failures of a classifier and the total number of evaluated samples. Figure 4.4 permits to see that both curves do not follow each other strictly. Local maxima and minima of the curves appear at different points. For instance, the best BER occurs when six metrics are used, whereas the error classification finds its minimum using five metrics.

Classifiers do not deal with the concept of balanced error because their goal is to minimize the overall error. This is a critical problem on highly unbalanced data such as these that are to be dealt with in anomaly detection. As a toy example, a typical learning algorithm prefers to correctly classify samples coming from the largest class at the price of misclassifying a few samples on the small class. Under such a situation, the classification error diminishes while the BER is seriously affected. Once again, we consider this problem can be aliviated through our formulated LPU with progressive increase of the training set (see Section 4.5).

According to error evolution the features composing the 5 selected metrics were used to define the training-testing sets of DLD-SVM and LPU algorithms.

Random forests

Figure 4.5 presents the obtained average classification error of each of the evaluated feature sets. Let us recall that the random forests approach works using backwards elimination described in Section 2.5.2. According to Figure 4.5, the subset that produces the smallest error is based on two metrics: `Concentric rings` and `Ball measure`.

SVM-RFE

Like Random forests, SVM-RFE selection strategy started with the individual ranking from Table 4.2) and iteratively eliminated the worst ranked metrics, according to the backwards elimination scheme described in Section 2.5.2. At each iteration, the error of each new subset was computed. Figure 4.5 presents the obtained average error at every step. Despite an initial ranking roughly similar to those obtained with other approaches, the classification error obtained with SVM-RFE vastly surpasses the classification error of Random forests. This is also true if compared with the classification error obtained with F-score. A detailed inspection of the classification results confirmed a very poor classification performance of SVM-RFE. It not only failed to detect lesions (in some cases $TP = 0$), but it also failed to correctly classify normal slices, thus generating many false positives. That justifies the high error rates.

The explanation for this behavior comes from the fact that SVM-RFE uses a linear kernel instead of a non-linear one. It has been established that nonlinear separating surfaces improve generalization over linear ones. This condition seems to have high incidence in our particular problem. In their seminal work Guyon et al. (2002) proposed

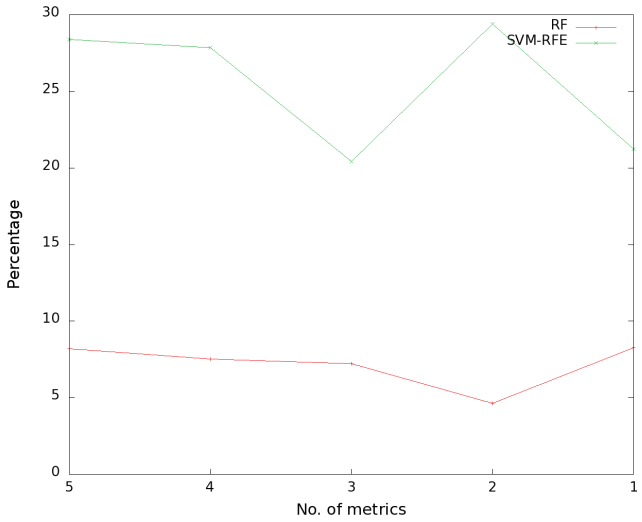


Figure 4.5: Feature selection error evolution - The curves represent the classification error evolution during feature selection using backwards elimination with random forests and SVM-RFE. At iteartion 1, five different metrics are used. On every new iteration a group of metrics is removed until only one is kept at the fourth iteration.

a version of the SVM-RFE algorithm using non-linear kernels. However, in the same work they state that the problem is computationally intractable and they only succeed in evaluating it on a toy example that learns the outputs of an XOR.

Based on these results, we consider that SVM-RFE is not a valid feature selection method for our problem. Selection based on a method that performs so badly in our data is not reliable.

4.3 Results of DLD-SVM and LPU Anomaly Detection with the Selected Best Feature Sets

In this section, we evaluate the subset of features that has been selected to each of the different approaches. Moreover, testing is also done using the subsets that were rejected by the each selection scheme. The goal is to assess whether or not the selected subset really outperforms the others (section 4.3.1). Once the optimal subsets are obtained, we apply the enhancement algorithm that modifies the bias of the decision function (see section 2.3.5 for its description) to determine whether or not the classification results can actually be improved.

4.3.1 Non-Enhanced Results

DLD-SVM and LPU. Figures 4.6(a)- (b) and Figures 4.7(a)-(b) show the obtained specificity, sensitivity and BER for each case.

The results of both classifiers using features extracted from 4 different metrics confirm that the selected features optimize the classification performance for both DLD-SVM and LPU. In general, the response of the selected subsets is as expected.

F-score. Figures 4.6(c)- (d) and Figures 4.7(c)-(d) present the obtained results.

In both cases, it can be seen that the best performance of DLD-SVM and LPU is not achieved using the number of metrics suggested by the selection based on the F-score. After the inclusion of `Steerable features`, the performance of both classifiers dramatically diminishes. Let us recall that this metric was poorly ranked according to the unsupervised selection strategies. When a fifth metric is included, namely `Concentric rings`, the LPU improves its performance but not significantly. The best response is obtained using 3 metrics in LPU. In DLD-SVM, it is arguable if the best performance is obtained with two or three metrics. The lesion-based BER in DLD-SVM with three metrics is lower than the one obtained with two. However, the learning algorithms receive slices as inputs and therefore the evaluation criterion should be slices. The lesion-based metric is an artificial grouping we use to evaluate the clinical relevance of the classification.

Random Forests. When evaluating the different candidate subsets with DLD-SVM (Figures 4.6(e)- (f)) and LPU (Figures 4.7(e)-(f)) a situation similar to the one that occurred with F-score: the best performance is not achieved with the selected subset. This situation can be explained by the fact that the `Ball measure` is included in the selected subset. `Ball measure` has a poor performance in DLD-SVM and LPU, which affects the performance of the classifiers. However, the inclusion of a sufficient amount of additional features diminishes the negative effect of `Ball measure`. As a consequence, the performance improves. In DLD-SVM it is necessary to add up three more metrics to see positive effects in performance, while LPU requires the inclusion of only two additional metrics.

The cancellation of the effect of `Ball measure` by the inclusion of additional metrics not only confirms what the literature reports on the robustness of SVM-based methods to the presence of spurious features (when a sufficient number of good features exist), but explains the low performance of DLD-SVM and LPU with the F-score selected subset. The size of the features obtained from `Steerable features` vastly overpasses the size of the features obtained from all the other remaining metrics. Therefore, adding up other features is not enough to cancel the negative effects of `Steerable features`.

Table 4.3 summarizes the performances of all the subsets selected using different selection strategies when they were evaluated using DLD-SVM and LPU. The performance is presented in terms of the slice-based BER. Here, it is possible to see that when using DLD-SVM and LPU as feature selection strategies, the selected optimal set coincides with the real optimal feature set. This does not happen with the other selection strategies. We consider this not to be surprising. Appropriate features for density level detection problems should follow certain characteristics that reflect the concentration of some samples and the low density regions. Therefore, the use of selec-

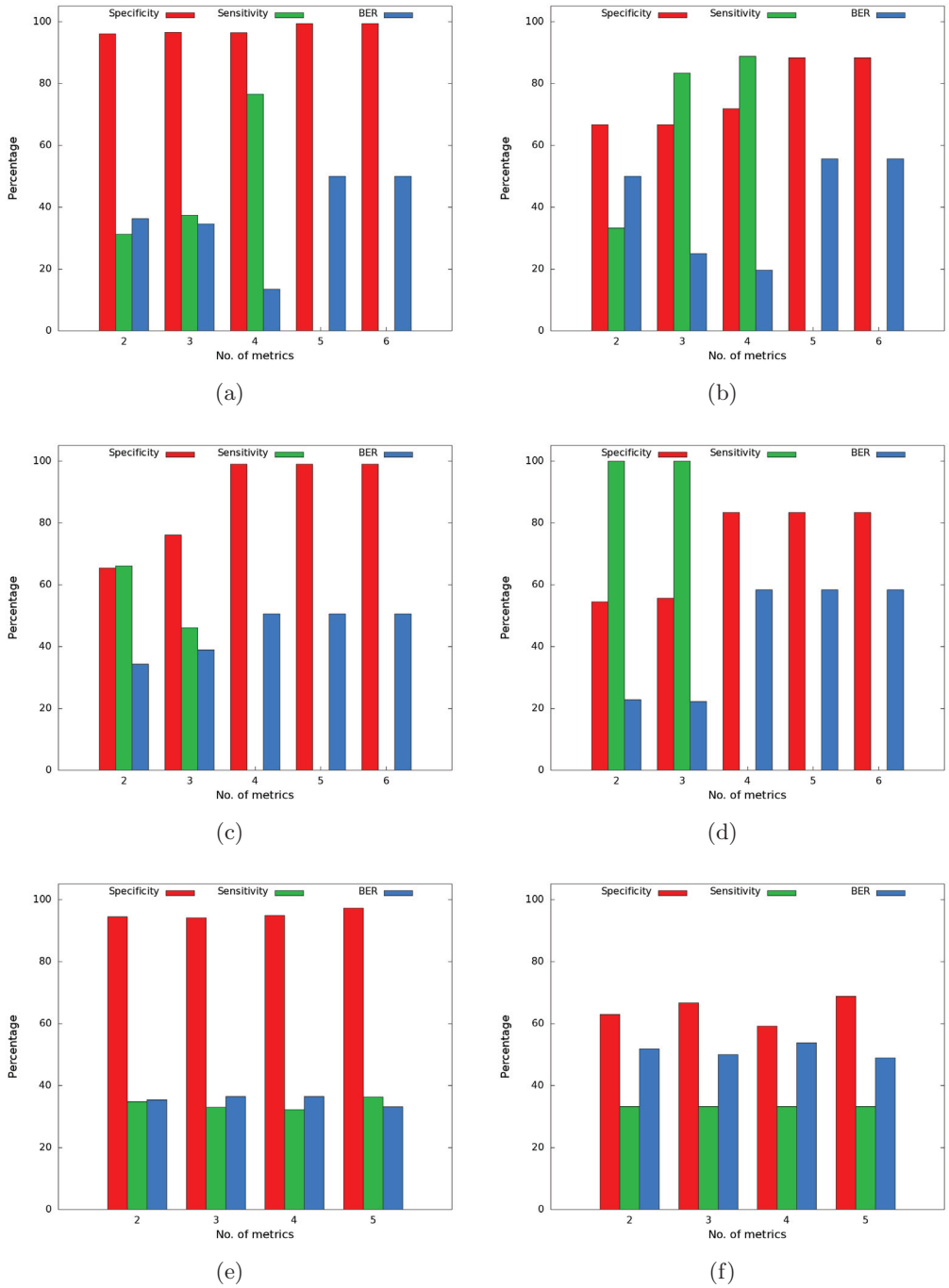


Figure 4.6: Feature selection evaluation using DLD-SVM - Unsupervised DLD-SVM, F-scores and random forest selection schemes are presented in a slice basis in (a), (c) and (e). Lesion-based results are presented in (b), (d) and (f).

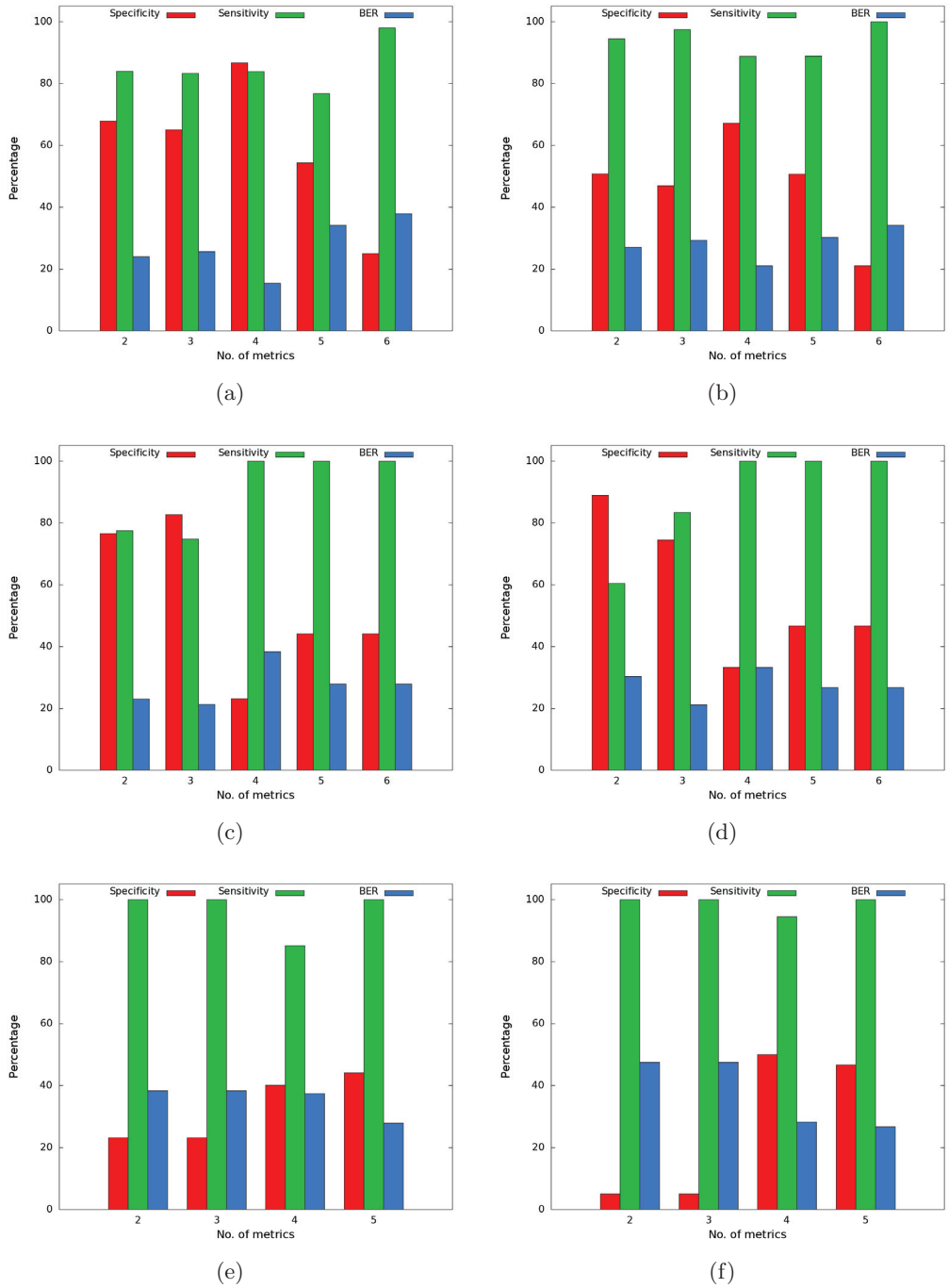


Figure 4.7: Feature selection evaluation using LPU - Unsupervised LPU, F-score and random forest selection schemes are presented in a slice basis in (a), (c) and (e). Lesion-based results are presented in (b), (d) and (f).

Table 4.3: Summary of different classification methods performance. The subsets selected through unsupervised and supervised methods are tested with DLD-SVM and LPU. The table presents the performance at the expected optimal feature set and the real one. For every case, the number of metrics used is shown in parentheses. Perfomance is presented in terms of BER.

Strategy	Expected		Real	
	DLD-SVM	LPU	DLD-SVM	LPU
DLD-SVM empirical risk	13.55 (3)	-	13.55 (3)	-
LPU empirical risk	-	15.39 (4)	-	15.39 (4)
F-score	50.50 (6)	27.92 (6)	34.28 (2)	21.31 (3)
Random Forests	35.43 (2)	38.39 (2)	33.19 (5)	27.92 (5)

tion schemes that do not work under the density level detection philosophy might lead to the selection of feature subset that does not contain high and low density regions of samples. The latter would explain why the subsets selected with F-score and random forests are not as performant as those selected with DLD-SVM and LPU.

4.3.2 Results Obtained with Bias Variation

Using the different subsets obtained, we have applied the post-enhancement algorithm previously introduced in Section 2.3.5. Let us recall, that the method consists simply in variating the term b on the SVM decision function with the guarantee that the error cannot worsen. Therefore, we have generated random values $\triangle b$ to adjust the original value of each of the trained models obtained with the different subsets (Table 4.3). A total of 100 values were tested for each of the models. The $\triangle b$ generating the best response was kept. Table 4.4 summarizes the obtained results.

From these results, it can be seen that it is possible to improve the final classification results by means of this simple enhancement algorithm. However, it should be remarked that there is no guarantee that the results will have a significant improvement. As an example, the result from LPU using the subset obtained with random forests (2 metrics) does not achieve any improvement. Nevertheless, models such as DLD-SVM trained with the set obtained from DLD-SVM empirical risk, LPU trained with the set obtained from LPU empirical risk and LPU trained with the set obtained from F-score (3 metrics) have significant improvement. Therefore, despite the lack of guarantee of performance improvement, we consider that this algorithm is a good alternative since the error cannot increase. Moreover, its computational overhead is negligible when compared to the time devoted to training.

4.3.3 Computational Times

Computational time is a critical factor in clinical practice. We provide an estimation of the time required to compute each of the proposed metrics (Figure 4.8) and the time devoted to training with each metric (Table 4.5).

Table 4.4: Effect of applying algorithm enhancement to the classification. The enhancement algorithm is applied to the best classification, in terms of BER, obtained with DLD-SVM and LPU when using the feature sets selected by different strategies.

Feature Selection Method	No Enhancement	With Enhancement
DLD-SVM		
DLD-SVM empirical risk	13.55	10.41
F-score 6 metrics	50.50	48.36
F-score 2 metrics	34.28	33.97
Random forests 2 metrics	35.43	35.41
Random forests 5 metrics	33.19	32.56
LPU		
LPU empirical risk	15.39	10.92
F-score 6 metrics	27.92	27.01
F-score 3 metrics	21.31	16.45
Random forests 2 metrics	38.39	38.39
Random forests 5 metrics	27.92	21.63

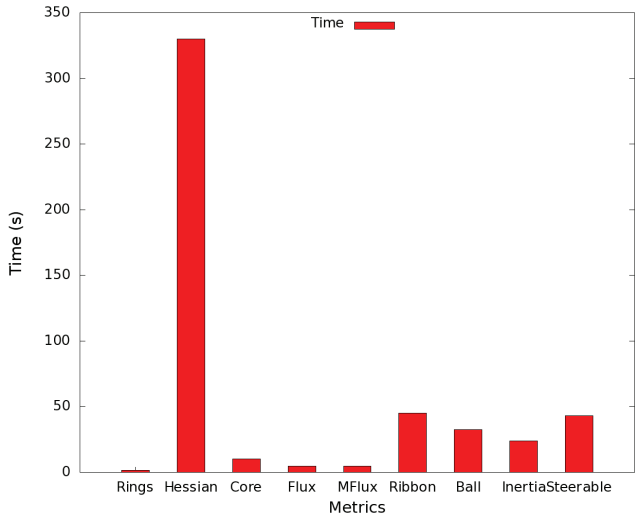


Figure 4.8: Computational time of each metric - The plot presents the average time (in seconds) required to compute a given metric over 300 slices.

Table 4.5: DLD-SVM and LPU training times using different feature sets. Training times of DLD-SVM and LPU using the different feature sets are presented in minutes. The computational time was measured when using cardiac set 2 for training.

Metric	DLD-SVM	LPU
Rings	75	21
Hessian	54.8	19
Core	55	10
Flux	54.1	11.3
MFlux	52.9	12.1
Ribbon	52.2	12.3
Ball	61.1	14.8
Inertia	52.8	15.1
Steerable	193	39

As can be seen from Figure 4.8, there is a significant difference between the time necessary to compute the Hessian eigenvalues and the other metrics. This is a consequence of the operations involved in Hessian eigenvalues computation that are highly time consuming. Nevertheless, it should be remarked that the provided times serve only to illustrate the computational burden since most of the metrics can be further optimized.

In what respects the training times, it can be seen DLD-SVM is more expensive than LPU. This is expected since DLD-SVM has always larger datasets due to the use of artificially generated data. Another important remark from the results presented in Table 4.5 relates to the effect that the inclusion of features has on the training time. It can be seen how a large number of features affects the computational time. As an example, steerable features always requires a larger amount of time than the other metrics. However, it should be noted that the same number of features does not imply the same computational time. Although the training times of metrics producing the same number of features are close, they are never the same.

Finally, let us recall the training stage is a process that should be performed only once or rarely, and off-line. Therefore, we consider that the time devoted to metric computation is more relevant than the training time.

4.4 Comparison With Other Classification Strategies

In the previous sections we used unsupervised, semi-supervised and supervised methods in order to select subsets of features supposed to have the best discriminative power. Then we evaluated the actual performance of each of thus selected subsets, in our anomaly detection scheme, by use of an unsupervised and semi-supervised method, namely DLD-SVM and LPU respectively. Hereafter, we compare the performance of our classification approaches, DLD-SVM and LPU, with well-known classifier schemes. We evaluate two supervised learning methods and one semi-supervised. The first one is the

classical soft margin SVM (Cortes and Vapnik, 1995), which is a very common method for classification. Then we evaluate the Random Forests (Breiman, 2001) method. The latter is not only a feature selection method. Nowadays, it is widely used in the medical imaging community also for classification. It might be expected that classification based on Random Forests is likely to perform particularly well with a feature set selected by a strategy also based on Random Forests. Finally, we evaluate one of the first methods designed for anomaly detection, the so-called one-class SVM (Schölkopf et al., 2001).

Different strategies could be followed to compare DLD-SVM and LPU with other classification approaches. Here we have compared the performance of each of the metrics individually by using them as input for each of the classifiers. Afterwards, we use the subsets selected by each of the feature selection strategies.

Individual performance

Figure 4.9 presents the obtained results using One-class SVM, Random forests and classical Soft-margin SVM.

As with DLD-SVM and LPU, the evaluated algorithms perform poorly when `Ball measure` and `Inertia moments` are used as inputs. `Steerable features`, which performs poorly with DLD-SVM and LPU, showed the best performance with the classical Soft-margin SVM. This behavior partially explains the fact that this metric has been used in works following the same goal that use supervised schemes (Tessmann et al., 2008, 2009). However, it shall be remarked that `Steerable features` perform poorly in Random forests which is another supervised scheme. Let us also note that the results of Soft-margin SVM with `Concentric rings` are almost the same as with `Steerable features`.

While the best performing metrics remain more or less the same, the key point that is to be underlined in this evaluation refers to the behavior of anomaly detection algorithms (DLD-SVM, LPU and One-class SVM) w.r.t. other classification schemes (Random forests and classical Soft-margin SVM). In general, anomaly detection methods have a higher sensitivity, *i.e.* they are better in detecting lesions than standard classification approaches. Both Random forests and classical SVM have very high specificities, but low sensitivities. This is, they succeed in classifying the learned classes (as new normal cases appear they correctly classify them), but when slightly different samples appear, which often occurs with lesions, they fail. The latter validates the formulation of our problem as an anomaly detection problem.

Table 4.6 compares the best performing metric of each method. Results show that DLD-SVM and LPU give the best results among all strategies.

Performance of selected feature sets

Figure 4.10 summarizes the evaluation results of the subsets selected by the different strategies. We also included the subsets that performed best, using DLD-SVM and LPU, when these were different from the selected ones.

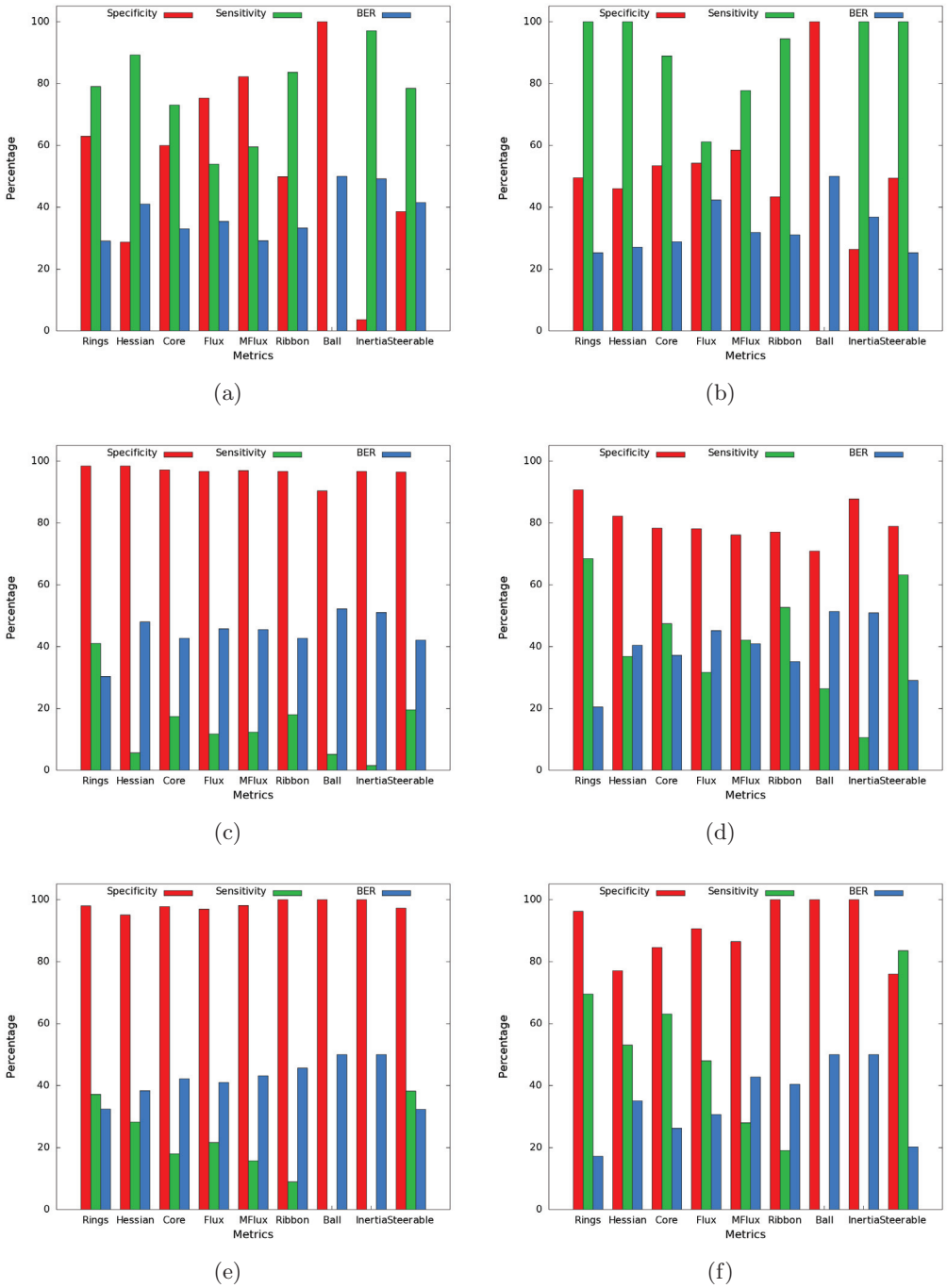


Figure 4.9: Performance of other classification strategies - (a), (c) and (e) present slice-based performance of One-class SVM, random forests and Soft margin SVM. (b), (d), (f) present lesion-based results.

Table 4.6: Summary of different classification methods performance. For each method, the results of the best performing metric are presented. These are Concentric Rings for the first three classifiers (DLD-SVM, One-class SVM and Random Forests), Cores for LPU, and Steerable Features for the classical Soft-margin SVM. The results appear on a slice-basis.

Classification Method	Specificity	Sensitivity	BER
DLD-SVM	87.56	83.14	14.65
LPU	75.35	83.84	20.39
One-class SVM	62.89	78.97	29.06
Random Forests	98.46	41.02	30.28
Soft margin SVM	97.24	38.19	32.28

As it occurred with the individual performances of the metrics, none of the evaluated classifiers performed better than DLD-SVM and LPU (in combination with their selected feature set). With this, we confirm the quality of our proposed selection scheme and also the superiority of DLD-SVM and LPU w.r.t. other classifiers in the anomaly detection problem.

An interesting observation can be made from Figure 4.10 and refers to the best performing subsets from each classifier. For One-class SVM, the best performing subsets are the ones obtained with DLD-SVM and LPU feature selection strategies. This makes sense since One-class SVM is also an anomaly detection approach. Additionally, it can be seen that Random forests classifier obtains its best response using the two metrics defined by the Random forests selection scheme and the Soft-margin SVM classifier obtains its best performance with six features selected by the F-score selection method. If we recall that F-score uses a Soft-margin SVM at final stage of the feature selection process, we can conclude that feature selection gives the best results when the used classifier is based on the same methodology, (*i.e.* Soft-margin SVM classifier also gives the best results with the subset selected by use of a method involving Soft-margin SVM. Nevertheless, let us recall that none of the proposed strategies outperforms DLD-SVM and LPU.

Results from the Soft-margin SVM, and especially from Random forests, show that their performances are poor, while the residual BER observed on the training set was very low (below 5%). A similar behavior was observed in the individual feature evaluation in the previous section. This means that the Soft-margin SVM and Random forests perform very well on the training sets, while they perform poorly on the testing sets, which demonstrates one of the principal weaknesses of supervised methods. Their accuracy is poor with new arriving data since they cannot adapt to the heterogeneity of anomalies.

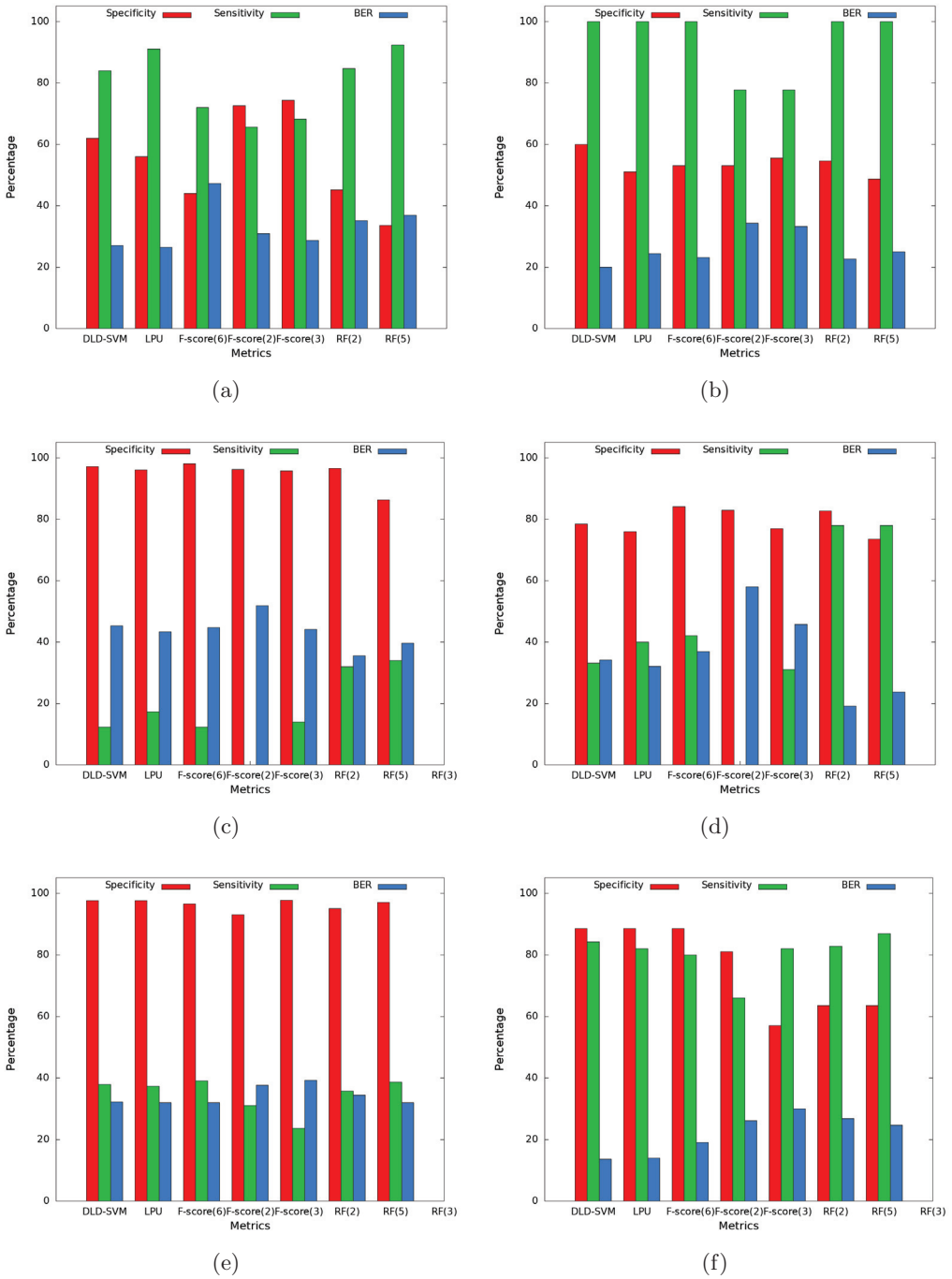


Figure 4.10: Performance of other classification strategies using optimal feature sets - (a), (c) and (e) present slice-based performance of one-class SVM, random forests and soft margin SVM. (b), (d), (f) present lesion-based results.

4.5 Preliminary results of LPU with progressive increase of the training set

This last section aims at the evaluation of our proposed LPU with progressive increase of the training set. From what has been presented, it is evident that our method requires the extraction of the vessel centerline axis. For our experimentations, manual axes have been traced so that we can evaluate the real performance of our method without the introduction of errors coming from other sources. However, this represents a limitation in the evaluation of the LPU with progressive increase of the training set approach on real data. Manual annotations are tedious and time consuming. Therefore, it is difficult to perform an evaluation over a large database of annotated vessels. Hence, the validation of this proposal is a preliminary experimentation over phantom data.

To validate the method, we have defined six different case scenarios. The first five cases simulate the incremental process described in section 2.3.4. The last case aims at illustrating the situation where the condition $|X| \gg |Q|$ is broken. The configuration of each case was as follows:

Case 1. Normal vessels,modeled as straight tubes are included in the Q set. No bifurcations or normal vessels with nearby structures are included. Sets sizes are: $|Q_1| = 500$, $|X_1| = 15000$.

Case 2. Bifurcations and normal vessels with nearby structures are included, so as to provide a larger set of examples. Sets sizes are: $|Q_2| = 1000$, $|X_2| = 19000$.

Case 3. After classification is performed in case 2, a number of random samples labeled as normal in X_2 are added to Q_2 and new unlabeled samples are added to X_2. Here, we aim at evaluating if random samples can improve the performance, as opposite to what was done in Case 2 where the new labeled samples were manually selected. Sets sizes are: $|Q_3| = 1200$, $|X_3| = 22000$.

Case 4. A total of 100 samples, labeled as normal in case 3 are added to Q_2. X is not modified. Sets sizes are thus: $|Q_4| = 1300$, $|X_4| = 22000$.

Case 5. A total of 300 samples are added to Q_4 containing standard normal vessels, bifurcations and vessels with nearby structures. X set is incremented with 1000 more samples. $|Q_5| = 1600$, $|X_5| = 23000$.

Case 6: Condition $|X| \gg |Q|$ is broken. We define $|Q_6| = 1200$, $|X_6| = 1600$.

Figure 4.11 shows the evolution of specificity, sensitivity and BER (on slice-basis) for every case, where the first five can be considered as successive iterations of this version of the LPU algorithm. From these results, it can be seen that LPU generally performs well in terms of sensitivity. As more normal samples are included to the Q set, the specificity starts to improve. Since the sensitivity remains unchanged, therefore, the BER decreases. Even for the 4^{th} case, where no samples are added to X, since the

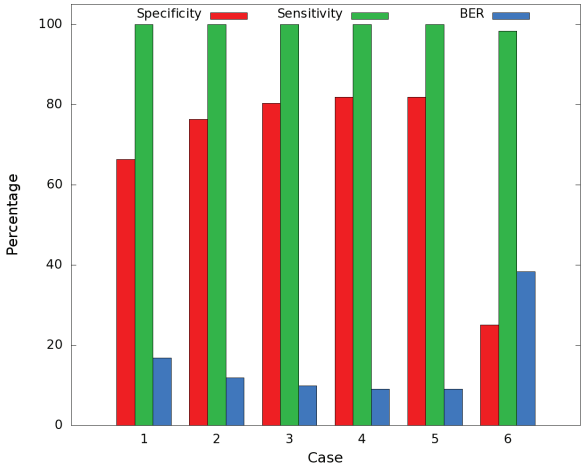


Figure 4.11: Effect of adding samples to Q and X datasets - The balanced error rate and the specificity tend to improve. When the condition $|X| \gg |Q|$ is broken the results are poor.

condition $|X| \gg |Q|$ holds, the performance of LPU still increases. Finally, on case 6, it is possible to see that the performance of LPU is poor.

Although the overall results show that the method permits improvements in the classification, two key issues need to be taken into account. First, an increase in the size of the sets Q and X also implies an increase in the computational time. Second, even at the infinite sample case the classifier is likely to make some errors. Therefore, a trade-off between computational time and capacity of improvement should always be evaluated.

4.6 Discussion

In this chapter, we have followed different strategies in order to find a subset of features that optimizes the performance of our classifiers. In order to perform feature selection, it was necessary to first rank individually the proposed set of metrics since all of our approaches require an inclusion/rejection criterion. We have used the empirical risk, in DLD-SVM and LPU strategies, the variable importance for Random forests and SVM-RFE, and the F-score itself as criteria.

Individual evaluation of metrics by using them with DLD-SVM and LPU algorithms demonstrated that our metric is suitable for the purpose of anomaly detection and, moreover, it outperformed all other evaluated metrics when used with DLD-SVM. Moreover, the feature sets selected using DLD-SVM and LPU both contained `Concentric rings` as part of the selected metrics. This confirms the potential of our metric in the vascular lesion detection problem. The results also show that first order-based metrics such as `MFlux`, `Flux` and `Cores` have a good performance.

Table 4.7: Individual vs. optimal subset performance. For each method, the BER of the best performing metric and the optimal subset are presented.

Configuration	DLD-SVM	LPU
Individual metric	14.65	20.39
Optimal subset	13.55	15.39

The evaluation of different feature selection strategies allowed us to see that our unsupervised and semi-supervised approaches based on the empirical risk of DLD-SVM and LPU are well suited for the anomaly detection problem. This was somehow expected. The features used under an anomaly detection approach should follow the formulated hypotheses of such a problem, *i.e.* anomalies should be represented as low density regions. Since DLD-SVM and LPU work under that principle, it is expected that they are able to find the features that best suit the formulation. At the same time, this explains why the set of features found by Random forests and F-score were not the optimal ones.

To validate the relevance of the selected classification framework and the anomaly detection formulation of our problem, we have compared DLD-SVM and LPU to other classification strategies. Two of them were classical classification schemes such as Random forests and Soft-margin SVM, while the third one was One-class SVM, which is a well-known anomaly detection method. One of the first conclusions that can be obtained from these comparative results is that the anomaly detection scheme outperforms the classical approaches. The three anomaly detection schemes, namely DLD-SVM, LPU and One-class SVM, have outperformed Random forests and Soft-margin SVM. Anomaly detection approaches have a higher capacity of detecting newly arriving patterns of anomalies. Additionally, DLD-SVM and LPU have shown to perform better than One-class SVM.

The evaluation of different feature subsets through One-class SVM, Random forests and Soft-margin SVM shown that, in general, a feature subset will perform better when the classifier in which it is evaluated is of the same nature as the algorithm used for feature selection. As an example, the feature subset selected through Random forests had the best performance when used with Random forests classifier. Similarly, the subset selected through F-score (in combination with Soft-margin SVM) had the best performance when evaluated with the Soft-margin SVM classifier. In what concerns One-class SVM, its best performance was achieved with the subsets obtained through DLD-SVM and LPU selection strategies. This can be explained by the fact that all of them are anomaly detection approaches. However, let us remark again that no other classifier showed a better performances than DLD-SVM or LPU either with individual metrics or feature subsets. Regarding the performance of the algorithms with a combination of metrics or with a single one, it is always arguable if it worths to include a feature selection step. Table 4.7 compares the results from DLD-SVM and LPU when using the best performing individual metric and the optimal feature subset. The results obtained from DLD-SVM might suggest that the feature selection step is

undesired since it implies a computational overhead while the results of using a single metric are comparable. Contrarily, a significant improvement can be seen for LPU. We consider that a feature selection step is always necessary even for cases such as the one of DLD-SVM. The main reason is that, in advance, it is impossible to know if such a situation can occur. A priori it is not possible to know which combination of features will have the best performance. The goal of the feature selection strategy is to provide an answer to this issue.

Although our results are promising there are still things that can be improved. The problem we are trying to solve is highly imbalanced and this can introduce difficulties in the classification. Particularly for LPU, the effects of finite samples can be significantly enhanced under a class imbalance situation. For this purpose, we proposed a variant of the LPU algorithm, which increases the size of the training set in order to reduce the finite samples effects. Preliminary results on phantom data show that this approach is promising. However, a validation on real data is required.

Finally, we have shown that the use of an additional step of the learning phase can improve the detection rate. We have used a simple approach that consists in varying the b parameter of the SVM decision function. Although results are not always improved, their quality never diminishes. Hence, it is recommended to use such an enhancement step, since it can ameliorate results and its computational overhead is minimal.

5

Inter-phase Vessel Segment Registration to Corroborate Anomaly Detection

In this chapter we present an approach that seeks to exploit information provided by multiple cardiac phases by means of a local registration of potential lesions. Here, we aim at simulating the procedure followed by a physician when trying to diagnose coronary artery disease. After detecting a potential lesion in a particular time-frame or cardiac phase, the physician often seeks the lesion in another time-frame (typically, these are end-diastole and end-systole) to confirm the initial diagnosis. In the previous chapters, we tackled the automatic detection of potential lesions. At this point, we want to tackle the second stage of the diagnosis, which consists in confronting the information found at one time-frame with another one. For this purpose, our goal is to define a registration framework that allows matching different cardiac time-frames without being computationally expensive. More particularly, we focus in designing a method that copes with the registration of an automatically selected volume of interest (VOI) surrounding the potential lesion, instead of the registration of the whole image.

We first define image registration and exhibit its principal components (Section 5.1). Subsequently, we present related works in literature dealing with vascular image registration (Section 5.2). In Section 5.3 we present our approach, which consists in an automatic VOI selection using *a priori* knowledge, followed by an intensity-based registration scheme. The evaluated similarity metrics are also presented. Feasibility of the approach is then demonstrated through the evaluation of the whole method (Sections 5.4 and 5.5). The obtained results are used in Section 5.6 in combination with anomaly detection problem to evaluate its applicability to vascular lesion detection. Finally, the chapter closes with a discussion of the proposed framework.

5.1 Image Registration: Definition

Image registration is the process of deforming one image, denoted the moving image $I_M(x)$, to fit another image, denoted the fixed image $I_F(x)$, by applying a transformation $\mathbf{T}(x)$ to $I_M(x)$. The registration problem is formulated as an optimization problem that seeks to minimize a metric $\mathcal{S}$ with respect to the transformation $\mathbf{T}(x)$:

$$\hat{\mathbf{T}} = \arg\min_{\mathbf{T}} \mathcal{S}(\mathbf{T}; I_F, I_M) \tag{5.1}$$

where $\hat{\mathbf{T}}$ is the optimal transformation.

In order to solve the optimization problem (Eq. 5.1), several aspects need to be addressed. First, a *transformation* $\mathbf{T}(x)$ determines the types of deformations that shall be applied to $I_M(x)$ in order to fit $I_F(x)$. An *interpolator* evaluates the transformed moving image at non-voxel positions. The *similarity metric* $\mathcal{S}$ evaluates the quality of the alignment between the images. Typically, $\mathcal{S}$ is obtained by evaluating all image voxels. However, a subset may suffice. In this case, a *sampler* component is necessary. Finally, an *optimizer* uses $\mathcal{S}$ as a criterion to obtain the optimal parameters of the *transformation*. Multiresolution strategies are often adopted to improve the results and / or speed-up the registration process. In this case, the registration starts with images of low complexity (smoothed or/and downsampled) that is increased gradually. The designed component for this approach is often called a *pyramid*. Figure 5.1 shows the general components of the registration framework as proposed by Klein et al. (2010).

For further details on the elements of a registration algorithm, the interested reader is referred to more extensive reviews (Ibánez et al., 2005; Klein et al., 2010; Maintz and Viergever, 1998).

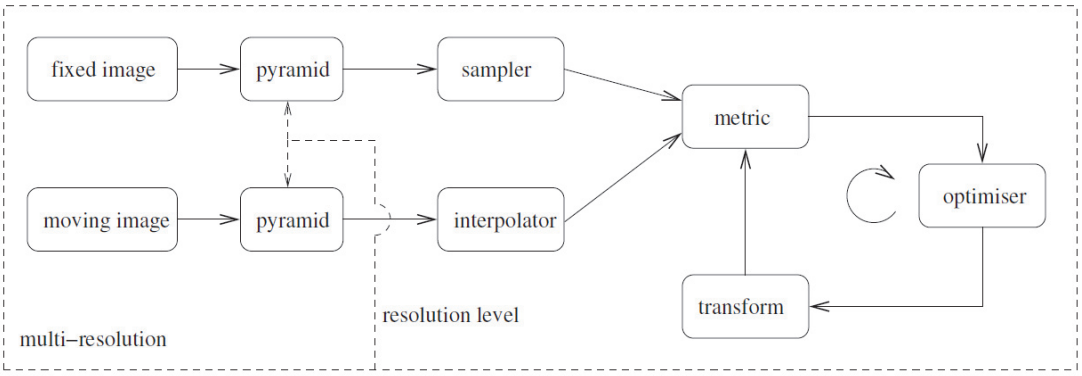


Figure 5.1: Basic registration framework components - Two input images $I_F(x)$ and $I_M(x)$, an interpolator, a transform, an optimizer and a similarity metric. Samplers and pyramids are optional. Adopted from Klein and Staring (2010).

5.2 Related Work

Most of the studies in vascular image registration have been done in the 2D-3D domain (Kita et al., 1998; Lange et al., 2003; Reinertsen et al., 2007; Ruijters et al., 2009; Turgeon et al., 2005). However, 3D-3D vascular registration is a challenging problem that poses additional difficulties. Correspondence and alignment ambiguities are inherent to the general task of automatically registering images of tubes (Aylward et al., 2003). The registration process is further complicated by the fact that vessels are typically surrounded by larger organs that swamp the similarity metric. In consequence, vascular registration has not been as widely studied as the registration of other organs. Extensive overviews (Hill et al., 2001; Maintz and Viergever, 1998) of medical image registration methods do not consider this domain.

Our bibliographic review is limited to 3D-3D vascular registration (which for simplicity we will refer as 3D registration). Methods that tackle this problem can be divided into two broad categories. First, there are methods that aim at registering static vascular images from longitudinal studies. On the other hand, there are algorithms for the registration of dynamic vascular images. These are images including movement. Dynamic vascular images imply additional challenges related to artifacts generated by movement.

5.2.1 Static Vascular Image Registration

To the best of our knowledge, the first work on 3D vascular image registration is the one from Aylward et al. (2003). In their approach, they rigidly register a vessel model, defined by a centerline and a radius, to a target image using an intensity-based metric (Aylward et al., 2001). For this purpose, vessel centerline needs to be previously extracted on the source image. Jomier and Aylward (2004) extend this work to include non-rigid transformations. Using this vascular model-to-image registration method, Chillet et al. (2003) developed a method to form vascular atlases by means of vascular distance maps. In a similar fashion, Cool et al. (2003) constructed a density map image of presegmented vessels which are then registered using an affine tissue-based mutual-information registration to create a vascular density atlas. Wong and Chung (2006) propose a method to detach abnormal regions from a segmented vessel. In the method, the initial axis of a tube model is set to the original centerline of a vessel segment. The model is then deformed such that its boundary is precisely registered onto only the vessel boundaries of normal regions. The registered tube model is then refined by a subsequent surface matching process so as to virtually generate an abnormality-free vessel. Using vessel matching, Charnoz et al. (2005) propose a tree matching algorithm for intra-patient hepatic vascular system registration. Vascular systems are segmented from CT images, and then modeled as trees. Starting from the tree root, edges and nodes are iteratively matched. The algorithm works on a set of match solutions that are updated to keep the best matches thanks to a quality criterion. Xue et al. (2006) perform extraction and matching of the neonatal cerebral vasculature from MRA-Time of Flight (MRA-TOF) images. After vasculature extraction, vessel segments are iter-

atively connected to compose a vessel tree that is automatically labeled. After this, an indirect vasculature registration method is used to recover global deformation between two vessel trees, and vessel matching is performed by comparing a cost function measuring average spatial distance between two vessel branches.

One of the main disadvantages of the above mentioned methods is that they strongly rely on an initial model or segmentation. A bad segmentation implies a bad result in the registration process. To overcome this, some other methods try to enhance vessels instead of explicitly segmenting them. As an example, Suh et al. (2010) generate a vesselness image (probability of having a vessel at any given voxel). The latter is used to construct a weighting factor that modifies an intensity-based metric so as to give preference to vascular structures. The vesselness image is used to perform non-rigid vascular registration within the context of a mutual information metric. Alternatively, Heldmann and Papenberg (2009) develop a combined two-dimensional morphological and Gaussian scale-space to allow the non-rigid image registration of tubular tree-like structures. The goal of their multilevel framework is to avoid locally ambiguous mappings between parts of the images by removing morphological details, but also finding a global optimal solution by spreading remaining local information using Gaussian scaling. Hameeteman et al. (2010) rely on image intensities to apply a non-rigid registration of the carotid artery. The resulting transformation is used to calculate the distensibility of the carotid artery at a given position.

5.2.2 Dynamic Vascular Image Registration

The registration of dynamic vascular images, such as the registration of coronary arteries, implies additional difficulties. The presence of motion can produce strong displacements of the vessels and it also causes image artifacts, introducing additional problems. In fact, dynamic cardiac (3D+t) registration is rather new (Lapp et al., 2004; Lombaert and Cheriet, 2010; Wilson et al., 2006). Most of the existing methods for dynamic coronary arteries registration rely on 2D techniques applied to coronary angiography (2D+t) (Ruijters et al., 2009; Turgeon et al., 2005). Only a few methods tackle the problem of 3D+t images.

One of the first attempts was the work from Laguitton et al. (2006), which simply tried to follow points placed at arteries locations through different time-frames. The local characteristics of the vessel are estimated on the first time-frame (3D volume) of the cardiac dataset and then used to track the vessel along the sequence. The correspondence between two volumes is solved through a region-matching based on a criterion of minimal distance combining moment-based descriptors with intensity information. Metz et al. (2009) adapted the non-rigid registration method proposed by Wierzbicki et al. (2004) to obtain a 4D deformation model of the coronary arteries on CTA images at different ECG phases. Similarly, Zhang et al. (2009b) include a non-rigid image registration using free-form deformation step for coronary artery segmentation and tracking through dynamic sequences. The goal of the registration step is to include prior probability information from previous time-frames to increase the robustness of the temporal tracking.

The same non-rigid approach was then used to model coronary artery motion (Zhang et al., 2009a). In this work, arteries are first enhanced using a multiscale Hessian-based vessel enhancement filter (Frangi et al., 1998) to facilitate the identification of coronaries. Recently, a study was performed to compare the method with a template matching scheme (Zhang et al., 2010).

All the previously described approaches tackle registration by using the information of the complete image. Contrarily, our proposal seeks to reduce the size of both I_F and I_M VOIs that exclusively contain structures of interest. In our case, we are exclusively interested in potential vascular lesions (*i.e.* plaque). Structures such as the ribs, the lungs or even healthy sections of the arteries do not provide useful information for coronary artery disease detection. Even more, cardiac CT images are typically large, which results in large computational times when the whole volume is used for registration. For this matter, our proposed algorithm focuses in selecting only candidate regions likely to have coronary disease, *i.e.* regions that have been detected as anomalous in a cardiac time-frame. Then, it uses *a priori* knowledge to define a region that matches, with high probability, the potential lesion zone in another time-frame. Once the two VOIs are identified, registration is performed among them so as to compare the information contained in both subvolumes.

5.3 Our Registration Approach

The proposed registration approach (Figure 5.2) performs three different steps to achieve final registration of potential vascular lesions: automatic definition of the volume of interest (VOI), coarse rigid registration and deformable registration. The first stage makes use of a priori knowledge to construct a VOI at two different cardiac phases that should be then registered. The second step aims at reducing the displacement between the structures from both images. This step is necessary to bring the vessels close enough. The obtained transform is used to initialize the deformable registration step. At this last step, actual deformations between the structures are captured. In practice, the first time-frame used typically corresponds to the diastole and the second one to the systole. This is respectively referred to as 75% and 40% of the heart beat in ECG-synchronized reconstructions.

5.3.1 Automatic VOI Definition

The aim of this stage is to reduce the size of both I_F and I_M by defining VOIs that exclusively contain potential lesions. According to Shechter et al. (2006), the maximum displacement magnitude that can be observed on a coronary branch, has been evaluated to be about 1 cm between two temporal instants. Therefore, we consider that if we identify a lesion at a time-frame, it is possible to find it, with high probability, at a second time-frame within a distance of 1cm from its original location. We use this information to define an algorithm that constructs two VOIs to be registered using the anomaly detection information (see Chapters 2 and 4). The main steps of the algorithm

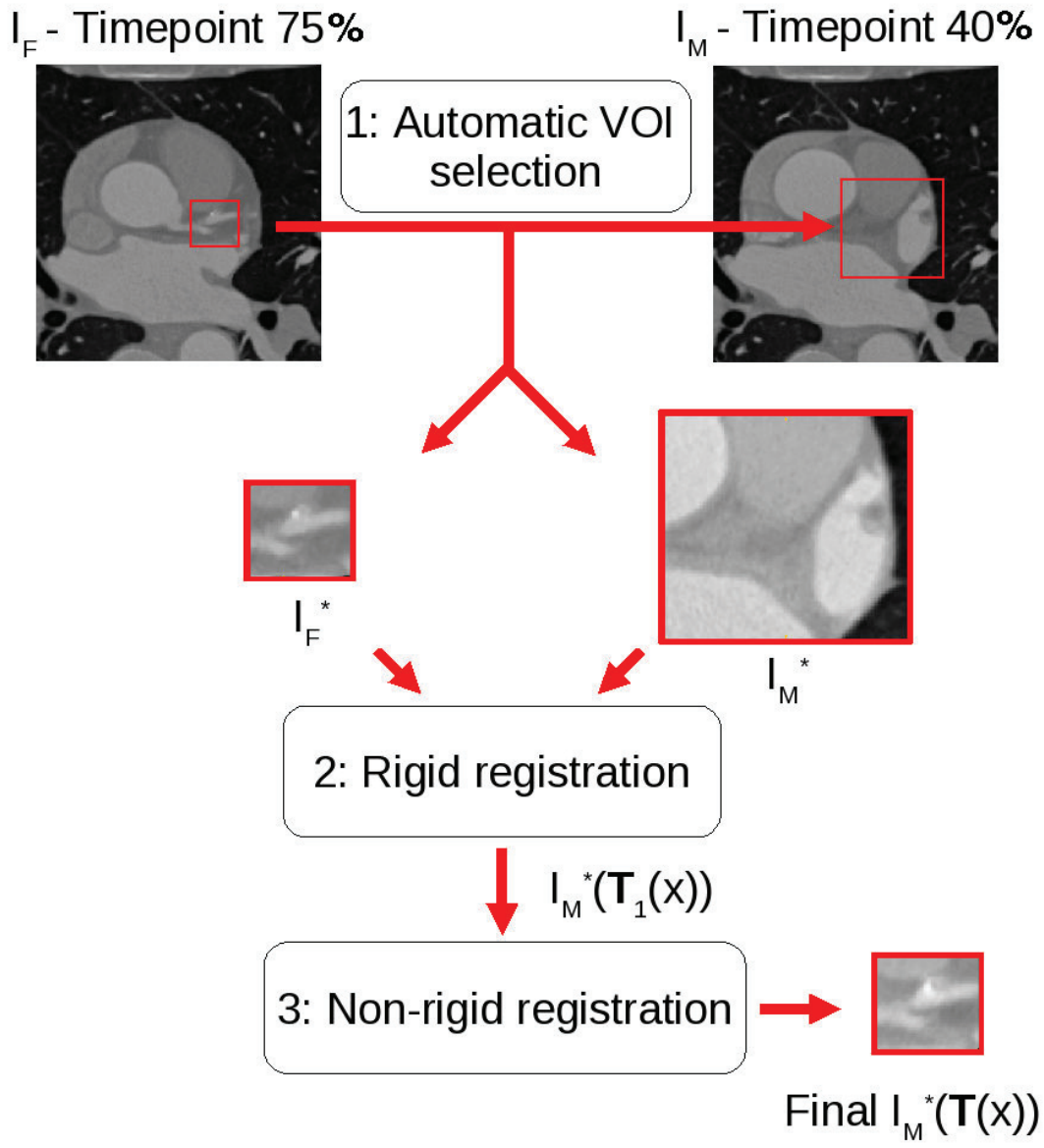


Figure 5.2: Overall registration process - Initial VOI selection (Section 5.3.1) is followed by a coarse rigid registration that brings the structures of interest closer (Section 5.3.2). Finally, a non-rigid registration is performed (Section 5.3.3). For simplicity, VOIs are presented in a 2D axial view. The final transformation is denoted as $\mathbf{T}(x)$ where $\mathbf{T}(x) = T_2(T_1(x))$, T_1 represents the coarse rigid registration and T_2 denotes the final non-rigid registration. Time-points usually correspond to diastole (75%) and systole (40%) respectively.

are described as follows (Figure 5.2):

1. Using the slices labeled as anomalous of a particular dataset and for **one time-frame** (see Chapters 2 and 4), potential lesions are identified. We define a potential lesion as a group of neighboring slices labeled as anomalous. We denote the subvolume of interest formed by the grouping as I_F^*.

2. Starting from the borders of I_F^*, a ring with thickness D_M is adjusted and mapped to the moving image I_M using world coordinates. The mapped VOI is denoted as I_M^*.

In practice, the construction of I_F^* requires an additional step. In many cases, the grouping of detected anomalous slices gives out a VOI that does not contain sufficient information that can be exploitable for the registration, leading to the so called aperture problem. In such a case, the reliability of the obtained registration would be very low. Therefore, it is necessary to add up a distance D_F around the detected potential lesion so that the information contained by I_F^* can avoid the aperture problem. The definition of values for D_M and D_F is discussed in Section 5.5.

5.3.2 Coarse Rigid Registration

The coarse rigid registration algorithm relies on a rigid body transformation

$$\mathbf{T}_1(\mathbf{x}) = \mathbf{R}(\mathbf{x} - \mathbf{c}) + \mathbf{t} + \mathbf{c} \tag{5.2}$$

where $\mathbf{R}$ denotes a rotation matrix, $\mathbf{c}$ the rotation center, $\mathbf{t}$ a translation and $\mathbf{x}$ represents a point in space. The image is treated as a rigid body that can be translated and rotated, but it cannot be scaled.

The registration is performed at a single resolution using an adaptive stochastic gradient descent optimizer (Klein et al., 2009) with a maximum of 500 iterations. In each iteration, 2048 samples are obtained from I_F^* using a random sampler (Klein et al., 2010) and a linear interpolator is used to evaluate the transformed I_M^*.

5.3.3 Deformable Registration

The deformable registration consists of a non-rigid transform using a free form deformation (FFD) model based on B-cubic splines (Rueckert et al., 1999):

$$\mathbf{T}_2(\mathbf{x}) = \sum_{l=0}^{3}\sum_{m=0}^{3}\sum_{n=0}^{3} B_l(u)B_m(v)B_n(w)\phi_{i+l,j+m,k+n} \tag{5.3}$$

where $\phi_{i+l,j+m,k+n}$ represent the control points of a mesh Φ and B_l, B_m, B_n represent the basis functions of the B-spline:

$$\begin{aligned} B_0(u) &= (1-u)^3/6 \\ B_1(u) &= (3u^3 - 6u^2 + 4)/6 \\ B_2(u) &= (-3u^3 + 3u^2 + 3u + 1)/6 \\ B_3(u) &= u^3/6 \end{aligned}$$

The registration process uses a coarse-to-fine multiresolution strategy. At each resolution, 500 iterations are, at most, executed by an adaptive stochastic gradient descent optimizer. In each iteration, 2048 samples are obtained from I_F^* using a random sampler and a linear interpolator is used to evaluate the transformed I_M^*. A 3^{rd} order B-spline interpolator is employed to apply the final deformation transform to I_M^*.

The registration stages were performed using `elastix` (Klein et al., 2010), a publicly available package for image registration [1].

5.3.4 Similarity Metrics

In this section we decribe a set of already existent similarity measures that permit to evaluate the quality of the alignment between the registered images. All the herein described candidate metrics were carefully evaluated (sec. 5.5.2) in order to select the best performing.

Mean Squared Differences (MSD). This metric computes the mean squared pixel-wise difference in intensity between I_F and a transformed version of I_M (here, we drop the asterisks in order to simplify notations of the sub-images). It is defined as:

$$MSD(\boldsymbol{\mu}, I_F, I_M) = \frac{1}{|\Omega_F|} \sum_{\mathbf{x}_i \in \Omega_F} (I_F(\mathbf{x}_i) - I_M(\mathbf{T}_{\boldsymbol{\mu}}(\mathbf{x}_i)))^2, \tag{5.4}$$

where $\boldsymbol{\mu}$ represents a vector containing the parameters associated to the transformation and Ω_F denotes the domain of I_F.

Normalized Correlation Coefficient (NCC). The metric computes the pixel-wise cross-correlation coefficient and normalizes it by the square root of the autocorrelation of the images:

$$NCC(\boldsymbol{\mu}, I_F, I_M) = \frac{\sum_{\mathbf{x}_i \in \Omega_F} \left(I_F(\mathbf{x}_i) - \bar{I}_F\right) \left(I_M\left(\mathbf{T}_{\boldsymbol{\mu}}(\mathbf{x}_i)\right) - \bar{I}_M\right)}{\sqrt{\sum_{\mathbf{x}_i \in \Omega_F} \left(I_F(\mathbf{x}_i) - \bar{I}_F\right)^2 \sum_{\mathbf{x}_i \in \Omega_F} \left(I_M\left(\mathbf{T}_{\boldsymbol{\mu}}(\mathbf{x}_i)\right) - \bar{I}_M\right)^2}}, \tag{5.5}$$

where $\bar{I}_F$ and $\bar{I}_M$ represent the average gray values of I_F and $I_M(\mathbf{T}_{\boldsymbol{\mu}}(\mathbf{x}))$ respectively.

Mutual Information (MI). Mutual information (Maes et al., 1997; Mattes et al., 2003; Viola and Wells, 1997) measures the information contributed to the overlapping volume by each image being registered together with the joint information. `elastix` uses the definition given by Thévenaz and Unser (2000):

$$MI(\boldsymbol{\mu}, I_F, I_M) = \sum_{m \in L_m} \sum_{f \in L_F} p(f, m; \boldsymbol{\mu}) \log_2 \left(\frac{p(f, m; \boldsymbol{\mu})}{p_F(f) p_M(m; \boldsymbol{\mu})} \right) \tag{5.6}$$

[1] http://elastix.isi.uu.nl

with L_F and L_M being sets of regularly spaced bin centers, p the discrete joint probability, and p_F and p_M the marginal discrete probabilities of I_F and I_M, obtained by summing p over m and f, respectively. The joint probabilities are estimated using B-spline Parzen windows.

Normalized Mutual Information (NMI). The normalized mutual information is defined as

$$NMI = \frac{(H(I_F) + H(I_M))}{H(I_F, I_M)} = 1 + \frac{MI(I_F, I_M)}{H(I_F, I_M)}, \tag{5.7}$$

where H denotes entropy.

5.4 Experimental Setup

The present section is devoted to describe all the required elements for the evaluation of the proposed registration methodology. First, we specify the characteristics of the data that was used in our experiments, followed by the performed experiments.

5.4.1 Data

A total of 40 CTA datasets were used along the evaluation. The datasets were acquired on a 64-row CT scanner (Brilliance 64 − Philips Healthcare, Cleveland, OH) with a standard scan protocol using the following parameters: 120 kV, 300 mAs, collimation 52×1.5 mm, rotation time 0.35 seconds and scan time 10-14 seconds. Image reconstructions were made with an in-plane pixel size of 0.37×0.37 mm^2, matrix size 512×512, slice thickness 0.9 mm, increment 0.45 mm, with an intermediate reconstruction kernel (B). 38 datasets contained 2 cardiac phases, namely, 40% and 75%. The remaining two contained 11 cardiac frames, corresponding to the following ECG phases:$\{0, 10, 20, 30, 40, 50, 60, 70, 75, 80, 90\}$ (here, we drop % for simplicity).

Out of the 40 datasets, 30 were selected exclusively for automatic VOI selection validation. More particularly, for the validation of the D_M parameter. The 10 remaining datasets were used to assess the registration stage. Moreover, 5 of the latter were used to determine the optimal value of D_F. It should be noted that only 10 datasets were used in the assessment of the registration stages because, as it will be seen in section 5.4.2, this assessment requires previously extracted vessel centerlines, which were only available for these 10 datasets.

Since image quality plays an important role in registration, the images used for this purpose were classified according to their quality as poor, moderate or good (Table 5.1). This information was used to assess the performance of the method when varying the image quality.

5.4.2 Experiments

In the following we describe the experiments that were performed to evaluate our approach. First, we focus in determining the optimal parameters for VOI selection.

Table 5.1: Cardiac CT datasets. Classification of the CT datasets used for evaluation, according to the number of timeframes and image quality.

Time-frames	Poor	Moderate	Good
$\{40, 75\}$	1	4	3
$\{0, 10, 20, 30, 40, 50, 60, 70, 75, 80, 90\}$	1	0	1

Then, using these parameters we register the subvolumes using the metrics described in 5.3.4 to determine the metric with the best response. Using the best configuration we compare our registration scheme with the one proposed by Zhang et al. (2009a, 2010) were cardiac phases are registered incrementally.

Automatic VOI selection

Subvolumes I_F^* were selected from the coronary arteries, namely, RCA, LAD and LCX on a diastolic time-frame (75%) of each dataset. At this stage, the potentially diseased arterial segment to be registered was manually selected by clicking at its endpoints. These segments were selected to make sure that the corresponding subvolumes contained a variety of situations that can result from the anomaly detection stage. This is, the resulting subvolumes should contain, among others, bifurcations, lesions and even normal sections (in a smaller proportion, in order to include possible false positives). Furthermore, for each artery, one volume should be in the proximal (P), another one in the distal (D) part of the artery, and the third one in-between (O).

Shechter et al. (2006) state that the maximum displacement magnitude that can be observed on a coronary branch between two time-frames is 1 cm, when the cardiac cycle is subdivided into 10 equally-spaced phases. However, our approach does not make use of subsequent time-frames. A set of different VOIs I_M^* were constructed using different values of D_M to determine the maximum displacement value between systolic and diastolic time-frames (40% and 75% respectively). Evaluated values were 1, 1.5, 2 and 2.5 cm.

Tuning of the parameter D_F was done after D_M was correctly parametrized. Different values of $D_F \in \{0.5, 1, 1.5\}$ cm were used to construct I_F^*. Using an already defined value for D_M, the two volumes I_F^* and I_M^* were registered and the obtained volume was evaluated.

The selection of optimal D_M and D_F was assessed through a visual inspection inspired from the work of Mattes et al. (2003) and by means of ITK-SNAP software (Yushkevich et al., 2006) (which allows spatial synchronization of volumes). For D_M definition, a VOI is denoted correct if the potential lesion in I_F was found in the subvolume $I_M^*(D_M)$. For D_F tuning, a VOI $I_F^*(D_F)$ is denoted correct if the resulting registered subvolume contains the same anatomic structures as the original subvolumes. The D_M and D_F values that lead to the maximum number of correct VOIs were kept.

Time-frame registration

Time-frame registration was first devoted to the selection of a similarity metric. Once a metric was selected, we performed an experiment to evaluate the pertinence of only using two time-frames for registration. Previous works from Metz et al. (2009); Zhang et al. (2010) have used an incremental registration approach, where time-frame t is registered to time-frame $t+1$ and the resulting transform is used to initialize the registration between frames t and $t+2$ and successively. In our approach, we avoid this incremental registration to reduce computational time. However, our experimentation aims to determine if there is a loss in quality when a direct registration between end-systole and end-diastole is done. To evaluate this aspect, we made use of the two datasets containing all the cardiac time-frames (see Table 5.1). We perform a direct registration between time-frames 40 and 75 and we compare it with an incremental registration.

Registration results assessment was done by evaluating the quality of the final registration result. Although our main interest are lesions, it is difficult to use them as landmarks to validate registration since it is not an easy task to determine with precision correspondence between spatial points at different temporal instants. To overcome this, we used vessel centerlines as an evaluation criterion. For this purpose, the centerlines of each artery were manually drawn and used as a reference.

For the selection of the similarity metric, ninety pairs of VOIs, I_F^* and I_M^* (3 VOIs on three arteries, namely, RCA, LCX and LAD, in 10 datasets), were generated through automatic VOI selection assuring each contained at least one of the arteries of interest. For incremental registration evaluation, the I_M^* VOIs were automatically generated at each of the intermediate phases used (70, 60, 50 and 40%).

After VOI generation, these were registered using the proposed approach. The obtained transformation was used to deform the centerline from I_M^* and the distance between the deformed centerline of I_M^* and the one from I_F^* was then measured as an indicator of vessel displacement. Let C_F be the vessel centerline from image I_F^* and C_M the deformed centerline from I_M^*, the average distance between centerlines proposed by Zhang et al. (2009a) was employed as a measure:

$$D(I_F^*, I_M^*, \mathbf{T}) = \frac{1}{N_F}\sum_{i=1}^{N_F}\|\mathbf{v}_i - l(\mathbf{v}_i, \mathbf{T}(I_M^*))\| + \frac{1}{N_M}\sum_{j=1}^{N_M}\|\mathbf{p}_j - l(\mathbf{p}_j, I_F^*)\|, \tag{5.8}$$

where N_F and N_M are the total number of vertices in C_F and C_M, respectively. For each vertex $\mathbf{v} \in C_F$, the function $l(\mathbf{v}, \mathbf{T}(I_M^*))$ calculates the closest vertex to $\mathbf{v}$ on a vessel C_M. Similarily, for each vertex $\mathbf{p} \in C_M$, the function $l(\mathbf{p}, (I_F^*))$ calculates the closest vertex to $\mathbf{p}$ on a vessel C_F.

Figure 5.3 (a) illustrates the way that the measurement $D(I_F^*, I_M^*, \mathbf{T})$ is computed given two axes of different lengths and how this is strongly penalized. In Figure 5.3 (b), we illustrate how this behavior affects the assessment of our registration process. Due to the different size of volumes I_F^*, I_M^*, the axis contained by one of the subvolumes

can be larger, resulting in a high distance score. To avoid this, we have selected to compute the metric in C_M, only between the points corresponding to the transformed extreme vertices of C_F.

5.5 Results

In this section we present the results obtained from automatic VOI definition and the registration assessment.

5.5.1 Automatic VOI Definition

Table 5.2 summarizes the results obtained from visual inspection of the generated VOIs I_M^*, when using different maximum displacement values D_M. Results show that 1 and 1.5 cm are not enough to capture the movement between end-diastole and end-systole. The highest number of errors occurs in the RCA. This is coherent with what has been reported in literature. The RCA is reported to be the artery with largest motion. VOIs formed by displacements of 2 and 2.5 cm succeeded in always containing the structure of interest. Although using 2.5 cm includes more information in the subvolume, our goal is to reduce the amount of processed information.

Table 5.2: Definition of D_M for automatic VOI selection. - For each evaluated displacement, the number of subvolumes I_M^* that contained the structures of interest contained in I_F^* is presented.

D_M	RCA			LAD			LCX		
	P	D	O	P	D	O	P	D	O
1 cm	15	0	11	30	21	30	22	17	21
1.5 cm	26	12	18	30	30	30	30	23	25
2 cm	30	30	30	30	30	30	30	30	30
2.5 cm	30	30	30	30	30	30	30	30	30

After defining the value for D_M, we proceeded to evaluate the required value of D_F. Table 5.3 summarizes the obtained results. As a reference, the table also presents the results of the registration when only the potential lesion is used and a larger VOI is not created (we denote it $D_F = 0$). These results allows us to confirm that the information contained by the VOI made up by the potential lesion is not sufficient to have good registration results.

From Table 5.3, it can be seen that even in some cases $D_F = 0.5$ cm, it is possible to obtain successful results. However, it is not always sufficient. Although $D_F = 1$ cm shows almost perfect results, we prefer to conservatively keep $D_F = 1.5$ cm for the construction of I_F^*.

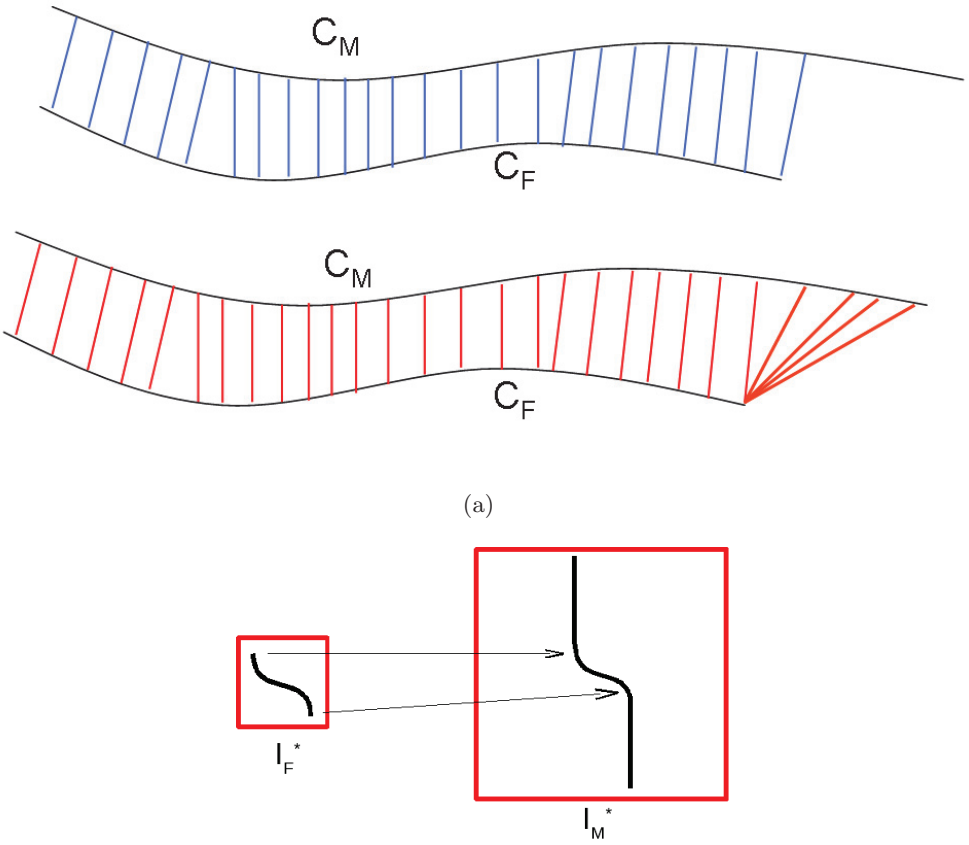


Figure 5.3: Registration Assessment Distance Measurement - (a) The distance measurement concept. At top, the blue lines represent the distance between a vertex $\mathbf{v} \in C_F$ and the the closest vertex in C_M found by function $l(\cdot)$. At the bottom, the red lines represent the distance between a vertex $\mathbf{p} \in C_M$ and the the closest vertex in C_F found by function $l(\cdot)$. As can be seen, from the right-most part of C_M, the fact that the axes are not of the same length is strongly penalized by the metric. Due to the different size of volumes I_F^*, I_M^*, the axis contained by one of the subvolumes can be larger, as shown in (b). To avoid strong penalization of the distance metric, the distance measure is only computed between the transformed extreme vertices of C_F, which are represented by two arrows in (b).

Table 5.3: Definition of D_F for automatic VOI selection. - For each evaluated distance, the number of correct registrations is presented.

D_F	RCA			LAD			LCX		
	P	D	O	P	D	O	P	D	O
0	0	0	0	0	0	0	0	0	0
0.5 cm	0	0	0	3	0	2	1	1	1
1 cm	5	4	5	5	5	5	5	5	5
1.5 cm	5	5	5	5	5	5	5	5	5

5.5.2 Time-frame Registration

Registration results in terms of residual centerline offset (Eq. 5.8) using different metrics are presented in Figure 5.4. As a reference, the initial displacement of the arteries (no registration) was also computed. For simplicity, results from each VOI are summarized on an artery-basis (instead of a segment-basis). Figure 5.5 illustrates an example of vessel centerline displacement correction. In general, the best results are obtained when using mutual information as a metric. In consequence, this metric was used in the subsequent experimentations. Figure 5.6 shows different examples of registered vessel segments using mutual information.

Although mutual information gives the best results in terms of final arterial displacement, from Figure 5.4 it can be seen that the residual offset at the RCA is somehow large. This can be explained by the fact that, for most cases, the initial displacement of the RCA between two time-frames is too large and the registration algorithm cannot recover from it. Actually, although I_M^* contains the structure of interest (potential lesion), as it was assessed when validating the value of D_M, it cannot completely include a subvolume equivalent to I_F^* and centered in the displaced pattern. Figure 5.7 illustrates the problem. Furthermore, an analysis of the final vessel displacement in terms of image quality (Figure 5.8) showed that the initial displacement has a higher incidence in the results than the image quality. The algorithm performs well on poor quality images but cannot recover excessively large displacements.

An evaluation of the residual offsets showed that the algorithm cannot sufficiently compensate initial displacements above 1.5 cm in both the RCA and LCX. The problem does not occur at the LAD. We consider that the registration does not compensate the initial displacement when the final distance between centerlines is above the average maximum vessel diameter. For the coronaries we consider this value to be 5 mm. The latter motivates our evaluation of an incremental approach that makes use of intermediate phases in order to register end-diastole to end-systole (Zhang et al., 2009a, 2010).

Table 5.4 compares the results obtained by direct and incremental frame registration on two datasets. The quality of dataset I was rated as poor, while the one from dataset II was classified as good. In 9 cases the direct frame registration outperforms incremental frame registration, while the latter is better in the remaining 9. Although the incremental approach performs better in most of the cases where the initial dis-

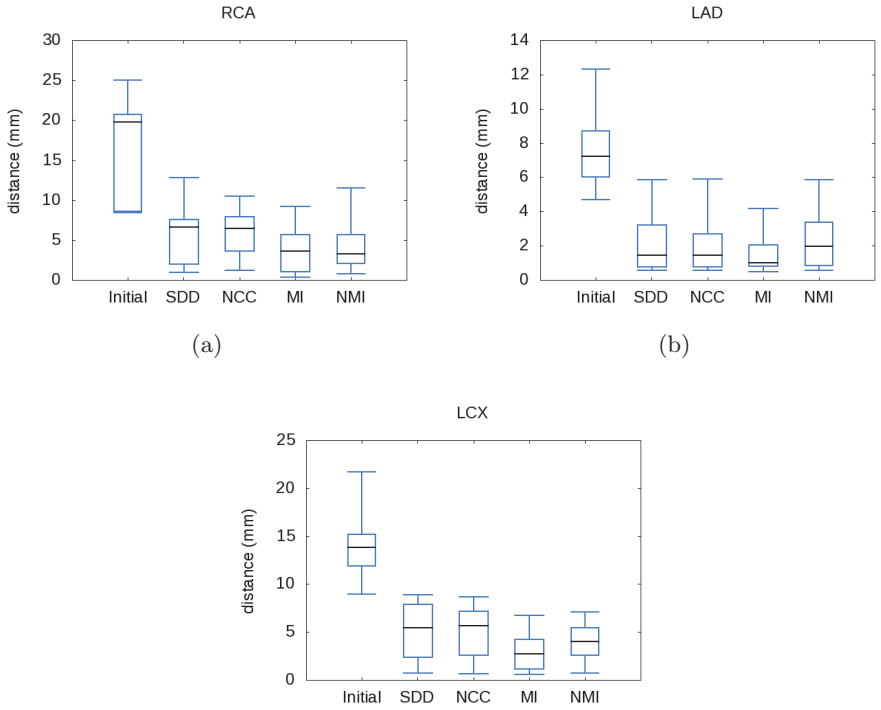


Figure 5.4: Coronary displacement measured through Eq. 5.8 - Initial coronary displacement and residual offset after non-rigid registration using SSD, NCC, MI and NMI metrics for (a) RCA, (b) LAD and (c) LCX arteries.

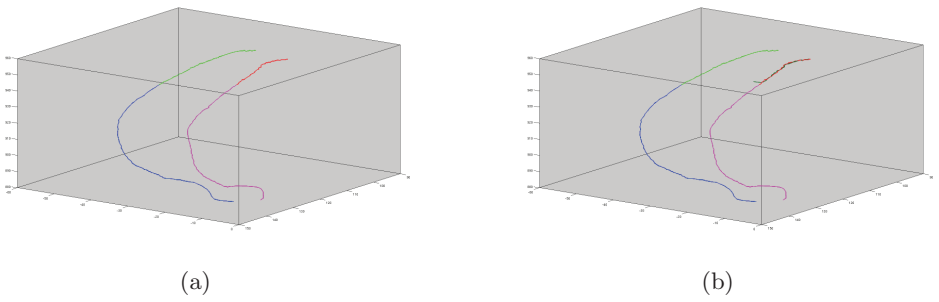
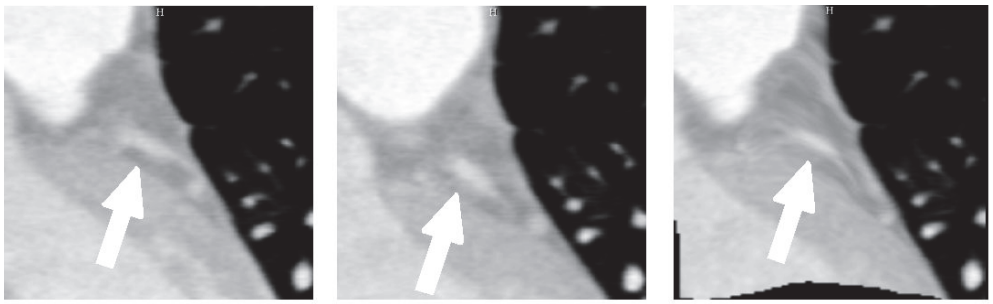
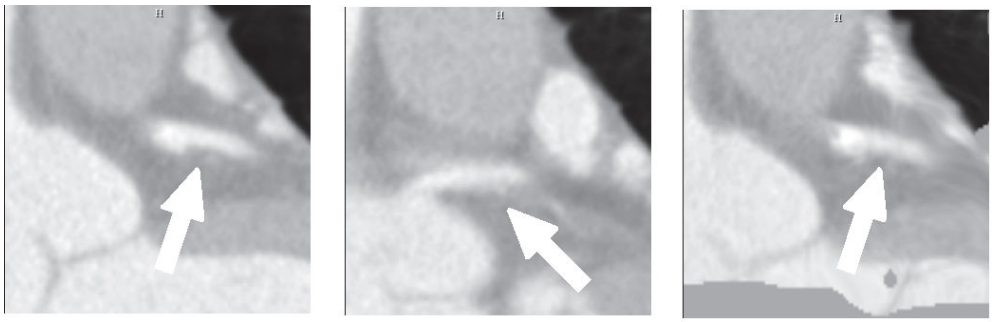


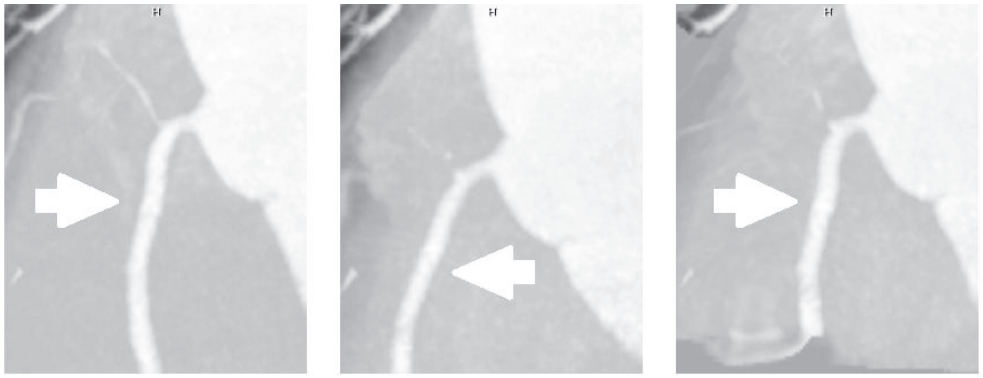
Figure 5.5: Example of coronary artery centerline displacement correction - (a) Initial displacement. The complete centerlines of the RCA at two time-frames are presented in blue and magenta, while green and red represent the respective centerlines segments falling within the VOI to be registered. (b) Dark green line (almost superimposed onto the red one) represents the result of applying the obtained transformation to the green centerline segment.



(a)

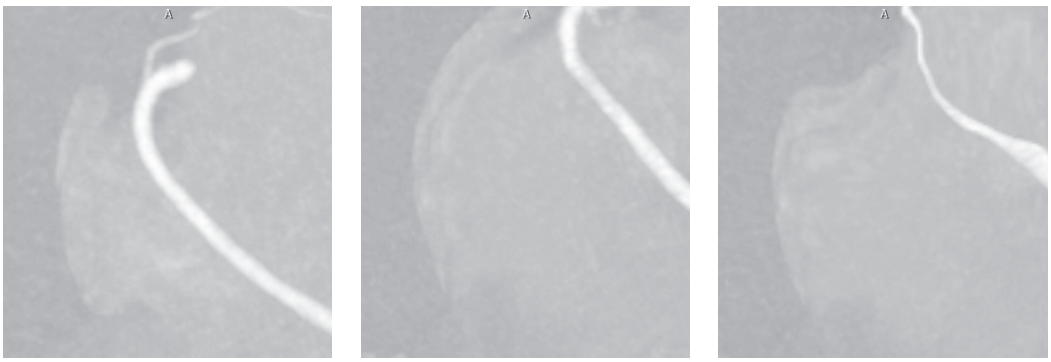


(b)



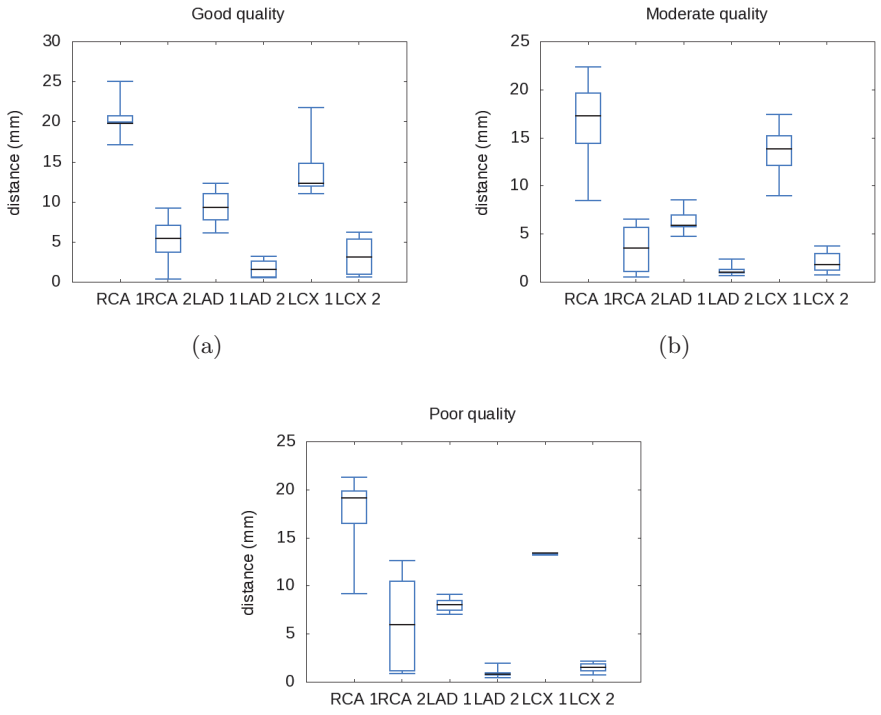
(c)

Figure 5.6: Vessel registration examples - From left to right images represent time-point 75, time-point 40 and registered image. Images have been cropped to the size of I_F^* to ease comparison. Arrows show the vessels of interest. (a) Coronal view of LAD segment registration (low quality image). Initial distance between centerlines is of 8.03 mm. After registration, the distance is reduced to 0.68 mm. (b) Coronal view of a LCX proximal segment (good quality image). Initial distance between segments is of 12.22 mm. After registration, the distance is reduced to 0.84 mm. (c) MIP view of a proximal RCA segment (good quality image) RCA at time-point 75 and RCA at time-point 40 have a separation of 24.98 mm. After the obtained registration, the distance is reduced to 0.55 mm.



(a) (b) (c)

Figure 5.7: MIP view of a misregistration problem - The initial distance between (a) the RCA segment at time-point 75 and (b) the RCA segment at time-point 40 is too large (35.03 mm). The obtained registration (c) cannot accomplish the structures alignment.



(a) (b) (c)

Figure 5.8: Final displacement comparison as a function of image quality - Initial and final displacements per artery when classifying images as (a) good, (b) moderate and (c) poor quality. On each plot, XXX 1 denotes initial displacement at artery XXX and XXX 2 the residual offset.

Table 5.4: Results of distance measuring using direct vs. incremental registration. Distance between the centerline at end-diastole (75%) of the cardiac cycle and the end-systole (40%) for three segments (proximal (P), distal (D) and other (O)) of the RCA, LAD and LCX. Results from two different datasets are presented.

Segment	Dataset I			Dataset II		
	Before transform	After direct transform	After incremental transform	Before transform	After direct transform	After incremental transform
RCA P	9.18	1.20	3.61	14.98	0.55	0.50
RCA D	5.42	1.78	1.26	18.25	5.01	4.60
RCA O	6.31	1.31	2.48	15.03	7.82	6.68
LAD P	9.08	0.44	0.55	10.90	0.49	0.75
LAD D	7.98	1.96	0.98	8.67	0.52	0.68
LAD O	8.03	0.68	1.00	10.52	0.49	0.83
LCX P	13.33	0.73	1.15	14.21	1.03	0.80
LCX D	13.27	2.16	0.87	14.93	3.30	2.92
LCX O	13.31	1.54	1.07	14.72	2.81	2.75

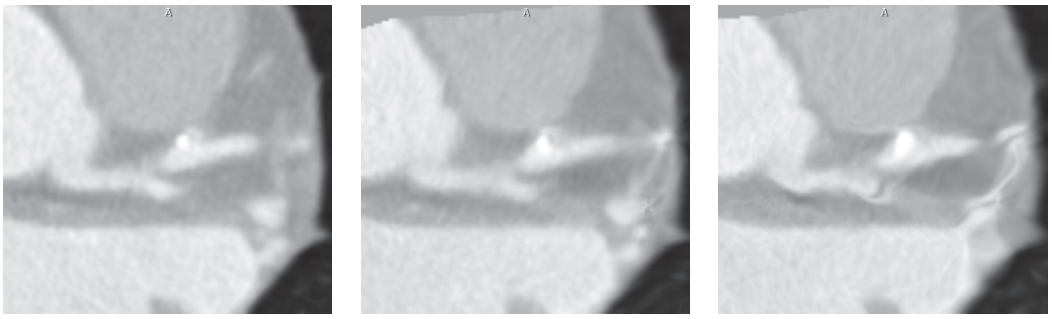
placement between vessels is large (as in RCA and LCX segments), the method does not perform well for very large distances, where the direct approach also fails. Previous works in literature (Zhang et al., 2009a, 2010) evaluating the incremental frame registration approach do not report such a large initial displacement. We believe that such a variability of the results can be influenced by the quality of images. Although a direct registration can hardly deal with a large displacement, the incremental approach can be affected by the residual errors that are propagated at every new registration step.

A comparison of the obtained images using each method, shows that the higher number of deformations applied in the incremental frame registration approach can cause excessive blurring in the final image and distortion of important structures such as calcified plaques. Figure 5.9 illustrates the problem. Although both approaches align the structures of interest (*i.e.* vessels and plaque components), the incremental frame approach tends to produce extremely smoothed images.

The overall process time for an artery segment was of 8 min in the direct approach and of 37 min for the incremental approach using a Pentium 4 with 3 GHz and 4 GB RAM. For this measurement, VOIs of similar sizes were selected ranging from 2.1 x 2.1 x 2.3 mm to 2.3 x 2.3 x 2.5 mm

5.6 Lesion Detection through Inter-phase Vessel Registration

In the previous sections we demonstrated the feasibility of our proposal by manually delimiting different VOIs, registering them and assessing the quality of the results through a distance measure. In the following, we connect the anomaly detection stage (Chapter 2) with the registration. More particularly, we replace the manually selected



(a) (b) (c)

Figure 5.9: Differences in the registration result using direct and incremental approaches - The calcification present in the fixed image (a), as well as the vessel general structure, are better conserved with the direct approach (b). The result obtained with the incremental approach (c) is more blurred.

VOIs by VOIs based on the anomaly detection stage. Our aim is to study how the information from the registered VOIs can be exploited to improve the detection.

Based on the procedure described in Section 5.3, we introduced some modifications in order to be able to compare the information from each of the time-frames. The final procedure is as follows:

1. The best performing configurations of DLD-SVM and LPU (Chapter 4) were applied to 10 datasets (see Section 5.4) at cardiac time-frame 75%. The latter is denoted the fixed image I_F.

2. The potential lesions detected on I_F were used to construct the different VOIs. These are denoted I_F^*.

3. The different I_F^* were used to obtain the corresponding I_M^* volumes using time-frame 40% (denoted I_M).

4. The different volumes, I_F^* and I_M^*, were then non-rigidly co-registered.

5. The best performing configurations of DLD-SVM and LPU were applied to all the obtained $\mathbf{T}(I_M^*)$.

6. Anomaly detection results from every I_F^* and I_M^* are confronted.

Table 5.5 summarizes the number of lesions found in the 10 images I_F as well as how many could be confirmed by the registration process.

Results from Table 5.5 show a pattern that has been seen at the classification stages: LPU has a better capacity to detect anomalies, while DLD-SVM is more conservative. Also, it is possible to see that in two cases (Datasets 04 and 05) an incorrectly detected lesion is rejected by the confrontation with the second time-frame. Lesion evaluation

Table 5.5: Anomaly detection confrontation. A confirmed anomaly means that at both time-frames an anomaly is detected.

Dataset	DLD-SVM		LPU	
	Detected	Confirmed	Detected	Confirmed
00	1	1	1	1
01	2	2	4	4
02	3	3	3	3
03	3	3	3	3
04	2	1	2	1
05	1	0	2	1
06	1	1	1	1
07	1	0	1	1
08	3	2	3	3
09	1	1	1	1

at two time-frames can help in the removal of false positives. However, the way the correspondence is formulated, it is not possible to detect initially undetected lesions.

Figure 5.10 shows three different examples of anomaly correspondence. In the first case, a detected anomaly in I_F is rejected at I_M (Figures 5.10 (a) and (b)), while in the other two cases it is confirmed.

An inconvenient of our approach is the lack of a higher precision in the registration. This is specially true for the right coronary artery (see Figures 5.10 (a), (b), (e) and (f)). In such a cases, the anomaly detection algorithm can be inaccurate owing to inaccuracies introduced by the registration process. In such a case, it is not useful to try to confirm / reject potential anomalies. However, in other cases, such as the one shown in Figures 5.10 (c), (d), the registration process can provide additional information. For instance, the moving image shows a better view of the calcification.

An interesting aspect that worths to be mentioned is related to bifurcations. In Table 5.5, we selected to remove from the statistics the bifurcations marked as anomalous by the DLD-SVM algorithm since we already know that the algorithm fails there. On the other hand, LPU does not detect bifurcations as anomalies but it tends to overestimate the lesions. Under such a condition, it could be desirable to combine the informations provided by the two methods using time-frame registration.

5.7 Discussion

In this chapter we have presented an approach for the registration of potential vascular anomalies. The method does not focus in the registration of the anomalies, as proposed in a previous work by Saur et al. (2008), but registers the potential anomaly and its surroundings to the (hopefully) corresponding fragment of the image reconstructed at a different time-frame. With this in mind, the proposed approach is more related to vascular registration than exclusively to anomaly registration.

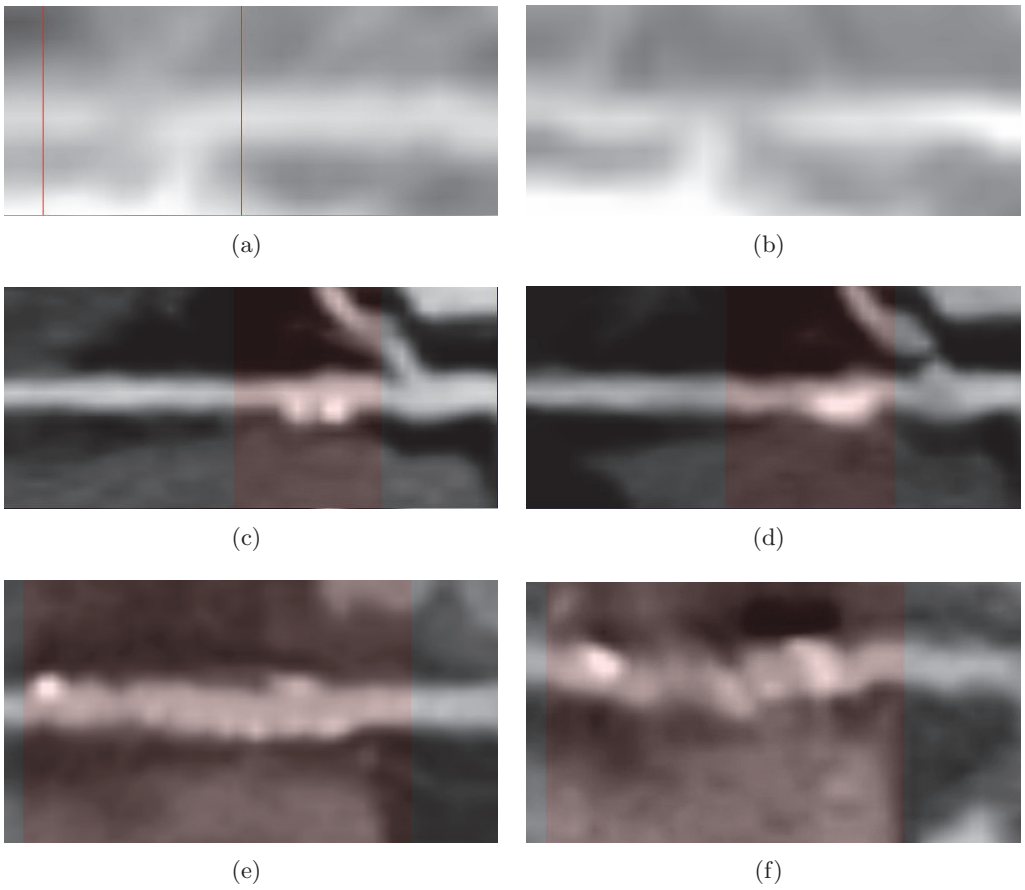


Figure 5.10: Examples of anomaly detection confrontation - Confronting anomaly detection results obtained through LPU. Left column represents time-frame 75% and the right one time-frame 40%. Anomalous sections are marked in translucent red, except in (a), where the anomaly corresponds to the section in-between the two red lines. (a) An anomaly is detected in the RCA, which is rejected in the second time-frame (b). A calcification is detected in the LAD (c) and confirmed in the second time-frame (d). Detection results in the RCA (e) are confirmed in the second time-frame (f). However, the registration result in the latter case is of poor quality.

The selected approach has the advantage that it relies solely on image intensities for registration. With this, it overcomes problems associated to vascular registration methods that require a good initial model/segmentation (Aylward et al., 2003; Chillet et al., 2003; Cool et al., 2003; Jomier and Aylward, 2004; Wong and Chung, 2006; Xue et al., 2006). In our application we are interested in regions where potential vascular lesions exist. Hence, we have proposed an automatic VOI definition that creates a region surrounding the possible lesion and maps the VOI to another time-frame. The advantages of defining a VOI for registration instead of using the whole image are two-fold. First, the computational time of the registration process is dramatically reduced. Second, a localized VOI avoids the interaction with large structures, such as the heart chambers, that usually interfere in vascular registration. In consequence, our method does not require to perform vascular enhancement before the registration as other methods do (Suh et al., 2010; Zhang et al., 2009a,b, 2010).

A total of 10 patients were used to optimize the automatic VOI selection. Then, 30 patients were used to validate the method showing a success rate of 100%. Additionally, 90 VOI pairs obtained from 10 different datasets were used to evaluate the registration stage. First, four similarity metrics were evaluated and optimized for the proposed task. Mutual information presented the best results over all the evaluated subvolumes. Using this metric, the method was compared with an incremental frame registration scheme that has been previously proposed in literature. Our results are comparable with this approach, while our method is less time consuming and requires less information (only two time-frames).

An interesting characteristic of this method relies in the fact that the image is subdivided into different VOIs which allows speeding up the process. Furthermore, the independence of each volume gives the possibility of developing a parallelized version of the method. This, combined with a GPU implementation of both rigid and non-rigid stages (Modat et al., 2010), could signify a reduction to the order of seconds of the overall registration stage, which would open perspectives for its application in clinical CAD. However, there is still one issue that needs to be solved in order to increase robustness of the method. The latter still fails to compensate large initial displacements. Although the incremental frame registration approach neither solves the problem, it shows better results in these cases. Therefore it seems as an alternative solution. An open possibility could be to use direct registration by default and switch to the incremental frame approach for large displacements.

The application of the method in order to confront vascular potential lesions so as to confirm / reject the diagnosis, showed that it can help to remove some false positives. However, we consider this part of our work as an interesting perspective that requires further investigations. First, the detection results obtained at the second time-frame can be seriously affected in the cases where the registration is of poor quality. Therefore, it is still necessary to improve the robustness of this stage. Second, the way the problem is formulated right now does not permit to correct false negatives. This would require a full anomaly detection process in both time-frames considered and an adequate way to combine the detections from both. We did not consider to perform

the full centerline extraction and anomaly detection at the second time-frame (systole) because, in general, the quality of the systolic image is not sufficient to extract a reliable centerline. Moreover, the lower quality of the second time-frame can introduce a large number of false positives. Therefore, in the systolic phase we only focus in a small portion around the suspected lesion. A possible path to improve both sensitivity and specificity may be oriented towards a combination of DLD-SVM and LPU, when using two time-frames, since these two methods have opposite behaviors in terms of sensitivity and specificity.

5. INTER-PHASE VESSEL SEGMENT REGISTRATION TO CORROBORATE ANOMALY DETECTION

6

Conclusions

In this work, we have discussed a methodological framework aimed at the detection of vascular lesions. Our main goal has been to propose a methodology that eases the diagnosis of vascular diseases and, more specifically, coronary heart disease (CHD). In the preceding chapters we have discussed the different elements that build up our framework and their issues when applying them in CHD diagnosis. Nearly each chapter has provided separate discussions about the achievements and the issues that can still be improved. In this chapter, we summarize the overall results and provide some general research perspectives.

6.1 Contributions

Figure 6.1 presents a scheme that summarizes the structure, as well as the main contributions of this work. The methodological contributions are attached to three main axes: the formulation of the lesion identification problem as an anomaly detection, the solution of such a problem through classification and the validation of potential lesions with inter-phase information.

Formulation of the Lesion Identification Task as an Anomaly Detection Problem

To tackle our particular problem we have chosen an approach different from the one followed in literature for lesion identification. We prefer to identify healthy vessel segments and to consider everything that deviates from thus identified normality as potentially diseased. Most of the approaches found in literature (see Section 2.2), tend to model the lesions. Since, by nature, lesions are heterogeneous, the approaches based on lesion modeling require user's manual interaction, the use of different algorithms, or different types of images for each type of lesions. Contrarily, our proposal is automatic, can handle different types of lesions and relies exclusively in computed tomography (CT) angiography images.

The implementation of our approach was based on a machine learning scheme to

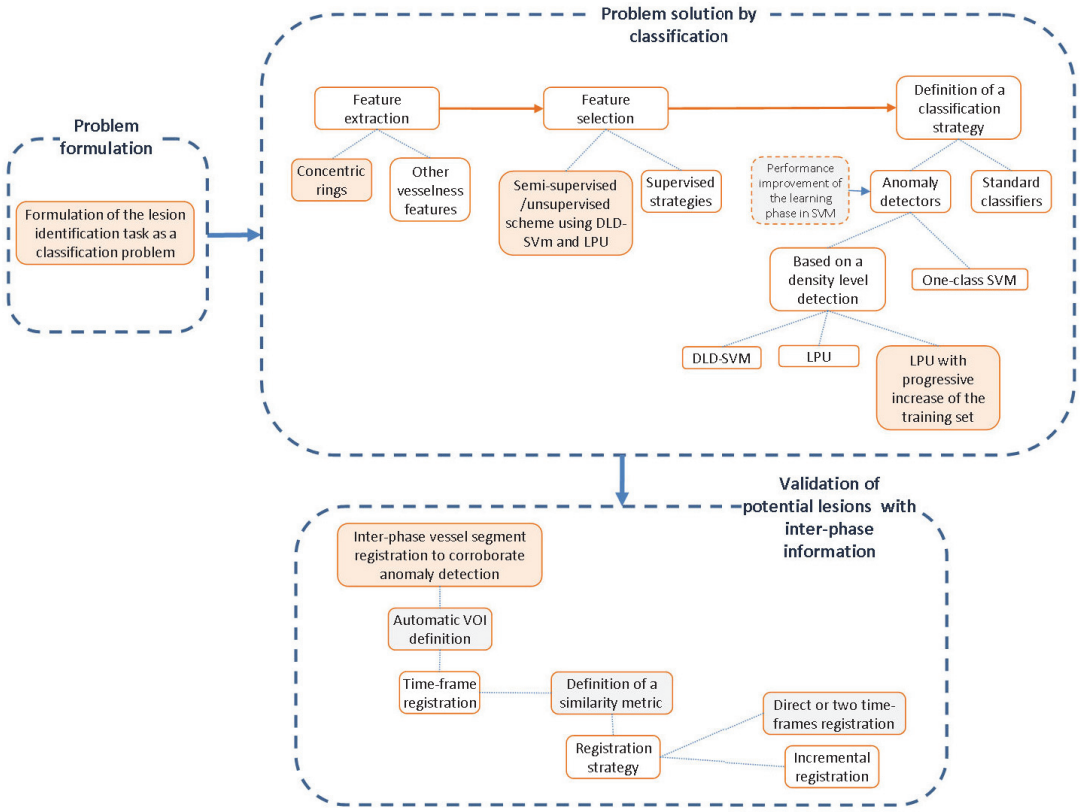


Figure 6.1: Summary of methodological contributions - The methodological contributions are attached to three main axes: the formulation of the lesion identification task as an anomaly detection problem, the solution of such a problem through classification and the validation of potential lesions with inter-phase information. The main contributions at each of the axis are marked in light red, while minor contributions are marked in gray.

classify normal and abnormal cross-sections orthogonal to a previously extracted centerline. Contrarily to the choice done by most of the existing ML-based approaches that follow a supervised classification scheme, we selected to avoid or minimize the use of labels. This decision was motivated by the difficulty to obtain accurate and representative labels. Moreover, supervised methods are incapable of detecting *new* emerging abnormalities. Since lesions are highly variable, unknown abnormal patterns can often appear. To overcome all of these difficulties, we formulated our problem as an anomaly detection problem (Chandola et al., 2007; Hodge and Austin, 2004; Markou and Singh, 2003a,b; Patcha and Park, 2007). To the best of our knowledge, such an approach is used for the first time used in this domain (Section 2.3). The selected perspective permitted us to focus in only one class (the normality). Everything that diverges from it is considered abnormal.

Solution of the Anomaly Detection Problem through Classification

Based on the definition of anomalies as low probability density regions, we made use of a density level detection (DLD) approach to formulate our anomaly detection problem. We selected to solve it through the use of support vector machines (SVM). The selection of a DLD approach was motivated by the possibility of avoiding or minimizing the dependence on labeled data. We assumed that two scenarios could occur: absolutely no information on the data was available (no labels) or labels of only one of the classes could be obtained. For the first case, we made use of the so called DLD-SVM algorithm (Steinwart et al., 2005a). For the second case, we employed a methodology called learning from positive and unlabeled samples (LPU) (Porter et al., 2009), which only required labels coming from one of the classes. For our particular case, we defined that these labels should come from normal cases (Section 2.3.1).

A classifier performance strongly depends on the quality of the input features. This work was also devoted to the study and selection of adequate features for our particular problem. Features for DLD problems should reflect the respectively high / low concentration regions of normal / abnormal samples. In Section 2.4.1 we introduced a novel intensity-based metric, denoted `Concentric rings`, that aimed to capture the intensity profiles and the axial symmetry of normal vessels, as well as the deviations from this behavior in diseased regions. In Chapter 3 we presented the experimental design of this metric through a series of experimentations on synthetic phantoms and its posterior validation on patients' CT data. Results showed good specificity and sensitivity values for both DLD-SVM and LPU. Moreover, the classification results with each algorithm demonstrated a substantial level of coincidence (Landis and Koch, 1977) w.r.t. observers' annotations when measured through the Kappa coefficient (Fleiss, 1971) (0.72% and 0.71%, respectively). The latter allows us to say that the metric is suitable for anomaly detection in the vascular domain.

To date, there is almost no state-of-the-art work on the selection of features for vascular lesion detection from a DLD point of view. Therefore, we also introduced a number of alternative candidate feature sets to be compared with our proposed metric. For this matter, we have proposed to use existent metrics that are commonly

used as vesselness criteria for vascular enhancement / segmentation. The metrics were parametrized such that they could be comparable (in terms of number of features, evaluated regions, among others) to the `Concentric rings` metric.

Since it is widely accepted that the combination of different types of features can improve the performance of classifiers, we derived an unsupervised/semi-supervised feature selection scheme that is based on the properties of our classifiers: DLD-SVM and LPU (Section 2.5.1). We made use of the empirical risk associated to these classifiers to define the optimal feature set criterion. The main advantage of this proposal was that it kept the unsupervised/semi-supervised nature of our method. Additionally, we investigated existing supervised feature selection methods with the perspective to compare their performance with our proposal.

Using the whole set of candidate features, in Chapter 4 we presented the results of feature selection through each of the proposed methods. The latter allows to say that the metric `Concentric rings` is suitable for anomaly detection. In DLD-SVM, it demonstrated to be the best performing, while it was ranked third with LPU not far from the best performing metrics. The results obtained with the unsupervised / semi-supervised proposal using DLD-SVM and LPU were the best and achieved specificities of 96.37 and 86.71%; and sensitivities of 76.52 and 83.84%, respectively. Moreover, the `Concentric rings` metric was always among the selected group of metrics, regardless the selection strategy. As can be seen from these results, as a general rule, we found that DLD-SVM tends to have better specificity than LPU, while LPU tends to have a higher sensitivity.

To further validate and justify our different methodological choices, we compared the performance of DLD-SVM and LPU with other existent state-of-the-art methods in classification. The evaluated methods consisted of two classical methods (Random forests and Soft-margin SVM) and a very popular anomaly detection method (One-class SVM). Our conclusion here was two-fold. First, that DLD-SVM and LPU outperform the other methods. Second, we could validate the relevance of the anomaly detection perspective. In every case, the three anomaly detection methods outperformed the conventional classification approaches.

One of the major concerns in the development of our framework was the additional difficulty that is imposed in anomaly related problems and which refers to class imbalance. In this type of problems, it is pretty common to have large databases of normal cases, while it is less common to find anomalous situations. This situation can affect negatively the performance of classifiers. For this matter, we introduced an algorithmic variation of the LPU implementation that defines a strategy to increment the training data (Section 2.3.4). The latter was assessed on phantom data and promising results were obtained (Section 4.5). Additionally, we have chosen to use an additional algorithm as the last step of the learning phase (Section 2.3.5). It is based on the variation of the b term of the SVM decision function and represents a negligible computational overhead. It is guaranteed that its use will not worsen the result, and in most cases a significant performance improvement was achieved.

Validation of Potential Lesions by Using Inter-Phase Information

The final element that builds up the core of our contribution is related to the validation of potential lesions by using information from multiple cardiac time-frames. Our proposal was inspired by the procedure followed by physicians in CHD diagnosis. Typically, they identify potential lesions at one time-frame and, afterwards, they confirm / reject it by inspecting a second time-frame. We followed this same workflow and use the anomalies detected at a particular time-frame to construct a volume of interest (VOI). Then, at a second time-frame a VOI with high probability of containing the same anatomic region was created. The two VOIs were then non-rigidly registered and presented to the observer so as to attract his/her attention and to aid the diagnosis by facilitating the local comparison between the images reconstructed at different time-points. We evaluated each of the stages of the method and the obtained results indicate that this approach is a promising perspective for CHD assessment.

Overall Evaluation

The implementation of the presented framework implied a large amount of work devoted to parameter tuning, validation and optimization of strategies. Figure 6.2 summarizes the tasks associated to the evaluation, validation and optimization of each of the elements constituting our work. Since a lot of work has been devoted to this task and, moreover, since this is a novel approach to handle vascular lesions, we believe that evaluation itself can be considered as an important contribution of this thesis.

To conclude we can say that the diagnosis of CHD is a complex task that involves many components. In this work we have proposed a methodological framework designed to ease this task and we have extensively evaluated it. We developed algorithms aiming CHD automation, while keeping a good level of specificity and sensitivity. However, there are issues that remain to be solved. This future work is discussed in the following section.

6.2 Perspectives

Several paths can be followed in order to improve our proposal. We consider that from the classifiers point of view it is difficult to introduce changes that can improve the performance of the method. As an example, it would be desirable that they manage the balanced error rate concept instead of the classification error. However, such a modification is not simple. There is a strong mathematical background supporting statistical learning and it is not an easy task to introduce a new concept. Therefore, to achieve improvements in the classification it is necessary to tackle other elements.

Perhaps the first direction that should be followed deals with the design of metrics to be used as input for the learning algorithms. One of the goals of our work has been to define a set of features that should be taken into account when designing a classifier for vascular lesion identification. Since the detection of lesions through ML techniques is a rather new approach, there is not a significant amount of literature that can guide

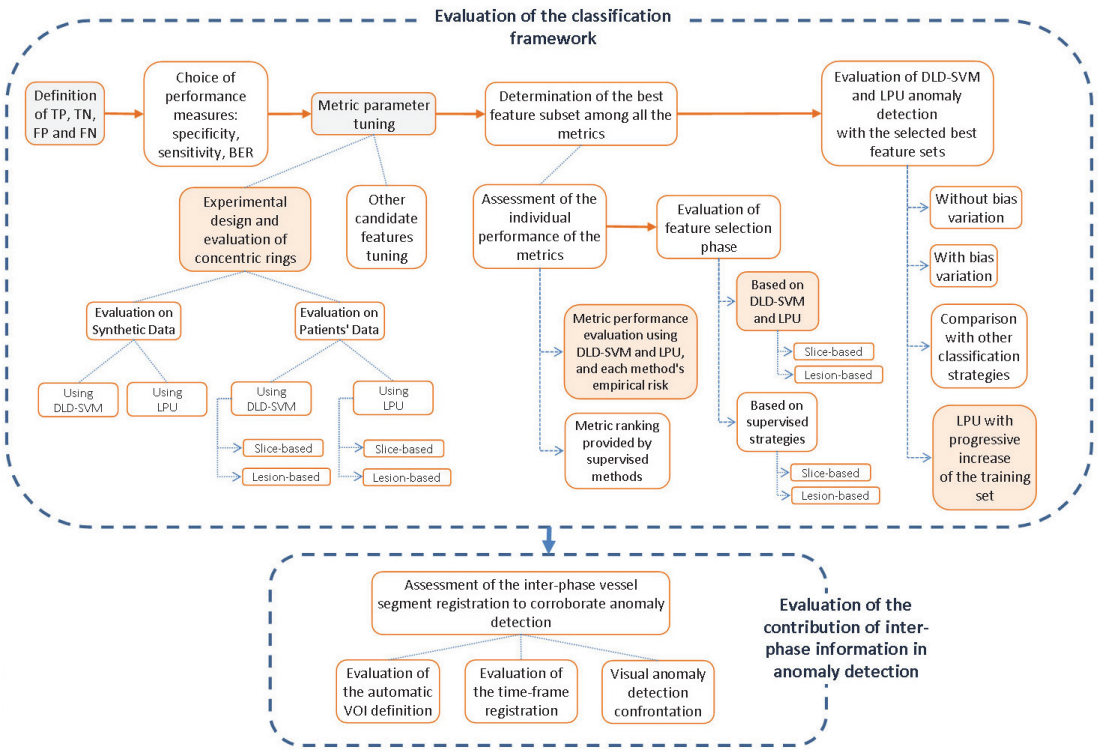


Figure 6.2: Evaluation of the different elements constituting the proposed methodological framework - Evaluation of the framework can be divided into two large albeit unequal groups: evaluation of the classification framework and evaluation of the use inter-phases information. The main contributions in the evaluation are highlighted in light red, while the minor ones appear in gray.

the selection of features. In this sense, our work can be considered as highly relevant. However, this should be just a starting point, and the possibility of evaluating other type of metrics is open. For instance, the possibility to include the extraction of 3D features is a natural extension of our approach.

Still in what concerns input features, we consider that the `Concentric rings` can be extended so that it takes into account the progressive narrowing of vessels. Up to now, the metric uses a fixed radius to extract features all along the vessel axis centerline. This can introduce undesired information that can affect the classifiers performance. Also, the metric can be extended to deal with 3D information in what could be considered as a concentric cylinder.

An additional way to improve performance could be the evaluation of further techniques of post-processing that enhance the resulting classification. As an example, the false positives introduced by bifurcations in DLD-SVM could be eliminated by including a post-processing stage that reviews the obtained labels. For this purpose, it would be necessary to develop a metric that can handle bifurcations. Generally speaking, bifurcations are a common problem in vascular literature. A similar post-processing approach could be thought for another common source of false positives, namely the vicinity of lesions.

The inherent weaknesses of the classifiers can be diminished by obtaining a large amount of training data. Although our proposal has the advantage of not requiring labeled data (or only partially labeled data), access to labeled data is still a major concern in medical image processing, at least for the purpose of validation. For this reason, we were only able to demonstrate the finite sample effects on synthetic data, but the lack of a sufficiently large database of patient data did not permit its evaluation on real cases. Therefore, we consider that initiatives such as those of Grand Challenges [1] should be highly encouraged.

From a more global view, our final aim is to assist clinicians in the diagnosis of vascular diseases. With this in mind, it is possible to say that our method is suitable for this purpose. However, there are still some issues that need to be solved so that it can be fully applicable to clinical routine. First of all, our method should be associated with a robust centerline extraction algorithm. Let us recall that our main target was not to develop a centerline extraction algorithm and that we wanted to evaluate the classification stage, while isolating the errors that can be introduced by an erroneous centerline. However, there are plenty of methods in literature that have a high performance (Friman et al., 2010; Lesage et al., 2009b) and that can be included in our framework. An advantage of our proposed framework is that its different elements are separated. These characteristic allows the possibility of choosing a state-of-the-art centerline extraction algorithm, and of replacing it in future by another more accurate one. A second issue that should be solved is related to the computational times of the classification framework. ML approaches are by nature computationally expensive. However, we consider that feature extraction, feature selection and the training stages are highly parallelizable. Therefore, a GPU implementation of these stages could be

[1] http://www.grand-challenge.org/

foreseen to reduce times.

As a final improvement, we consider that further work should be carried out to integrate the information obtained in the registration stage. Up-to-now we perform a registration that allows a fast comparison of locations at two different time-points. This proposal has been inspired from the way clinicians evaluate cardiac images. However, our method only to reduces the time required to find the different "potential lesion" locations and possibly reject false positives. The final diagnostic is given by the clinician. Although we consider the latter should remain that way, there is still information that can be exploited. First, the opposite behavior of DLD-SVM and LPU may be exploited: a good alternative could be to combine them by means of the time-frames in order to improve the precision in the detection. Second, information from the registration, *e.g.* its reliability quantified by the resemblance between the VOIs, amount of deformation, etc., may be exploited to provide more assessment elements to the clinician.

6.3 Publications

International Journal Articles

- K. Hameeteman, **M.A. Zuluaga**, M. Freiman, L. Joskowicz, W.J. Niessen, T. van Walsum *et al.* Evaluation Framework for Carotid Bifurcation Lumen Segmentation and Stenosis Grading. Submitted to: *Medical Image Analysis.* 2010. **Accepted for publication.**

- **MA. Zuluaga**, IE. Magnin, M. Hernández Hoyos, EJF. Delgado, F. Lozano and M. Orkisz. Automatic detection of abnormal vascular cross-sections based on Density Level Detection and Support Vector Machines. *International Journal Of Computer Assisted Radiology And Surgery.* 2010. **In press.**

National Journal Articles

- **MA. Zuluaga**, M. Orkisz, E. J. F. Delgado Leyton, V. Doré, A. Morales and M. Hernández Hoyos. Adaptación del algoritmo MARACAS para la segmentación de la bifurcación de la arteria carótida y cuantificación de estenosis en imágenes TAC. *Actas Biol Colombianas.* 2010. **In press.**

- **MA. Zuluaga**, M. Orkisz, and M. Hernández Hoyos. Mejoramiento de la segmentación arterial coronaria a través del modelamiento del volumen parcial en imágenes TAC. *Revista Colombiana de Radiología*, vol. 20, no. 3, pp. 2702-2707, 2009.

International Conference Articles

- K. Hameeteman, **MA. Zuluaga**, M. Freiman, A. van der Lugt, and T. van Walsum. An standardized evaluation framework for automated carotid bifurcation lumen segmentation and stenosis grading methods. In: *Radiological Society of*

North America, 96th Annual Meeting, Chicago, USA, pages in-press, November 2010. (abstract)

- **MA. Zuluaga**, E. J. F. Delgado Leyton, M. Hernández Hoyos, and M. Orkisz. Feature selection for SVM-based vascular anomaly detection. In: *MICCAI Workshop on Medical Computer Vision*, volume LNCS, Beijing, China, pages in-press, September 2010.

- **MA. Zuluaga**, E. J. F. Delgado Leyton, M. Hernández Hoyos, and M. Orkisz. Can the coronary artery centerline extraction in computed tomography images be improved by use of a partial volume model?. In: *Int Conf Comput Vision Graphics*, volume LNCS 6375, Warsaw, Poland, pages 385-392, September 2010. Springer.

- **MA. Zuluaga**, M. Orkisz, M. Hernández Hoyos, F. Lozano and IE. Magnin. Towards vascular abnormality detection: An SVM Approach. In: *Proceedings of Computer Assisted Radiology And Surgery.* p. S355-6, Genève, June 2010. (abstract)

- **MA. Zuluaga**, M.Orkisz, EJF. Delgado, V. Doré, A. Morales and M. Hernández Hoyos. Adaptations of MARACAS algorithm to the segmentation of the carotid bifurcation and stenosis quantification in CTA images. In: *MICCAI Workshop - 3D Segmentation in the Clinic: A Grand Challenge III.* London, Midas Journal, pp. 7 on-line, 2009. http://hdl.handle.net/10380/3102.

- K. Hameeteman, M. Freiman, **MA. Zuluaga**, L. Joskowicz, S. Rozie, MJ. Van Gils, L. Van den Borne, J. Sosna, P. Berman, N. Cohen, P. Douek, I. Sanchez, M. Aissat, A. Van der Lugt, GP. Krestin, W. Niessen and T. Van Walsum. Editorial: 3D Segmentation in the Clinic: A Grand Challenge III - Carotid Lumen Segmentation and Stenosis Grading Challenge In: *MICCAI Workshop - 3D Segmentation in the Clinic: A Grand Challenge III.* London, Midas Journal, pp. 15, 2009. http://hdl.handle.net/10380/3128

- **MA. Zuluaga**, M. Hernández Hoyos and M.Orkisz. Evaluation of partial volume effects in computer tomography for the improvement of coronary artery segmentation. In: *Proceedings of Computer Assisted Radiology And Surgery.* pp. S40-1, Berlin, June 2009. (abstract)

- M. Hernández Hoyos, **MA. Zuluaga**, M. Lozano, JC. Prieto, PC. Douek, IE. Magnin and M.Orkisz. Coronary centerline tracking in CT images with use of an elastic model and image moments. In: *3D Segmentation in the Clinic: a Grand Challenge II: MICCAI 2008 workshop proceedings*, New York, Midas Journal, pp. 8, 2008. http://hdl.handle.net/10380/1401.

National Conference Articles

- **MA. Zuluaga**, M. Orkisz, E. J. F. Delgado Leyton, V. Doré, A. Morales Pinzón, and M. Hernández Hoyos. Adaptación del algoritmo MARACAS para la segmentación de la bifurcación de la arteria carótida y cuantificación de estenosis en imágenes TAC. *Quinto Seminario de Ingeniería Biomédica y Quinto Seminario Internacional de Procesamiento de y Análisis de Imágenes Médicas*, Bogota, Colombia, pp. 5 in electronic proc., November, 2009.

- **MA. Zuluaga**, M. Hernández Hoyos, and M. Orkisz. Evaluación del efecto de volumen parcial en imágenes TAC para el mejoramiento de la segmentación arterial coronaria. *Quinto Seminario de Ingeniería Biomédica y Quinto Seminario Internacional de Procesamiento de y Análisis de Imágenes Médicas*, Bogota, Colombia, pp. 4 in electronic proc., November, 2009.

- F. Benmansour, L.D. Cohen, E.E. Dávila-Serrano, P.C. Douek, M. Orkisz, and **MA. Zuluaga**. New Interactive Methods for Tubular Structure Segmentation on Medical Images. In: *12th ISRACAS (Israeli Symposium on Computer-Aided Surgery, Medical Robotics, and Medical Imaging)*, Tel-Aviv, Israel, pp. 10 in CD, May, 2009.

References

S. Achenbach. Cardiac CT: State of the Art for the Detection of Coronary Arterial Stenosis. *Journal of Cardiovascular Computed Tomography*, 1:3–20, 2007. 4, 5, 6, 9

S. Achenbach, F. Moselewski, D. Ropers, M. Ferencik, U. Hoffman, B. MacNeill, K. Pohhle, U. Baum, K. Anders, I. kyung Jang, W. G. Daniel, and T. J. Brady. Detection of Calcified and Noncalcified Coronary Atherosclerotic Plaque by Contrast-Enhanced, Submillimeter Multidetector Spiral Computed Tomography: A Segment-Based Comparison With Intravascular Ultrasound. *Circulation*, 109:14–17, 2004. 11

I. Adame, R. van der Geest, B. Wasserman, M. Mohamed, J. Reiber, and B. Lelieveldt. Automatic Plaque Characterization and Vessel Wall Segmentation in Magnetic Resonance Images of Atherosclerotic Carotid Arteries. In J. M. Fitzpatrick and M. Sonka, editors, *Proceedings of SPIE Medical Imaging 2004: Image processing*, volume 5370, pages 265–273, San Diego, CA, February 2004. 21

E. Amaldi and V. Kann. On the Approximation of Minimizing Non Zero Variables or Unsatisfied Relations in Linear Systems. *Theoretical Computer Science*, 209:237–260, 1998. 43

E. Arnoldi, M. Gebregziabher, U. J. Schoepf, R. Goldenberg, L. Ramos-Duran, P. L. Zwerner, K. Nikolaou, M. F. Reiser, P. Costello, and C. Thilo. Automated Computer-aided Stenosis Detection at Coronary CT Angiography: Initial Experience. *European Radiology*, 20(5):1160–1167, 2010. 23, 60, 61

S. Aylward and E. Bullit. Initialization, Noise, Singularities, and Scale in Height Ridge Traversal for Tubular Object Centerline Extraction. *IEEE Transactions on Medical Imaging*, 21:61–75, 2002. 37

S. R. Aylward, S. Weeks, and E. Bullit. Analysis of the Parameter Space of a Metric for Registering 3D Vascular Images. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI*, volume 2208 of *LNCS*, pages 932–939, 2001. 113

S. R. Aylward, J. Jomier, S. Weeks, and E. Bullit. Registration and Analysis of Vascular Images. *International Journal of Computer Vision*, 55(2):123–138, 2003. 113, 132

V. Barnett and T. Lewis. *Outliers in Statistical Data*. John Wiley & Sons, 3^{rd} edition, 1994. 19, 24

REFERENCES

R. Bellman. *Adaptive Control Processes*. Princeton University Press, 1961. 42

A. Ben-Hur and I. Guyon. Detecting Stable Clusters Using Principal Component Analysis. In M. Brownstein and A. Kohudursky, editors, *Functional Genomics: Methods and Protocols*, pages 159–182, 2003. 44

H. Bennink, H. van Assen, G. Streekstra, R. ter Wee, J. Spaan, and B. ter Haar Romeny. A Novel 3D Multi-scale Lineness Filter for Vessel Detection. In *Medical Image Computing and Computer-Assited Intervention - MICCAI*, volume Part II of *LNCS 4792*, pages 436–443, 2007. 37

C. Boldak, Y. Rolland, C. Toumoulin, and J. Coatrieux. An Improved Model-based Vessel Tracking Algortihm with Application to Computed Tomography Angiography. *Journal of Biocybernetics and Biomedcal Engineering*, 23:41–63, 2003. 41

L. Breiman. Random Forests. *Machine Learning*, 45(1):5–32, 2001. 46, 47, 102

L. Breiman. *Manual on Setting Up, Using, and Understanding Random Forests*, 3.1 edition, 2002. 48

H. Brodoefel, A. Reimann, M. Heuschmid, I. Tsiflikas, A. F. Kopp, S. Schroeder, C. D. Claussen, M. E. Clouse, and C. Burgstahler. Characterization of Coronary Atherosclerosis by Dual-source Computed Tomography and HU-based Color Mapping: a Pilot Study. *European Radiology*, 18(11):2466–2474, 2008. 8

G. Brunner, U. Kurkure, D. R. Chittajallu, R. P. Yalamanchili, and I. A. Kakadiaris. Toward Unsupervised Classification of Calcified Arterial Lesions. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI*, volume 5241 of *LNCS*, pages 144–152, New York, USA, 2008. 21, 24, 44

C. J. Burges. A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery*, 2:121–167, 1998. 160

P. Carrascosa, C. Capunay, J. Parodi, L. Padilla, P. Johnson, J. Carrascosa, S. Chandra, D. Smith, and J. Belardi. General Utilities of Multislice Tomography in the Cardiac Field. *Herz*, 28(1):44–51, 2003. 8

C. Caussin, A. Ohanessian, S. Ghostine, L. Jacq, B. Lancelin, G. Dambrin, A. Sigal-Cinqualbre, C.-Y. Angel, and J. F. Paul. Characterization of Vulnerable Nonstenotic Plaque with 16-Slice Computed Tomography Compared With Intravascular Ultrasound. *The American Journal of Cardiology*, 94:99–104, July 2004. 11

V. Chandola, A. Banerjee, and V. Kumar. Anomaly Detection: A Survey. Technical Report TR 07-017, University of Minnesota, August 2007. 19, 24, 25, 137

C.-C. Chang and C.-J. Lin. *LIBSVM: a library for support vector machines*, 2001. URL `http://www.csie.ntu.edu.tw/~cjlin/libsvm`. 55, 56, 57, 58, 59

A. Charnoz, V. Agnus, G. Malandain, S. Nicolau, M. Tajine, and L. Soler. Design of Robust Vascular Tree Matching: Validation on Liver. In *Information Processing in Medical Imaging (IPMI)*, pages 443–455, 2005. 113

N. V. Chawla, A. Lazarevic, L. O. Hall, and K. W. Bowyer. SMOTEBoost: Improving Prediction of the Minority Class in Boosting. In *Proceedings of the Principles of Knowledge Discovery in Databases, PKDD-2003*, pages 107–119, 2003. 50

Y.-W. Chen and C.-J. Lin. Combining SVMs with Various Feature Selection Strategies. In I. Guyon, M. Nikravesh, S. Gunn, and L. Zadeh, editors, *Feature Extraction*, Studies in Fuzziness and Soft Computing, pages 315–324. Springer Berlin / Heidelberg, 2006. 46, 47

D. Chillet, J. Jomier, D. Cool, and S. Aylward. Vascular Atlas Formation Using a Vessel-to-Image Affine Registration Method. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI*, volume 2878 of *LNCS*, pages 335–342, 2003. 113, 132

J. Cohen. A Coefficient of Agreement for Nominal Scales. *Educational and physiological measurement*, XX(1):37–46, 1960. 64

D. Cool, D. Chillet, J. Kim, J.-P. Guyon, M. Foskey, and S. Aylward. Tissue-Based Affine Registration of Brain Images to Form a Vascular Density Atlas. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI*, volume 2879 of *LNCS*, pages 9–15, 2003. 113, 132

C. Cortes and V. Vapnik. Support-Vector Networks. *Machine Learning*, 20(3):273–297, 1995. 28, 58, 102, 161

T. de Weert, C. de Monyé, E. Meijering, R. Booij, W. J. Niessen, D. W. J. Dippel, and A. van der Lugt. Assessment of Atherosclerotic Carotid Plaque Volume with Multidetector Computed Tomography Angiography. *International Journal of Cardiovascular Imaging*, 24:751–759, 2008. 7

J. Dehmeshki, X. Ye, H. Amin, M. Abaei, X. Y. Lin, and S. D. Qanadli. Volumetric Quantification of Atherosclerotic Plaque in CT Considering Partial Volume Effect. *IEEE Transactions on Medical Imaging*, 26(3):273–282, March 2007. 21

T. Dietterich. Overfitting and Undercomputing in Machine Learning. *Computing Surveys*, 27:326–327, 1995. 43

J. L. Fleiss. Measuring Nominal Scale Agreement Among Many Raters. *Physiological Bulletin*, 76(5):378–382, 1971. 65, 137

C. Florin, N. Paragios, and J. Williams. Particle Filters, a Quasi-Monte Carlo Solution for Segmentation of Coronaries. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI*, volume 3749 of *LNCS*, pages 246–253, 2005. 39, 40

C. Florin, N. Paragios, and J. Williams. Globally Optimal Active Contours, Sequential Monte Carlo and On-line Learning for Vessel Sementation. In *Computer Vision - ECCV 2006*, pages 476–489, 2006. 39

A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever. Multiscale Vessel Enhancement Filtering. In W. Wells, A.Colchester, and S.L.Delp, editors, *Medical Image Computing and Computer-Assisted Intervention - MICCAI*, volume 1496 of *LNCS*, pages 130–137, 1998. 37, 115

Y. Freund and R. Schapire. A Decision-theoretic Generalization of On-line Learning and an Application to Boosting. In *Proceedings of the Second European Conference on Computational Learning Theory*, pages 23–37, 1995. 22

Y. Fridman. *Extracting Branching Object Geometry via Cores*. PhD thesis, University of North Carolina at Chapel Hill, 2004. 37, 78

Y. Fridman, S. M. Pizer, S. Aylward, and E. Bullitt. Segmenting 3D Branching Tubular Structures Using Cores. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI*, volume 2879 of *LNCS*, pages 570–577, 2003. 38

O. Friman, M. Hindennach, C. Kühnel, and H.-O. Peitgen. Multiple Hypothesis Template Tracking of Small 3D Vessel Structures. *Medical Image Analysis*, 14(2):160–171, April 2010. 141

K. Fukunaga. *Introduction to Statistical Pattern Recognition*. Academic Press, 2^{nd} edition, 1972. 32

N. Funabashi, Y. Kobayashi, M. Perlroth, and G. D. Rubin. Coronary Artery: Quantitative Evaluation of Normal Diameter Determined with Electron-Beam CT Compared with Cine Coronary Angiography - Initial Experience. *Radiology*, 226(1):263–271, 2003. 65

H. Gray. *Anatomy of the Human Body*. Philadelphia: Lea & Febiger, 1918, 20^{th} edition, 2000. URL `www.bartleby.com/107`. 2

I. Guyon and A. Elisseeff. An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research*, 3:1157–1182, 2003. 43, 44, 45, 49, 77

I. Guyon, J. Weston, S. Barnhil, and V. Vapnik. Gene Selection for Cancer Classification using Support Vector Machines. *Machine Learning*, 46(1-3):389–422, 2002. 46, 49, 94

K. Hameeteman, S. Rozie, C. Metz, S. Klein, T. van Walsum, A. van der Lugt, and W. Niessen. Automated Carotid Artery Distensibility Measurements from CTA Using Nonrigid Registration. In *IEEE International Symposium on Biomedical Imaging - ISBI*, pages 13–16, 2010. 114

X. He, D. Cai, and P. Niyogi. Laplacian Score for Feature Selection. In Y. Weiss, Schölkopf, and J. Platt, editors, *Advances in Neural Information Processing Systems*, Cambridge, MA, USA, 2005. 45

P. Hein, J. Mews, and P. Rogalla. Cardiac MSCT Study Assessment with Sure Plaque-Software. *Toshiba Medial Systems Journal - Visions*, 2:22–26, 2007. 8

S. Heldmann and N. Papenberg. A Scale-Space Approach for Image Registration of Vessel Structures. In *Bildverarbeitung für die Medizin 2009*, pages 137–141, Heidelberg, 2009. 114

M. Hernández Hoyos, J. Serfaty, A. Maghiar, C. Mansard, M. Orkisz, I. Magnin, and P. Douek. Evaluation of Semi-Automatic Arterial Stenosis Quantification. *International Journal of Computer Assisted Radiology and Surgery*, 1(3):167–175, 2006. 41

D. L. G. Hill, P. G. Batchelor, M. Holden, and D. J. Hawkes. Medical Image Registration. *Physics in Medicine and Biology*, 46:R1–R45, 2001. 113

V. J. Hodge and J. Austin. A Survey of Outlier Detection Methodologies. *Artificial Intelligence*, 22:85–126, 2004. 24, 25, 137

D. Hush, P. Kelly, C. Scovel, and I. Steinwart. Provably Fast Algorithms for Anomaly Detection. Technical Report LA-UR-05-4367, Los Alamos National Laboratory, June 2005. 25, 26, 45

D. Hush, C. Scovel, and I. Steinwart. Stability of Unstable Learning Algorithms. *Machine Learning*, 67:197–206, 2007. 32

L. Husmann, S. Leschka, L. Desbiolles, T. Schepis, O. Gaemperli, B. Seifert, P. Cattin, T. Frauenfelder, T. G. Flohr, B. Marincek, P. A. Kaufmann, and H. Alkadhi. Coronary Artery Motion and Cardiac Phases: Dependency on Heart Rate-Implications for CT Image Reconstruction. *Radiology*, 245(2):567–476, 2007. 7

L. Ibánez, W. Schroeder, L. Ng, J. Cates, and the Insight Software Consortium. *The ITK Software Guide*. Kitware, second edition edition, 2005. 112

I. Išgum, A. Rutten, M. Prokop, and B. van Ginneken. Detection of Coronary Calcifications from Computed Tomography Scans for Automated Risk Assessment of Coronary Artery Disease. *Medical Physics*, 34(4):1450–1461, 2007. 21

J. Jomier and S. R. Aylward. Rigid and Deformable Vasculature-to-Image Registration: A Hierarchical Approach. In *Medical Image Computing and Computer-Assited Intervention - MICCAI*, volume 3216 of *LNCS*, pages 829–836, St. Malo, 2004. 113, 132

D.-G. Kang, D. C. Suh, and J. B. Ra. Three-Dimensional Blood Vessel Quantification via Centerline Deformation. *IEEE Transactions on Medical Imaging*, 28(3):405–414, 2009. 20

REFERENCES

A. Kanitsar, D. Fleischmann, R. Wegenkittl, P. Felkel, and M. E. Gröller. CPR - Curved Planar Reformation. Technical Report TR-186-2-02-06, Institute of Computer Graphics and Algorithms, Vienna University of Technology, 2002. 13

Y. Kita, D. Wilson, and J. Noble. Real-time Registration of 3D Cerebral Vessels to X-ray Angiograms. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI 1998*, pages 1125–1133, 1998. 113

S. Klein and M. Staring. *elastix: The manual*, 2010. 112

S. Klein, J. Pluim, M. Staring, and M. Viergever. Adaptive Stochastic Gradient Descent Optimisation for Image Registration. *International Journal of Computer Vision*, 81 (3):227–239, 2009. 117

S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. W. Pluim. elastix: A Toolbox for Intensity-Based Medical Image Registration. *IEEE Transactions on Medical Imaging*, 29:196–205, January 2010. 112, 117, 118

R. Kohavi and G. H. John. Wrappers for Feature Subset Selection. *Artificial Intelligence*, 97(1-2):273–324, 1997. 43

S. Komatsu, A. Hirayama, Y. Omori, Y. Ueda, I. Mizote, Y. Fujisawa, M. Kiyomoto, T. Higashide, and K. Kodama. Detection of Coronary Plaque by Computed Tomography With a Novel Plaque Analysis System, 'Plaque Map', and Comparison With Intravascular Ultrasound and Angioscopy. *Circulation Journal*, 69:72–77, 2005. 21

K. Krissian, G. Maladain, and N. Ayache. Model Based Detection of Tubular Structures in 3D Images. *Computer Vision and Image Understanding*, 80:130–171, 2000. 33, 52

U. Kurkure, D. Chittajallu, G. Brunner, R. Yalamanchili, and I. Kakadiaris. Detection of Coronary Calcifications Using Supervised Hierarchical Classification. In G. Unal, I. A. Kakadiaris, N. Navab, and M. Sonka, editors, *Second International Workshop on Computer Vision for Intravascular and Intracardiac Imaging (CVII)*, pages 45–52, New York, USA, 2008. 22

S. Laguitton, C. Boldak, A. Bousse, G. Yang, and C. Toumoulin. Temporal Tracking of Coronaries in MSCTA by Means of 3D Geometrical Moments. In *Proceedings of the 28th IEEE EMBS Annual International Conference*, pages 924–927, New York, 2006. 114

J. Landis and G. Koch. The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1):159–174, March 1977. 70, 137

T. Lange, S. Eulensteina, M. Hünerbeina, and P.-M. Schlag. Vessel-Based Non-Rigid Registration of MR/CT and 3D Ultrasound for Navigation in Liver Surgery. *Computer Aided Surgery*, 8:228–240, 5 2003. 113

S. Lankton, A. Stillman, P. Raggi, and A. Tannenbaum. Soft Plaque Detection and Automatic Vessel Segmentation. In *Probabilistic Models For Medical Image Analysis 2009*, pages 25–33, London, 2009. 22

R. M. Lapp, M. Lorenzo-Valdés, and D. Rueckert. 3D/4D Cardiac Segmentation Using Active Appearance Models, Non-rigid Registration, and the Insight Toolkit. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI*, 2004. 114

A. W. Leber, A. Knez, A. Becker, C. Becker, F. von Ziegler, K. Nikolau, C. Rist, M. Reiser, C. White, G. Steinbeck, and P. Boekstegers. Accuracy of Multidetector Spiral Computed Tomography in Identifying and Differentiating the Composition of Coronary Atherosclerotic Plaques: A Comparative Study with Intracoronary Ultrasound. *Journal of the American College of Cardiology*, 43(7):1241–1247, 2004. 4, 8, 11

K. Lekadir and G.-Z. Yang. Carotid Artery Segmentation Using an Outlier Immune 3D Active Shape Models Framework. In *Medical Image Computing and Computer-Assisted Intervetion - MICCAI*, volume 4190, 2006. 21

D. Lesage. *Models, Features and Extraction Schemes for Vascular Segmentation: Application to the Delineation of Coronary Arteries from 3D Computed Tomography Data*. PhD thesis, Ecole Nationale Supérieure des Télécommunications, October 2009. 6, 38, 40, 77

D. Lesage, E. Angelini, I. Bloch, and G. Funka-Lea. Design and Study of Flux-based Features for 3D Vascular Tracking. In *IEEE International Symposium on Biomedical Imaging: From Nano to Macro - ISBI 2009*, pages 286–289, Boston, USA, 2009a. 38, 39

D. Lesage, E. D. Angelini, I. Bloch, and G. Funka-Lea. Bayesian Maximal Paths for Coronary Artery Segmentation from 3D CT Angiograms. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI*, volume 5761 of *LNCS*, pages 222–229, 2009b. 141

Q. Li, H. Arimura, and K. Doi. Selective Enhancement Filters for Nodules, Vessels, and Airway Walls in Two- and Threedimensional CT Scans. *Medical Physics*, 30: 2040–2051, 2003. 37

T. Lindeberg. Edge Detection and Ridge Detection with Automatic Scale Selection. *International Journal of Computer Vision*, 30(2):77–116, 1996. 37

H. Liu and L. Yu. Toward Integrating Feature Selection Algorithms for Classification and Clustering. *IEEE Transactions on Knowledge and Data Engineering*, 17(4): 491–502, 2005. 44

H. Liu, H. Motoda, R. Setiono, and Z. Zhao, editors. *Proceedings of the Fourth International Workshop on Feature Selection in Data Mining*, volume 10, Hyderabad, India, June 2010. 43

H. Lombaert and F. Cheriet. Spatio-temporal Segmentation of the Heart in 4D MRI Images Using Graph Cuts with Motion Cues. In *IEEE International Symposium on Biomedical Imaging - ISBI*, pages 492–495, 2010. 114

C. Lorenz, I.-C. Carlsen, T. Buzug, C. Fassnacht, and J. Weese. Multi-scale Line Segmentation with Automatic Estimation of Width, Contrast and Tangential Direction in 2D and 3D Medical Images. In *CVRMed-MRCAS97*, volume 1205, 1997. 37

B. Lu, S. Mao, N. Zhuang, H. Bakhsheshi, H. Yamamoto, J. Takasu, S. Liu, and M. Budoff. Coronary Artery Motion During the Cardiac Cycle and Optimal ECG Triggering for Coronary Artery Imaging. *Investigative Radiology*, 36(5):250–256, 2001. 7

L. Lu, L. Wanyu, and S. Xiaoming. Automated Detection of Pulmonary Nodules in CT Images with Support Vector Machines. In *Proc. of SPIE Fifth International Symposium on Instrumentation Science and Technology*, volume Vol. 7133, pages 713326–1 – 713326–6, 2009. 13

F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality Image Registration by Maximization of Mutual Information. *IEEE Transactions on Medical Imaging*, 16(2):187–198, 1997. 118

J. B. A. Maintz and M. A. Viergever. A Survey of Medical Image Registration. *Medical Image Analysis*, 2:1–36, 1 1998. 112, 113

S. Makrogiannis, R. Bhotika, J. V. Miller, J. Skinner Jr., , and M. Vass. Nonparametric Intensity Priors for Level Set Segmentation of Low Contrast Structures. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2009*, volume 5761 of *LNCS*, pages 239–246, London, 2009. 22

L. M. Manevitz and M. Yousef. One-class SVMs for Document Classification. *Journal of Machine Learning Research*, 2:139–154, 2001. 162

M. Markou and S. Singh. Novelty Detection: A Review - Part 1: Statistical Approaches. *Signal Processing*, 83:2481 – 2497, 2003a. 24, 25, 137

M. Markou and S. Singh. Novelty Detection: A Review - Part 2: Neural Network Based Approaches. *Signal Processing*, 83:2499 – 2521, 2003b. 24, 25, 137

D. Mattes, D. R. Haynor, H. Vesselle, T. K. Lewellen, and W. Eubank. PET-CT Image Registration in the Chest Using Free-form Deformations. *IEEE Transactions on Medical Imaging*, 22(1):120–128, 2003. 118, 120

C. T. Metz, M. Schaap, S. Klein, L. A. Neefjes, E. Capuano, C. Schultz, R. J. van Geuns, P. W. Serruys, T. van Walsum, and W. J. Niessen. Patient Specific 4D Coronary Models from ECG-gated CTA Data for Intra-operative Dynamic Alignment of CTA with X-ray Images. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI*, volume 5761 of *LNCS*, pages 369–376, London, 2009. 114, 121

T. Mitchell. *Machine Learning*. McGraw Hill, 1997. 20

M. Modat, G. R. Ridgway, Z. A. Taylor, M. Lehmann, J. Barnes, N. C. Fox, D. J. Hawkes, and S. Ourselin. Fast Free-form Deformation Using Graphics Processing Units. *Computer Methods and Programs in Biomedicine*, 98:278–284, 2010. 132

S. Motoyama, T. Kondo, M. Sarai, A. Sugiura, H. Harigaya, T. Sato, K. Inoue, M. Okumura, J. Ishii, H. Anno, R. Virmani, Y. Ozaki, H. Hishida, and J. Narula. Multislice Computed Tomographic Characteristics of Coronary Lesions in Acute Coronary Syndromes. *Journal of the American College of Cardiology*, 50(4):319–326, July 2007. 8, 11

C. Murray. The physiological principle of minimum work: I the vascular system and the cost of blood volume. *National Academy of Sciences*, 12(3):207–214, 1926. 52

D. Nain, A. Yezzi, and G. Turk. Vessel Segmentation Using a Shape Driven Flow. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI'2004*, volume 3216 of *LNCS*, pages 51–59, 2004. 40

S. Oka and M. Nakai. Optimality principle in vascular bifurcation. *Biorheology*, 24(6): 737–751, 1987. 52

A. Patcha and J.-M. Park. An Overview of Anomaly Detection Techniques: Existing Solutions and Latest Technological Trends. *Computer Networks*, 51:3448–3470, 2007. 24, 25, 137

H. Peng, F. Long, and C. Ding. Feature Selection Based on Mutual information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy. *IEEE Transactions on Pattern Analysis and Machine Learning*, 27(8):1226–1238, August 2005. 44

K. Pohle, S. Achenbach, B. MacNeill, D. Ropers, M. Ferencik, F. Moselewski, U. Hoffmann, T. J. Brady, I. kyung Jang, and W. G. Daniel. Characterization of Noncalcified Coronary Atherosclerotic Plaque by Multi-detector Row CT: Comparison to IVUS. *Atherosclerosis*, 190:174–180, 2007. 8, 11

R. Porter, C. Ruggiero, and D. Hush. Density-Based Similarity Measures for Content Based Search. In *Proc. 43rd Asilomar Conference on Signals, Systems, and Computers*, 2009. 25, 137

F. Pugliese, M. G. M. Hunink, K. Gruszczynska, F. Alberghina, R. Malagó, N. van Pelt, N. R. Mollet, F. Cademartiri, A. C. Weustink, W. B. Meijboom, C. L. M. Witteman, M. Pim J. de Feyter, and G. P. Krestin. Learning Curve for Coronary CT Angiography: What Constitutes Sufficient Training? *Radiology*, 251:359–368, 2009. 14, 20

A. J. Reimann, I. Tsiflikas, H. Brodoefel, M. Scheuering, D. Rinck, A. F. Kopp, C. D. Claussen, and M. Heuschmid. Efficacy of Computer Aided Analysis in Detection

of Significant Coronary Artery Stenosis in Cardiac Using Dual Source Computed Tomography. *International Journal of Cardiovascular Imaging*, 25:195–203, 2009. 60, 61

I. Reinertsen, M. Descoteaux, K. Siddiqi, and D. Collins. Validation of Vessel-based Registration for Correction of Brain Shift. *Medical Image Analysis*, 11:374–388, 2007. 113

F. Renard and Y. Yang. Image Analysis for Detection of Coronary Artery Soft Plaques in MDCT Images. In *2008 IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI 2008)*, pages 25–28, Paris, France, May 14-17 2008. 22

D. Rinck, S. Krüger, Reimann, and M. Scheuering. Shape-based Segmentation and Visualization Techniques for Evaluation of Atherosclerotic Plaques in Coronary Artery Disease. In R. L. G. J. Kevin R. Cleary, editor, *Proceedings of SPIE Medical Imaging 2006: Visualization, Image-Guided Procedures and Display.*, volume 6141, pages 124–132, 2006. 22

D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes. Non-rigid Registration Using Free-form Deformations: Application to Breast MR Images. *IEEE Transactions on Medical Imaging*, 18(8):712–721, 1999. 117

D. Ruijters, B. ter Haar Romeny, and P. Suetens. Vesselness-based 2D-3D Registration of the Coronary Arteries. *International Journal of Computer Assisted Radiology and Surgery*, 4(4):391–397, 2009. ISSN 1861-6410. 113, 114

O. Salvado. *Characterization Of Atherosclerosis With Magnetic Resonance Imaging, Challenges and Validation*. PhD thesis, Case Western Reserve University, August 2006. 21

Y. Sato, S. Nakajima, N. Shiraga, H. Atsumi, S. Yoshida, T. Koller, G. Gerig, and R. Kikinis. Three-dimensional Multi-scale Line Filter for Segmentation and Visualization of Curvilinear Structures in Medical Images. *Medical Image Analysis*, 2(2): 143–168, 1998. 37

S. Saur. *Quantitative Assessment of Atherosclerosis in Coronary Arteries*. PhD thesis, ETH Zurich, 2009. 7, 8, 23

S. C. Saur, H. Alkadhi, L. Desbiolles, G. Székely, and P. C. Cattin. Automatic Detection of Calcified Coronary Plaques in Computed Tomography Data Sets. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2008*, volume 5241 of *LNCS*, pages 170–177, New York, USA, 2008. 22, 130

M. Schaap, C. Metz, T. van Walsum, A. G. van der Giessen, A. C. Weustink, N. R. Mollet, C. Bauer, H. Bogunovic, C. Castro, X. Deng, E. Dikici, T. O'Donnell, M. Frenay, O. Friman, M. Hernández Hoyos, P. H. Kitslaar, K. Krissian, C. Kühnel, M. A. Luengo-Oroz, M. Orkisz, Ö. Smedby, M. Styner, A. Szymczak, H. Tek, C. Wang,

S. K. Warfield, S. Zambal, Y. Zhang, G. P. Krestin, and W. J. Niessen. Standardized Evaluation Methodology and Reference Database for Evaluating Coronary Artery Centerline Extraction Algorithms. *Medical Image Analysis*, 13(5):701–714, 2009. 53, 54

P. Schoenhagen, S. S. Haliburton, A. E. Stillman, S. A. Kuzmiak, S. E. Nissen, E. M. Tuzcu, and R. D. White. Noninvasive Imaging of Coronary Arteries: Current and Future Role of Multi-Detector Row CT. *Radiology*, 232(July):7–17, 2004. 10

U. J. Schoepf, C. R. Becker, B. M. Ohnesorge, and E. K. Yucel. CT of Coronary Artery Disease. *Radiology*, 232:18–37, July 2004. 4

B. Schölkopf and A. J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. The MIT press, 2001. 49

B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson. Estimating the Support of a High-Dimensional Distribution. *Neural Computation*, 13: 1443–1471, 2001. 25, 102, 161, 162

S. Schroeder, A. Kopp, A. Baumbach, C. Meisner, A. Kuettner, C. Georg, B. Ohnesorge, C. Herdeg, C. Claussen, and K. Karsch. Noninvasive Detection and Evaluation of Atherosclerotic Coronary Plaques with Multislice Computed Tomography. *Journal of the American College of Cardiology*, 37:1430 1435, 2001. 4, 8

G. Shechter, J. R. Resar, and E. R. McVeigh. Displacement and Velocity of the Coronary Arteries: Cardiac and Respiratory Motion. *IEEE Transactions on Medical Imaging*, 23(3):369–375, 2006. doi: 10.1109/TMI.2005.862752. 115, 120

N. Søndberg-Madsen, C. Thomsen, and J. M. Peña. Unsupervised Feature Subset Selection. In *Proceedings of the Workshop on Probabilistic Graphical Models for Classification*, pages 71–82, 2003. 44

H. Stary. Natural history and histological classification of atherosclerotic lesions : An update. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 20(5):1177–1178, 2000. 3, 4

I. Steinwart, D. Hush, and C. Scovel. A Classification Framework for Anomaly Detection. *Journal of Machine Learning Research*, 6:211–232, 2005a. 26, 27, 28, 29, 44, 46, 55, 137

I. Steinwart, D. Hush, and C. Scovel. Density Level Detection is Classification. *Neural Information Processing Systems*, 17:1337–1344, 2005b. 26, 28, 44, 55

J. Suh, D. Scheinost, X. Qian, A. Sinusas, C. Breuer, and X. Papademetris. Serial Nonrigid Vascular Registration Using Weighted Normalized Mutual Information. In *IEEE International Symposium on Biomedical Imaging - ISBI*, pages 25–28, 2010. 114, 132

J. Sun, Z. Zhang, B. Lu, W. Yu, Y. Yang, Y. Zhou, Y.Wang, and Z. Fan. Identification and Quantification of Coronary Atherosclerotic Plaques: A Comparison of 64-MDCT and Intravascular Ultrasound. *American Journal of Roentgenology*, 190(3):748–754, March 2008. 8

V. Svetnik, A. Liaw, C. Tong, and T. Wang. Application of Breiman's Random Forest to Modeling Structure-Activity Relationships of Pharmaceutical Molecules. In *Multiple Classifier Systems*, volume 3077 of *Lecture Notes in Computer Science*, pages 334–343. 2004. 48, 59

M. Tessmann, F. Vega-Higuera, D. Fritz, M. Scheuering, and G. Greiner. Learning-Based Detection of Stenotic Lesions in Coronary CT Data. In *Proceedings of Vision, Modeling, and Visualization 2008*, pages 189–198, 2008. 22, 23, 37, 60, 61, 102

M. Tessmann, F. Vega-Higuera, D. Fritz, M. Scheuering, and G. Greiner. Multi-scale Feature Extraction for Learning-Based Classification of Coronary Artery Stenosis. In *Proceedings of SPIE Medical Imaging 2009*, pages 1–8, Lake Buena Vista, Florida, 2009. 22, 23, 60, 61, 73, 102

P. Thévenaz and M. Unser. Optimization of Mutual Information for Multiresolution Image Registration. *IEEE Transactions on Image Processing*, 9(12):2083–2099, 2000. 118

C. Toumoulin, C. Boldak, M. Garreau, and D. Boulmier. Coronary Characterization in Multi-slice Computed Tomography. In *Computers in Cardiology*, pages 749–752, 2003. 21

G.-A. Turgeon, G. Lehmann, G. Guiraudon, M. Drangova, D. Holdsworth, and T. Peters. 2D-3D Registration of Coronary Angiograms for Cardiac Procedure Planning and Guidance. *Medical Physics*, 32(12):3737–3749, 2005. 113, 114

B. van Ginneken. Computer-aided Diagnosis in Chest Imaging: How to Improve Performance and Avoid Reinventing the Wheel. In *IEEE International Symposium on Biomedical Imaging - ISBI*, page 274, 2010. 13

P. Viola and W. M. Wells. Alignment by Maximization of Mutual Information. *International Journal of Computer Vision*, 24(2):137–154, 1997. 118

D. Vukadinovic, T. van Walsum, R. Manniesing, R. Hameeteman, S. Rozie, T. de Weert, A. van der Lugt, and W. Niessen. Segmentation of the Outer Vessel Wall of the Common Carotid Artery in CTA. *IEEE Transactions on Medical Imaging*, 29(1): 65–76, January 2010. 21

S. Wesarg, M. F. Khan, and E. A. Firle. Localizing Calcifications in Cardiac CT Data Sets Using a New Vessel Segmentation Approach. *Journal of Digital Imaging*, 19(3): 249–257, September 2006. 21

A. C. Weustink, N. R. Mollet, F. Pugliese, W. B. Meijboom, K. Nieman, M. H. Heijenbrok-Kal, T. G. Flohr, L. A. E. Neefjes, F. Cademartiri, P. J. de Feyter, and G. P. Krestin. Optimal Electrocardiographic Pulsing Windows and Heart Rate: Effect on Image Quality and Radiation Exposure at Dual-source Coronary CT Angiography. *Radiology*, 248(3):792–798, 2008. 53

M. Wierzbicki, M. Drangova, G. Guiraudon, and T. Peters. Validation of Dynamic Heart Models Obtained Using Non-linear Registration for Virtual Reality Training, Planning, and Guidance of Minimally Invasive Cardiac Surgeries. *Medical Image Analysis*, 8:387–401, 2004. doi: doi:10.1016/j.media.2004.06.014. 114

K. Wilson, G. Guiraudon, D. Jones, and T. M. Peters. 4D Shape Registration for Dynamic Electrophysiological Cardiac Mapping. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI*, LNCS 4191, pages 520–527, 2006. 114

W. C. K. Wong and A. C. S. Chung. Augmented Vessels for Quantitative Analysis of Vascular Abnormalities and Endovascular Treatment Planning. *IEEE Transactions on Medical Imaging*, 25(6):665–684, 2006. 20, 113, 132

World Health Organization. The top ten causes of death - Fact sheet N310, October 2008. 1

S. Wörz and K. Rohr. Segmentation and Quantification of Human Vessels Using a 3D Cylindrical Intensity Model. *IEEE Transactions on Image Processing*, 16(8): 1994–2004, August 2007. 33

E. P. Xing and R. M. Karp. CLIFF: Clustering of High-dimensional Microarray Data via Iterative Feature Filtering Using Normalized Cuts. *Bioinformatics*, 17(Suppl. 1): S306–S315, 2001. 44

H. Xue, C. Malamateniou, J. Allsop, L. Srinivasan, J. V. Hajnal, and D. Rueckert. Automatic Extraction and Matching of Neonatal Cerebral Vasculature. In *3rd IEEE International Symposium on Biomedical Imaging: Nano to Macro, 2006.*, pages 125–128, Arlington, VA, USA, April 2006. 113, 132

X. Yang, L. Jia, H. Qingmao, C. Zhijun, D. Xiaohua, and H. Pheng Ann. F-score Feature Selection Method May Improve Texture-based Liver Segmentation Strategies. In *2008 IEEE International Symposium on IT in Medicine and Education*, pages 697–702, 2008. 46

P. A. Yushkevich, J. Piven, C. Heather, R. Gimpel Smith, S. Ho, J. C. Gee, and G. Gerig. User-Guided 3D Active Contour Segmentation of Anatomical Structures: Significantly Improved Efficiency and Reliability. *Neuroimage*, 31(3):1116–1128, 2006. 120

P. Zanzonico, L. N. Rothenberg, and H. W. Strauss. Radiation Exposure of Computed Tomography and Direct Intracoronary Angiography: Risk has its Reward. *Journal of the American College of Cardiology*, 47(9):1846–1849, 2006. 4

D. P. Zhang, E. Edwards, L. Mei, and D. Rueckert. 4D Motion Modeling of the Coronary Arteries from CT Images for Robotic Assisted Minimally Invasive Surgery. In *Proc. of SPIE Medical Imaging 2009: Image Processing*, volume 7259, 2009a. 115, 120, 121, 124, 128, 132

D. P. Zhang, O. Pedro, K. Mori, P. Edwards, and D. Rueckert. Coronary Artery Tracking from Dynamic Cardiac CT Sequences. In *Medical Image Understanding and Analysis 2009*, pages 22–26, 2009b. 114, 132

D. P. Zhang, L. Risser, C. Metz, L. Neefjes, N. Mollet, W. Niessen, and D. Ruecker. Coronary Artery Motion Modeling from 3D Cardiac CT Sequences Using Template Matching and Graph Search. In *IEEE International Symposium on Biomedical Imaging - ISBI*, pages 1053–1056, Rotterdam, The Netherlands, April 2010. 115, 120, 121, 124, 128, 132

Z. Zhao and H. Liu. Spectral Feature Selection for Supervised and Unsupervised Learning. In *Proceedings of the 24th International Conference on Machine Learning*, Corvallis, OR, USA, 2007. 45

Z. Zhao, F. Morstatter, S. Sharma, S. Alelyani, A. Anand, and H. Liu. Advancing Feature Selection Research - ASU Feature Selection Repository. Technical Report TR-10-007, School of Computing, Informatics, and Decision Systems Engineering, Arizona State University, Tempe, AZ, USA, 2010. 43

Y. Zheng, A. Barbu, B. Georgescu, M. Scheuering, and D. Comaniciu. Fast Automatic Heart Chamber Segmentation from 3D CT Data Using Marginal Space Learning and Steerable Features. In *Int. Conf on Computer Vision*, pages 1–8, 2007. 42, 72

Y. Zheng, A. Barbu, B. Georgescu, M. Scheuering, and D. Comaniciu. Four-Chamber Heart Modeling and Automatic Segmentation for 3-D Cardiac CT Volumes Using Marginal Space Learning and Steerable Features. *IEEE Transactions on Medical Imaging*, 27(11):1668–1681, November 2008. 42

Appendix A

SVM Formulation

The following Appendix provides further details on two SVM-based methods that were employed along this thesis. These are: Soft Margin SVM (Section A.1) and One-class SVM (Section A.2).

A.1 Formulation of Hard Margin and Soft Margin SVM

Let $\mathcal{D}$ denote a training set of sample-label pairs $(\mathbf{x}_i, y_i)$, $i = 1, ..., N$, where $\mathbf{x}_i \in \mathbb{R}^p$ and $y_i \in \{1, -1\}$. Suppose that $\mathcal{D}$ is linearly separable, so there is an hyperplane H that separates positive from negative samples. The points $\mathbf{x}$ lying on a hyperplane satisfy:

$$\langle \mathbf{w}, \mathbf{x} \rangle + b = 0, \tag{A.1}$$

where $\mathbf{w}$ is a vector normal to the hyperplane, $\langle \cdot, \cdot \rangle$ represents the scalar product, $\frac{b}{\|\mathbf{w}\|}$ denotes the distance from the hyperplane to the origin and $\|\mathbf{w}\|$ is the Euclidean norm of $\mathbf{w}$. The goal is to find the hyperplane H, *i.e.* the pair $\mathbf{w}, b$, that maximizes the margin defined as a sum of the shortest distances separating H from the closest positive and negative sample, respectively. Actually, thus separated data should be located beyond two parallel hyperplanes H_1, H_2 equidistant from H. The respective subspaces containing the positively and negatively labeled data can be formulated as follows:

$$\langle \mathbf{x}_i, \mathbf{w} \rangle + b \geq 1 \qquad \text{for } y_i = 1, \tag{A.2}$$

$$\langle \mathbf{x}_i, \mathbf{w} \rangle + b \leq -1 \qquad \text{for } y_i = -1, \tag{A.3}$$

which combined give:

$$y_i \left(\langle \mathbf{x}_i, \mathbf{w} \rangle + b \right) - 1 \geq 0 \qquad \forall i. \tag{A.4}$$

The hyperplanes H_1 and H_2 respectively can be described by:

$$\begin{array}{ll} H_1: & \langle \mathbf{x}_i, \mathbf{w} \rangle + b = 1 \\ \text{with distance to origin} & \frac{|1-b|}{\|\mathbf{w}\|} \\ H_2: & \langle \mathbf{x}_i, \mathbf{w} \rangle + b = -1 \\ \text{with distance to origin} & \frac{|-1-b|}{\|\mathbf{w}\|} \end{array} \tag{A.5}$$

They define the margin $\frac{2}{\|\mathbf{w}\|}$. It is possible to find H_1 and H_2, which give the maximum margin, by minimizing $\|\mathbf{w}\|$, subject to constraint from Eq. A.4. To make the optimization problem easier to solve, the minimization of $\|\mathbf{w}\|$ is replaced by the minimization of:

$$\frac{1}{2}\|\mathbf{w}\|^2 \tag{A.6}$$

under the same constraint (Eq. A.4). Figure A.1 illustrates the previously formulated problem in a two-dimensional case. Training points $\mathbf{x}_i$, for which Eq. A.5 is satisfied, belong either to H_1 or to H_2d, and are called support vectors. The maximization of the margin implies defining support vectors as far as possible from the hyperplane.

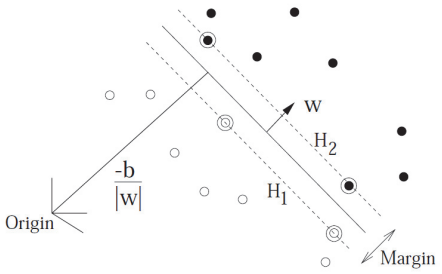


Figure A.1: Linear separating hyperplanes - The optimal hyperplane is shown in a solid line, while support vectors are circled. Illustration from (Burges, 1998).

Eq A.6 is called an objective function, while Eq. A.4 represents the inequality constraint. Both, form a constrained optimization problem. This type of problems are solved by the introduction of Lagrange multipliers $\alpha_i \geq 0$ and a Lagrangian

$$L_P \equiv \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{N} \alpha_i (y_i \left(\langle \mathbf{x}_i, \mathbf{w}\rangle + b\right) - 1) \tag{A.7}$$

The Lagrangian has to be minimized w.r.t the primal variables $\mathbf{w}$ and b and maximized w.r.t the dual variables α_i. This is a convex quadratic programming problem. Therefore, it can be solved through its dual problem: maximizing L_P, subject to the constraints imposed by both primal and dual variables ($\alpha_i \geq 0$).

Requiring that the derivatives of L_P w.r.t. $\mathbf{w}$ and b must vanish, gives

$$\sum_i \alpha_i y_i = 0 \tag{A.8}$$

and

$$\mathbf{w} = \sum_i \alpha_i y_i \mathbf{x}_i \tag{A.9}$$

Replacing Eqs. A.8 and A.9 in A.7, it is possible to eliminate the primal variables $\mathbf{w}$ and b, arriving to the so-called dual optimization problem which is the one solved in

practice:

$$\max \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} \alpha_i \alpha_j y_i y_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle \tag{A.10}$$

subject to $\alpha_i \geq 0 \; \forall i$ and $\sum_i \alpha_i y_i = 0$.

As so far formulated, the SVM assumes linearly separable data. Cortes and Vapnik (1995) formulated the case where some of the data could be misclassified so that linear separability is not required. This formulation is commonly refered to as *soft margin SVM*.

The formulation of a soft margin requires to relax the constraints A.2 and A.3 but, only when necessary. Then a further cost needs to be introduced. This is done by introducing a positive slack variable $\xi_i, i = 1...N$ into the constraints. A.4 becomes:

$$y_i \left(\langle \mathbf{x}_i, \mathbf{w} \rangle + b \right) \geq 1 - \xi_i \qquad \forall i \tag{A.11}$$

with

$$\xi_i \geq 0 \qquad \forall i \tag{A.12}$$

The realization of the so formulated soft margin classifier is obtained by minimizing the function

$$\frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^{N} \xi_i, \tag{A.13}$$

subject to constraints A.11 and A.12, where $C > 0$ is a constant that determines the trade-off between margin error maximization and training error minimization. Rewriting in terms of the Lagrange multipliers, leads again to the problem of maximizing Eq. A.10, subject to the constraints

$$0 \leq \alpha_i \leq C \qquad \forall i \tag{A.14}$$

and

$$\sum_{i=1}^{N} \alpha_i y_i = 0. \tag{A.15}$$

A.2 The One Class SVM

Schölkopf et al. (2001) proposed a method to adapt the the standard SVM methodology (Section A.1) to what they call the one-class classification problem. Their proposal makes use of an algorithm that computes a binary function that is supposed to capture regions in input space where the probability density lives. A function which is nonzero in a region where most of the data is located. The problem is formulated as follows:

Suppose there is some data drawn from an underlying probability P. The goal is to find a subset S on the feature space such that the probability that a point drawn from P lies outside S equals some *a priori* specified $\upsilon \in (0, 1)$. The problem is solved by

mapping data into a feature space H using an appropriate kernel function and finding a function $f(x)$ that is positive inside S and negative elsewhere.

In order to make use of SVM's it is necessary to define two clasess. Therefore, Schölkopf et al. (2001) treat the origin as the only member of the second class. Under this formulation, the goal of the algorithm is to find the hyperplane to separates the mapped vectors from the origin with maximum margin (Figure A.2)

$$f(x) = \begin{cases} +1 & \text{if } x \in S \\ -1 & \text{if } x \in \bar{S} \end{cases} \tag{A.16}$$

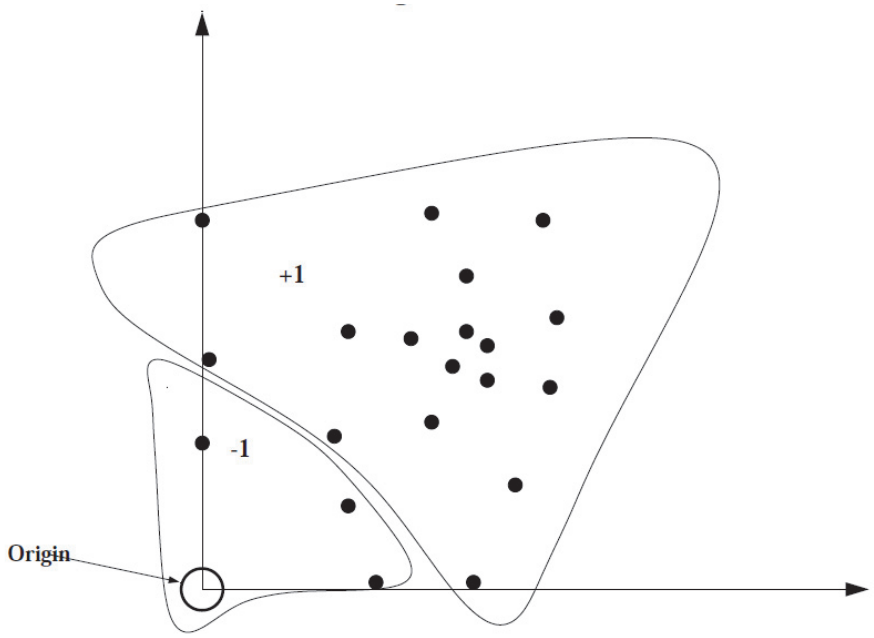


Figure A.2: One-class SVM illustration - The origin is the only original member of the second class. Circles surrounding data points represent the optimal data partition found by the algorithm.

Manevitz and Yousef (2001) Based on A.13, the following quadratic program, is defined in order to separate the data from the origin:

$$\min \frac{1}{2}\|\mathbf{w}\|^2 + \frac{1}{\upsilon N}\sum_{i=1}\xi_i - b \tag{A.17}$$

subject to the restrictions imposed by Eqs. A.11, A.12 and where $\mathbf{w}$, b and N have the same significance as in the previous section. The difference introduced in this formulation is the presence of υ.

Since nonzero slack variables ξ_i are penalized in the objective function, it is expected that if $\mathbf{w}$ and b solve the problem, then the decision function

$$f(x) = sign(\langle \Phi(\mathbf{x}_i), \mathbf{w} \rangle - b) \tag{A.18}$$

will be positive for most examples $\mathbf{x}_i$ contained in the training set, while the SV type regularization term $|\mathbf{w}\|$ will still be small. The actual trade-off between these two goals is controlled by v. It should be noted that the notation $\Phi(\mathbf{x}_i)$ is introduced to remark that the input data is mapped to the feature space by a kernel function Φ.

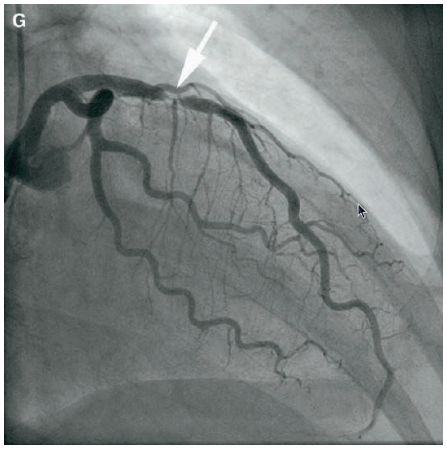
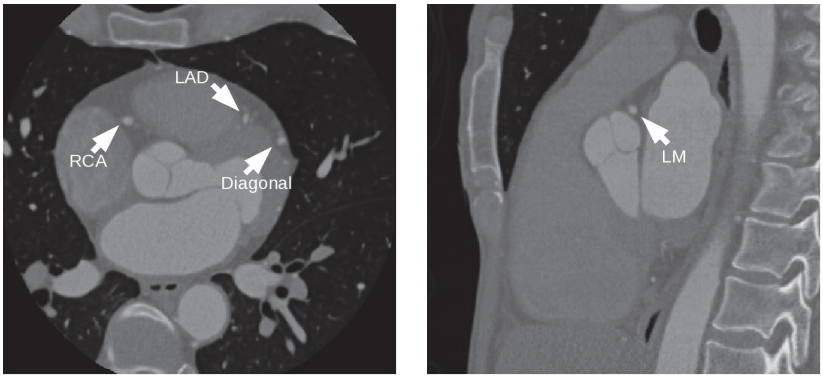
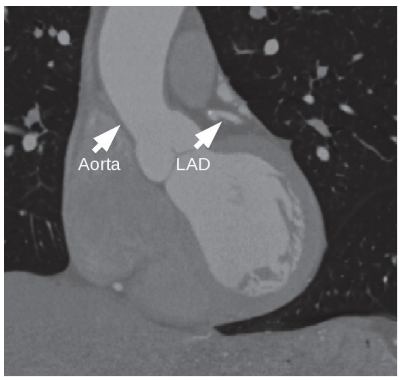


Figure 1.3: Conventional Coronary Angiography - The arrow shows a detected stenosis. Source: (Achenbach, 2007))



(a) (b)



(c)

Figure 1.4: Cardiac CT - Multi-planar reformation (MPR) view of a 3D cardiac data set. (a) axial view, (b) coronal view and (c) sagittal view.

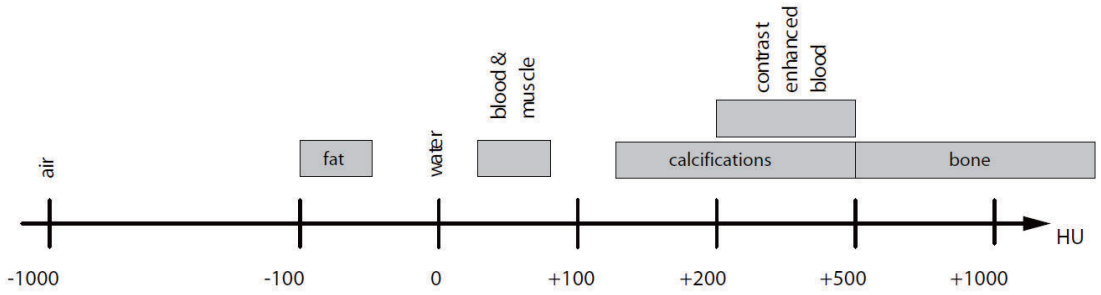


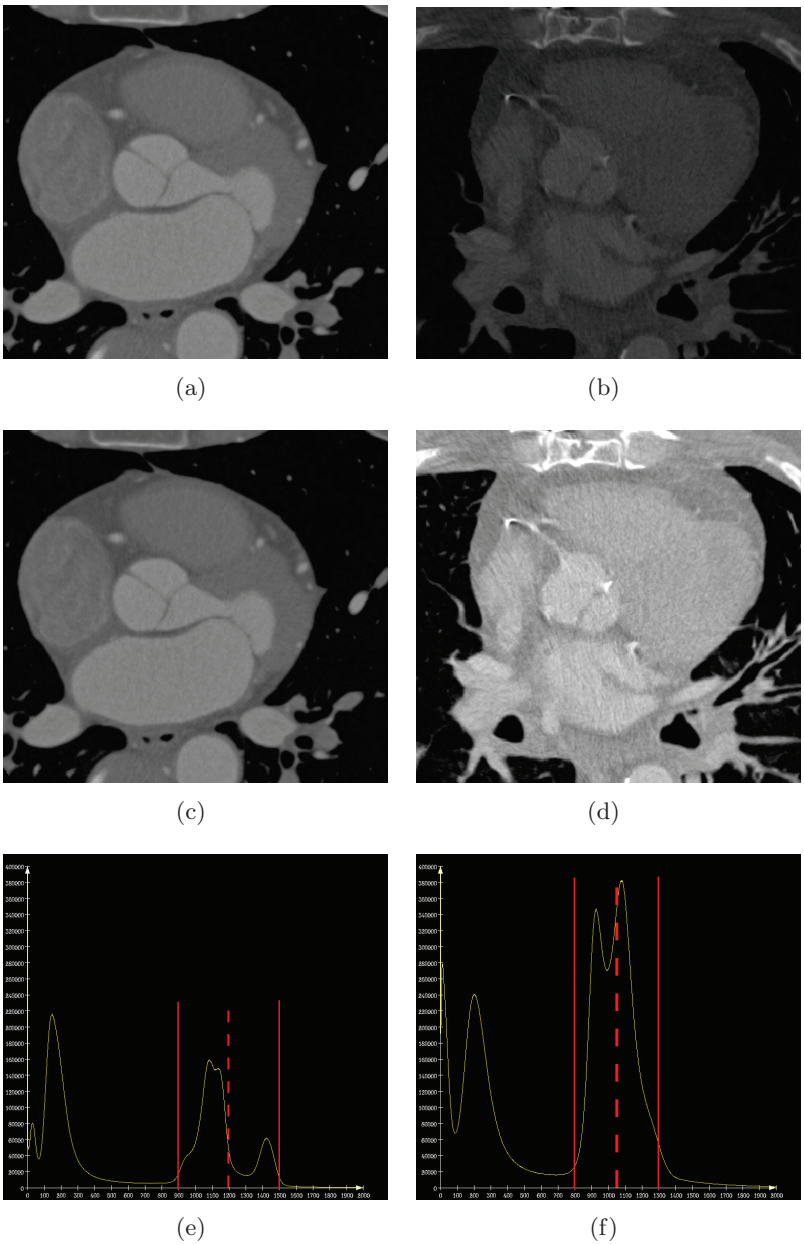
Figure 1.6: Example of HU ranges - Example of HU ranges for selected tissues. Blood is meant with contrast-enhancing agent injected. Source: (Saur, 2009).

Although one can find in literature HU ranges for certain tissues (Figure 1.6), there is no consensus on their actual bounds. More particularly, neighboring ranges partially overlap. As an example, in literature it is not possible to find clearly defined thresholds that differentiate plaque components as a function of the Hounsfield Units. Saur (2009) summarized the ranges proposed by different studies on CT images aiming at the detection of coronary plaques. Results demonstrate that there is not a general consensus (Table 1.2). Moreover, additional difficulties arise from the fact that some of these values overlap with the intensity ranges that are proposed for vessel lumen. This mainly comes from the use of the contrast material that modifies the radio-opacity of the blood in order to permit the visualization of the lumen. As an example, let us cite de Weert et al. (2008) who used a threshold of 130 HU for calcifications and had to resort to manual separation between them and contrast-enhanced arterial blood, since the latter reached on average 217.4 ± 36.9 HU, which clearly overlaps with the calcium range. Although the quoted work reports results in carotid arteries, the same can be observed in coronaries.

Temporal and spatial resolution

Coronary arteries undergo heterogeneous movement and deformations throughout the cardiac cycle causing motion artifacts. ECG synchronization is performed to reduce cardiac motion and minimize artifacts. Retrospective ECG gating is a method that is based on the simultaneous acquisition of CT data and the ECG signal. The acquired projected raw data are selected for image reconstruction with respect to a predefined cardiac phase (also denoted time-frame or time-point). However, when the velocity exceeds the temporal resolution of the scanner (Husmann et al., 2007) motion artifacts cannot be avoided. Moreover, frequently there are diverging motion patterns between the left and right coronary systems and often also between proximal and distal portions of the same coronary artery (Lu et al., 2001) leading again to motion correction problems.

Coronary arteries are relatively small and thin. Typically, they have radii in a range



(a) (b) (c) (d) (e) (f)

Figure 1.9: Quality variation in cardiac CT images - Two CT images acquired at the same medical center. (a) and (b) show the original window-level configuration. (c) A high quality image where the vessels are well contrasted (d) Bad quality image with low contrast and image artifacts. The corresponding image histograms are shown in (e) and (f). The red lines denote the limits of the selected window on each image to visualize arteries and the dashed line represents its center.

Two types of visualization are common in CHD assessment. The first one is multi-planar reformation (MPR), where the radiologist scrolls through the 2D slices of axial, coronal and sagittal views (Figure 1.4). The second one is curved planar reformation (CPR) (Kanitsar et al., 2002). The goal of CPR visualization (*e.g.* Figure 1.5) is to make a tubular structure visible in its entire length within one single image (Kanitsar et al., 2002), so that the radiologists can sweep through the vessel searching for possible lesions. The CPR display mode requires a previous extraction of the centerline of the tubular structure of interest.

1.2 Computer-Aided Diagnosis (CAD)

Computer-aided diagnosis (CAD) is a rapidly expanding area of medical image analysis that aims to help clinicians in making a proper diagnosis by, for example, detecting abnormalities, classifying and diagnosing these, and quantifying the spread of disease (van Ginneken, 2010). In general, CAD systems aid the radiologist in order to enable a faster or more accurate and reproducible diagnosis.

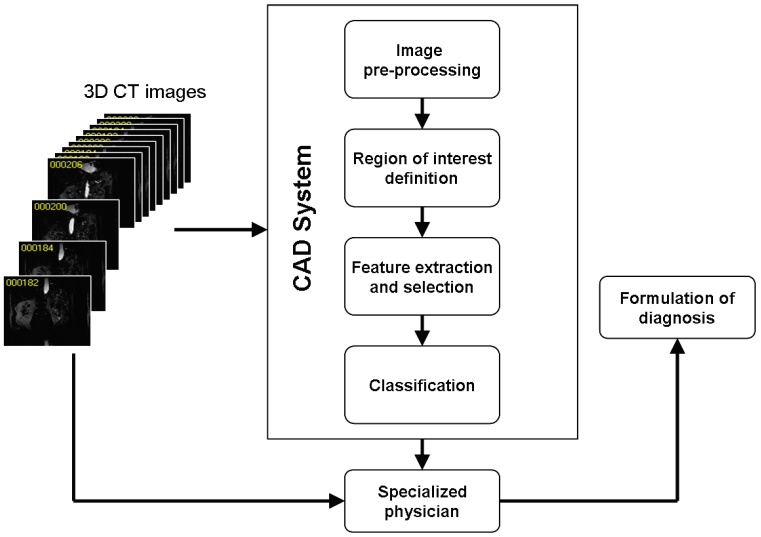
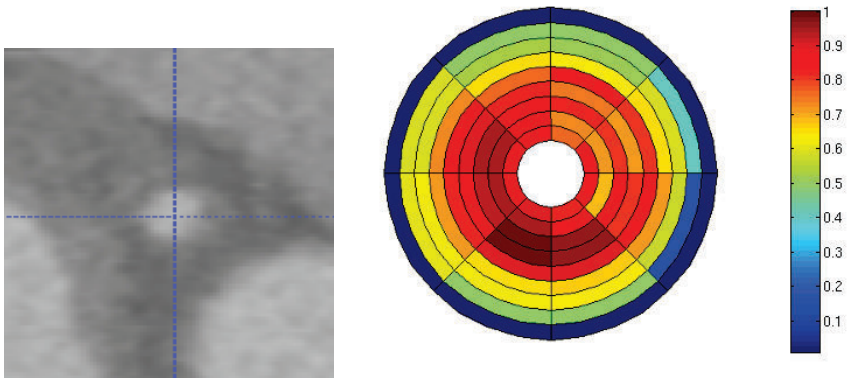


Figure 1.10: CAD system - Image-based CAD system architecture and its incorporation into medical diagnosis. Adapted from (Lu et al., 2009).

Typically, image-based CAD systems consist of four main modules (Lu et al., 2009) (Figure 1.10).

- *Image pre-processing* stage aims at improving image quality by denoising and artifact reduction. It can also handle data standardization that allows comparison of images obtained under different conditions.
- *Region of interest (ROI) definition* seeks to reduce the size of the analyzed data by defining local regions where a lesion can be found. ROI definition can be done



(a) (b)

Figure 1.11: An example of the proposed Concentric rings metric - (a) Healthy coronary cross-section (b) Color-coded visual representation of the metric response. The colors encode the normalized values of intensity integrals within each sector of each ring.

Chapter 2 describes the proposed methodological framework and all the necessary steps required to construct it. First, we discuss the hypothesis over which our proposal is based and then we describe each of the elements that compose it. The framework is made up of three main elements that are largely described in the chapter. These are: the classifiers, the metrics and the feature selection strategy.

Chapter 3 is devoted to the experimental design of the metric we have proposed to use in our framework. Once the metric is properly tuned up, the remaining candidate metrics are also parametrized, so that they can be directly compared to our approach. Additionally, the evaluation protocol followed in the different experiments is described.

Chapter 4 is devoted to the selection of an optimal feature set based on the different sets of candidate metrics that were defined in **Chapter 2** and parametrized in **Chapter 3**. The performance of the metrics is first individually evaluated and this information is used to define in which order they will be evaluated by the different feature selection algorithms. The quality of the selected group of metrics is assessed through its use as inputs for the selected classifiers.

Chapter 5 introduces a methodology that eases the task of comparing different phases to formulate a diagnosis of CHD. The method constructs a ROI surrounding potential lesions and maps the ROI to a different time-point. The two ROI are then non-rigidly registered. The two registered ROIs are compared so as to confirm / reject potential lesions.

A final chapter is devoted to give overall conclusions and point out research perspectives.

where $\triangle$ denotes the symmetric difference (Figure 2.1[1]).

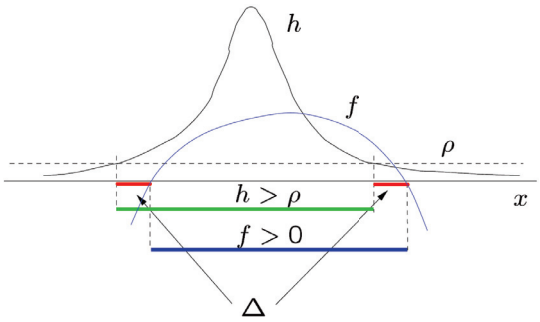


Figure 2.1: Performance of function f evaluated through the symmetric difference - The symmetric difference, defined as $A \triangle B = A \cup B \setminus A \cap B$, represents the number of cases where the function f falsely predicts a sample as belonging to the high density region, plus the number of cases where the function f falsely predicts a sample as not belonging to the high density region.

A DLD algorithm should seek $\hat{f}$ such that $\mathcal{S}(\hat{f})$ is close to zero. Unfortunately, there is no known and computationally feasible method to estimate it from a training set with guaranteed accuracy. However, the problem can be reformulated by use of a risk function.

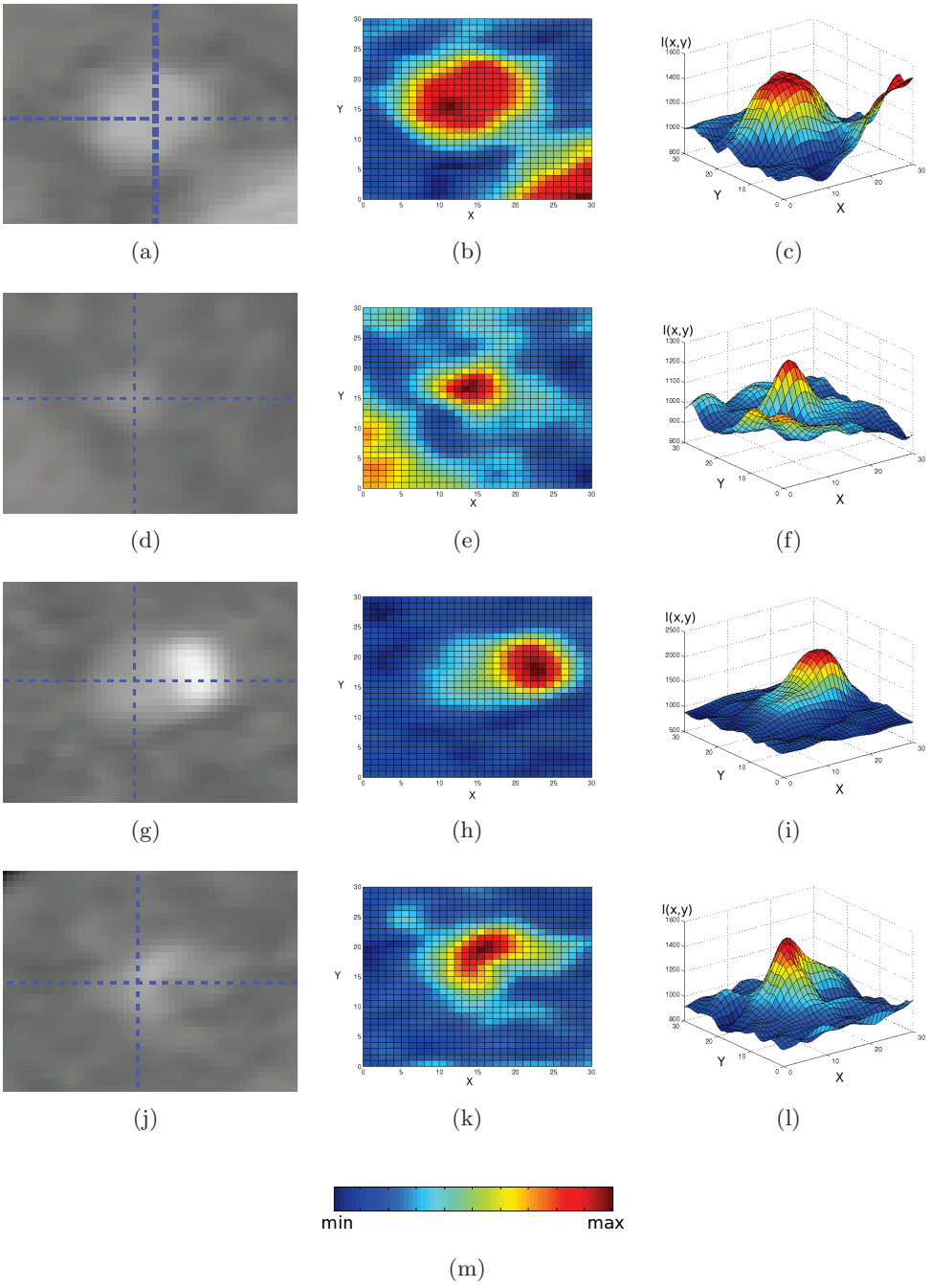
Let now μ be a known probability measure on $\mathcal{X}$, the risk defined as:

$$\mathcal{R}(f) := \frac{1}{1+\rho} P_Q(f \leq 0) + \frac{\rho}{1+\rho} \mu(f > 0), \tag{2.5}$$

Steinwart et al. (2005a) demonstrated that a function f minimizing $\mathcal{R}$ also minimizes $\mathcal{S}$ and that there is a tight relationship between them. Thus $\mathcal{R}$ is a performance measure for the DLD problem, and choosing $\hat{f}$ that minimizes $\mathcal{S}$ can be done by choosing $\hat{f}$ that minimizes $\mathcal{R}$. Moreover, unlike $\mathcal{S}$, the risk $\mathcal{R}$ can be empirically estimated from sample data.

The definition of $\mathcal{R}$ as a performance measure also permits the interpretation of the DLD problem as a supervised binary classification problem. Given the label set $Y := \{-1, 1\}$, let $x \in \mathcal{X}$ and $y \in Y$ respectively denote the values of random variables $\mathbf{x}$ and $\mathbf{y}$. The binary classification problem is formed by identifying P_Q and μ with the conditional distributions $P_{x|y=1}$ and $P_{x|y=-1}$ respectively and defining the class marginals $P(y = 1) := 1/(1+\rho)$ and $P(y = -1) := \rho/(1+\rho)$. It should be remarked that DLD-SVM, in principle, is unsupervised. The only available data comes from the unlabeled set Q. Therefore, to solve the supervised binary classification problem surrogate labeled sets need to be created. The training set $\mathcal{T}$ is built as follows. First, a set T is formed such that $T = Q$, and all its elements are assigned a label $y = 1$. For the sake of coherence, from now on we denote $|Q| = |T| = n_1$. Then, a set T' of n_{-1}

[1]Illustration resulting from discussions with Dr. Don Hush, researcher in Machine Learning and pattern recognition at Los Alamos National Library (Los Alamos, USA)



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

(j) (k) (l)

(m)

Figure 2.3: Intensity plots of orthogonal 2-D slices in a coronary artery - First column presents normal cross-sections of the proximal left coronary artery (LCA) and of the distal LCA, a cross-section with a calcification and a cross-section with a stenosis. Second column shows 2D views of pseudo-color intensity plots, and the third column shows a 3D view of the same intensity plots. Color map (m) encodes the intensities of the 2D cross-section normalized between its minimum and maximum values.

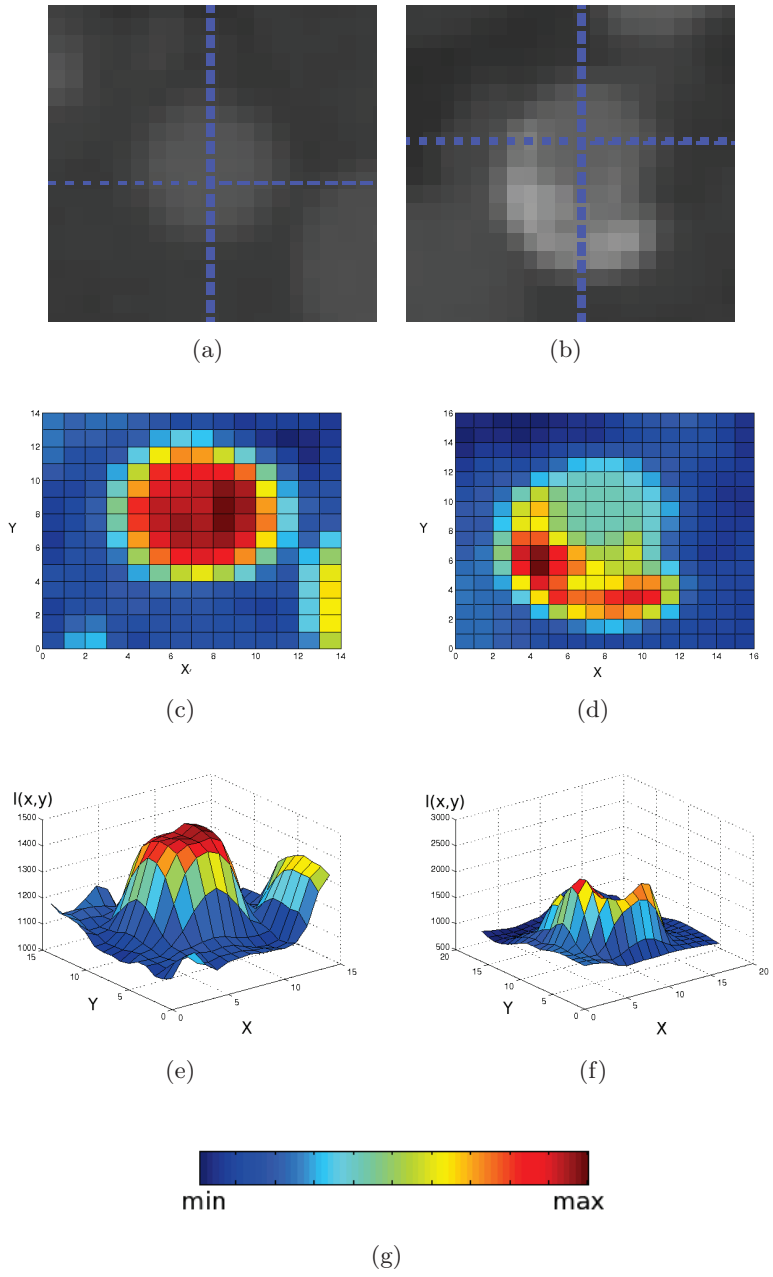
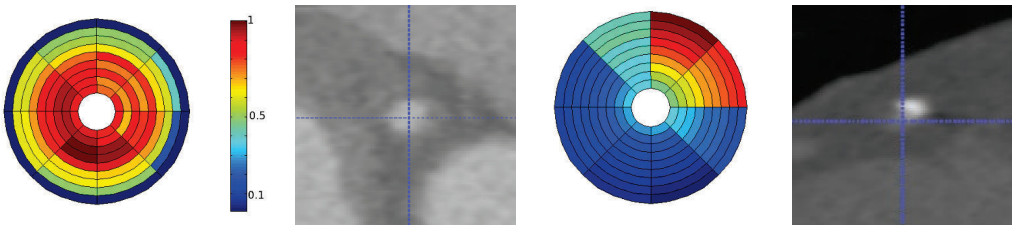


Figure 2.4: Intensity plots of orthogonal 2-D slices in a carotid artery - (a) normal carotid artery, (c) 2D representation of the intensity profile and (e) 3D representation of the intensity profile. In (b), calcified carotid artery and the 2D (d) and 3D (f) intensity plots. Colors encode the intensities of the 2D cross-section normalized between its minimum and maximum values. (g) Color map.



(a) (b) (c) (d)

Figure 2.5: Concentric rings metric representation using the Bull's Eye plot - In a normal case (a), high values are obtained in the inner rings, while low values are obtained in the outer rings. In an anomalous case (c), the pattern varies. Corresponding healthy (b) and a calcified slice (d) are shown (the calcification is the bright spot, top right with regard to the cross-hair indicating the center of the lumen). The colors encode the normalized values of intensity integrals within each sector of each ring.

2.4.2 Other Candidate Features

Apart from our previously defined metric, we are willing to evaluate the performance of other metrics. Since the detection of vascular anomalies through ML techniques is rather new, it is not possible to establish a state-of-the-art set of features for this purpose. Therefore, we make use of global features commonly used in vascular enhancement and/or segmentation to define an initial set of candidates.

Our criterion to select vascular enhancement/segmentation features comes from the fact that these features are supposed to give good responses at lumen positions in normal regions, while their responses are likely to deviate from the normality in the presence of a lesion. As an example, it is well-known that Hessian eigenvalues have poor responses (deviating from the normality) in the vicinity of calcifications. What represents a disadvantage for lumen enhancement or segmentation, can be exploited in our framework to allow anomaly detection.

An additional criterion in the selection of the alternative candidate features is motivated by the possibility of comparing their performance to our metric. Therefore, we keep the slice-by-slice strategy and our features have to be restricted to 2D features.

The candidate features exploit various image properties and different mathematical mechanisms. Their presentation hereafter follows the order from derivative- to integral-based. We thus begin with a metric based on second order derivatives, namely `Hessian eigenvalues`, which seeks normality by evaluating the symmetry of the cross-sectional shape through the eigenvalues of the Hessian matrix. Then we describe measures that exploit the medialness of normal vessels by evaluating first order derivatives at radial positions. These are: `Cores, Flux` and `MFlux` measures. We follow with the `Ribbon` metric that combines integration and differentiation in order to evaluate overall contrast between lumen and background. The `Ball` measure can be considered as integral-based, as it calculates intersection area to evaluate how well the shape of the lumen fits a disc. Another example of integral features are `Inertia moments`

tip of each spoke, where the derivative is taken in the direction of the spoke (Fridman et al., 2003). Mathematically, this is expressed as:

$$\mathcal{C}(\mathbf{p}, \sigma_r) = \sum_{i=1}^{N} D_{i,\sigma_r}(\mathbf{x}_i), \tag{2.18}$$

where $\mathbf{x}_i$ is the tip location of the i-th spoke, $D_{i,\sigma_r}(\mathbf{x})$ is the value, at location $\mathbf{x}$, of the directional image derivative along the i-th spoke, *i.e.* along the vector $\mathbf{x}_i - \mathbf{p}$, and σ_r is the scale of interrogation. The authors recommend $\sigma_r = \|\mathbf{x}_i - \mathbf{p}\|/4$, where $\mathbf{p}$ is the center of the medial atom, in our case a centerline point.

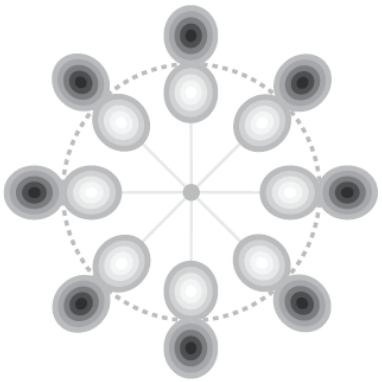


Figure 2.6: The Cores measure - Core evaluation pattern (medial atom) using Gaussian first order derivatives at the tips of the spokes (rays). Illustration obtained from (Lesage, 2009)

Thus defined `Cores` feature is closely related with `Flux` (Lesage et al., 2009a) but differs in the way the derivatives are calculated and in the scales selection.

Flux and MFlux

Flux is a gradient-based measure that exploits the orientation of gradient vectors by computing the gradient flux through the surface of an object. The flux $F(S)$ through a surface S is defined as:

$$F(S) = \int_S \langle \nabla I, \mathbf{n} \rangle \, ds \tag{2.19}$$

where ds is an infinitesimal surface patch, and $\langle \nabla I, \mathbf{n} \rangle$ denotes the scalar product of the image intensity gradient vector ∇I with the inward surface normal $\mathbf{n}$.

We apply a discrete version of the `Flux` metric, based on the implementation by Lesage et al. (2009a). In their proposal, after an equi-angular discretization of the cross-sectional circular contour into N points $\mathbf{x}_i$, `Flux` is defined as:

$$\text{Flux}(\mathbf{p}, r) = \frac{1}{N} \sum_{i=1}^{N} \langle \nabla I(\mathbf{x}_i), \mathbf{n}_i \rangle \tag{2.20}$$

Steerable features

Steerable features were proposed by Zheng et al. (2007, 2008) to capture the orientation and scale of an object to be segmented, which in their case were the heart chambers. Steerable features sample points from the evaluated volume according to a sampling pattern (Figure 2.8). Afterwards, local features are extracted at each sample point, such as: image intensity, its derivatives, etc.

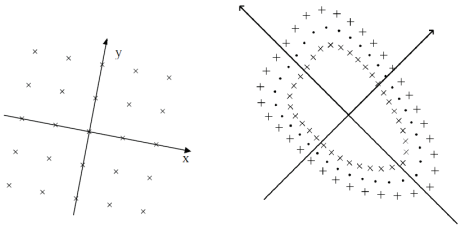


Figure 2.8: Sampling pattern examples - 2D sampling patterns of steerable features. Left, a regular sampling pattern. Right, sampling pattern around an expected shape boundary. Illustration from (Zheng et al., 2007)

In (Zheng et al., 2007), 24 different features are proposed for 3D datasets. Since our features are extracted on a 2D-basis, we eliminate the derivatives along the z direction, which reduces the number of features to 20. Given a point (x,y) with intensity I, the extracted features are denoted: I, $\sqrt{I}$, I^2, I^3, $\log I$, I_x, I_y, $|\nabla I|$, $\sqrt{|\nabla I|}$, ..., I_{xx}, I_{yy}, I_{xy}, $|\mathbf{H}|_F$, $\sqrt{|\mathbf{H}|_F}$...etc. Here, $|\mathbf{H}|_F$ denotes the Frobenius norm of the Hessian. To exploit the symmetrical characteristics of a normal vessel a circular pattern is used.

2.5 Feature Selection: Methodology

An individual set of features, *i.e.* variables extracted from a particular metric, can be used to train a classifier, while obtaining adequate results. Chapter 4 devotes a section to present the individual performance of features extracted from the metrics defined in section 2.4. The performance of individual feature sets rises the idea of combining some of them to compose a larger set that would outperform the classifiers trained with a single feature set.

Our goal is to construct an optimal feature set by combining the metrics defined in section 2.4. In practice, however, this is not a simple task and several issues arise when constructing a feature set. For instance, larger feature sets do not imply higher accuracy of the classifier. Moreover, large feature sets are computationally expensive.A high number of features can induce the so-called curse of dimensionality Bellman (1961) problem during the SVM training, which is related to the exponential growth of the data volume when adding new dimensions (here features). Thus, combining all possible features is not an adequate solution and a subset of features needs to be selected following a predefined criterion. The latter is not obvious. In fact, the best individual

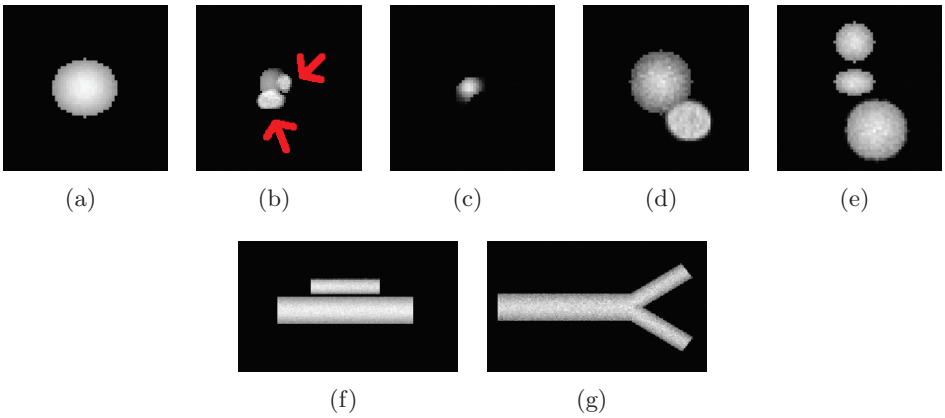
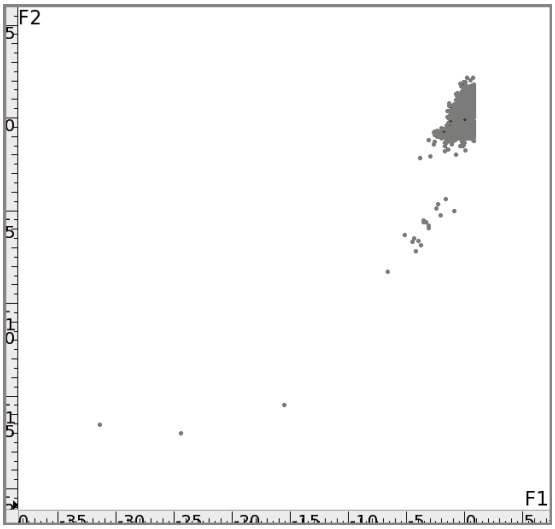


Figure 3.1: Examples of phantoms - (a) normal elliptical cross-section, (b) calcified-plaque (pointed with the arrows) and (c) a stenosis resulting from a soft-plaque. (d) A cross-section with an aneurysm, (e) cross-section with adjacent structures (the vessel of interest is located in the middle of the image), (f) sagittal view of a vessel with an adjacent structure and (g) a sagittal view of a bifurcation.

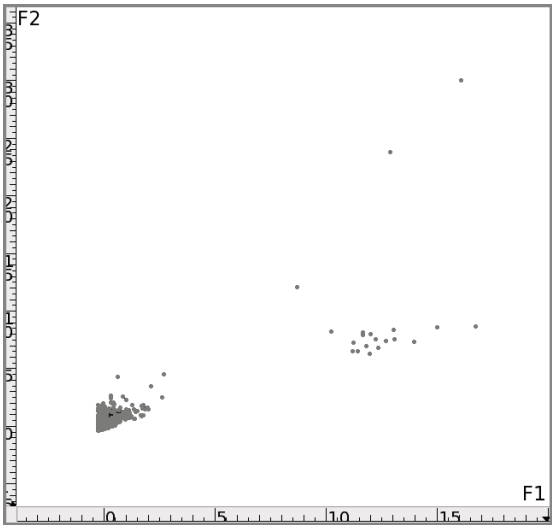
Set 1. Rotterdam Coronary Artery Algorithm Evaluation Framework. The Rotterdam Coronary Artery Algorithm Evaluation Framework (Schaap et al., 2009) aims at evaluating and comparing different algorithms for coronary artery centerline extraction from CT angiography (CTA) data. For this purpose, thirty-two datasets are provided. The acquisition protocol is as follows:

The CTA data was acquired in the Erasmus MC, University Medical Center Rotterdam, The Netherlands. Thirty-two datasets were randomly selected from a series of patients who underwent a cardiac CTA examination between June 2005 and June 2006. Twenty datasets were acquired with a 64-slice CT scanner and twelve datasets with a dual-source CT scanner (Sensation 64 and Somatom Definition, Siemens Medical Solutions, Forchheim, Germany). A tube voltage of 120 kV was used for both scanners. All datasets were acquired with ECG-pulsing (Weustink et al., 2008). The maximum current (625 mA for the dual-source scanner and 900 mA for the 64-slice scanner) was used in the window from 25% to 70% of the R-R interval and outside this window the tube current was reduced to 20% of the maximum current.
Both scanners operated with a detector width of 0.6 mm. The image data was acquired with a table feed of 3.8 mm per rotation (64-slice datasets) or 3.8 mm to 10 mm, individually adapted to the patient's heart rate (dual-source datasets).
Diastolic reconstructions were used, with reconstruction intervals varying from 250 ms to 400 ms before the R-peak. Three datasets were reconstructed using a sharp (B46f) kernel, all others were reconstructed using a medium-to-smooth (B30f) kernel. The mean voxel size of the datasets is $0.32 \times 0.32 \times 0.4$ mm (Schaap et al., 2009).

From the thirty-two available datasets, we have selected eight cases which correspond to the training set of the Rotterdam Coronary Artery Algorithm Evaluation Framework.



(a)



(b)

Figure 3.4: Illustration of bifurcation misdetection problem in DLD-SVM - Bifurcation samples are seen as outlying points with respect to the vast majority of concentrated points. As an example, the behavior is reflected by (a) plotting two features from the Concentric rings metric and (b) plotting of λ_1 vs. λ_2 parameters from the Hessian metric at a given scale.

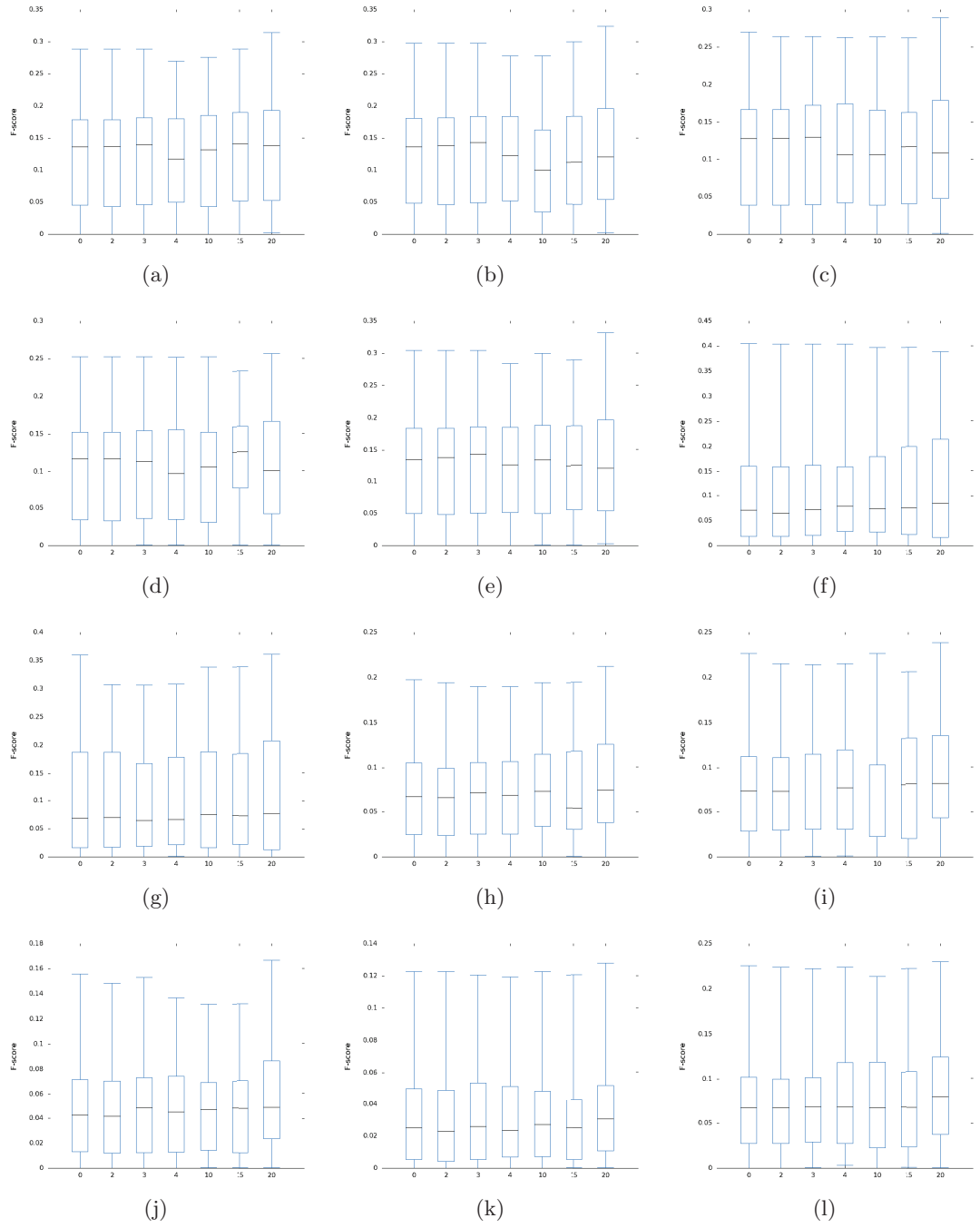


Figure 3.6: F-score as function of spatial locations sub-sampling in zero and first order steerable features. (a) I, (b) $\sqrt{I}$, (c) I^2, (d) I^3, (e) $\log I$, (f) I_x, (g) I_y, (h) $|\nabla I|$, (i) $\sqrt{|\nabla I|}$, (j) $|\nabla I|^2$, (k) $|\nabla I|^3$, (l) $\log |\nabla I|$.

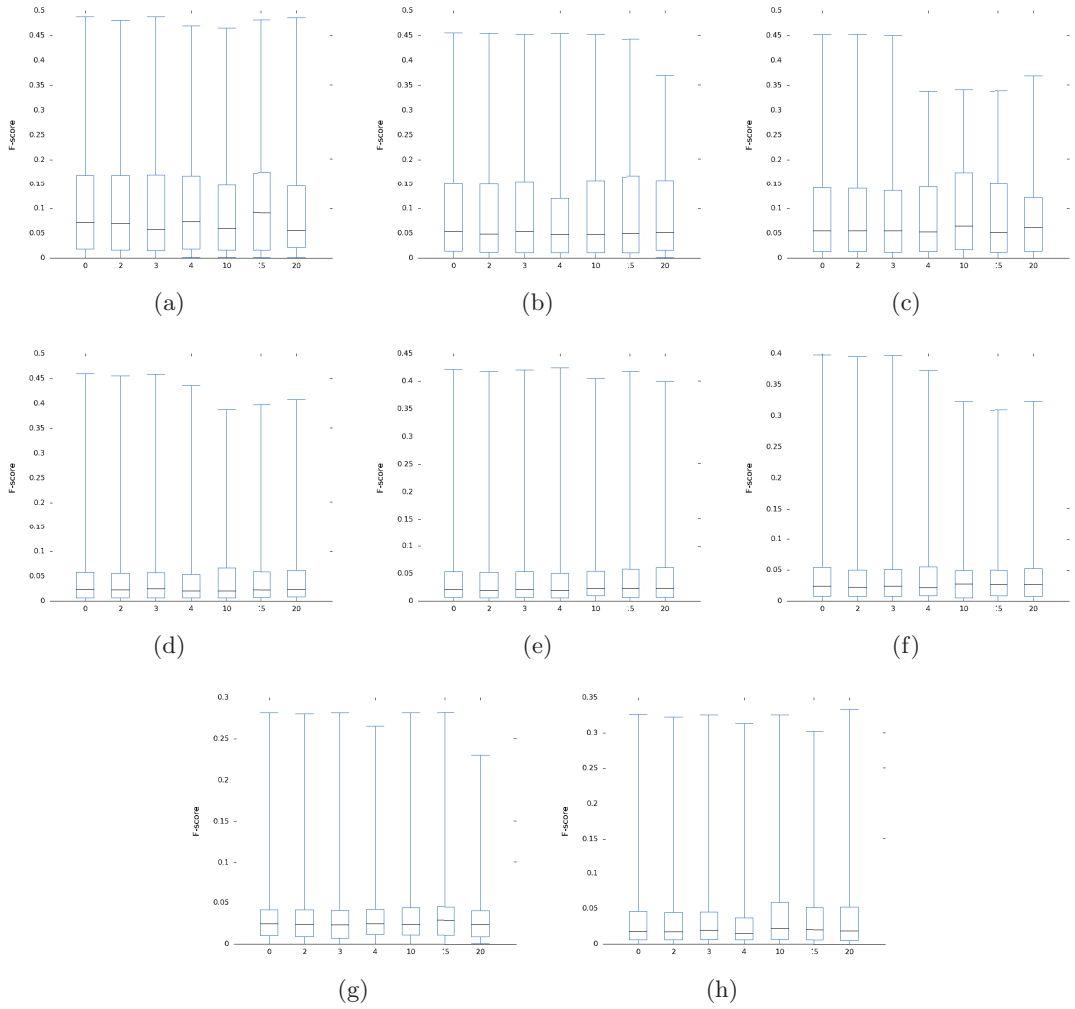


Figure 3.7: F-score as function of spatial locations sub-sampling in second order steerable features. (a) I_{xx}, (b) I_{yy}, (c) I_{xy}, (d) $|\mathbf{H}|_F$, (e) $\sqrt{|\mathbf{H}|_F}$, (f) $|\mathbf{H}|_F^2$, (g) $|\mathbf{H}|_F^3$ and (h) $\log|\mathbf{H}|_F$.

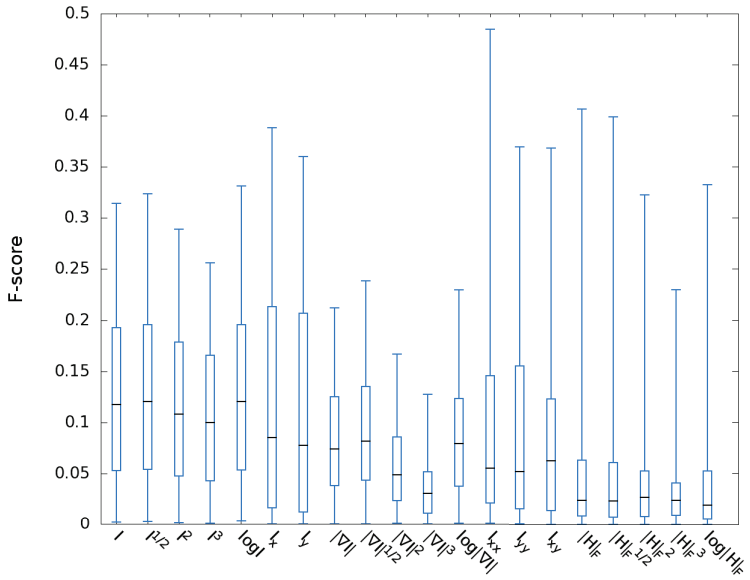


Figure 3.8: F-score summary for steerable features - Features of order zero (I), one (I_x, I_y and $|\nabla I|$) and two (I_{xx}, I_{yy}, I_{xy} and $|\mathbf{H}|_F$) are presented.

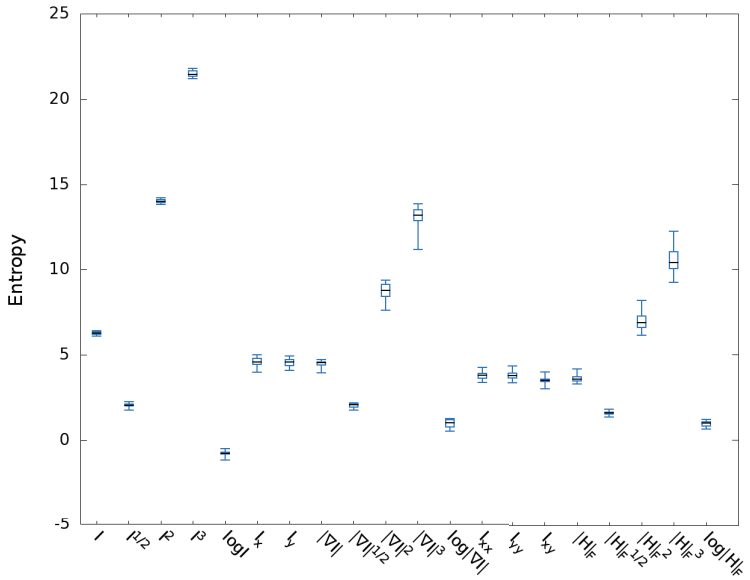
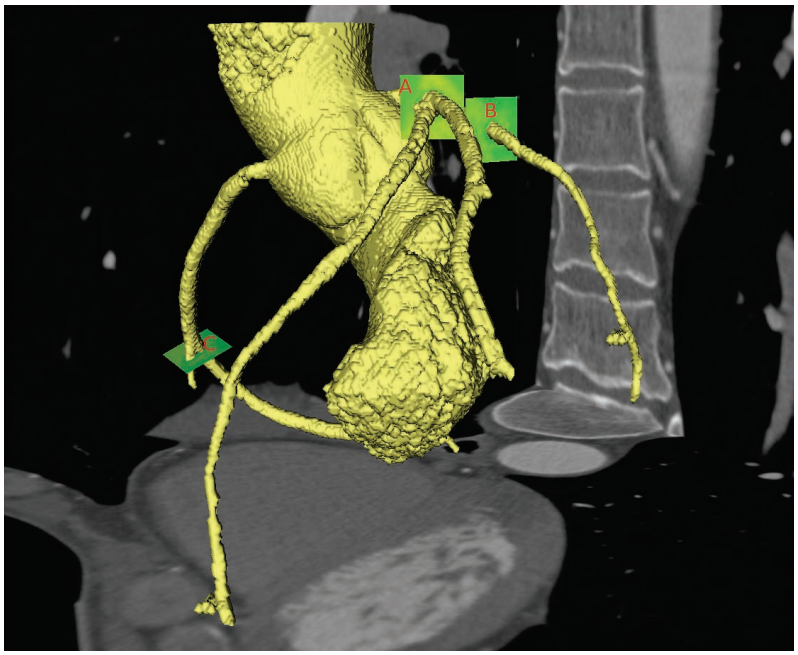
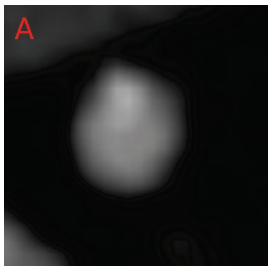


Figure 3.9: Entropy summary for steerable features - Features of order zero (I), one (I_x, I_y and $|\nabla I|$) and two (I_{xx}, I_{yy}, I_{xy} and $||\mathbf{H}||_F$) are presented. Opposite to F-score, best ranked features when using entropy are the ones with the lowest entropy.

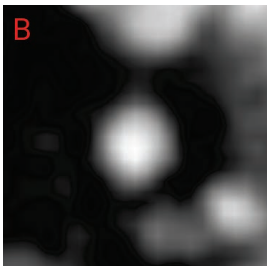
Once again, we should recall that the goal of our proposal is to call the physician's attention by marking potential lesions. Figure 3.10 shows an example of the obtained results in a real dataset, how we foresee it as a clinical application. In Figures 3.11 and 3.12 several examples of labeled vessels using both DLD-SVM and LPU are presented.



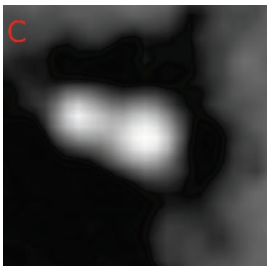
(a)



(b)



(c)



(d)

Figure 3.10: Examples of anomaly detection results using Concentric rings. (a) 3D view of the coronary tree showing the locations of three slices: A containing a calcification, B normal and C containing a bifurcation. Slice A was correctly detected as anomalous by both LPU and DLD-SVM, while slice B was correctly classified as normal. In slice C, LPU correctly detected the bifurcation as normal, and DLD-SVM marks it as anomalous. (b), (c) and (d) respectively represent these three cross-sections.

The final part of this chapter has been devoted to the parametrization of the remaining sets of candidate metrics. We have determined them following the guidelines provided by the parameters defined for the `Concentric rings` metric, so that a fair

5.1 Image Registration: Definition

Image registration is the process of deforming one image, denoted the moving image $I_M(x)$, to fit another image, denoted the fixed image $I_F(x)$, by applying a transformation $\mathbf{T}(x)$ to $I_M(x)$. The registration problem is formulated as an optimization problem that seeks to minimize a metric $\mathcal{S}$ with respect to the transformation $\mathbf{T}(x)$:

$$\hat{\mathbf{T}} = \arg\min_{\mathbf{T}} \mathcal{S}(\mathbf{T}; I_F, I_M) \tag{5.1}$$

where $\hat{\mathbf{T}}$ is the optimal transformation.

In order to solve the optimization problem (Eq. 5.1), several aspects need to be addressed. First, a *transformation* $\mathbf{T}(x)$ determines the types of deformations that shall be applied to $I_M(x)$ in order to fit $I_F(x)$. An *interpolator* evaluates the transformed moving image at non-voxel positions. The *similarity metric* $\mathcal{S}$ evaluates the quality of the alignment between the images. Typically, $\mathcal{S}$ is obtained by evaluating all image voxels. However, a subset may suffice. In this case, a *sampler* component is necessary. Finally, an *optimizer* uses $\mathcal{S}$ as a criterion to obtain the optimal parameters of the *transformation*. Multiresolution strategies are often adopted to improve the results and / or speed-up the registration process. In this case, the registration starts with images of low complexity (smoothed or/and downsampled) that is increased gradually. The designed component for this approach is often called a *pyramid*. Figure 5.1 shows the general components of the registration framework as proposed by Klein et al. (2010).

For further details on the elements of a registration algorithm, the interested reader is referred to more extensive reviews (Ibánez et al., 2005; Klein et al., 2010; Maintz and Viergever, 1998).

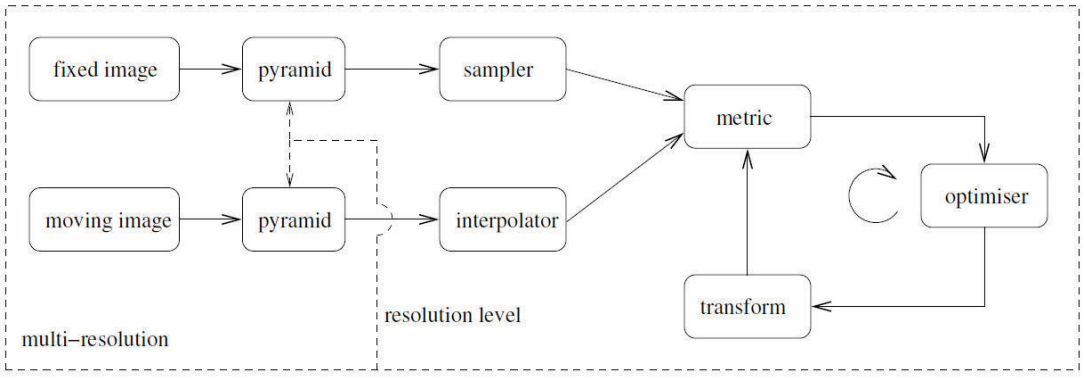


Figure 5.1: Basic registration framework components - Two input images $I_F(x)$ and $I_M(x)$, an interpolator, a transform, an optimizer and a similarity metric. Samplers and pyramids are optional. Adopted from Klein and Staring (2010).

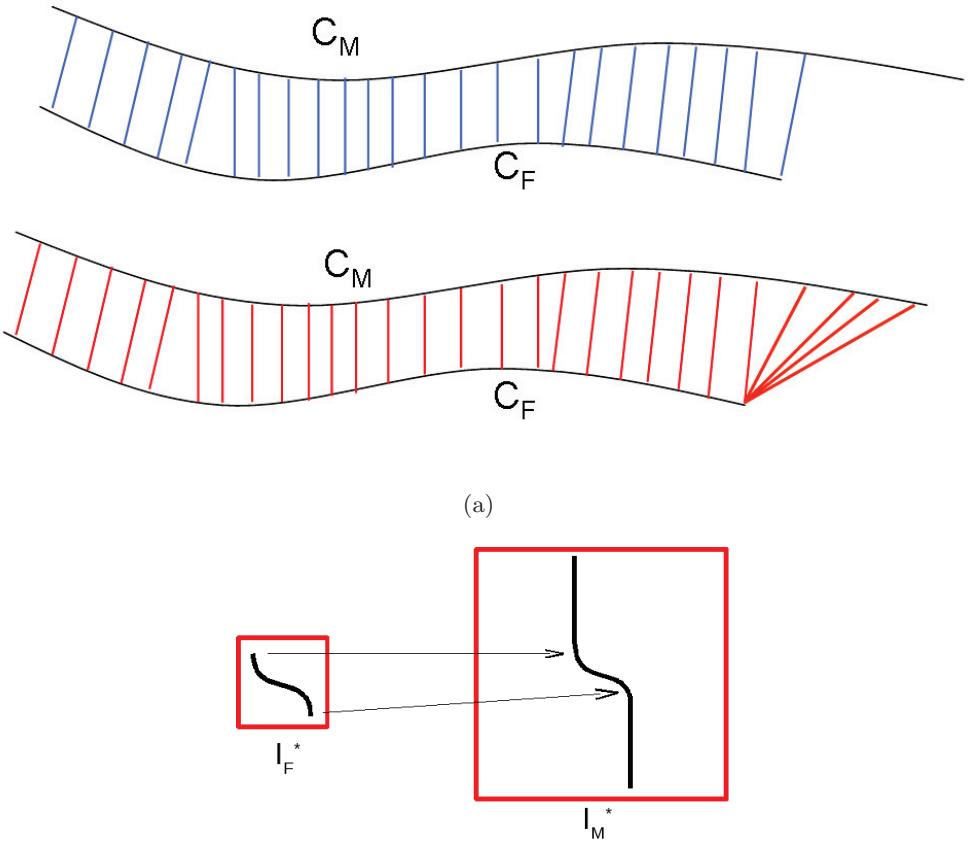


Figure 5.3: Registration Assessment Distance Measurement - (a) The distance measurement concept. At top, the blue lines represent the distance between a vertex $\mathbf{v} \in C_F$ and the the closest vertex in C_M found by function $l(\cdot)$. At the bottom, the red lines represent the distance between a vertex $\mathbf{p} \in C_M$ and the the closest vertex in C_F found by function $l(\cdot)$. As can be seen, from the right-most part of C_M, the fact that the axes are not of the same length is strongly penalized by the metric. Due to the different size of volumes I_F^*, I_M^*, the axis contained by one of the subvolumes can be larger, as shown in (b). To avoid strong penalization of the distance metric, the distance measure is only computed between the transformed extreme vertices of C_F, which are represented by two arrows in (b).

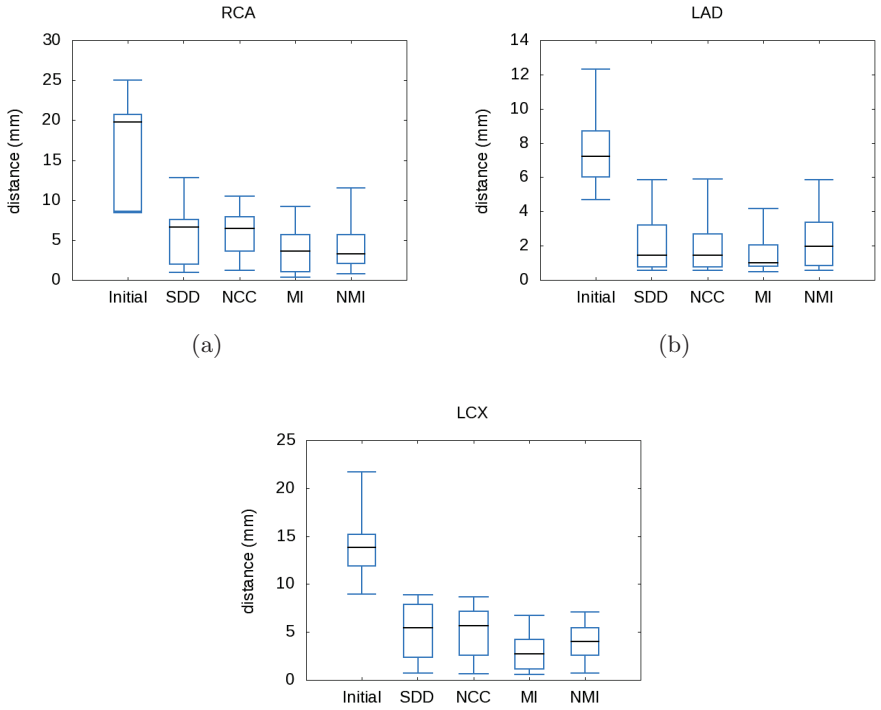


Figure 5.4: Coronary displacement measured through Eq. 5.8 - Initial coronary displacement and residual offset after non-rigid registration using SSD, NCC, MI and NMI metrics for (a) RCA, (b) LAD and (c) LCX arteries.

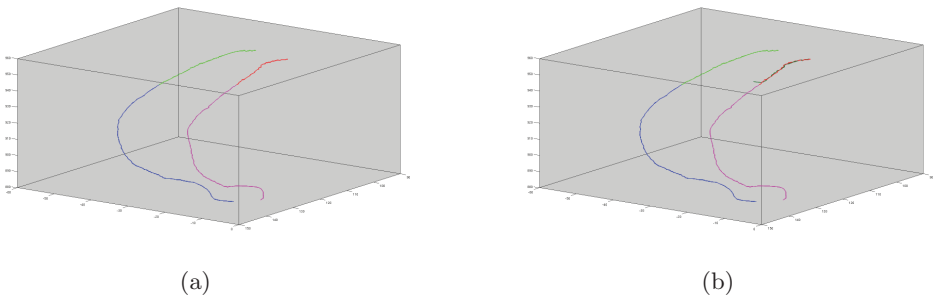
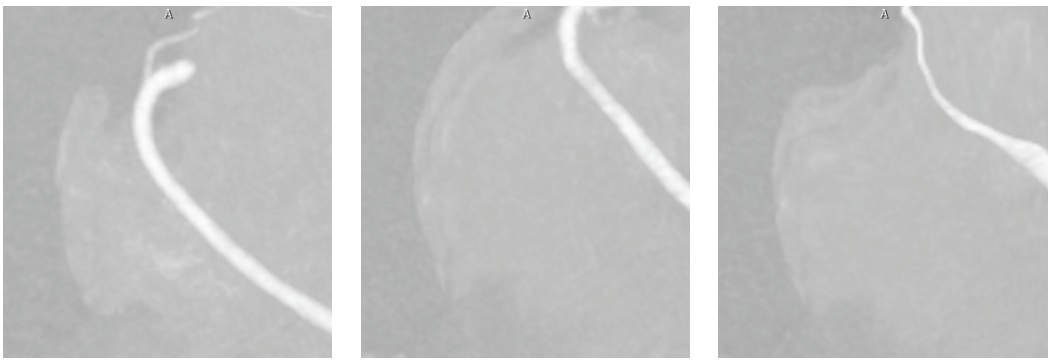
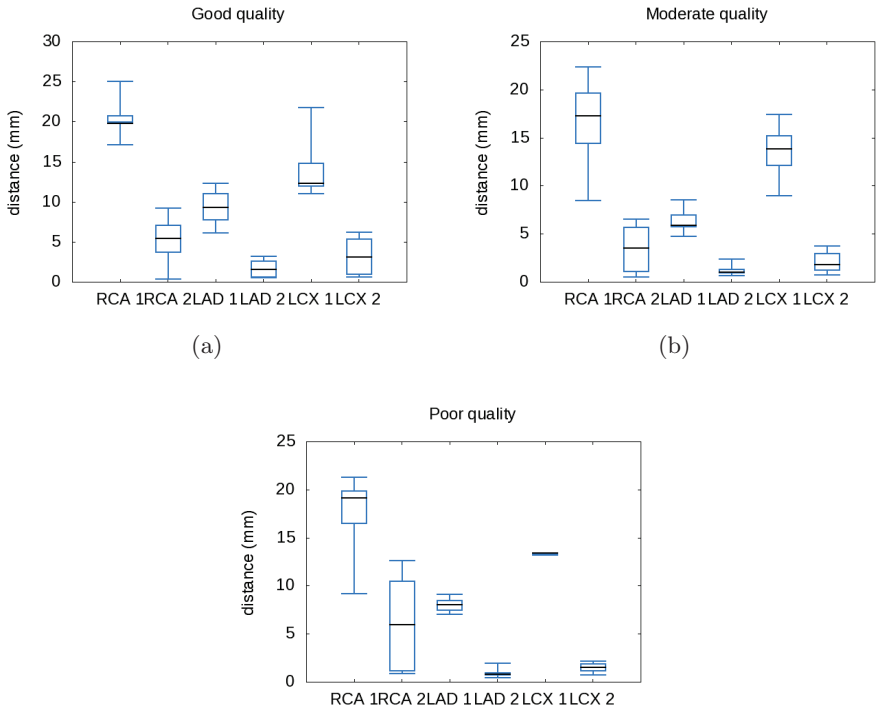


Figure 5.5: Example of coronary artery centerline displacement correction - (a) Initial displacement. The complete centerlines of the RCA at two time-frames are presented in blue and magenta, while green and red represent the respective centerlines segments falling within the VOI to be registered. (b) Dark green line (almost superimposed onto the red one) represents the result of applying the obtained transformation to the green centerline segment.



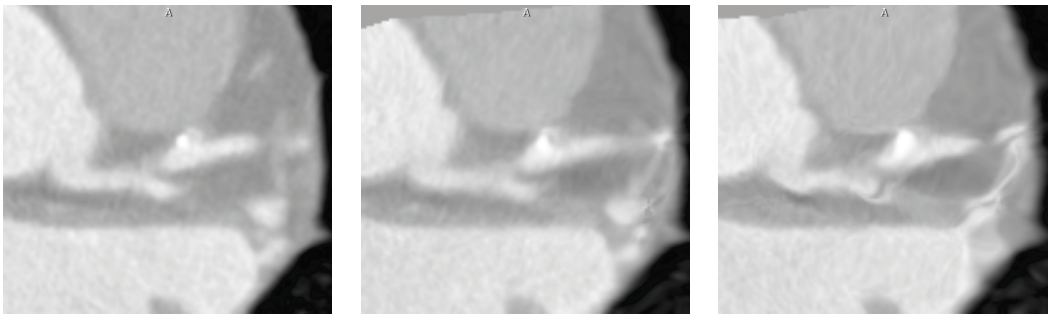
(a) (b) (c)

Figure 5.7: MIP view of a misregistration problem - The initial distance between (a) the RCA segment at time-point 75 and (b) the RCA segment at time-point 40 is too large (35.03 mm). The obtained registration (c) cannot accomplish the structures alignment.



(a) (b) (c)

Figure 5.8: Final displacement comparison as a function of image quality - Initial and final displacements per artery when classifying images as (a) good, (b) moderate and (c) poor quality. On each plot, XXX 1 denotes initial displacement at artery XXX and XXX 2 the residual offset.



(a) (b) (c)

Figure 5.9: Differences in the registration result using direct and incremental approaches - The calcification present in the fixed image (a), as well as the vessel general structure, are better conserved with the direct approach (b). The result obtained with the incremental approach (c) is more blurred.

VOIs by VOIs based on the anomaly detection stage. Our aim is to study how the information from the registered VOIs can be exploited to improve the detection.

Based on the procedure described in Section 5.3, we introduced some modifications in order to be able to compare the information from each of the time-frames. The final procedure is as follows:

1. The best performing configurations of DLD-SVM and LPU (Chapter 4) were applied to 10 datasets (see Section 5.4) at cardiac time-frame 75%. The latter is denoted the fixed image I_F.

2. The potential lesions detected on I_F were used to construct the different VOIs. These are denoted I_F^*.

3. The different I_F^* were used to obtain the corresponding I_M^* volumes using time-frame 40% (denoted I_M).

4. The different volumes, I_F^* and I_M^*, were then non-rigidly co-registered.

5. The best performing configurations of DLD-SVM and LPU were applied to all the obtained $\mathbf{T}(I_M^*)$.

6. Anomaly detection results from every I_F^* and I_M^* are confronted.

Table 5.5 summarizes the number of lesions found in the 10 images I_F as well as how many could be confirmed by the registration process.

Results from Table 5.5 show a pattern that has been seen at the classification stages: LPU has a better capacity to detect anomalies, while DLD-SVM is more conservative. Also, it is possible to see that in two cases (Datasets 04 and 05) an incorrectly detected lesion is rejected by the confrontation with the second time-frame. Lesion evaluation

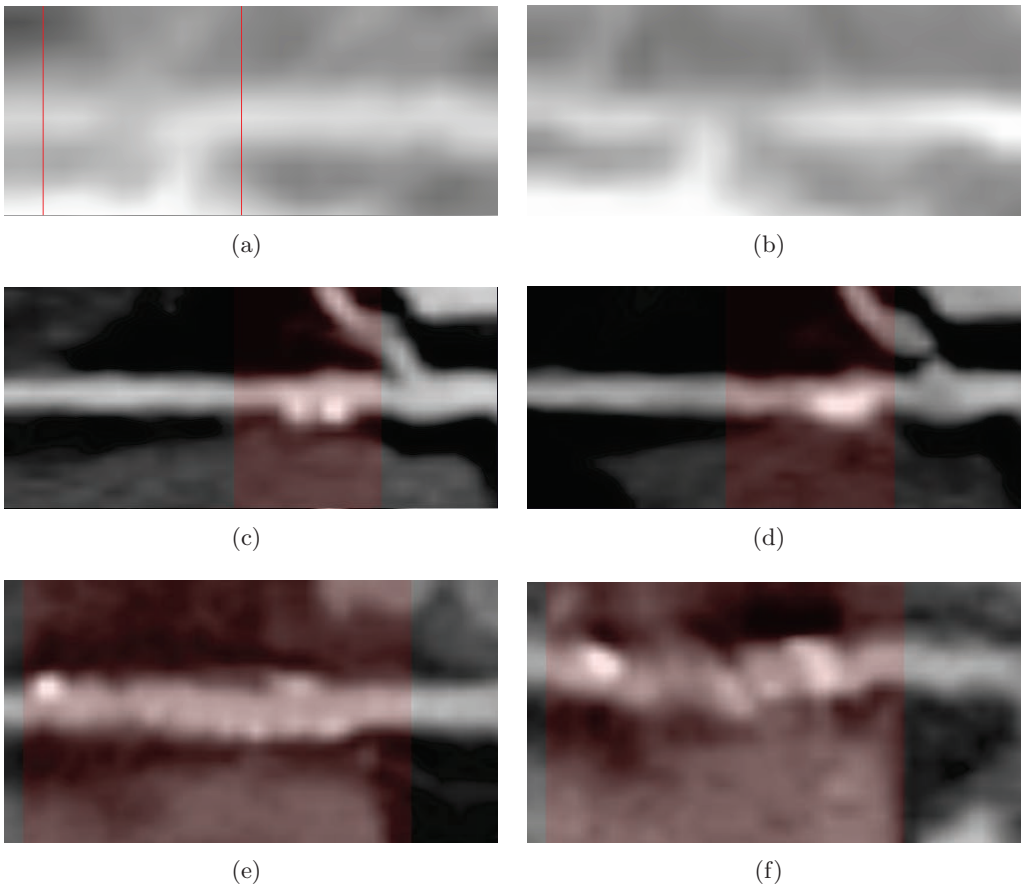


Figure 5.10: Examples of anomaly detection confrontation - Confronting anomaly detection results obtained through LPU. Left column represents time-frame 75% and the right one time-frame 40%. Anomalous sections are marked in translucent red, except in (a), where the anomaly corresponds to the section in-between the two red lines. (a) An anomaly is detected in the RCA, which is rejected in the second time-frame (b). A calcification is detected in the LAD (c) and confirmed in the second time-frame (d). Detection results in the RCA (e) are confirmed in the second time-frame (f). However, the registration result in the latter case is of poor quality.

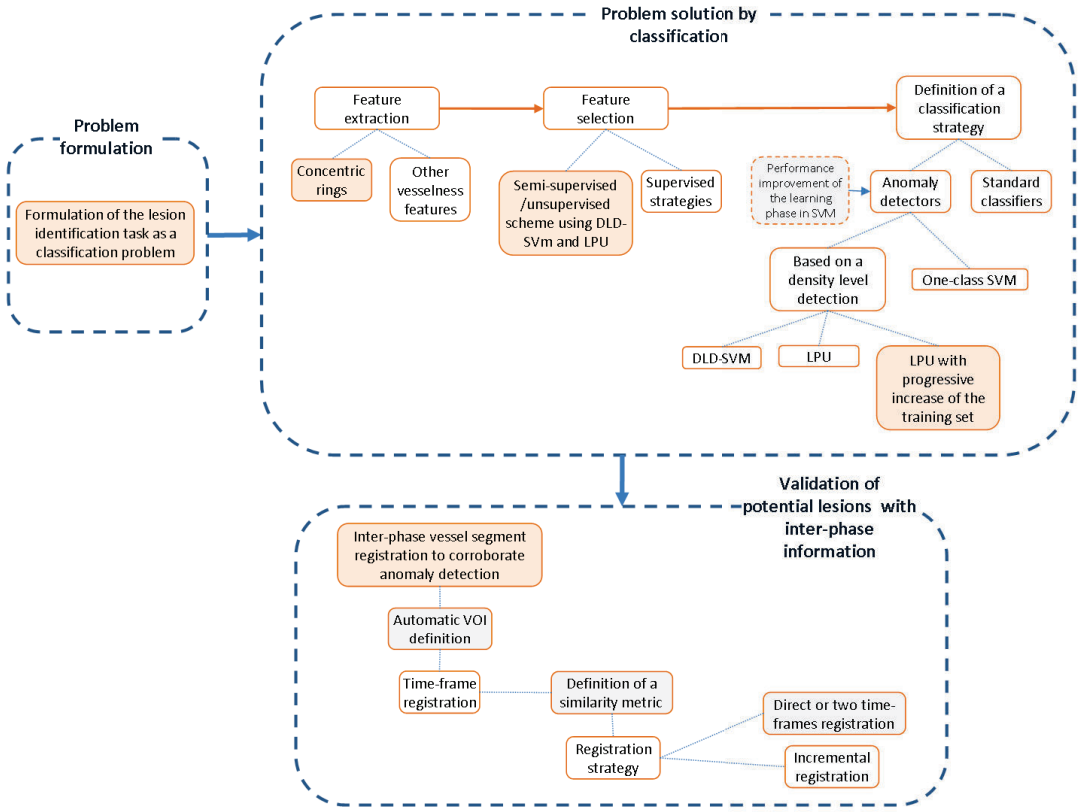


Figure 6.1: Summary of methodological contributions - The methodological contributions are attached to three main axes: the formulation of the lesion identification task as an anomaly detection problem, the solution of such a problem through classification and the validation of potential lesions with inter-phase information. The main contributions at each of the axis are marked in light red, while minor contributions are marked in gray.

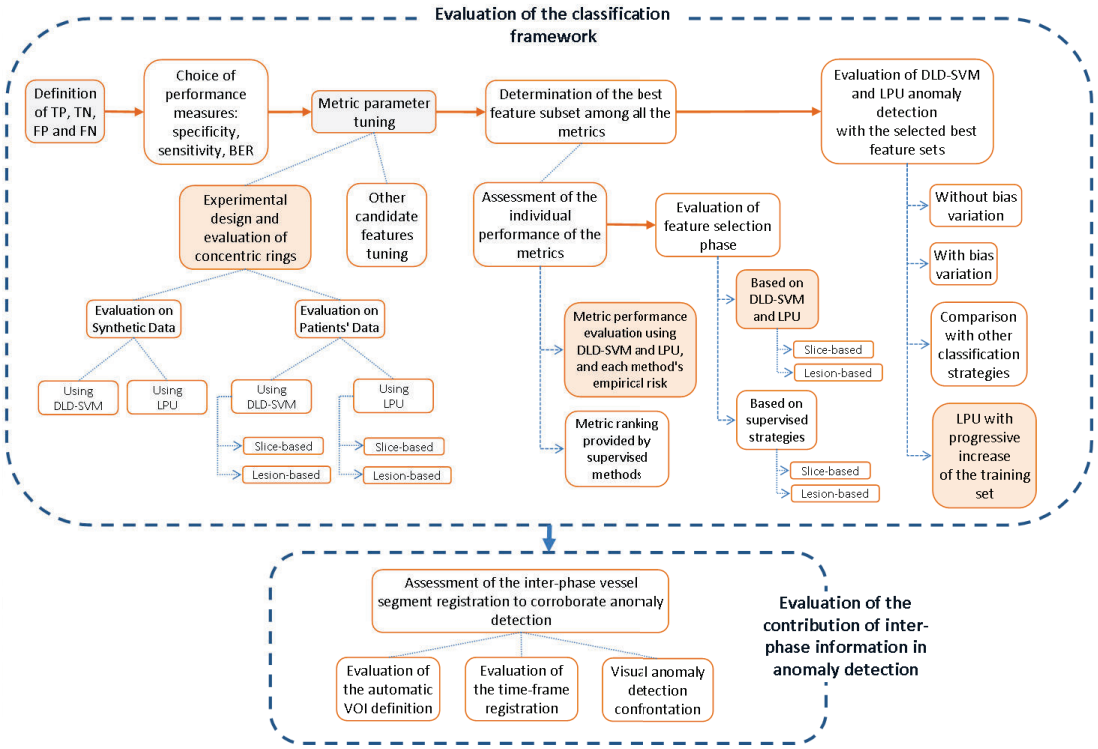


Figure 6.2: Evaluation of the different elements constituting the proposed methodological framework - Evaluation of the framework can be divided into two large albeit unequal groups: evaluation of the classification framework and evaluation of the use inter-phases information. The main contributions in the evaluation are highlighted in light red, while the minor ones appear in gray.

They define the margin $\frac{2}{\|\mathbf{w}\|}$. It is possible to find H_1 and H_2, which give the maximum margin, by minimizing $\|\mathbf{w}\|$, subject to constraint from Eq. A.4. To make the optimization problem easier to solve, the minimization of $\|\mathbf{w}\|$ is replaced by the minimization of:

$$\frac{1}{2}\|\mathbf{w}\|^2 \tag{A.6}$$

under the same constraint (Eq. A.4). Figure A.1 illustrates the previously formulated problem in a two-dimensional case. Training points $\mathbf{x}_i$, for which Eq. A.5 is satisfied, belong either to H_1 or to H_2d, and are called support vectors. The maximization of the margin implies defining support vectors as far as possible from the hyperplane.

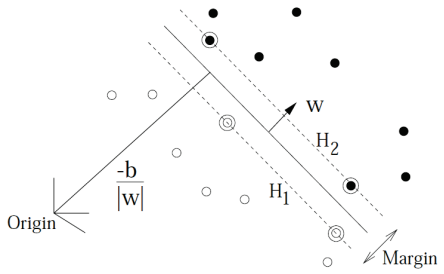


Figure A.1: Linear separating hyperplanes - The optimal hyperplane is shown in a solid line, while support vectors are circled. Illustration from (Burges, 1998).

Eq A.6 is called an objective function, while Eq. A.4 represents the inequality constraint. Both, form a constrained optimization problem. This type of problems are solved by the introduction of Lagrange multipliers $\alpha_i \geq 0$ and a Lagrangian

$$L_P \equiv \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{N} \alpha_i (y_i \left(\langle \mathbf{x}_i, \mathbf{w}\rangle + b\right) - 1) \tag{A.7}$$

The Lagrangian has to be minimized w.r.t the primal variables $\mathbf{w}$ and b and maximized w.r.t the dual variables α_i. This is a convex quadratic programming problem. Therefore, it can be solved through its dual problem: maximizing L_P, subject to the constraints imposed by both primal and dual variables ($\alpha_i \geq 0$).

Requiring that the derivatives of L_P w.r.t. $\mathbf{w}$ and b must vanish, gives

$$\sum_i \alpha_i y_i = 0 \tag{A.8}$$

and

$$\mathbf{w} = \sum_i \alpha_i y_i \mathbf{x}_i \tag{A.9}$$

Replacing Eqs. A.8 and A.9 in A.7, it is possible to eliminate the primal variables $\mathbf{w}$ and b, arriving to the so-called dual optimization problem which is the one solved in

mapping data into a feature space H using an appropriate kernel function and finding a function $f(x)$ that is positive inside S and negative elsewhere.

In order to make use of SVM's it is necessary to define two clasess. Therefore, Schölkopf et al. (2001) treat the origin as the only member of the second class. Under this formulation, the goal of the algorithm is to find the hyperplane to separates the mapped vectors from the origin with maximum margin (Figure A.2)

$$f(x) = \begin{cases} +1 & \text{if } x \in S \\ -1 & \text{if } x \in \bar{S} \end{cases} \tag{A.16}$$

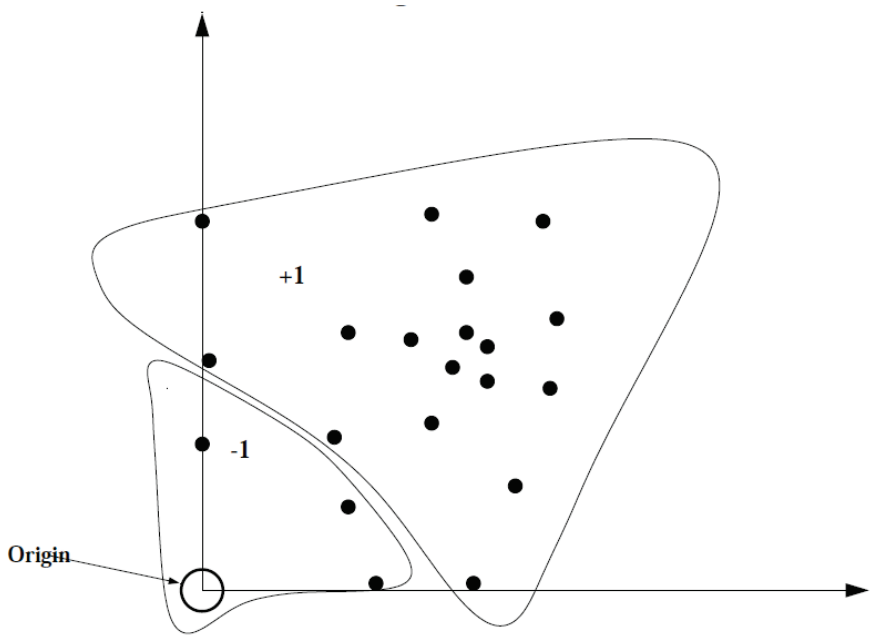


Figure A.2: One-class SVM illustration - The origin is the only original member of the second class. Circles surrounding data points represent the optimal data partition found by the algorithm.

Manevitz and Yousef (2001) Based on A.13, the following quadratic program, is defined in order to separate the data from the origin:

$$\min \frac{1}{2}\|\mathbf{w}\|^2 + \frac{1}{\upsilon N}\sum_{i=1}\xi_i - b \tag{A.17}$$

subject to the restrictions imposed by Eqs. A.11, A.12 and where $\mathbf{w}$, b and N have the same significance as in the previous section. The difference introduced in this formulation is the presence of υ.

Since nonzero slack variables ξ_i are penalized in the objective function, it is expected that if $\mathbf{w}$ and b solve the problem, then the decision function

$$f(x) = sign(\langle \Phi(\mathbf{x}_i), \mathbf{w} \rangle - b) \tag{A.18}$$