Diagnosis and prognosis of cardiovascular diseases by means of texture analysis in magnetic resonance imaging

VNIVERSITAT DE VALÈNCIA

Andrés Martín Larroza Santacruz

DOCTORAL THESIS

Supervisors:
Prof. Dr. Vicente Bodí Peris
Prof. Dr. David Moratal Pérez

Programa de Doctorado en Ingeniería Electrónica
Departamento de Ingeniería Electrónica
Universitat de València
September 2017

Diagnosis and prognosis of cardiovascular diseases by means of texture analysis in magnetic resonance imaging

by

Andrés Martín Larroza Santacruz

DOCTORAL THESIS

Supervisors:

Prof. Dr. Vicente Bodí Peris
Department of Medicine
Universitat de València

Prof. Dr. David Moratal Pérez
Center for Biomaterials and Tissue Engineering
Universitat Politècnica de València

VNIVERSITAT DE VALÈNCIA

Programa de Doctorado en Ingeniería Electrónica
Departamento de Ingeniería Electrónica
Universitat de València
September 2017

VNIVERSITAT DE VALÈNCIA

Escola Tècnica Superior d'Enginyeria

Departamento de Ingeniería Electrónica
Escuela Técnica Superior de Ingeniería

D. VICENTE BODÍ PERIS, Doctor en Medicina, Profesor Titular del Departamento de Medicina de la Universitat de València, y

D. DAVID MORATAL PÉREZ, Doctor en Ingeniería de Telecomunicación, Profesor Titular del Departamento de Ingeniería Electrónica de la Universitat Politècnica de València,

HACEMOS CONSTAR QUE:

Andrés Martín Larroza Santacruz ha realizado bajo nuestra dirección el trabajo de investigación titulado "Diagnosis and prognosis of cardiovascular diseases by means of texture analysis in magnetic resonance imaging", que se presenta en esta memoria para optar al grado de Doctor.

Y para que así conste a los efectos oportunos, firmamos el presente certificado, en Valencia, a _____.

_____ Vicente Bodí Peris

_____ David Moratal Pérez

Contents

Contents i

Declarations v

Abbreviations and Acronyms vii

Resumen ix

Abstract xxi

Chapter 1

Introduction 1

1.1 Motivation 1

1.2 Related Work 3

1.3 Objectives 5

1.4 Contributions to Knowledge 5

1.5 Thesis Structure 6

Chapter 2

Background on Cardiac MRI 9

2.1 Introduction 9

2.2 Heart Anatomy and Function 9

2.2.1 Basic Principles 9

2.2.2 Myocardial Infarction 11

2.2.3 Cardiac Imaging Planes 12

2.2.4 Cardiac Segments 14

2.2.5 Cardiac Function Parameters 15

2.3 MRI Principles 16

2.3.1 MRI Physics 16

2.3.2 The MRI system 18

2.3.3 Pulse Sequences 19

2.4 Cardiac MRI Techniques 20

2.4.1 Cine CMR 20

2.4.2 Late Gadolinium Enhancement CMR 21

2.4.3 Edema CMR 22

2.4.4 Perfusion CMR 23

Chapter 3

Texture Analysis in MRI 25

3.1 Introduction 25

3.2 MRI Acquisition 26

3.2.1 Sequences 26
3.2.2 Influence of Spatial Resolution and SNR 27
3.2.3 Influence of Field Strength 27
3.2.4 Multicenter Studies 28

3.3 Region of Interest Definition 29

3.3.1 Size of the Region of Interest 30
3.3.2 Feature Maps 30

3.4 Region of Interest Preprocessing 31

3.4.1 Interpolation 31
3.4.2 Normalization 32
3.4.3 Inhomogeneity Correction 33
3.4.4 Quantization of Gray-levels 33

3.5 Texture Feature Extraction 34

3.5.1 Histogram 34
3.5.2 Absolute Gradient 36
3.5.3 Gray-Level Co-occurrence Matrix 37
3.5.4 Gray-level Run-length Matrix 40
3.5.5 Gray-level Size Zone Matrix 43
3.5.6 Neighborhood Gray-tone Difference Matrix 44
3.5.7 Autoregressive models 45
3.5.8 Wavelets 47
3.5.9 Local Binary Patterns 48
3.5.10 Volumetric Texture Feature Extraction 49
3.5.11 Feature Extraction Tools 50

Chapter 4

Overview on Machine Learning 51

4.1 Introduction 51

4.2 Feature Selection 52

4.2.1 Filter Methods 52
4.2.2 Wrapper Methods 53

4.3 Classification Methods 55
4.3.1 K-Nearest Neighbors 56
4.3.2 Artificial Neural Network 57
4.3.3 Random Forest 58
4.3.4 Support Vector Machines 60

4.4 Model Validation and Evaluation 64
4.4.1 Over-fitting and Bias-variance Trade-off 64
4.4.2 Model Tuning 64
4.4.3 Resampling techniques 65
4.4.4 Measures of Classification Performance 66

Chapter 5

Segmentation of Infarcted Myocardium in LGE CMR. A preliminary multicenter evaluation 69

5.1 Introduction 69

5.2 Material and Methods 70
5.2.1 MRI Data 70
5.2.2 Region Labeling 71
5.2.3 Feature Extraction 71
5.2.4 Feature Selection 71
5.2.5 SVM Training 72

5.3 Experiments and Results 73

5.4 Conclusion 76

Chapter 6

Differentiation between Acute and Chronic Myocardial Infarction 77

6.1 Introduction 77

6.2 Materials and Methods 78
6.2.1 Study Group and Imaging Protocol 78
6.2.2 Region of Interest Definition 80
6.2.3 Region of Interest Preprocessing 81
6.2.4 Texture Feature Extraction 81
6.2.5 Feature Selection and Classification 82
6.2.6 Model Evaluation 83

6.3 Results 85
6.3.1 LGE CMR 85
6.3.2 Cine CMR 92

6.4 Discussion **99**

6.5 Conclusion **100**

Chapter 7

Detection of Infarcted Myocardial Segments in Cine CMR **101**

7.1 Introduction **101**

7.2 Materials and Methods **102**

7.2.1 Study Group and Imaging Protocol 102

7.2.2 Region of Interest Definition 103

7.2.3 Region of Interest Preprocessing 105

7.2.4 Texture Feature Extraction 105

7.2.5 Classification 107

7.2.6 Model Evaluation 107

7.3 Results **108**

7.4 Discussion **118**

7.5 Conclusion **119**

Chapter 8

Conclusions **121**

Publications **123**

References **125**

Declarations

- This thesis was supported by grant FPU12/01140 from the Spanish Ministerio de Educación, Cultura y Deporte (MECD).

- All patients gave written informed consent and the study protocol was approved by the institutional committee on human research and conforms to the ethical guidelines of the 1975 Declaration of Helsinki.

The work presented in this thesis, including data analysis, was carried out by the author except in the cases outlined below:

- MRI images were collected from the Hospital Clínico Universitario of Valencia: The cardiologist staff of ERESA was responsible for collecting magnetic resonance imaging data and for diagnosis of each case.
- The manual segmentations for ROI definition previous to texture analysis were performed with assistance of Dr. María Pilar López Lereu and Dr. José Vicente Monmeneu Menadas, from ERESA.

Abbreviations and Acronyms

AHA = American Heart Association
AMI = Acute Myocardial Infarction
ANN = Artificial Neural Network
AUC = Area under the ROC Curve
CMI = Chronic Myocardial Infarction
CMR = Cardiovascular Magnetic Resonance
ECG = Electrocardiography
FLAIR = Fluid Attenuated Inversion Recovery
FN = False Negatives
FOV = Field of View
FP = False Positives
FWHM = Full Width Half Maximum
GLCM = Gray-level Co-occurrence Matrix
GLN = Gray-level Nonuniformity
GLRLM = Gray-level Run-length Matrix
GLSZM = Gray-level Size Zone Matrix
GLV = Gray-level Variance
HCM = Hypertrophic Cardiomyopathy
HGRE = High Gray-level Run Emphasis
KNN = k-Nearest Neighbors
LBP = Local Binary Patterns
LDA = Linear Discriminant Analysis
LGE = Late Gadolinium Enhancement
LGRE = Low Gray-level Run Emphasis
LRE = Long Run Emphasis
LRHGE = Long Rung High Gray-level Emphasis
LRLGE = Long Run Low Gray-level Emphasis
LV = Left Ventricle
LZLGE = Large Zone Low Gray-level Emphasis
MI = Myocardial Infarction
MR = Magnetic Resonance

MRI = Magnetic Resonance Imaging
MVO = Microvascular Obstruction
NGTDM = Neighborhood Gray-tone Difference Matrix
NPV = Negative Predictive Value
PCA = Principal Component Analysis
PET = Positron Emission Tomography
PPV = Positive Predictive Value
RF = Radiofrequency, Random Forest
RFE = Recursive Feature Elimination
RLN = Run-length Nonuniformity
RLV = Run-length Variance
ROC = Receiver Operating Characteristics
ROI = Region of Interest
RP = Run Percentage
SD = Standard deviation
SNR = Signal to Noise Ratio
SRE = Short Run Emphasis
SRHGE = Short Run High Gray-level Emphasis
SRLGE = Short Run Low Gray-level Emphasis
SSFP = Steady State Free Precession
SVM = Support Vector Machine
SZHGE = Small Zone High Gray-level Emphasis
T1 = Longitudinal Relaxation Time
T2 = Transverse Relaxation Time
TE = Echo Time
TI = Inversion Time
TN = True Negatives
TP = True Positives
TR = Repetition Time
VOI = Volume of Interest
ZP = Zone Percentage
ZSN = Zone-size Nonuniformity
ZSV = Zone-size Variance

Resumen

Introducción

Las enfermedades cardiovasculares constituyen la principal causa de morbilidad y mortalidad. Como consecuencia del aumento de la esperanza de vida esta tendencia será aún más acusada en los próximos años. Es por ello que este grupo de enfermedades está entre las entidades que demandan una mayor atención en nuestro sistema sanitario. La disposición de herramientas fiables y económicas que permitan un diagnóstico rápido y dinámico de los pacientes con estas patologías será crucial para evitar el consumo de recursos innecesarios en pruebas diagnósticas complementarias, y así disponer de información pronostica fiable para el adecuado tratamiento de los pacientes.

En esta tesis doctoral se aborda principalmente uno de los síndromes cardiovasculares que con más frecuencia motivan la atención de los pacientes en las instituciones sanitarias, el infarto agudo de miocardio. Cuando esta entidad se presenta parte del miocardio afectado sufre necrosis. El miocardio viable es aquel miocardio no necrótico cuya capacidad contráctil queda disminuida, pero es potencialmente recuperable. La determinación de la viabilidad miocárdica es de utilidad para predecir la función sistólica posterior a un infarto agudo de miocardio, cuya importancia es vital para determinar el consecuente tratamiento del paciente.

La resonancia magnética cardíaca (CMR) se considera, en la actualidad, el método de imagen diagnóstica de referencia dado que permite explorar la anatomía del corazón de forma no invasiva y valorar su utilidad, no sólo de forma cualitativa, sino también cuantitativa. Con ella es posible analizar múltiples variables para predecir la función sistólica tardía. La modalidad rutinaria de CMR es la denominada cine CMR, pero en esta modalidad no es posible visualizar el miocardio infartado. Para poder visualizarlo, se utiliza la modalidad CMR con realce tardío de gadolinio (LGE). Existen técnicas de procesamiento de imágenes con las que se podrían extraer parámetros cuantitativos adicionales facilitando y mejorando el diagnóstico y pronóstico. Una de estas técnicas es el análisis de texturas que hasta la fecha ha sido poco estudiada en CMR.

La textura de una imagen puede ser descrita con palabras como: finura, aspereza, irregularidad y suavidad por mencionar algunas. El análisis de texturas es una técnica que permite cuantificar la textura de una imagen mediante diversos métodos de cálculo, obteniéndose de esta manera los denominados parámetros texturales. El análisis de texturas ha sido utilizado en varias aplicaciones médicas permitiendo la clasificación de tejidos y el diagnóstico de patologías. Con esta técnica es posible extraer información no apreciable visualmente en imágenes rutinarias con lo que se puede evitar el uso de técnicas diagnósticas más complejas. Debido al elevado número de técnicas de análisis de texturas, es posible extraer cientos de parámetros texturales, por lo que la correcta aplicación de técnicas de aprendizaje máquina como la selección de características y clasificadores es un requisito fundamental para la consecución de resultados satisfactorios.

El trabajo de investigación en el que se enmarca este proyecto de tesis se basa en la aplicación del análisis de texturas en CMR para la clasificación y detección de miocardio infartado. La hipótesis de partida es que el tejido cardiaco presenta textura diferente según su afección, la cual muchas veces no se aprecia visualmente, y que podría ser detectada mediante la correcta aplicación del análisis de texturas. La relevancia de este proyecto se basa en la posibilidad de detectar el miocardio infartado en imágenes cine CMR, reduciendo al máximo la necesidad de realizar adquisiciones en la modalidad LGE, lo que implica una reducción en los costos sanitarios y en el tiempo de diagnóstico. Además, los mayores beneficiados serían aquellos pacientes con contraindicaciones al contraste gadolinio. Por lo tanto, el propósito de esta tesis es aplicar el análisis de texturas en imágenes convencionales de CMR para la evaluación de pacientes con infarto de miocardio, como alternativa a métodos existentes.

En esta tesis se presentan tres aplicaciones del análisis de texturas en imágenes de resonancia magnética para la evaluación de pacientes con infarto de miocardio. La aplicación de estos estudios experimentales se basó en tres técnicas fundamentales que son presentadas en el documento de tesis como introducción teórica en los capítulos 2, 3 y 4 que describen respectivamente: la resonancia magnética cardíaca (CMR), el análisis de texturas, y las técnicas de aprendizaje máquina.

Objetivos

El objetivo principal de esta tesis doctoral se centra en la aplicación del análisis de texturas en imágenes convencionales de CMR, como método alternativo a las técnicas actuales para la valoración del infarto de miocardio. Para ello se proponen los siguientes objetivos específicos:

a) Explorar la capacidad de los parámetros texturales para diferenciar el miocardio infartado del miocardio remoto en LGE CMR, y evaluar un método de segmentación basado en texturas en un estudio multicentro preliminar.
b) Investigar la capacidad del análisis de texturas de imágenes LGE CMR para diferenciar infartos de miocardio en estado agudo y crónico, y estudiar la posibilidad de solucionar este problema usando únicamente imágenes cine CMR en las cuales el infarto es visualmente imperceptible.
c) Detectar los segmentos infartados no-viables aplicando el análisis de texturas en cine CMR, como potencial técnica alternativa libre del contraste gadolinio.

Metodología

Se han estudiado, implementado y analizado tres aplicaciones del análisis de texturas y aprendizaje máquina en CMR. Dichos estudios buscan cumplir los objetivos específicos mencionados y son presentados en el documento de la tesis doctoral como tres estudios independientes descritos en los capítulos 5, 6 y 7.

1. Detección del infarto de miocardio en imágenes de realce tardío de gadolinio (LGE) CMR

La segmentación del miocardio infartado se realiza rutinariamente usando valores umbrales de intensidad. No obstante, aunque exista un consenso en la práctica clínica para el uso de la técnica que etiqueta como tejido infartado a aquellas zonas del miocardio que presenten 5 desviaciones estándar por encima del miocardio remoto, o la técnica denominada FWHM (full width and half máximum), todavía existen limitaciones. Por este motivo, en este primer estudio experimental hemos decidido utilizar parámetros texturales y técnicas de aprendizaje máquina para proponer un método alternativo para la segmentación del miocardio infartado.

En este estudio se han utilizado imágenes LGE CMR provenientes de diez pacientes masculinos con infarto crónico de miocardio. Éstas imágenes fueron adquiridas con un equipo de resonancia magnética de 1,5T (Sonata Magnetom,

Siemens, Erlangen, Alemania) y fueron las imágenes utilizadas para la etapa de entrenamiento del modelo predictivo utilizado posteriormente para la segmentación de imágenes en datos de prueba. Los datos de prueba corresponden a imágenes LGE CMR de cinco pacientes con infarto crónico de miocardio adquiridos con un equipo de resonancia magnética de 1,5T (Achieva, Philips, Best, Holanda).

La etapa de entrenamiento consistió en extraer los parámetros de texturas de las regiones de interés (ROI), en este caso el miocardio infartado previamente segmentado con ayuda de cardiólogos expertos. La extracción de parámetros texturales se realizó con el software MaZda, versión 4.6 (Instituto de Electrónica, Universidad Politécnica de Lodz, Lodz, Polonia). Con dicho software se computaron un total de 122 parámetros texturales derivados de cuatro métodos: histograma (9 parámetros), matriz de co-ocurrencia (88 parámetros), matriz run-length (20 parámetros), y modelo autoregresivo (5 parámetros). Los parámetros texturales fueron calculados en ventanas de 5 x 5 píxeles dentro de cada ROI en 2D, por lo que en total se obtuvo un vector de datos de 2976 muestras. Este vector de datos fue separado en entrenamiento (50%) y validación (50%). El modelo seleccionado para entrenar los datos fue un support vector machine (SVM) con kernel radial, junto con un método de selección de características basado en el método de búsqueda de eliminación de parámetros recursivo (RFE). El modelo final consistió en 17 parámetros texturales, que usados en conjunto con el SVM otorgaron un área bajo la curva (AUC) ROC de 0,944 en los datos de validación. Este modelo final fue el utilizado para hacer la segmentación en los datos de prueba.

La segmentación en los datos de prueba requiere la delineación previa del miocardio, ya que el algoritmo de segmentación extrae los parámetros texturales dentro de esa región. Es decir, de cada pixel dentro del miocardio, se calculan los parámetros texturales tomando en cuenta los píxeles vecinos dentro de una ventana de 5 x 5 píxeles. La etapa de entrenamiento fue implementada en lenguaje R versión 3.0.1 (R Development Core Team, Viena, Austria), y el algoritmo de segmentación en Matlab 2014b (MathWorks Inc., Natick, MA).

Para evaluar la calidad de segmentación se utilizó el coeficiente Dice que compara la segmentación del método propuesto con la segmentación verdadera que fue obtenida junto con las imágenes de prueba de la base de datos de la STACOM challenge MICCAI 2012 (http://stacom.cardiacatlas.org/ventricular-infarction-

challenge/). Los resultados obtenidos se presentan en la tabla 1:

Tabla 1. *Coeficientes Dice en los casos de prueba.*

Caso	1	2	3	4	5	Total
Promedio	0,69	0,59	0,59	0,76	0,72	0,71
Desviación Estándar	(0,23)	(0,06)	(0,11)	(0,04)	(0,08)	(0,12)

En general el coeficiente Dice obtenido fue de 0,71; el cual es aceptable ya que un índice de 1 indica segmentación perfecta y un índice de 0 indica pésima segmentación. Lo más representativo de este primer estudio experimental es que se comprobó la transferibilidad del análisis de texturas, ya que los datos de prueba fueron obtenidos a partir de imágenes adquiridas con un equipo de resonancia magnética completamente diferente a aquél usado para adquirir las imágenes de entrenamiento. Esto es muy relevante para la práctica clínica y fue el estudio que sirvió de motivación para los dos siguientes que pretenden dar soluciones más innovadoras a las existentes en la práctica cardiológica.

2. Diferenciación entre infarto de miocardio agudo y crónico

Si bien las imágenes LGE CMR realzan y hacen posible la detección del miocardio infartado, no es posible hacer una discriminación visual del estado de la lesión, ya sea reciente (agudo) o avanzado (crónico). La identificación del estado del infarto es de especial importancia cuando ambas entidades coexisten, es decir el paciente presenta más de un infarto, ya que el tratamiento dependerá del tipo de lesión. En este estudio, se ha implementado el análisis de texturas para la diferenciación de infartos agudos de crónicos.

Se incluyeron 44 casos: 22 pacientes con infarto agudo y 22 pacientes con infarto crónico. Fueron utilizadas imágenes cine CMR y LGE CMR adquiridas con un equipo de resonancia magnética de 1,5T (Sonata Magnetom, Siemens, Erlangen, Alemania). Las imágenes fueron adquiridas a la primera semana (agudo) y al sexto mes (crónico) del episodio de infarto. Se han analizado independientemente las imágenes LGE CMR y cine CMR. Incluimos el análisis independiente de cine CMR partiendo de la hipótesis de que el análisis de texturas permite evidenciar la información no visible, ya que el infarto no se visualiza en las imágenes cine CMR.

Las regiones de interés (ROIs) utilizadas para extraer los parámetros texturales fueron: el miocardio infartado en las imágenes LGE CMR, y el miocardio completo en cine CMR. Los parámetros texturales fueron computados en 2D para cada corte con el software MaZda, versión 4.6 (Instituto de Electrónica, Universidad Politécnica de Lodz, Lodz, Polonia). En total se calcularon 279 parámetros texturales para cada ROI, derivados de 6 métodos: histograma, gradiente absoluto, matriz de co-ocurrencia, matriz run-length, modelo autoregresivo, y transformadas Wavelets.

Dos métodos de selección de características fueron implementados, uno tipo filtro basado en el coeficiente Fisher y el otro tipo *wrapper:* el SVM-RFE. Con ambos métodos se obtienen un ranking de parámetros texturales. Los parámetros del más al menos importante fueron agregados uno por uno y utilizados como entrada para entrenar distintos modelos clasificadores. Los modelos clasificadores utilizados fueron: k-nearest neighbors (kNN), red neuronal artificial (ANN), random forest (RF); y support vector machine (SVM) con kernel lineal, radial y polinomial. Los modelos fueron evaluados con validación cruzada del tipo anidado, es decir, la selección de características fue incluida dentro del bucle de validación para evitar una sobreestimación de los resultados. No se encontraron diferencias estadísticamente significativas al comparar los diferentes modelos, esto tomando siempre el subconjunto de parámetros óptimo según el método de selección de características.

En el caso de las imágenes LGE CMR, el mejor modelo fue el SVM polinomial usando los mejores 99 parámetros texturales según el ranking otorgado por el SVM-RFE (AUC = 0,873; IC: 0,85 - 0,88). Para las imágenes cine CMR, el mejor modelo fue el SVM lineal con 22 parámetros texturales según el ranking SVM-RFE (AUC = 0,831; IC: 0,80 - 0,85). El resumen de resultados se muestra en las tablas 2 y 3.

Con los resultados obtenidos se pudo comprobar el potencial del análisis de texturas para la diferenciación de infartos agudos de crónicos usando imágenes LGE CMR e incluso cine CMR, modalidad en la que el miocardio infartado no se aprecia visualmente. Sin embargo, la correcta clasificación de ambas entidades no es directa, sino que se logra mediante la correcta aplicación de algún método de selección de características en conjunto con un modelo predictivo. En ese sentido, el presente estudio también permitió corroborar la importancia de la selección de características para obtener óptimos resultados cuando se tratan datos de alta dimensionalidad.

Tabla 2. *Clasificación en LGE CMR.*

Fisher - LGE CMR						
Model	**AUC**	**IC**	**Sensibilidad**	**IC**	**Especificidad**	**IC**
KNN	0,831	(0,80 – 0,85)	0,732	(0,69 – 0,76)	0,848	(0,81 – 0,88)
ANN	0,851	(0,83 – 0,86)	0,819	(0,79 – 0,84)	0,826	(0,80 – 0,85)
RF	0,862	(0,84 – 0,88)	0,819	(0,79 -0,84)	0,861	(0,83 – 0,89)
SVM-linear	0,867	(0,85 – 0,88)	0,846	(0,81 – 0,87)	0,84	(0,81 – 0,86)
SVM-radial	0,854	(0,83 – 0,87)	0,789	(0,75 – 0,82)	0.865	(0,83 – 0,89)
SVM-poly	0,865	(0,84 – 0,88)	0,831	(0,80 – 0,85)	0,851	(0,82 – 0,87)
SVMRFE - LGE CMR						
Model	**AUC**	**IC**	**Sensibilidad**	**IC**	**Especificidad**	**IC**
KNN	0,821	(0,79 – 0,84)	0,73	(0,69 – 0,76)	0.835	(0,80 – 0,86)
ANN	0,851	(0,83 – 0,86)	0,789	(0,75 – 0,81)	0.859	(0,83 – 0,88)
RF	0,858	(0,83 – 0,87)	0,8	(0,77 – 0,82)	0.865	(0,84 – 0,88)
SVM-linear	0,857	(0,84 – 0,87)	0,811	(0,78 – 0,83)	0.829	(0,80 – 0,84)
SVM-radial	0,834	(0,81 – 0,85)	0,795	(0,76 – 0,82)	0.866	(0,84 – 0,88)
SVM-poly	0,873	(0,85 – 0,88)	0,82	(0,79 – 0,84)	0.848	(0,82 – 0,87)

Tabla 3. *Clasificación en cine CMR.*

Fisher - Cine CMR						
Model	**AUC**	**CI**	**Sensitivity**	**CI**	**Specificity**	**CI**
KNN	0,711	(0,67 – 0,74)	0,697	(0,65 – 0,73)	0,671	(0,63 – 0,70)
ANN	0,797	(0,77 – 0,82	0,770	(0,74 – 0,79)	0,804	(0,77 – 0,83)
RF	0,797	(0,77 – 0,81)	0,773	(0,75 – 0,79)	0,767	(0,73 – 0,79)
SVM-linear	0,797	(0,77 – 0,81)	0,747	(0,72 – 0,77)	0,822	(0,79 – 0,85)
SVM-radial	0,737	(0,71 – 0,76)	0,722	(0,68 – 0,75)	0,760	(0,73 – 0,79)
SVM-poly	0,802	(0,78 – 0,82)	0,769	(0,74 – 0,79)	0,790	(0,76 – 0,81)
SVMRFE - Cine CMR						
Model	**AUC**	**CI**	**Sensitivity**	**CI**	**Specificity**	**CI**
KNN	0,700	(0,67 – 0,72)	0,650	(0,61 – 0,68)	0,690	(0,65 – 0,72)
ANN	0,828	(0,80 – 0,84)	0,792	(0,76 – 0,82)	0,816	(0,78 – 0,84)
RF	0,796	(0,77 – 0,81)	0,749	(0,72 – 0,77)	0,788	(0,75 – 0,81)
SVM-linear	0,831	(0,80 – 0,85)	0,812	(0,77 – 0,84)	0,803	(0,77 – 0,83)
SVM-radial	0,783	(0,75 – 0,81)	0,762	(0,73 – 0,79)	0,753	(0,72 – 0,78)
SVM-poly	0,827	(0,80 – 0,84)	0,790	(0,76 – 0,81)	0,806	(0,77 – 0,83)

El resultado más novedoso de este estudio fue la posibilidad de diferenciar imágenes con infarto agudo de crónico en imágenes convencionales de cine CMR, las cuales se obtienen sin contraste. Para tal efecto se utilizó como región de interés para extraer los parámetros texturales el miocardio delineado en su totalidad, ya que en esta modalidad el infarto es imperceptible. La metodología empleada no es directamente aplicable en la práctica clínica debido a que en estos casos la importancia radica en diferenciar infartos agudos de crónicos cuando ambas entidades coexisten, pero los resultados son motivadores ya que pueden ser usados como hipótesis para futuras aplicaciones que no necesiten la administración del contraste gadolinio. De hecho, estos resultados motivaron el siguiente estudio experimental.

3. Detección de segmentos miocárdicos infartados en cine CMR

Este estudio experimental estuvo motivado por los resultados del estudio previo en donde se encontró que es posible diferenciar infartos agudos de crónicos usando solamente imágenes cine CMR sin la necesidad de inyectar contraste. En el presente estudio se emplearon imágenes cine CMR para clasificar segmentos infartados en pacientes crónicos. El grupo de estudio consistió de 50 casos (edad media, 61; rango, 23 – 80 años de edad). Fueron adquiridas imágenes cine CMR y LGE CMR con un equipo de resonancia magnética de 1,5 T (Sonata Magnetom, Siemens, Erlangen, Alemania).

El miocardio de cada imagen fue manualmente delineado corte por corte y luego se utilizó la división por segmentos según la recomendación de la American Heart Association (AHA). Las imágenes LGE CMR fueron tomadas como referencias para identificar y etiquetar los segmentos según el porcentaje de infarto presente:

- Segmentos no-viables: aquellos que presentan infarto transmural ≥ 50%.
- Segmentos viables: aquellos que presentan infarto transmural 0 < LGE < 50%.
- Segmentos remotos: aquellos que no muestran porcentaje de masa infartada.

Los parámetros texturales fueron calculados para cada segmento, pero solamente utilizando las imágenes cine CMR, en las cuales no se puede identificar directamente si existe o no infarto. A diferencia de los dos estudios anteriores, en este hemos decidido utilizar Matlab 2015b (MathWorks Inc., Natick, MA) para el cálculo de parámetros y los mismos fueron analizados en 2D y 2D + *t*. Se calcularon 75

parámetros en 2D y 87 en 2D + *t* utilizando los siguientes métodos: matriz de co-ocurrencia, matriz run-length, matriz size-zone, matriz de diferencia de tonos de grises, y *local binary patterns* (LBP).

Análisis 2D

Dado que las imágenes cine CMR corresponden a una secuencia temporal que abarca un ciclo cardíaco, para el análisis 2D se seleccionó el instante temporal correspondiente al fin de diástole. Por lo tanto, cada corte en este instante temporal fue usado para extraer parámetros texturales de cada segmento miocárdico previamente identificado.

Análisis 2D + *t*

Aprovechando la naturaleza temporal de las secuencias cine CMR, el análisis 2D + *t* consistió en tomar cada ROI como un volumen en donde la tercera dimensión corresponde al tiempo. De esta manera fue posible calcular los parámetros de texturas en su variante 3D teniendo en cuenta la heterogeneidad en las tres dimensiones. Lo ideal hubiera sido incluir la dimensión *z* de cada volumen cardíaco, pero esto no supuso resultados alentadores con nuestras imágenes ya que la distancia entre cortes es muy grande (del orden de los 7 mm) por lo que no es posible extraer mucha información en esa dimensión.

Para la clasificación de segmentos se utilizaron los siguientes subconjuntos de datos, independientemente para 2D y para 2D + *t*:

- Parámetros texturales basados en matrices
- Parámetros LBP
- Parámetros basados en matrices + LBP
- Parámetros seleccionados por el algoritmo SVM-RFE del total de parámetros texturales.

Cada subconjunto de datos fue utilizado para entrenar un SVM con kernel radial. Para la fase de entrenamiento se emplearon 30 de los 50 pacientes disponibles y la evaluación previa se hizo mediante validación cruzada del tipo anidado. Dado que este problema de clasificación consiste de tres clases, el entrenamiento se hizo con la técnica uno contra todos, de manera que la evaluación final se hizo tomando el AUC

promediado de cada clasificador: no-viable contra viable, no-viable contra remoto, y viable contra remoto.

Notablemente para todos los subconjutos de datos, los AUCs obtenidos para el análisis 2D + *t* fueron estadísticamente superiores ($p < 0,01$) que los obtenidos en el análisis 2D. Estos resultados indican la importancia de captar la heterogeneidad en varias dimensionas ya que es posible obtener mayor información que repercute en los resultados. El mejor modelo se obtuvo con los parámetros LBP en 2D + *t* (AUC = 0.849 usando los 20 pacientes de validación). Los valores de AUC específicos para cada clase fueron de 0,935; 0,819 y 0,794 para segmentos no-viables, viables y remotos respectivamente.

La formación de infarto crónico está asociada con disminución del espesor del miocardio. Por este motivo, el engrosamiento de miocardio también fue utilizado individualmente para entrenar el modelo clasificador. Este parámetro sirve como punto de referencia para la comparación con los obtenidos con los parámetros texturales. Con el engrosamiento de miocardio se consiguió un AUC de 0,561, valor muy por debajo de los obtenidos con texturas. Los principales resultados se muestran en la tabla 4.

Tabla 4. *Clasificación en datos de validación.*

	Subconjunto	**N parámetros**	**AUC**
	Engrosamiento pared	1	0,561
2D	Matriz	39	0,753
	LBP	29	0,727
	Matriz + LBP	68	0,737
	SVM-RFE	67	0,755
	Engrosamiento + SVM-RFE	68	0,756
2D + t	Matriz	39	0,811
	LBP	48	0,849
	Matriz + LBP	87	0,822
	SVM-RFE	57	0,807
	Engrosamiento + SVM-RFE	58	0,820

En los últimos años, la resonancia magnética se ha convertido en la técnica no invasiva de referencia para la evaluación de las consecuencias estructurales del infarto de miocardio. Las secuencias cine CMR son empleadas en la actualidad para cuantificar parámetros de relevancia en pacientes post-infarto como la fracción de eyección o volúmenes ventriculares, en ese sentido las imágenes LGE CMR surgieron como la única técnica empleada para calcular la extensión de miocardio infartado. Esta variable ha demostrado ser decisiva para la predicción del remodelado del ventrículo izquierdo, recuperación sistólica tardía y evolución del paciente. Sin embargo, el uso de LGE CMR implica que los estudios sean más prolongados además de la administración del contraste gadolinio. Esto trae como consecuencia ciertas limitaciones para un grupo selecto de pacientes y para los laboratorios de CMR. i) un número significativo de pacientes post-infarto son clínicamente inestables al momento del estudio, y en consecuencia, no pueden tolerar estudios prolongados. ii) el uso de gadolinio puede inducir ciertos efectos, especialmente a aquéllos con un cierto grado de insuficiencia renal. iii) el número de estudios por turnos en los laboratorios debe ser reducido debido a la prolongación de estudios en donde se requiere la adquisición de secuencias LGE CMR.

Si bien los resultados obtenidos son alentadores ya que indican que el análisis de texturas puede ser utilizado para diferenciar segmentos infartados de los no infartados en una modalidad de imagen en la que el infarto no es visualmente perceptible, esto todavía no tiene una aplicación clínica directa, pero sirve de hipótesis para futuros trabajos. La aplicación clínica de utilidad consistiría en poder detectar las zonas de infarto, además de la masa infartada, utilizando solamente las imágenes cine CMR. Para ello habría que hacer un análisis micro-textural para poder identificar las zonas a nivel de píxel.

Conclusiones

La presente tesis doctoral proporciona tres conclusiones principales derivadas de los tres estudios experimentales:

1. Los parámetros de texturas pueden ser utilizados en combinación con un clasificador SVM para la segmentación del miocardio infartado en imágenes LGE CMR. Se comprobó la transferibilidad de esta aplicación a imágenes adquiridas con un equipo de resonancia magnética diferente a aquel utilizada para los datos

de entrenamiento. Sin embargo, son necesarios estudios multicentros de mayor volumen para la correcta validación de la técnica. La naturaleza preliminar de este estudio motivó la realización de los siguientes estudios presentados.

2. La diferenciación entre infarto agudo e infarto crónico es posible utilizando parámetros texturales y técnicas de aprendizaje máquina tanto en LGE CMR y en imágenes convencionales pre-contraste de cine CMR. La elección del clasificador resultó no ser relevante siempre y cuando se incluya un método de selección de características dentro del proceso. El método cuantitativo propuesto es una alternativa a la técnica más utilizada en la actualidad que implica la valoración visual de la presencia o no de edema en una modalidad de resonancia magnética específica para tal efecto.
3. Diferencias implícitas entre segmentos no-viables, viables y remotos están presentes en imágenes cine CMR y pueden ser detectadas mediante la aplicación del análisis de texturas. Estos resultados sirven como punto de partida para el desarrollo de diversas aplicaciones, incluyendo la detección del miocardio infartado en cine CMR, aplicación que tendría un gran impacto si permite la cuantificación de la extensión del infarto en las secuencias cine CMR sin la necesidad de la administración de gadolinio para la adquisición de imágenes LGE CMR.

Los tres estudios experimentales propuestos fueron ejecutados con éxito obteniendo resultados prometedores demostrando la utilidad del análisis de texturas para la evaluación de pacientes con infarto de miocardio. De esta manera se puede decir que el análisis de texturas es una potencial herramienta cuantitativa en comparación con métodos existentes. La contribución principal de esta tesis fue la posibilidad de clasificar segmentos infartados usando solamente imágenes cine CMR. Estos resultados abren una línea de investigación que apunta a delinear el miocardio infartado sin la necesidad de administrar el contraste gadolinio. Como nota final se puede decir que el análisis de texturas permite resaltar detalles “invisibles” presentes en las imágenes.

Abstract

Cardiovascular diseases constitute the leading global cause of morbidity and mortality. Magnetic resonance imaging (MRI) has become the gold standard technique for the assessment of patients with myocardial infarction. However, limitations still exist thus new alternatives are open to investigation. Texture analysis is a technique that aims to quantify the texture of the images that are not always perceptible by the human eye. It has been successfully applied in medical imaging but applications to cardiac MRI (CMR) are still scarce. Therefore, the purpose of this thesis was to apply texture analysis in conventional CMR images for the assessment of patients with myocardial infarction, as an alternative to current methods.

Three applications of texture analysis and machine learning techniques were studied:

i) Detection of infarcted myocardium in late gadolinium enhancement (LGE) CMR. Segmentation of the infarcted myocardium is routinely performed using image intensity thresholds. The inclusion of texture features to aid the segmentation was analyzed obtaining overall good results. The method was developed using 10 LGE CMR datasets and tested on a separate dataset comprising 5 cases that were acquired with a completely different scanner than that used for training. Therefore, this preliminary study showed the transferability of texture analysis which is important for clinical applicability.

ii) Differentiation of acute and chronic myocardial infarction using LGE CMR and standard pre-contrast cine CMR. In this study, two different feature selection techniques and six different machine learning classifiers were studied and compared. The best classification was achieved using a polynomial SVM obtaining an overall AUC of 0.87 ± 0.06 in LGE CMR. Interestingly, results on cine CMR in which infarctions are visually imperceptible in most cases were also good (AUC = 0.83 ± 0.08).

iii) Detection of infarcted non-viable segments in cine CMR. This study was motivated by the findings of the previous one. It demonstrated that texture analysis can be used to distinguish non-viable, viable and remote segments using standard pre-contrast cine CMR solely. This was the most relevant contribution of this thesis as it can be used as hypothesis for future work aiming to accurately delineate the infarcted myocardium as a gadolinium-free alternative that will

have potential advantages.

The three proposed applications were successfully performed obtaining promising results. In conclusion, texture analysis can be successfully applied to conventional CMR images and provides a potential quantitative alternative to existing methods.

Chapter 1

Introduction

1.1 Motivation

Ischemic heart disease is the most frequent etiology of cardiovascular diseases and constitute the leading global cause of morbidity and mortality. *Myocardial infarction* following a coronary occlusion is the primary cause of ischemic cardiomyopathy. Each year, more than 1.9 million people are affected by cardiovascular diseases in the European Union [1]. Myocardial infarction is the irreversible death (necrosis) of heart muscle secondary to prolonged lack of oxygen supply (ischemia) [2]. Early detection and accurate monitoring are essential to guide optimal patient treatment and to assess the patient's prognosis.

Magnetic resonance imaging (MRI) has become a powerful diagnostic tool by providing high quality images thanks to new advances in technology. MRI offers excellent anatomic details due to its high soft-tissue contrast and the possibility to enhance different types of tissues using different imaging techniques. Cardiovascular magnetic resonance (CMR), which entered the arena of noninvasive cardiovascular imaging over the past two decades, became a very important imaging modality, mainly due to its unique versatility [3]. CMR constitutes the gold standard technique for evaluation of myocardial infarction and several CMR image sequences have been developed in order to depict patterns that are relevant for diagnosis [4]. Nevertheless, some cases remain challenging due to technical limitations and the restricted ability of the human eye to detect intrinsic, heterogeneous characteristics of certain tissues or lesions. For example, the well-established technique to assess the extent of myocardial infarction is the so-called late gadolinium enhancement (LGE) CMR. This image modality requires the administration of gadolinium as contrast agent to enhance the infarcted region, but there are several contraindications to gadolinium that makes it inappropriate for certain group of patients, i.e. those with renal dysfunction. Therefore, alternatives to enhance the infarcted region without the use of gadolinium

are open to investigation [5].

In the past years, *texture analysis* has become the focus of interest in the field of medical imaging. Image texture can be defined as the spatial variation of pixel intensities that provides the visual appearance of coarseness, randomness, smoothness, etc. Figure 1.1 shows examples of visually perceptible textures that can be found in nature. However, in medical imaging the visual appearance of texture is usually imperceptible. It has been shown that different image areas exhibit different textural patterns that are beyond the limits of human visual perception [6]. Texture analysis describes a wide range of techniques for quantification of gray-level patterns and pixel inter-relationships within an image providing a measure of heterogeneity [7].

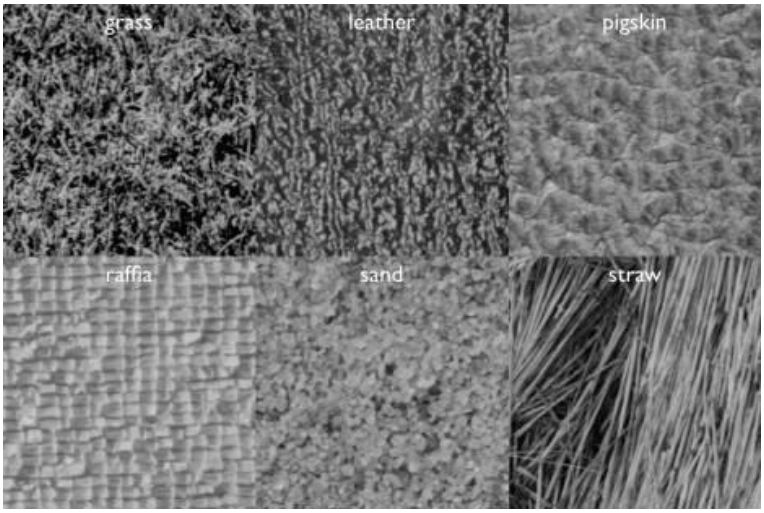

Figure 1.1. *Examples of visually perceptible textures found in nature* [8].

The term *Radiomics* has been recently introduced after successful applications of texture analysis methods in medical imaging problems [9]. The idea behind Radiomics is that one can extract an infinitive number of features or characteristics from the image, such as shape, color, texture, etc. The vast variety of features that can be obtained are not easy to understand but we can take advantage of *machine learning* techniques in order to exploit all the available information. Computer aided diagnostic tools assist the radiologist in the diagnosis by providing quantitative

measures of morphology, function, and other biomarkers in different tissues. Combining texture analysis and machine learning techniques can provide a powerful computer aided diagnostic tool [10].

Applications of texture analysis in medical imaging include classification and segmentation of tissues and lesions. Some applications include discrimination between different brain tumors [P01][11], [12], classification of diseases like Alzheimer's [13] or Friedreich ataxia [14], and brain segmentation [15], [16]. Applications in liver, breast, and prostate are also found in the literature [17]–[19]. To the best of the author's knowledge, studies reporting the application of texture analysis in CMR are very limited. The main studies found in the literature are cited in the following section.

The reported success of texture analysis in several applications, added to the need to overcome technical limitations and contraindications of CMR by providing alternatives that would permit faster and cost-effective procedures, motivated the present thesis.

1.2 Related Work

This section presents a review of the studies regarding texture analysis in CMR that are mostly related to this thesis.

Texture analysis was used by Kotu et al. [20] for segmentation of scarred myocardium in LGE CMR. Their method included dictionary learning and sparse representation of textures in combination with a maximum likelihood Bayes classifier. They showed that texture analysis aided with intensity values provides segmentation of scar with sensitivity and specificity values that are comparable to manual segmentations performed by expert cardiologists. Chapter 5 of this thesis introduces a similar approach for segmentation of left ventricular infarction in LGE CMR. Further work of the same group aimed to identify patients with myocardial infarction with high and low risk of arrhythmias that could benefit from implantable cardioverter defibrillator (ICD). They obtained good classification results using gray-level co-occurrence matrix [21] and local binary pattern (LBP) features [22] extracted from LGE CMR. In a subsequent work they developed an interesting method to obtain probability maps of the scarred myocardium by using dictionary-based texture and intensity features [23]. The probability maps aid the visual inspection of the

myocardial tissue giving information of diagnostic importance, like core and border zone in the scarred myocardium.

An implementation of 3D texture analysis using high resolution (7T scanner) LGE CMR images was presented by [24]. They computed feature maps of the GLCM *contrast* feature aiming to detect diffuse myocardial fibrosis in aging rats. They found a significant increase of myocardial fibrosis in the aged compared to the young rats, and fibrosis detected in texture feature maps correlated with histology. The discrimination between the elderly and young rats was also improved using texture maps in comparison to LGE CMR solely. A similar study from the same group found significant differences between acute and chronic myocardial infarction using three GLCM texture features in 3D [25]. Chapter 6 of this thesis delves into a similar idea in order to differentiate both infarction stages.

Thornhill et al. [26] compared fibrotic and non-fibrotic segments of patients with hypertrophic cardiomyopathy (HCM) and healthy controls. Two texture features derived from the gray-level run-length matrix (GLRLM) were extracted from cardiac segments in LGE CMR. They found that both features were greater in patients with HCM in comparison to healthy controls, even in non-hypertrophic segments. Significant statistical differences were also found between non-hypertrophic, non-fibrotic segments of HCM patients and controls.

A recent study [27] used texture analysis in standard pre-contrast cine CMR to study different etiologies of left ventricular hypertrophy: HCM, amyloid and aortic stenosis. They implemented a scale filtration to extract features corresponding to fine, medium and coarse textures, and six histogram features were used to quantify the pre-filtered images. Statistical differences were found predominantly at the fine and medium texture scales. The novelty of this study was the application of texture analysis in conventional cine CMR as the visual appearance of the diseases cases are very similar in this image modality. This latter study follows our motivation for the experimental study described in Chapter 7, in which cine CMR images were used to detect infarcted segments.

1.3 Objectives

The main objective of this thesis was to study the application of texture analysis in conventional CMR images, as an alternative to current methods for the assessment of myocardial infarction.

Three specific objectives were defined as follows:

- To explore the capability of texture features to distinguish infarcted from healthy myocardium in LGE CMR, and to test a segmentation method in a preliminary multicenter evaluation.
- To investigate the capability of texture analysis of LGE CMR to differentiate acute from chronic myocardial infarction, and to study the possibility to address this problem using standard cine CMR solely in which infarctions are usually imperceptible.
- To detect infarcted non-viable segments using cine CMR solely and texture analysis, as a potential gadolinium-free alternative.

1.4 Contributions to Knowledge

This thesis offers three novel contributions for the assessment of patients with myocardial infarction by the application of texture analysis and machine learning in conventional CMR images.

The first contribution is that texture analysis and intensity features are useful for detection of the infarcted myocardium in LGE CMR, and this can be used as an alternative to existing intensity threshold methods for segmentation of the infarcted myocardium. The segmentation method performed well in a small group of images acquired with a completely different MRI scanner, which indicates the transferability of texture analysis for this application.

The second contribution is that texture analysis can be used to distinguish between acute and chronic myocardial infarctions in LGE CMR, in which the infarction appears hyperenhanced but the infarction's age is not visually distinguishable. The quantitative nature of this approach is an advantage over existing methods that relies on the visual assessment of different image sequences. This problem was also assessed using cine CMR solely, in which the infarctions are imperceptible in most cases, obtaining promising results that indicate that texture

analysis may be useful to enhance the infarctions areas in conventional pre-contrast cine CMR.

The latter motivated the hypothesis for the last contribution of this thesis, which purpose was to detect infarcted non-viable segments using cine CMR solely. This contribution opens a promising area of research aiming to enhance the infarcted regions in conventional cine CMR as a gadolinium-free alternative.

1.5 Thesis Structure

This thesis is structured in 8 chapters that are self-contained and can be read independently. Chapters 2 to 4 present the theoretical background that is directly relevant for the understanding of the experimental studies. Chapters 5 to 7 are the experimental studies performed and represent the main contributions. Chapter 8 is an overall discussion of the experimental results.

A summary of the remaining chapters of the thesis is introduced below:

Chapter 2: Background on cardiac MRI

This chapter gives a background on the principles of cardiac MRI with a focus on the techniques used to assess patients with myocardial infarction. The chapter begins with an introduction of the physiological principles of the heart and myocardial infarction, followed by an introduction of the general principles of MRI physics.

Chapter 3: Texture analysis in MRI

This chapter explains the process to undertake in order to perform texture analysis of MR images. The texture outcome can be considerably affected depending on the methodology used throughout the process. It is presented as a review of previous studies and the considerations to take in order to successfully apply texture analysis in MRI. The texture analysis methods that were used in the experimental studies are also described.

Chapter 4: Overview on machine learning

Feature selection and classification are the machine learning techniques mainly involved in texture analysis applications. This chapter introduces the methods used in the experimental studies. It also provides a summary regarding the evaluation of

model performance.

Chapter 5: Segmentation of infarcted myocardium in LGE CMR. A preliminary multicenter evaluation

This is the first experimental study that aimed to detect the infarcted myocardium in LGE CMR using texture analysis and histogram intensity features. A method for segmentation of the infarcted myocardium was developed and tested on a small sample data acquired with a completely different scanner than that used for training.

Chapter 6: Differentiation between acute and chronic myocardial infarction

This study aimed to distinguish acute and chronic myocardial infarction using quantitative texture analysis of LGE CMR and also using conventional pre-contrast cine CMR solely. The performance of six machine learning classification algorithms and two feature selection techniques were analyzed and compared.

Chapter 7: Detection of infarcted myocardial segments in cine CMR

This study aimed to classify infarcted non-viable segments using cine CMR solely as a gadolinium-free alternative to LGE CMR. The performance of 2D and 2D + t texture features were studied. 2D + t features include the temporal dimension of cine sequences in the analysis.

Chapter 8 presents an overall conclusion.

Chapter 2

Background on Cardiac MRI

2.1 Introduction

Cardiac MRI, also referred as CMR, is the well-established technique for noninvasive characterization of cardiac anatomy and function. It has been widely used to accurately measure left and right ventricular volumes, ventricular wall thickness, mass, and to characterize myocardial viability [4]. This chapter begins with a basic description of the heart anatomy and function, followed by the MRI principles, and ends with an introduction to the CMR techniques mostly employed for assessment of myocardial infarction.

2.2 Heart Anatomy and Function

2.2.1 Basic Principles

The heart is the main organ of the cardiovascular system and supplies blood rich on oxygen and other nutrients to the body tissues. The heart wall is comprised of three layers: epicardium, myocardium and endocardium. The *myocardium* is the middle layer of muscular tissue composed of contracting cardiac fibers that allows contraction of the heart. The myocardium is surrounded by the *epicardium* (outer layer of the heart wall) and the *endocardium* (inner layer) [28].

The heart is composed of four cavities: right atrium, right ventricle, left atrium and left ventricle. *Atriums* are the receiving chambers; they receive blood back to the heart. *Ventricles* are the discharging chambers, and represent the actual pump of the heart. Thus, the ventricular walls are more massive than the atrial walls. Moreover, the myocardium is thickest in the left ventricle as it has to create a lot of pressure to pump blood throughout the body. The heart wall that separates the left from the right cavities is called *septum.* Atrium and ventricles are separated by the atrioventricular valves: the *tricuspid* valve separates the right atrium from the right ventricle, whereas the *mitral* valve separates the left atrium from the left ventricle. These valves are responsible of controlling the blood flow through the chambers in only one direction:

from atriums to ventricles. The *papillary muscles* are projected into the ventricular cavities and play a role in controlling the atrioventricular valves (Figure 2.1).

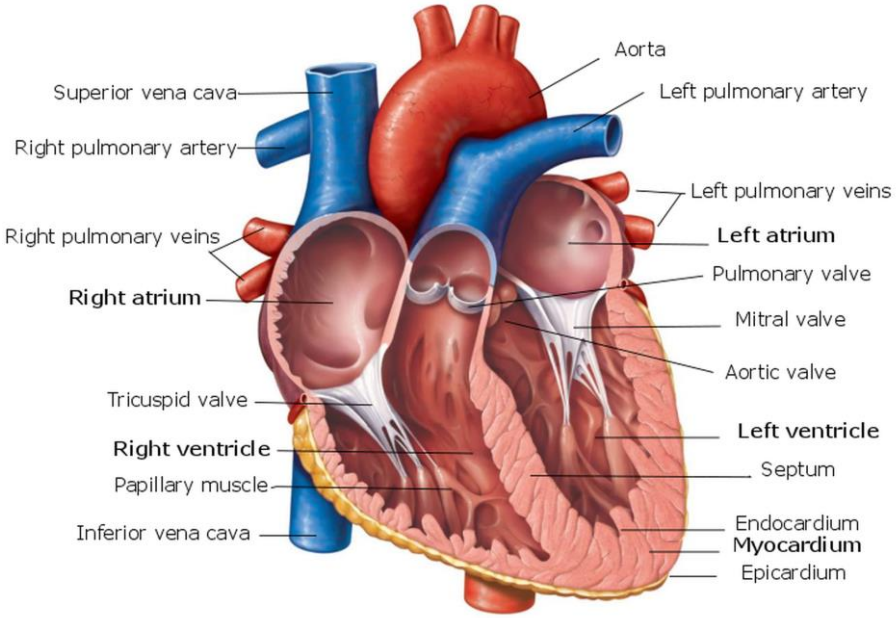

Figure 2.1. *Anatomy of the heart. Frontal section showing the four cardiac chambers: atriums and ventricles (modified from* [28]*).*

The cardiac cycle includes all events associated with the blood flow through the heart during one heartbeat. The right atrium collects blood from the body through the cava veins. This blood then flows into the right ventricle, which pumps deoxygenated blood through the pulmonary artery to the lungs where gas exchange occurs. The pulmonary veins transport blood from the lungs back to the heart into the left atrium. Blood flows then to the left ventricle, which pumps blood into the aorta and from there blood is supplied to the rest of the body. The semilunar valves, aortic and pulmonary, are responsible of controlling the blood flow from ventricles to arteries. In the cardiac cycle, the term *systole* refers to the moments of heart contraction, and *diastole* refers to the moments of relaxation [28].

2.2.2 Myocardial Infarction

The cardiac muscle, or myocardium, also needs nutritive blood supply in order to function normally. The *coronary arteries* are responsible of providing blood to the myocardium. When the heart muscle is deprived of oxygen and fail to function normally it suffers *ischemia.* Ischemic cardiomyopathy is a disease caused by inadequate blood supply to the myocardium. The coronary arteries are tiny blood vessels and may easily become obstructed by, for example, an arteriosclerotic plaque. Atherosclerotic plaques are deposits of fats, calcium, or other substances inside the blood vessels that either block or partially block the blood flow. Immediately after an acute coronary occlusion, the area of myocardium that receives little or zero blood flow cannot maintain its normal function, and is said to be *infarcted*. The overall process is called *myocardial infarction* (Figure 2.2) [29].

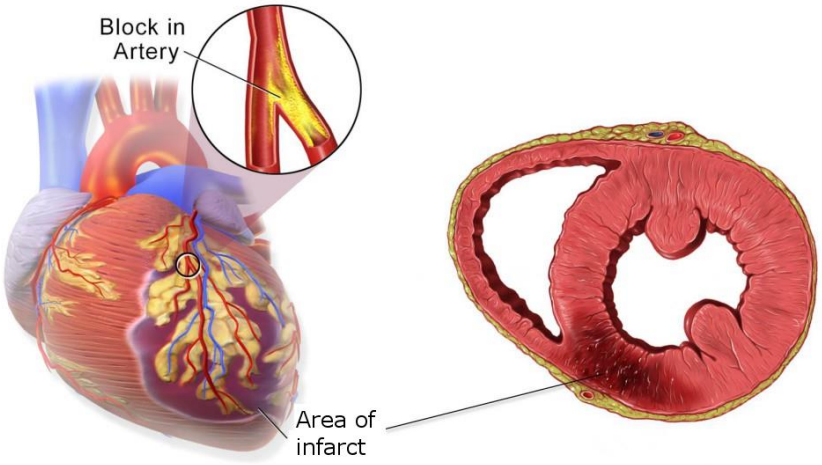

***Figure 2.2**. Myocardial infarction due to an obstruction in a coronary artery* [30].

Myocardial infarction is the most developed manifestation of ischemic heart disease. It mainly affects the left ventricle causing a decrease in its contractibility as a result of necrosis, or cell death, due to the prolonged ischemia. The presence of ST-segment elevation in the electrocardiogram (ECG) denotes total occlusion of a coronary artery. This type of myocardial infarction, known as ST-segment elevation myocardial infarction (STEMI), is the most severe type as virtually all the heart muscle

being supplied by the affected artery becomes necrotic [31].

Shortly after acute myocardial infarction (AMI), the muscle fibers in the center of the ischemic area suffers necrosis. The affected areas undergo a process of change in their composition within the first days and weeks. The process includes resorption and scar tissue formation with the possibility of ventricular remodeling that may deteriorate the systolic function. In AMI, infarct areas exhibit necrosis, considerable myocardial edema, and normally microvascular obstruction (MVO) resulting from capillary obstruction. In the setting of chronic myocardial infarction (CMI), the initially necrotic myocardium is replaced by scar tissue [32].

Viable myocardium is the myocardium that due to ischemia does not contract normally but has the potential to recover [33]. Hence, viability can be defined as the capability of the myocardium to improve its contractile function after revascularization. Reperfusion techniques include primary angioplasty and thrombolytic therapy with fibrinolytic drugs [34]. Myocardial viability is of importance to determine late systolic function after an event of ischemia. Necrosis of the myocardial tissue is inversely related to its viability, and can be detected with techniques like ECG, echocardiography, or positron emission tomography (PET). However, MRI is the well-established technique for a complete assessment of the cardiovascular function after AMI [35]. Different MRI modalities were developed for this purpose and are described is Section 2.4.

2.2.3 Cardiac Imaging Planes

Body planes are oriented orthogonal to the long axis of the body and consist of axial, sagittal, and coronal views. Cardiac scans are usually acquired in planes different to the body planes in order to show the cavities (atriums and ventricles) of the heart. Standard cardiac planes include vertical long-axis (2-chamber) view, horizontal long-axis (4-chamber) view, and short-axis view. These planes are set along a line that extends from the apex to the center of the mitral valve (long-axis of the heart). The short-axis view extends perpendicular to the long axis of the heart at the level of the mid left ventricle. The horizontal long-axis is set by selecting the horizontal plane that is perpendicular to the short-axis, whereas the vertical long-axis is set along a vertical plane orthogonal to the short-axis plane (Figure 2.3) [36].

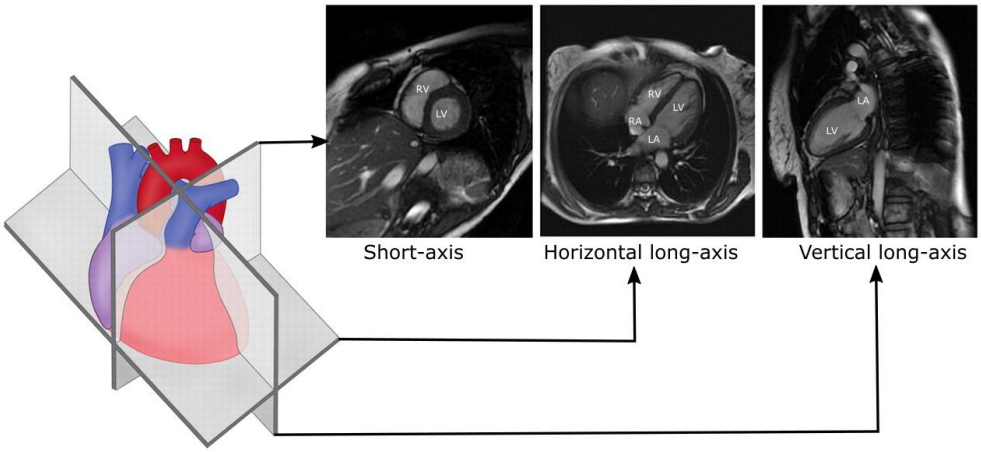

Figure 2.3. *Orientation of the major cardiac planes with respect to the heart and their appearance on MRI. The cardiac chambers are indicated in each view. RV: right ventricle, LV: left ventricle, RA: right atrium, LA: left atrium (modified from* [36]*).*

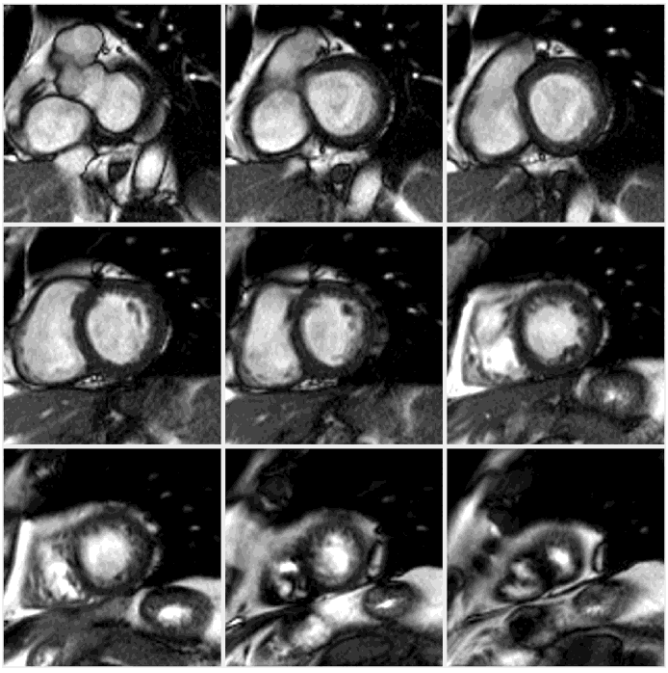

Figure 2.4. *Multi-slice cardiac MRI stack in short-axis view at end-diastole. The top-left image represents the most basal slice.*

The vertical long-axis (2-chamber) view exhibits the left atrium and left ventricle, whereas the horizontal long-axis (4-chamber) view shows the atrium and ventricle of both sides of the heart. The left and right ventricles are better visualized in short-axis views, which are usually acquired in form of a multi-slice stack (Figure 2.4). Hence, short-axis views are routinely used to perform measurements of the left ventricle [37].

2.2.4 Cardiac Segments

Evaluation of the left ventricular function is normally performed by dividing the myocardium into a specific number of segments. The standard method for regional analysis was proposed by the American Heart Association (AHA) and it is known as the 17-segment model [38].

In the AHA model, the heart is sectioned into *basal*, *mid-ventricular*, and *apical* thirds perpendicular to the long-axis. Then, each section is divided into segments at circumferential locations. The basal and mid-ventricular sections are divided into 6 segments of 60° each, while the apical section is divided into 4 segments of 90° each. The last segment of the model is the apex, which is the area of myocardium beyond the end of the left ventricular cavity. Myocardial segments are named and localized with reference to the long axes of the left ventricle and the 360° locations on the short-axis views. Segment locations and nomenclature are illustrated in Figure 2.5.

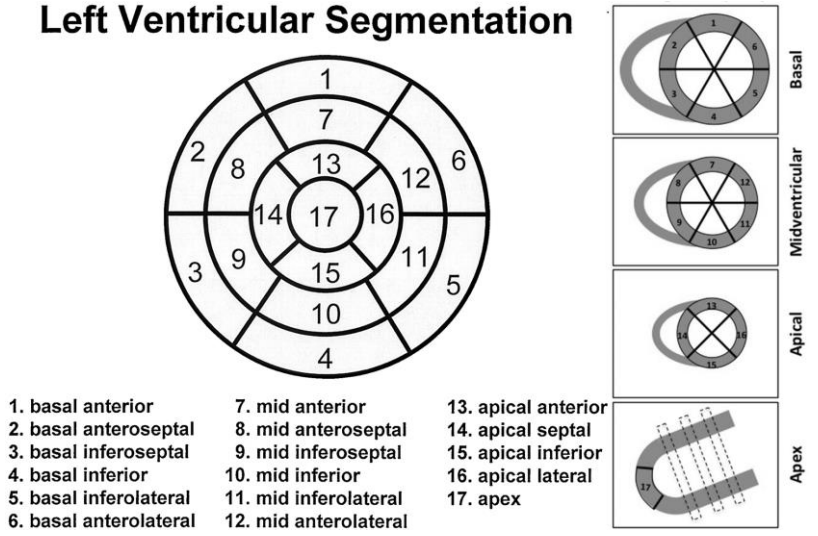

Figure 2.5. *Segment locations and nomenclature of the 17-segment model* [38].

To identify and separate the septum from the left ventricular anterior and inferior walls, the attachment of the right ventricular wall to the left ventricular is used. Representative basal, mid-cavity, and apical slices from the short-axis views should be selected for analysis. Alternatively, various slices can be aggregated to create just three thick short-axis sections. Only slices containing myocardium in all 360° should be selected. The latter is especially important at the basal slices where myocardium and connective tissue forms a complex mixing.

2.2.5 Cardiac Function Parameters

Because the function of the heart is to pump blood to the body, the left ventricle is used to measure the cardiac function according to one cardiac cycle. Ventricular volume is measured at end-diastole, the end of filling when volume is maximum, and at end-systole, the end of contraction when volume is minimum. Cardiac MRI can be used to compute the following parameters:

- **Stroke volume:** Is the volume of blood pumped out by one ventricle with each beat. It correlates with the force of ventricular contraction. Normal resting value for stroke volume is around 70 ml/beat [28]. It is computed by subtracting the ventricular volume at end-systole by the volume at end-diastole in one cardiac cycle.

$$Stroke\ volume\ (ml) = end\ diastolic\ volume - end\ systolic\ volume \quad (2.1)$$

- **Cardiac output:** Is the amount of blood pumped out by each ventricle per minute. It is the product of stroke volume and heart rate. Considering that normal values of heart rate are around 75 beats/min, the average adult cardiac output will be 5250 ml/min [28].

$$Cardiac\ output\ (ml/min) = stroke\ volume \times heart\ rate \quad (2.2)$$

- **Ejection fraction:** Is the fraction of the end-diastolic volume that is ejected in one beat – usually equal to about 60% [29].

$$Ejection\ fraction\ (\%) = \frac{stroke\ volume \times 100}{end\ diastolic\ volume} \quad (2.3)$$

- **Volume indices:** The volume measurements vary with the body size. Therefore, ventricular volumes are expressed as index using the body surface area (BSA): $BSA = \sqrt{((weight \times height)/3600)}$ [39].

$$End\ diastolic\ volume\ index\ (ml/m^2) = \frac{end\ diastolic\ volume}{BSA} \quad (2.4)$$

$$End\ systolic\ volume\ index\ (ml/m^2) = \frac{end\ sytolic\ volume}{BSA} \quad (2.5)$$

- **Wall thickening:** The thickening of the myocardium perceived during wall motion, between end-diastole (WTh_{ED}) and end-systole (WTh_{ES}). In regional analysis, a segment with wall thickening < 2mm is considered affected [40].

$$Wall\ thickening\ (mm) = WTh_{ES} - WTh_{ED} \quad (2.6)$$

- **Left ventricular mass:** The mass of the left ventricle is computed at end-diastole by multiplying the end-diastolic volume index with the myocardial density. The myocardial density is normally assumed to be 1.05 g/ml [41].

2.3 MRI Principles

2.3.1 MRI Physics

Magnetic resonance (MR) is a physical phenomenon in which atom nuclei with uneven number of protons, or neutrons, possess a spin and thereby a magnetic moment. The hydrogen atom (^{1}H) is widely found in the human body due to its abundance in water and fat. Therefore, ^{1}H is the most used nucleus in clinical MRI. The magnetic moments from protons are randomly oriented but in the presence of a strong magnetic field B_0, they will align with the field and start to precess at a specific angular frequency (Figure 2.6).

In the presence of a magnetic field, protons will take two distinct energy levels. This energy difference is the basis for generating a MR signal and causes a magnetization vector M_0 along the z axis in the direction of the magnetic field. The number of spins aligned with the magnetic field is proportional to its strength, thus

higher field strengths yield a stronger MR signal. A spin can jump to the level of high energy if it receives an external energy equal or greater than the difference of both energy levels. This process is called *excitation* and it is achieved with *radiofrequency* (RF) pulses (Figure 2.7).

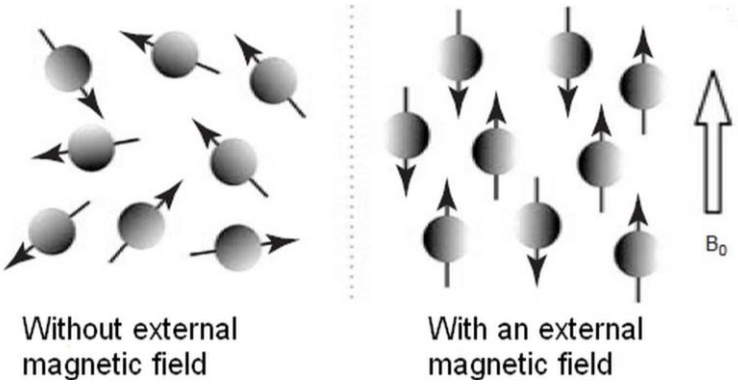

Figure 2.6. *Magnetic moments of hydrogen protons are normally oriented at random but they align at two different orientations (energy levels) in presence of an external magnetic field* B_0 [42].

After receiving the external energy, the spin will try to return to the lower energy level, emitting the previously absorbed energy that is then detected and used to computationally reconstruct images. The latter is the *relaxation* process. There are two types of relaxation: longitudinal relaxation, restoration of longitudinal magnetization M_0 (direction of the main field strength) to its equilibrium value; and transversal relaxation, the net magnetization leaving the transverse plane M_{xy}. Longitudinal relaxation is defined by a time T1, which is the duration of time it takes for the magnetization M_0 to recover to 63% of its equilibrium value. The transversal relaxation is defined by a time T2, which is the time it takes for the transverse magnetization M_{xy} to decrease to 37% of its initial value directly after an RF pulse [43].

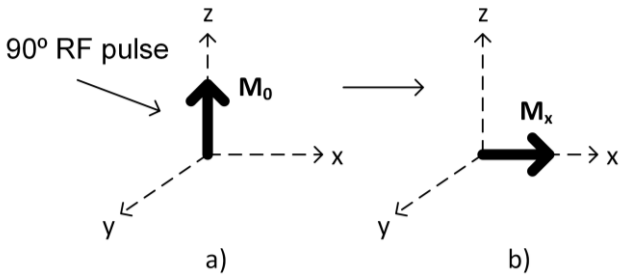

Figure 2.7. *a) The net magnetization vector M_0 in the longitudinal plane. b) After the application of a RF pulse with 90º flip angle, the magnetization vector is tilted to the xy plane.*

2.3.2 The MRI system

An MRI system consists of three main components: a *magnet* that produces a strong, constant magnetic field; radiofrequency *transmit* and *receive coils*, which excite and detect the magnetic resonance signal; and magnetic field *gradients*, which enable the spatial localization the MR signals [43].

The *magnet* is the main part of the MRI system and it generates the magnetic field which is measured in units of Tesla (T). Typical field strengths for cardiac imaging are 1.5T or 3T. Higher magnetic fields produce stronger MR signals thus increasing image quality and contrast. The majority of systems for clinical use are of 1.5T whereas 3T systems are more commonly used in research.

The MR signals are produced within the patient's tissue in response to RF pulses that are generated by a *transmitter coil*. A body coil is usually built into the construction of the magnet. Smaller transmitter coils are necessary for imaging the head or extremities, and special ones are used for cardiac MRI. The MR signals produced in the body are detected using a *receiver coil*. Special shielding is built into the MRI system room to minimize interference, as the MR signals are very weak and very sensitive to electrical interference.

To localize the MR signals in the body to produce images, it is necessary to generate short-term spatial variations in magnetic field strength across the patient. These are commonly referred as *gradients*. Stronger gradients permit smaller anatomical features to be seen in the images. A gradient field is produced by using three sets of gradient coils, one for each direction, through which large electrical currents are applied repeatedly in a controlled *pulse sequence*.

2.3.3 Pulse Sequences

Pulse sequences describe a temporal succession of RF pulses that allows to produce a wide range of image contrast. The difference in relaxation times between different tissues is exploited to generate specific contrast in the images according to the tissues of interest. For example, T1 and T2 vary according to the type of tissue in which the nucleus is located, and are longer in free water than in bound water. Pulse sequence parameters, i.e. timing and amplitude of the RF and/or gradient pulses, can be manipulated to elicit predominantly T1-weighted or T2-weighted image contrast. There are literally hundreds of pulse sequences and every year new ones are launched. Still, pulse sequences can be divided in two main groups: *spin-echo* and *gradient-echo* sequences. These are based on the way the echo of the MR signal is formed: rephasing 180º RF pulses, or rephasing gradients [42], [43].

In a conventional spin-echo sequence (Figure 2.8), the equilibrium magnetization is tilted to the transverse plane after a 90º RF pulse so the spins dephase naturally for a certain time. Then, an inversion pulse of 180º is applied to refocus spin dephasing. After a time equal to the delay between the 90º and 180º pulse, all the spins will come back into phase along the *xy* plane forming the spin echo. The second pulse is transmitted at half the time of the echo acquisition which is referred as the echo time (TE). The process is repeated at a time referred as the repetition time (TR).

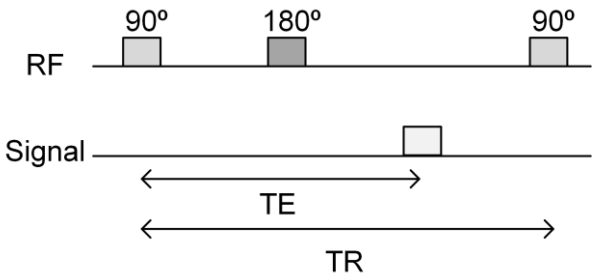

Figure 2.8. *Pulse sequence timings for a spin-echo sequence. The time to echo formation is the echo time (TE), while the time between successive excitations is the repetition time (TR).*

The conventional spin-echo sequence requires acquisition times of the order of minutes since the 90º and 180º pulses take time and the magnetization need to be recovered between each TR. Reduction of acquisition times is an important consideration for clinical applicability. In the gradient-echo sequence, an initial RF pulse with a flip angle < 90º is applied and the measurement is performed while the spins are dephasing. To boost the normal dephasing process, a gradient pulse is sent and then the gradient is reversed to allow the spins to rephase thus generating an echo. Signals can be detected with significantly reduced times in comparison to spin-echo sequences.

2.4 Cardiac MRI Techniques

Different MRI modalities have been developed in order to depict cardiac anatomy and physiology. This section focuses on CMR modalities used to assess myocardial infarction and viability. Pulse sequence techniques receive different names depending on the manufacturer. In the following lines the simplified terms proposed by Friedrich et al. [44] are given.

2.4.1 Cine CMR

Cardiac motion and contractile function are studied using cine imaging. These are short movies that show heart motion through the cardiac cycle. The heart is constantly moving due to respiration and intrinsic cardiac motion. Image acquisition is synchronized with the cardiac cycle using ECG-gating to alleviate the intrinsic movement. Usually this is done retrospectively which involves continuous and simultaneous acquisition of the ECG and MR signal. The lungs are also moving during the respiratory cycle and this motion is usually alleviated with breath holding during imaging. To fully cover the heart, multiple breath-holds are often needed.

The most used sequence for cine imaging is balanced steady state free precession (SSFP) which is a modification of the gradient-echo sequence. The signal in these images utilizes the ratio between T2 and T1. Cine SSFP images provide high contrast between the myocardium (appears dark) and the blood pool (appears bright) without the use of contrast. These sequences are used for the evaluation of wall motion and volumetric measurement, due to their clear delineation between myocardium and blood pool. Quantification is usually performed at end-diastole and end-systole [45].

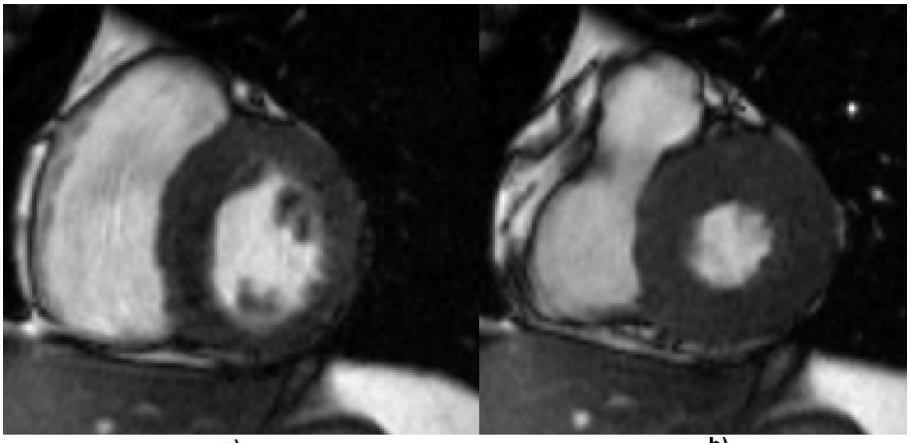

Figure 2.9. *Short-axis views of cine CMR at end-diastole (a) and end-systole (b). The left ventricular myocardium appears dark while the blood pool appears bright.*

2.4.2 Late Gadolinium Enhancement CMR

Delayed enhancement is a post-contrast sequence used to evaluate the extent of myocardial scar or fibrosis. The technique involves T1-weighted inversion recovery (IR) approximately 10 min after intravenous administration of contrast media. The contrast agents used in MRI are mostly paramagnetic based on gadolinium, hence this technique is referred as late gadolinium enhancement (LGE) CMR. The mechanism of LGE relies upon the assumption that the gadolinium distribution volume is small and tissue concentration is low in normal myocardium, whereas cell membrane rupture in acute necrosis or chronic scar tissue allows gadolinium to diffuse into myocytes leading to increased gadolinium concentration, shortened T1, and thus hyperenhancement. Normal myocardium enhances and washes out, while ischemic myocardium enhances late. LGE images are usually acquired during repeated breath holds in long axis views and in a stack of short axis slices covering the left ventricle.

Inversion recovery (IR) sequences are used to null the signal from normal myocardium to accentuate surrounding pathology. IR sequences have a special parameter known as inversion time (TI). An appropriate TI to null normal myocardium must be found at each acquisition as this parameter varies from person to person. With appropriate settings, normal myocardium appears black or nulled, whereas infarcted regions appear bright or hyperenhanced (Figure 2.10). LGE CMR is

considered the reference method for assessing myocardial viability. It depicts irreversibly damaged myocardium that is present both acutely as necrosis, and chronically as fibrosis. The enhancement produced by necrosis or fibrosis is the same, therefore other techniques are necessary to differentiate AMI from CMI [3].

Quantification of the infarct using LGE can be performed using different techniques that include manual delineation of the lesion, thresholding by 2, 3, 4, 5, or 6 SD above remote myocardium, and the full width and half maximum (FWHM) technique. In the latter approach, half of the maximum intensity within a user-selected hyperenhanced region is selected as intensity threshold. Using this threshold, a region-growing algorithm is used to identify the infarct boundaries from user-selected seeds. It has been shown that the FWHM technique for LGE quantification gives quantification results similar to manual delineation and it is statistically the most reproducible. Infarct size is usually assessed by reporting the percentage of the left ventricular mass showing LGE [46], [47].

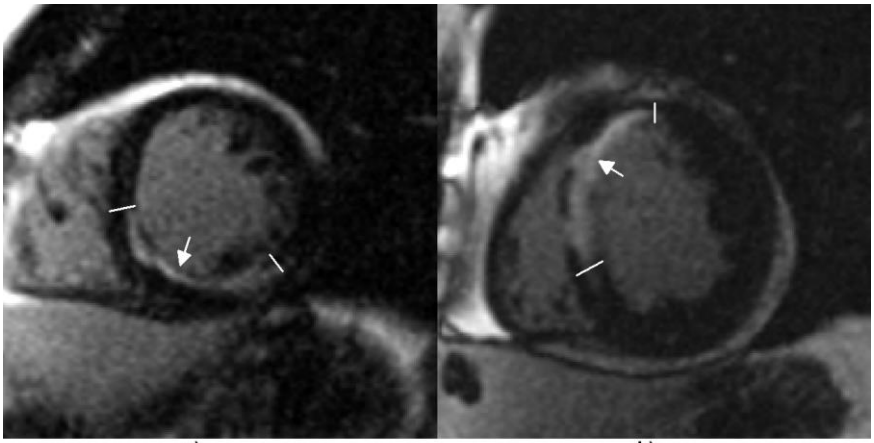

Figure 2.10. *Short-axis views of LGE CMR. The signal from normal myocardium is nulled and appears dark, whereas regions of myocardial infarction (arrowed) appear hyperenhanced: inferior (a) and anteroseptal (b). The limits of infarcted areas are indicated with white lines.*

2.4.3 Edema CMR

Black-blood T2-weighted short time inversion recovery (STIR) imaging is a sequence sensitive to increased myocardial water content, thus allowing the

enhancement of areas of myocardial edema. Increased water content associated with edema appears hyperenhanced on T2-weighted STIR images (Figure 2.11). Myocardial edema is present in the acute stage of myocardial injury and represents the area that is compromised from ischemia. Edema is a transitorily phenomenon that disappears before the sixth month after infarction, thus T2-weighted images can be used to differentiate acute from chronic MI. This technique also allows detection of the salvaged myocardium, which correspond to the myocardium at risk that does not exhibit necrosis [45], [48].

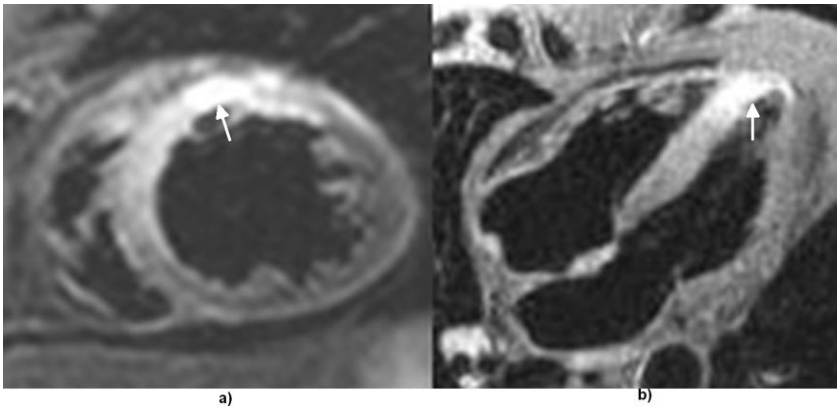

Figure 2.11. *Short-axis (a) and horizontal long-axis (b) views of edema CMR. Edema appears hyperenhanced (arrowed) in relation to normal myocardium.*

2.4.4 Perfusion CMR

Regional myocardial perfusion can be assessed by measuring the inflow of a contrast agent, like gadolinium, during the first passage through the myocardial vasculature. Perfusion CMR consists of a series of ECG-gated T1-weighted images (using gradient-echo or SSFP acquisition) with every one to two heartbeats during the first pass of an intravenously administered bolus of contrast media during one breath-hold. The contrast agent is used to visualize the delay in perfusion in ischemic myocardium compared to healthy myocardium. In the presence of a microvascular disease or ischemia, perfusion will decrease to that part of the myocardium. Regions with low or no perfusion will appear hypoenhanced (Figure 2.12). Apart from visualizing the regional perfusion, some indices can also be calculated to quantitatively assess perfusion of ischemic myocardium [3], [49].

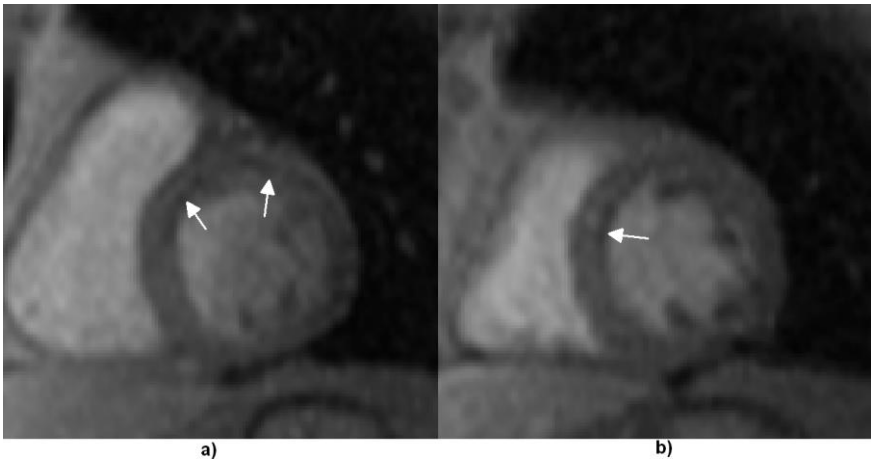

Figure 2.12. *Short-axis views of perfusion CMR. Regions with low or no perfusion will appear hypoenhanced, anterior (a) and anteroseptal (b), in relation to normal myocardium.*

Chapter 3

Texture Analysis in MRI

Part of this chapter has been previously published in:

Larroza A, Bodí V, Moratal D.
Texture Analysis in Magnetic Resonance Imaging: Review and considerations for future applications. In: *Assessment of Cellular and Organ Function and Dysfunction using Direct and Derived MRI Methodologies*, Christakis Constantinides (Editor)
InTech, Rijeka, Croatia, 2016.

3.1 Introduction

Texture analysis applications in MRI involve a general process that consists of six steps: MRI acquisition, region of interest (ROI) definition, ROI preprocessing, texture feature extraction, feature selection, and classification (Figure 3.1). None of these steps is specific, and the methods have to be chosen according to the application. The texture outcome can be considerably affected depending on the methodology used throughout the process. Here we present a condensed description of the possibilities at each step of the texture analysis process. The feature selection and classification steps will be described in detail in chapter 4, as these are part of the machine learning theory.

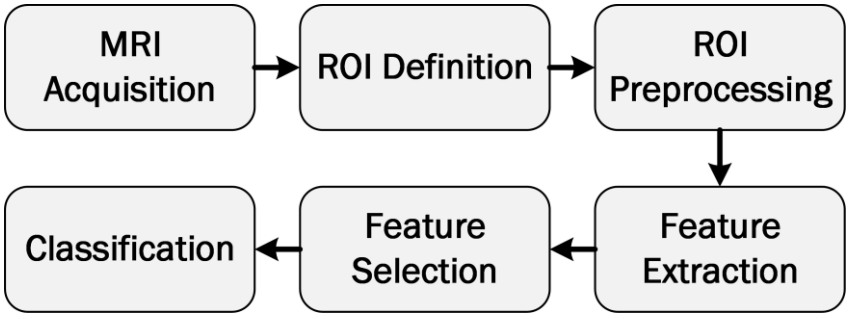

Figure 3.1. *Texture analysis process for MRI classification. ROI: Region of Interest.*

3.2 MRI Acquisition

As explained in Chapter 2, MRI gives us the possibility to enhance specific tissues by varying the acquisition sequence parameters. In this respect, the outcome of texture analysis strongly relies on the image acquisition protocols, and these should be carefully selected in order to obtain the maximum accuracy and reproducibility. Different measuring techniques produce different patterns in texture and these may vary among centers and manufacturers [7]. Texture analysis can be used reliably at one center with a specific imaging protocol but this does not mean that the same methodology can be directly applied to images acquired at different centers with different protocols [50].

3.2.1 Sequences

Repetition time (TR), bandwidth/echo time (BW/TE), and flip angle are the properties that are most likely altered in a clinical setting. Variation of these parameters enhances different tissues thus obtaining different image texture. Repetition time had the biggest impact when comparing different foam phantoms using clinical breast MRI protocols, whereby better texture discrimination was elicited at higher TR [51].

The choice of the MRI sequence for texture analysis depends on the application. Contrast-enhanced T1-weighted images is the current standard MRI protocol used by clinicians to assess brain tumors and was used for texture analysis in [P01], [11], [52]. Some studies compared different modalities obtaining diverse results. In the study of Tiwari et al. [53], contrast-enhanced T1-weighted provided better performance over T2-weighted and fluid-attenuated inversion recovery (FLAIR) images for discrimination of recurrent brain tumors from radiation induced lesions. T1-weighted MRI was also notably better than FLAIR images for dementia classification [54]. T2-weigthed images were more suitable for differentiation between benign and malignant tumors [55], [56], and for discrimination of posterior fossa tumors in children [57]. Texture analysis applied to diffusion-weighted MR images also proved to be efficient for cancer classification [58]–[60]. Nevertheless, texture features used in these studies differ from each other, so a definite assumption of which MRI sequence is better cannot be made. Texture analysis applications in CMR were mainly applied to LGE CMR [20]–[26] .

3.2.2 Influence of Spatial Resolution and SNR

Spatial resolution and signal to noise ratio (SNR) have been reported to be the most influential factors for texture analysis [7], [61], [62]. Image resolution is defined by slice thickness, field of view (FOV) and matrix size. Signal-to-noise ratio is defined as the coefficient between the mean signal over a homogeneous region of a tissue of interest and the standard deviation of the background noise. Texture discrimination improves with higher levels of SNR and it has been reported that a SNR > 4 is necessary to measure the real textural behavior of the human brain [63]. Discrimination based on texture analysis also improves with higher spatial resolution [7]. Texture analysis fails if the image resolution is insufficient since the finest textural details cannot be spotted, but at higher spatial resolution it is more sensitive to variations in the acquisition parameters [62].

Although current routine MRI scanners can produce high resolution images, these are susceptible to motion artifacts given the long scan times and are not widespread in clinical practice. A strong correlation between 3D structural indices and 3D texture features in trabecular bone in osteoporosis was found by [64], using routine images with standard resolution (0.7×0.7 mm^2); indicating that these can be used to quantify the bone architecture without the need of higher resolution images. These previous results show that even if high-resolution images provide better texture discrimination, its application in clinical practice is far complicated as no good reproducibility among centers is expected. Apparently the slice thickness does not influence significantly the outcome of 2D texture analysis according to [65], who found only moderate differences between 1 mm and 3 mm thickness to separate white matter tissue and multiple sclerosis plaques.

3.2.3 Influence of Field Strength

One important difference among MRI scanners is the field strength of the magnet, being the most common values in clinical routine nowadays those of 1.5 T and 3 T. Scanners with higher field strength provide more SNR, thus increasing spatial and temporal resolution. In counterpart, artefacts resulting from breathing or any other type of body motion are more prominent on 3 T than on 1.5 T scanners, but these are generally compensated using some techniques offered by manufacturers [66]. Better texture-based discrimination is expected on the higher quality images acquired on 3 T

scanners as it was reported for liver fibrosis [67] and breast cancer classification [68]. In [56], they found significant differences between 1.5 T and 3 T when squamous cell carcinoma tumors on head and neck were compared. However, their results are in contrast with previous evidence [67], [68] since benign versus malignant tumor discrimination was better on 1.5 T. In the study of Waugh et al. [51], texture discrimination of foam phantoms using different clinical breast MRI protocols was in general improved when a 3 T scanner was used but changes in the imaging parameters at 1.5 T had less influence on the texture outcome.

3.2.4 Multicenter Studies

Few multicenter studies regarding the application of texture analysis in MRI have been published. In [55], they concluded that texture analysis of MRI can discriminate between different brain tissues obtained in routine procedures at three different centers. In [50], they compared the classification performance to discriminate between bone marrow and fat tissue on T1-weighted MRI of knees from 63 patients obtained from three centers with two different field strength MRI scanners: two centers at 1T and one at 3T. Texture information was extracted from two centers and was used to predict tissue on data from the third center, concluding that feature sets from one center may be used for tissue discrimination in data from other centers. In a very large multicenter study, Karimaghaloo et al. [15] analyzed 2380 scans from 247 different centers for segmentation of multiple sclerosis lesions achieving an overall sensitivity of 95% on a separate dataset of 120 scans from 24 centers. The promising results of this study may be the consequence of extracting texture features from different MRI protocols (T1, T2, proton density, and FLAIR) and using them in combination when modeling the classifier. It should be also noted that images were corrected for non-uniformity effects and were normalized into a common spatial and intensity space, thus reducing the possible differences among multicenter scans. Opposite conclusions were reached by [56], as they stated that texture analysis is useful for discrimination of benign and malignant tumors when using one scanner with the same protocol but it is not recommended for multicenter studies. However, they did not mention any image normalization or inhomogeneity correction that could somehow have affected their results, as it is discussed in Section 3.4.

3.3 Region of Interest Definition

Texture features are computed inside a predefined region of interest (ROI), or volume or interest (VOI) in the case of 3D texture analysis, and are usually placed over a homogeneous tissue or lesion area. Manual definition of ROIs is still considered the gold standard in many applications, and it is the chosen option over automatic methods [69]–[72]. Different approaches have been used to define ROIs that are also extended to 3D texture analysis. One approach for ROI definition is the positioning of squares [73] or circles [74] of predefined sizes over the tissue to be analyzed. Using this approach, only information of the underlying tissue is captured but some texture details can be lost because the ROI does not cover the entire area of interest. Another alternative is to use a bounding box defined as the smallest enclosing rectangular area of the tissue of interest [75], [76]. The latter approach has the advantage that the ROI covers the complete tissue or lesion, but it also includes information from adjacent parts that can affect the texture quantification. Although delineation of the entire tissue or lesion can be tedious, it is a better approach since the entire area of interest is included [57], [77]. It has been shown that for both 2D and 3D texture analysis, delineation of the entire lesion provides better accuracy than the bounding box approach [78]. Figure 3.2 shows examples of the three aforementioned ROI definition approaches.

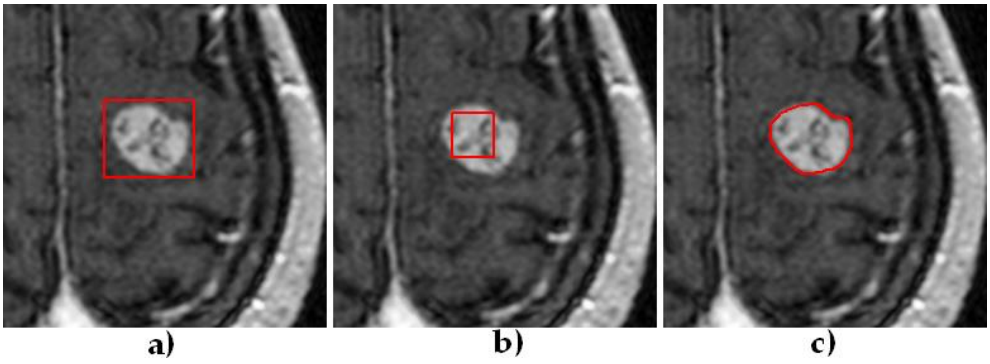

Figure 3.2. *Approaches for defining a region of interest (ROI) over a brain tumor. The use of a bounding box that covers the entire lesion (a), or a small square inside the tumor (b) can be defined quickly and easily, but the delineation of the entire lesion (c) is preferred in order to capture the maximum texture information only within the area of interest.*

3.3.1 Size of the Region of Interest

The size of the ROI should be sufficiently large to capture the texture information thereby eliciting statistical significance [7]. In [79], they studied the effect of ROI size on various texture features extracted from circular ROIs of ten different sizes on brain MRI of healthy adults. They concluded that ROI size has a significant effect on the computed value of several texture features, and that comparisons of different ROI sizes will lead to false results.

A good methodology to avoid possible influences of ROI size might be to use squares and circles of the same size among all samples but as we mentioned before, the complete delineation of the ROI might offer better results. It is recommended the use of the ROI delineation approach when the range of lesion sizes among samples is not significantly broad and when the employed texture features are not affected between this range, otherwise ROIs of fixed size might be a better approach.

3.3.2 Feature Maps

Texture feature maps can be computed by defining ROIs as sliding blocks of $n \times n$ pixels centered at each pixel on the image, so for each pixel a specific texture feature value is computed including its surrounding neighborhood. The block size should be large enough to capture sufficient texture information from each pixel neighborhood but small enough to capture more local characteristics allowing finer detection of regions [7]. Figure 3.3 shows examples of texture maps computed for sliding blocks of different sizes. Texture maps can reveal some characteristics that are not visible on the original image and are mainly used for segmentation tasks [80]. Computing features over texture maps can lead to better results than using the original MR images [81].

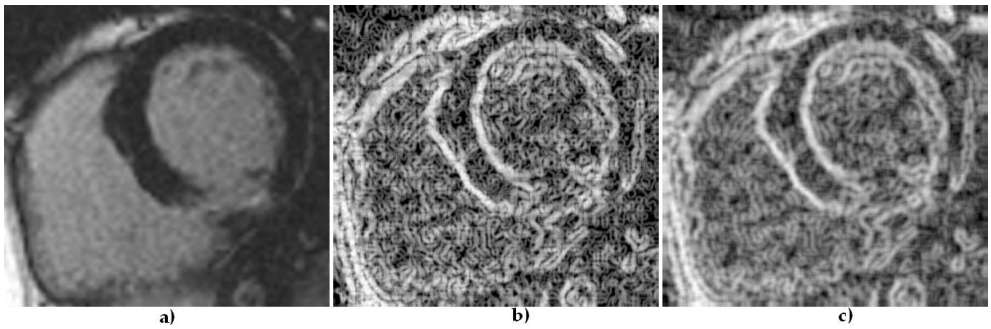

Figure 3.3. *Texture feature maps of a CMR image: a) original image, b) entropy feature map computed with a sliding block with a size of 5 × 5 pixels, and c) entropy feature map computed with a sliding block of 9 × 9 pixels.*

3.4 Region of Interest Preprocessing

It is clear from Section 3.2 that MRI acquisition protocols are relevant for texture analysis. Several preprocessing techniques have been proposed in order to minimize the effects of acquisition protocols and are especially important when dealing with multicenter studies. The main purpose of these preprocessing techniques is to put all ROIs in the same condition so features extracted from them represent essentially the texture being examined. Some preprocessing techniques also aim to improve the texture discrimination.

3.4.1 Interpolation

Image spatial resolution is one of the most influential factors in texture analysis, and it was demonstrated that higher resolutions tends to improve texture-based classification but high-resolution images are not usually available in clinical routine [7], [64]. Image interpolation is an option to enhance images with a low spatial resolution. The effect of image interpolation on texture features was analyzed by [82], who compared three interpolation methods applied to T2-weighted images acquired at five different resolutions. They concluded that MR image interpolation has the potential to improve the results of texture-based classification, recommending a maximum interpolation factor of four. In their study, the most considerable improvements were found when images with original resolution of 0.94 × 0.94 mm^2 and 0.47 × 0.47 mm^2, were interpolated by factors of two or four using the zero-fill

interpolation technique at the *k-space* domain. Image interpolation is of special interest when dealing with 3D texture analysis because in most MRI sequences the slice thickness is larger than the in-plane resolution. Re-slicing all images to obtain isotropic image resolution is required for computing texture features to retain the scales and directions in all three dimensions [83].

3.4.2 Normalization

Even if the same scanner and protocol are used, they may produce MR images with different intensities. Certainly, they may depend on the acquisition conditions such as room temperature and hygrometry, calibration adjustment, slice location, etc. [84]. It was demonstrated that some features are not only dependent on texture, but also on other ROI properties, such as the mean intensity and variance [55]. To avoid the influence of such factors, ROI normalization is a recommended preprocessing step. In [84], they studied the effects of ROI normalization on texture classification of T2-weighted images and demonstrated that classification errors were dependent on the MR acquisition protocols if no normalization was applied. They compared three techniques, and the one that yielded the best results is known as "$\pm 3\sigma$" normalization. In this method, image intensities are normalized between $\mu \pm 3\sigma$, where μ is the mean value of gray-levels inside the ROI, and σ is the standard deviation, so that gray-levels located outside the range $[\mu - 3\sigma, \mu + 3\sigma]$ are not considered for further analysis. The "$\pm 3\sigma$" normalization technique has become the most popular and preferred choice in most publications [85]–[87]. An example of this method is shown in Figure 3.4.

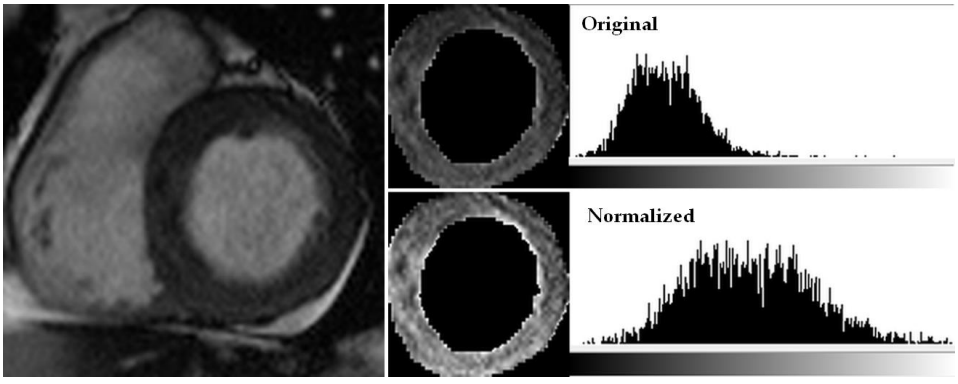

Figure 3.4. *Example of region of interest (ROI) normalization of a CMR image. The extracted ROI is shown in the original histogram and after normalization using the "$\pm 3\sigma$" method.*

3.4.3 Inhomogeneity Correction

There is still another residual effect that is not eliminated by ROI normalization, which are the variations of intensity present in MR images; mainly caused by static magnetic field inhomogeneity and imperfections of the radiofrequency coils [63]. Texture features depend on local average image intensity and are therefore affected by image inhomogeneity. Correction of non-uniformity artifacts in MRI is recommended as a preprocessing step prior to ROI normalization and especially for large ROIs [88]. Figure 3.5 shows an CMR image with inhomogeneity and the corrected image within the myocardium ROI. A review of methods for MRI inhomogeneity correction is available in [89], being the most popular method found in texture literature [90]–[93], the so-called *N3* algorithm [94].

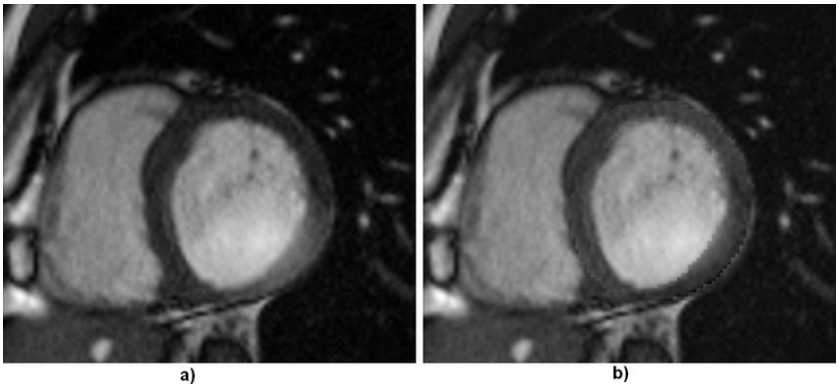

Figure 3.5. *Example of a CMR image with inhomogeneity (a), the average local image intensity of the lower right part is brighter than the upper part. The myocardium region of interest was corrected and appears homogeneous in (b).*

3.4.4 Quantization of Gray-levels

Texture analysis methods based on matrix computation, e.g., co-occurrence and run-length matrices, require quantization of gray-levels. A typical MR image is represented by 10 or 12 bits per pixel, that is, 1024 or 4096 levels of gray. So, in MRI texture analysis, quantization will refer to the reduction of levels of gray used to represent the image. Typical number of gray-levels used for texture feature computation are 16, 32, 64, 128, and 256. Reducing the number of gray-levels improves SNR and the counting statistics inherent in the matrix-based texture analysis methods

at the expense of discriminatory power [95]. Some studies reported that no significant effects were found when different numbers of gray levels were tested [85], [96] while in the study of Chen et al. [78], a gray-level of 32 was reported to be an optimal choice for breast MRI. A specific study regarding the impact of the number of gray-levels on co-occurrence matrix texture features was carried out by [97]. They concluded that the number of gray levels, or dynamic range, has a significant influence on the classification of brain white matter, obtaining an optimal number of 128 levels for both 2D and 3D texture analysis approaches. It is recommended to optimize the number of gray levels for each specific application.

3.5 Texture Feature Extraction

Feature extraction is the main and specific step in the texture analysis process and implies the computation of texture features from predefined ROIs. Many approaches have been proposed in order to quantify the texture of an image allowing the computation of numerous features. In this section, we briefly describe the most popular texture analysis methods that were implemented in the experimental studies of this thesis (chapters 5-7). Extensive reviews of existing feature extraction methods are given in [6], [83], [98], [99].

3.5.1 Histogram

The simplest method for feature extraction is by using the image histogram, which represents the counts of pixels that shares a given intensity value. Gray-scale MR images are normally represented by 12 bits per pixel, representing a total of 4096 gray-levels. Pixels with low gray-level values appear darker the those with higher levels. Histograms for two different ROIs are shown in Figure 3.6.

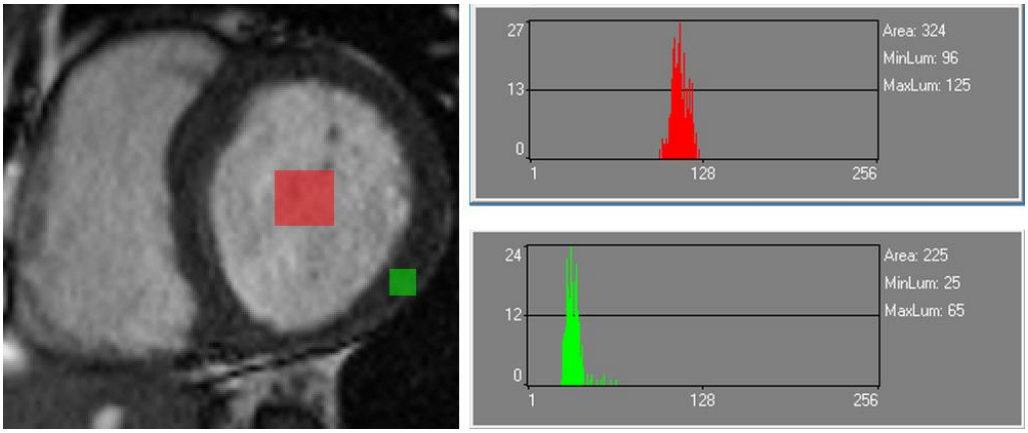

Figure 3.6. *Histograms of two different ROIs inside a CMR image. The blood pool appears brighter (red ROI) than the myocardium (green ROI) in cine CMR and this situation is reflected in the histogram. The image in the example was quantized to 256 gray-levels.*

Histogram-based features do not really describe the actual texture of the image or ROI being analyzed as they do not take into account spatial neighborhood relationships or dependencies among pixels. Nevertheless, they are normally included in combination with other feature extraction methods as they may improve the texture-based classification [100].

Assuming that N_g is the number of distinct gray-levels, and $p(i)$ is the normalized histogram vector (histogram whose entries are divided by the total number of pixels in the ROI), features derived from the histogram are:

- **Mean:** The average gray-level value.

$$\mu = \sum_{i=1}^{N_g} ip(i) \tag{3.1}$$

- **Variance:** Indicates how far the gray-levels are from the mean thus giving and idea of the homogeneity of the pixel distribution.

$$\sigma^2 = \sum_{i=1}^{N_g} (i-\mu)^2 p(i) \tag{3.2}$$

- **Skewness:** Measures the lack of symmetry. The histogram is symmetric when the left and right distributions of the center point are similar.

$$\mu_3 = \sigma^{-3} \sum_{i=1}^{N_g} (i - \mu)^3 p(i) \tag{3.3}$$

- **Kurtosis:** Indicates if the histogram is peaked or flat in relation to the normal distribution.

$$\mu_4 = \sigma^{-4} \sum_{i=1}^{N_g} (i - \mu)^4 p(i) - 3 \tag{3.4}$$

- **Percentiles:** A percentile gives the highest gray-level value under which a given percentage of the ROI pixels are contained. Percentiles are normally calculated at 1, 10, 50, 90, and 99%. For example, if the 90% percentile of a 256 gray-levels ROI is 100, the 90% of the pixels in that ROI has a gray-level value from 0 to 99.

3.5.2 Absolute Gradient

The gradient measures the spatial variation of gray-level values across the image. Therefore, if the gray-level varies abruptly from black to white, a high gradient value is obtained; but if the variation is smooth the gradient value will be low. The gradient may be positive or negative, depending on whether the gray-level variation is from dark to bright or in the opposite way. The absolute gradient does not take the sign into consideration, as the interest is to know if the intensity variation is abrupt or smooth. To obtain the absolute gradient histogram, each pixel $X(i,j)$ is assigned a value that is obtained considering its neighborhood as shown in Equation 3.5. Features computed from the absolute gradient histogram are: mean, variance, skewness, kurtosis, and the percentage of non-zero elements [6].

A	B	C
D	$X(i,j)$	F
G	H	I

$$AbsGrX(i,j) = \sqrt{(H - B)^2 + (F - D)^2} \tag{3.5}$$

3.5.3 Gray-Level Co-occurrence Matrix

The gray-level co-occurrence matrix (GLCM) is a feature extraction method based on second-order statistics. Pairs of pixels separated by a predefined distance and direction are counted and the resulting values are allocated in the GLCM. The count is based on the number of pairs of pixels that have the same distribution of gray-level values [7]. Normally, GLCMs are computed in four directions (horizontal, vertical, 45°, 135°) for 2D, and in 13 directions for 3D approaches [83], using different pixel or voxel separations. Figure 3.7 shows an example for computing a GLCM. The pixel distance has to be chosen according to the application: a larger distance will allow detection of coarse areas but care must be taken not to overstep the size of the ROIs.

One main concern about GLCMs is their dependence on direction, so different values may be obtained if the image is rotated. This situation may be unacceptable for texture characterization on MRI since images from different patients may have different orientations. Rotation invariant features can be achieved by averaging, or simultaneously adding up, the matrix values over all directions [101], [102] or by averaging the statistical features derived from the matrices [103].

Several texture features can be computed from each GLCM, aiming to quantify the homogeneity (smoothness) or heterogeneity (coarseness) of the image. The most popular features computed from GLCMs are those introduced by Haralick et al. [101]. Here we define the features used in this thesis together with their formulas.

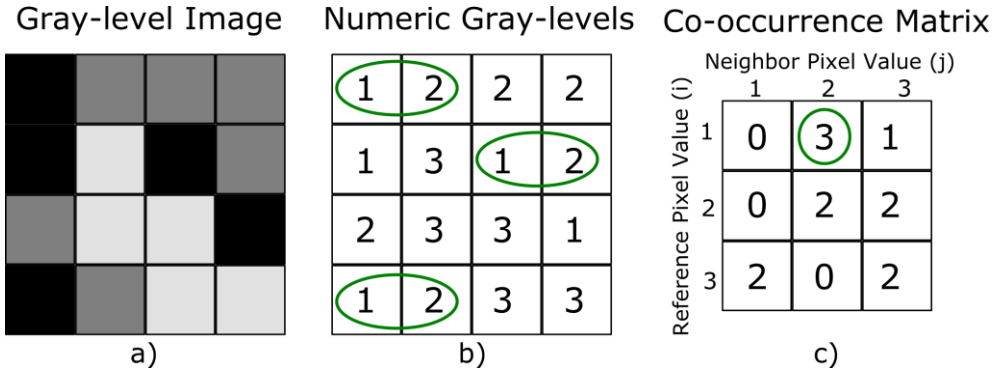

Figure 3.7. *Computation of a gray-level co-occurrence matrix (GLCM) for a given 4 × 4 pixel image (a) with three gray-levels (b). In this example the matrix is computed in horizontal direction for one-pixel separation. The number of transitions of gray-levels is counted and allocated in the co-occurrence matrix (c). The marked values indicate that there are three*

transitions from one to two gray-levels and this count is allocated in the corresponding position in the co-occurrence matrix.

Assuming that N_g is the number of distinct gray-levels, and $p(i,j)$ is the $p(i,j)^{th}$ entry in a normalized GLCM. The mean of rows and column sums of the GLCM are represented by $\mu_x, \mu_y, \sigma_x, \sigma_y$; these are related to the marginal distributions $p_x(i) = \sum_{i=1}^{N_{rows}} p(i,j)$, and $p_y(j) = \sum_{j=1}^{N_{column}} p(i,j)$.

- **Angular Second Moment:** Also known as *Energy* or *Uniformity*. High values indicate good homogeneity due to the presence of few entries $p(i,j)$, but with relatively high values.

$$Energy = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}[p(i,j)]^2 \tag{3.6}$$

- **Contrast:** Refers to how much difference there is between gray-level values of different objects within the image. Low values indicate high homogeneity.

$$Contrast = \sum_{n=0}^{N_g-1}(i-j)^2 \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i,j) \tag{3.7}$$

- **Correlation:** This feature measures how correlated a pixel is to its neighbors. It is a measure of gray-level linear dependencies within the image.

$$Correlation = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\frac{ijp(i,j) - \mu_x\mu_y}{\sigma_x\sigma_y} \tag{3.8}$$

- **Entropy:** Measures the randomness of the pixel distribution; the amount of disorder within the image. It takes high values for a more random distribution.

$$Entropy = -\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i,j)log\big(p(i,j)\big) \tag{3.9}$$

- **Sum of Squares:** It is the *variance* computed from the GLCM. It is also measure of randomness.

$$SumSq = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}(1-\mu_x)^2 p(i,j) \tag{3.10}$$

- **Inverse Difference Moment:** It is a measure of *homogeneity*. High values indicate smooth texture.

$$InvDiffMom = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\frac{1}{1+(i-j)^2}p(i,j) \tag{3.11}$$

- **Sum Average:** Measures the mean of p_{x+y}.

$$Sum\ average = \sum_{i=1}^{2N_g} i p_{x+y}(i) \tag{3.12}$$

- **Sum Variance:** Measures the variance of p_{x+y}.

$$Sum\ variance = \sum_{i=1}^{2N_g}(i - SumAverg)^2 p_{x+y}(i) \tag{3.13}$$

- **Sum Entropy:** Measures the entropy of p_{x+y}.

$$Sum\ entropy = -\sum_{i=1}^{2N_g} p_{x+y}(i)\log(p_{x+y}(i)) \tag{3.14}$$

- **Difference Variance:** Measures the variance of p_{x-y}.

$$DifVarnc = \sum_{i=0}^{N_g-1}(i-\mu_{x-y})^2 p_{x-y}(i) \tag{3.15}$$

- **Difference Entropy:** Measures the entropy of p_{x-y}.

$$DifEntrp = -\sum_{i=1}^{N_g} p_{x-y}(i)\log(p_{x-y}(i)) \tag{3.16}$$

- **Dissimilarity** [85]**:** Measures the dissimilarity between pair of pixels.

$$Dissimilarity = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}|i-j|p(i,j) \tag{3.17}$$

3.5.4 Gray-level Run-length Matrix

Gray-level run-length matrices (GLRLMs) consider higher-order statistical information in comparison with GLCMs. Runs of a specific gray-level are counted for a chosen direction. For example, three consecutive pixels with the same gray-level value along the horizontal direction constitutes one run of length three. Computation of a simple run-length matrix is shown in Figure 3.8. Fine textures will be dominated by short runs whereas coarse textures will include longer runs [98]. Rotation invariance can be achieved by averaging over all directions, as previously mentioned for the co-occurrence matrix method.

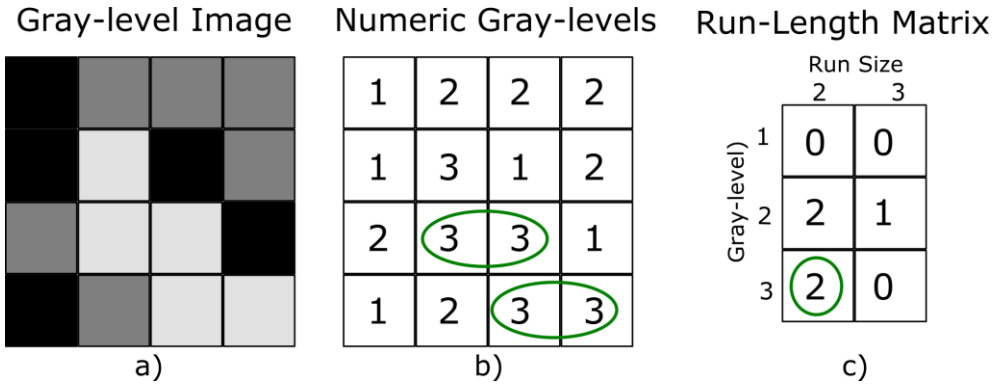

Figure 3.8. *Computation of a gray-level run-length matrix (GLRLM) for a given 4 × 4 pixel image (a) with three different gray-levels (b). The number of runs for each gray-level is allocated in the run-length matrix (c). For example, the marked values indicate that there are two runs of size two for gray-level 3.*

The features computed from GLRLMs implemented in this thesis are defined in the following lines. Assuming that N_g is the number of gray-levels and N_r is the number of runs, $p(i,j)$ is the number of times there is a run of length j having gray-level i, and $C = \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} p(i,j)$.

- **Short Run Emphasis** [104]**:** It divides each run-length value by the length of the run squared thus emphasizing short runs.

$$SRE = \frac{1}{C}\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}\frac{p(i,j)}{j^2} \tag{3.18}$$

- **Long Run Emphasis** [104]**:** It multiplies each run-length value by the length of the run squared thus emphasizing long runs.

$$LRE = \frac{1}{C}\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} j^2 p(i,j) \tag{3.19}$$

- **Gray-level Nonuniformity** [104]**:** Measures the gray-level nonuniformity. It takes low values when runs are equally distributed throughout the gray-levels.

$$GLN = \frac{1}{C}\sum_{i=1}^{N_g}(\sum_{j=1}^{N_r} p(i,j))^2 \tag{3.20}$$

- **Run-length Nonuniformity** [104]**:** Measures the nonuniformity of the run-lengths taking low values when the runs are equally distributed throughout the lengths.

$$RLN = \frac{1}{C}\sum_{j=1}^{N_r}(\sum_{i=1}^{N_g} p(i,j))^2 \tag{3.21}$$

- **Run Percentage** [104]**:** It is the fraction of the total number of runs to the total number of possible runs if all of them had a length of one. It takes low values for images with linear structures.

$$RP = \sum_{i=1}^{N_g}\sum_{j=1}^{N_r} p(i,j) / \sum_{i=1}^{N_g}\sum_{j=1}^{N_r} jp(i,j) \tag{3.22}$$

- **Low Gray-level Run Emphasis** [105]**:** Analogous to *SRE* but it makes use of the distribution of gray-levels of runs.

$$LGRE = \frac{1}{C}\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}\frac{p(i,j)}{i^2} \tag{3.23}$$

- **High Gray-level Run Emphasis** [105]**:** Analogous to *LRE* but it makes use of the distribution of gray-levels of runs.

$$HGRE = \frac{1}{C}\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} i^2 p(i,j) \tag{3.24}$$

- **Short Run Low Gray-level Emphasis** [106]**:** Emphasizes the joint distribution properties of the run-lengths and gray levels instead of individual ones separately.

$$SRLGE = \frac{1}{C}\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} \frac{p(i,j)}{i^2 j^2} \tag{3.25}$$

- **Short Run High Gray-level Emphasis** [106]**:** Similar to *SRLGE*.

$$SRHGE = \frac{1}{C}\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} \frac{i^2 p(i,j)}{j^2} \tag{3.26}$$

- **Long Run Low Gray-level Emphasis** [106]**:** Similar to *SRLGE*.

$$LRLGE = \frac{1}{C}\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} \frac{j^2 p(i,j)}{i^2} \tag{3.27}$$

- **Long Run High Gray-level Emphasis** [106]**:** Similar to *SRLGE*.

$$LRHGE = \frac{1}{C}\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} i^2 j^2 p(i,j) \tag{3.28}$$

- **Gray-level Variance** [107]**:** Variance of the gray-level distribution.

$$GLV = \frac{1}{N_g \times N_r}\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} (ip(i,j) - \sum_{i=1}^{N_g} i \sum_{j=1}^{N_r} p(i,j))^2 \tag{3.29}$$

- **Run-length Variance** [107]**:** Variance of the run-length distribution.

$$RLV = \frac{1}{N_g \times N_r}\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} (jp(i,j) - \sum_{j=1}^{N_r} j \sum_{i=1}^{N_g} p(i,j))^2 \tag{3.30}$$

3.5.5 Gray-level Size Zone Matrix

The gray-level size zone matrix (GLSZM) takes into account the size of areas with pixels of the same intensity level. It is calculated using the same principle as the GLRLM, but considering areas instead of runs. However, it is not necessary to compute matrices in several directions so this matrix is invariant to image rotation [107]. Figure 3.9 shows and example for computation of a GLSZM.

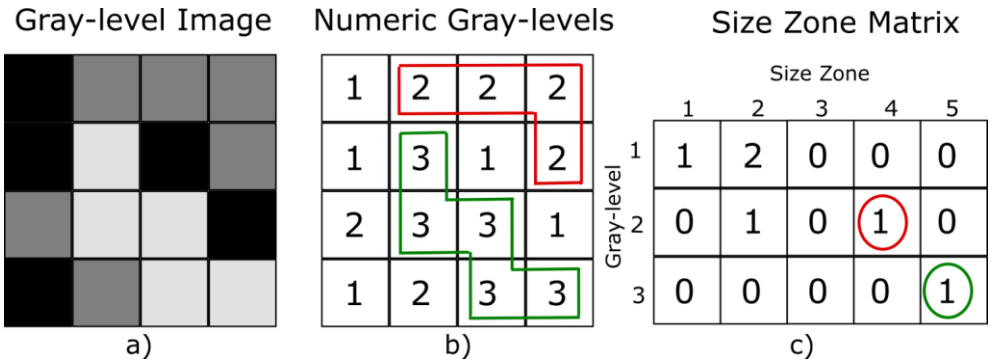

Figure 3.9. *Computation of a gray-level size zone matrix (GLSZM) for a given 4 × 4 pixel image (a) with three different gray-levels (b). The area size of each gray-level is allocated in the size zone matrix (c). For example, the marked values indicate that there is one area of size 5 for gray-level 3 (green), and one area of size 4 for gray-level 2 (red).*

Features calculated from the GLSZM are equivalent to those of the GLRLM (Equations 3.18 to 3.30) but replacing the number of runs (N_r) with the number of zones (N_z). The corresponding names when the features are derived from the GLSZM are: small zone emphasis (SZE), large zone emphasis (LZE), gray-level nonuniformity (GLN), zone-size nonuniformity (ZSN), zone percentage (ZP), low gray-level zone emphasis (LGZE), high gray-level zone emphasis (HGZE), small zone low gray-level emphasis (SZLGE), small zone high gray-level emphasis (SZHGE), large zone low gray-level emphasis (LZLGE), large zone high gray-level emphasis (LZHGE), gray-level variance (GLV), zone-size variance (ZSV) [107].

3.5.6 Neighborhood Gray-tone Difference Matrix

The neighborhood gray-tone difference matrix (NGTDM) was introduced by Amadasun and King [108] aiming to describe the visual properties of textures. To compute a NGTDM from an image, let $P(i)$ represent the summation of gray-level differences among all pixels with gray-level i. N_g represents the highest gray-level value present in the image, and $(N_g)_{eff}$ is the total number of different gray-levels in the image. The i^{th} entry of the NGTDM is defined as:

$$P(i) = \begin{cases} \sum_{0} |i - \overline{A}_i| & for\ i\ \in N_i\ if\ N_i \neq 0, \\ & otherwise. \end{cases} \tag{3.31}$$

Where N_i is the set of all pixels with gray-level i in the image excluding the peripheral region of width d, and $\overline{A}_i$ is the average gray-level of the neighbors around a center pixel with gray-level i located at position (k, l) so:

$$\overline{A_i} = \overline{A}(k,l) = \frac{1}{W-1}\left[\sum_{m=-d}^{d}\sum_{n=-d}^{d} f(k+m, l+n)\right], \qquad (m,n) \neq (0,0)$$

Where d specifies the neighborhood size and $W = (2d+1)^2$. Figure 3.10 shows an example for computing a NGTDM with $d = 1$.

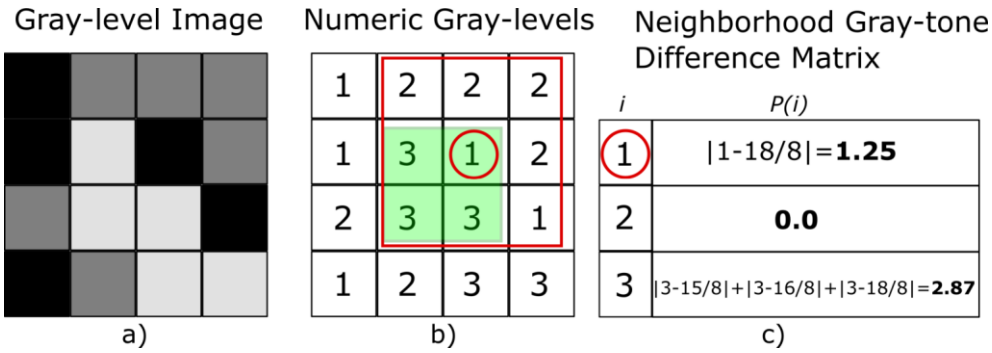

Figure 3.10. *Computation of a neighborhood gray-tone difference matrix (NGTDM) for a given 4 × 4 pixel image with three different gray-levels. Specifying a distance d = 1, results in a 3 × 3 neighborhood. This neighborhood can only be centered on pixels within the green square while the rest of the pixels are considered as being in the periphery of the image. So, for the centered pixels, the P(i) value is computed using Equation 3.31 and allocated in the NGTDM.*

For an $N \times N$ image, the probability of occurrence of the gray-level value i is defined as $n_i = \frac{N_i}{n^2}$, where $n^2 = N - 2d$. The NGTDM features are defined as:

- **Coarseness:** Coarse textures are represented by high values of this feature where gray-level differences are small. The constant ϵ in the formula is a small number to prevent *coarseness* becoming infinite.

$$Coarseness = \left[\epsilon + \sum_{i=1}^{N_g} n_i P(i)\right]^{-1} \tag{3.32}$$

- **Contrast:** High contrast means that the intensity difference between neighboring regions is large.

$$Contrast = \left[\frac{1}{(N_g)_{eff}[(N_g)_{eff} - 1]} \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} n_i n_j (i-j)^2\right]\left[\frac{1}{n^2}\sum_{i=1}^{N_g} P(i)\right] \tag{3.33}$$

- **Busyness:** A busy texture is one in which the spatial frequency of intensity changes is very high.

$$Busyness = \frac{\sum_{i=1}^{N_g} n_i P(i)}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}(in_i - jn_j)}, n_i \neq 0, n_j \neq 0 \tag{3.34}$$

- **Complexity:** Textures with high information content are considered complex, for example when many patches or primitives are present.

$$Complexity = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} \frac{|i-j|\left[n_i P(i) + n_j P(j)\right]}{n^2(n_i + n_j)}, n_i \neq 0, n_j \neq 0 \tag{3.35}$$

- **Strength:** Strong textures are comprised of primitives that are easily definable and clearly visible. The constant ϵ in the formula is a small number to prevent *strength* becoming infinite.

$$Strength = \frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}(n_i + n_j)(i-j)^2}{\left[\epsilon + \sum_{i=1}^{N_g} P(i)\right]}, n_i \neq 0, n_j \neq 0 \tag{3.36}$$

3.5.7 Autoregressive models

The autoregressive models assume a local interaction among pixels by considering the pixel gray-level as a weighted sum of its neighbors. Assuming f is a zero mean random field, an autoregressive causal model is defined by:

$$f_s = \sum_{r \in N_s} \theta_r f_r + e_s \tag{3.37}$$

Where f_s is the image intensity at site s and e_s denotes an identically distributed noise.

N_s is a neighborhood of s and θ is a vector of model parameters. An example of a local neighborhood for a causal model is shown in Figure 3.11

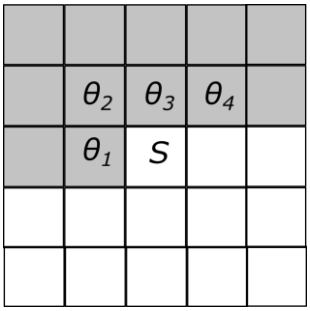

Figure 3.11. *Local neighborhood of image element fs. Shaded area indicates the region where valid causal half-plane autoregressive model neighborhood may be located. The parameters θ can be interpreted as measures of statistical similarity between intensities of pixel S and its neighbors, in four directions (vertical, horizontal and two diagonal).*

To use the autoregressive model for image texture classification, it is necessary to identify the model parameters for a given image region. In the case of the example in Figure 3.11, there are five unknown model parameters: the standard deviation σ of the noise e_s and the model parameter vector $\theta = [\theta_1, \theta_2, \theta_3, \theta_4]$. These parameters can be estimated through Equations 3.38 and 3.39 by previously minimizing the sum of the squared error: $\sum_s e^2 = \sum_s (f_s - \hat{\theta} w_s)^2$.

$$\hat{\theta} = \left(\sum_s w_s w_s^T \right)^{-1} \left(\sum_s w_s f_s \right) \tag{3.38}$$

$$\sigma^2 = R^{-1} \sum_s \left(f_s - \hat{\theta} w_s \right)^2 \tag{3.39}$$

Where $w_s = col[f_b \; i \in N_s]$, and R is the number of pixels inside ROI such that for the point s moved to a pixel location, all the four immediate neighbors of s will be placed inside the ROI as well.

The autoregressive model gives an indication on how smooth or coarse the texture is because it assumes a local interaction among local pixels, so it finds relations among groups of neighboring pixels. The parameters θ are measures of statistical similarity between intensities of the pixel s and its neighbors. Therefore, the parameters θ will

widely vary for coarse textures whereas they will be similar to each other for smooth textures. High values of the parameter σ indicates high image similarity to white random noise with no regular structure in it, and small values of σ suggests presence of regular textures [7].

3.5.8 Wavelets

The wavelet transform is a technique that analyzes the frequency content of an image within different scales and frequency directions. The frequency is directly proportional to the gray-level variations within the image. Wavelet coefficients corresponding to different frequency scales and directions can be obtained to describe a given image. Wavelet coefficients are associated to each pixel in an image to characterize the frequency content at that point over different scales [99]. Figure 3.12 shows an example of a Wavelet transform applied to an image at one scale.

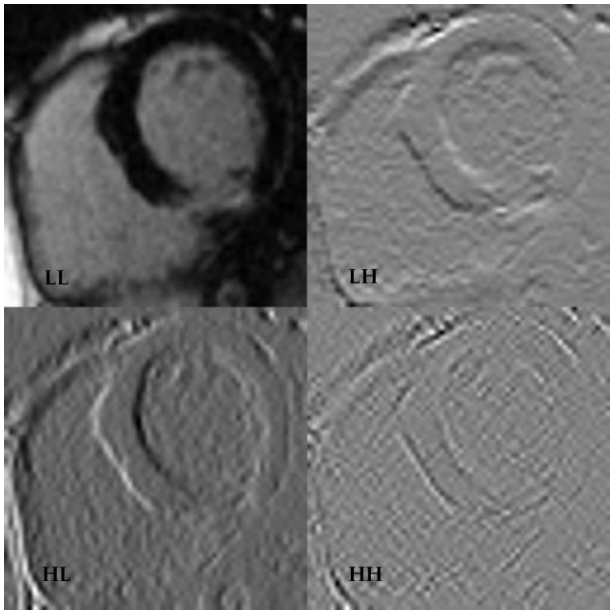

Figure 3.12. *Wavelet transform of a CMR image at one-scale decomposition. The high-high (HH) subimage represents diagonal high frequencies, high-low (HL) extracts the horizontal high frequencies, low-high (LH) vertical high frequencies and the subimage low-low (LL) represents the lowest frequencies.*

Wavelet transform methods are popular because they offer some advantages, such as: variation of the spatial resolution to represent textures at the most appropriate scale, and the wide range of choices for the wavelet function that can be adjusted for specific applications [100]. The Wavelet *energy* can be computed at any considered scale and sub-band within a marked ROI.

$$Energy = \frac{\sum_{x,y \in ROI} \left(d_{x,y}^{subband}\right)^2}{n} \tag{3.40}$$

Where n is the number of pixels within the ROI of subimage d at given scale and sub-band. ROIs need to be reduced in size at successive scales in order to correspond to sub-band image dimensions [109].

3.5.9 Local Binary Patterns

The local binary pattern (LBP) is a texture descriptor introduced by Ojala et al. [110] and it became very popular thanks to its simplicity and high-discriminative power. The LBP descriptor labels each pixel in an image by comparing its gray-level with the surrounding pixels and then assigning a binary number. A value of unity is assigned to the surrounding neighbors with gray-level greater than the central pixel in a predefined patch and a value of zero otherwise. A binary number is then obtained and assigned to the central pixel. The original LBP operator considers a 3 × 3 patch, so the surrounding pixels form a binary number of 8 digits. After labeling all pixels in an image, a LBP feature map is obtained as well as a histogram that consists of 256 bins when considering 3 × 3 patch. Figure 3.13 summarizes the described procedure.

The LBP histogram can be used as feature vector for classification where each bin represents one feature. Another approach is to compute new features from the LBP map as carried out by [54], [111]. Uniform LBPs have been proposed to reduce the length of the feature histogram. A LBP binary code is uniform if it contains at most two transitions from 0 to 1 or vice-versa. Examples of uniform patterns are: 0000000 (no transitions), 00001111 (one transition), and 10001111 (two transitions). Patterns with more than two transitions are labeled as non-uniform, and distinct labels are assigned to each uniform pattern. For a 3 × 3 patch, the number of bins on the uniform histogram is reduced to 59 instead of the original 256. Uniform LBP patterns function as templates for microstructures, such as spots, edges, corners, etc.

The original LBP descriptor defined for a 3 × 3 patch was extended to include

more neighbors by adding two parameters: the parameter *P* that defines the number of neighbors, and the radius *R* that determines the spatial resolution. Quantification at different resolutions can be obtained by varying these two parameters. Rotation invariance is achieved by performing a bitwise shift operation on the binary pattern *P-1*, times and assigning the LBP value that is the smallest. It has been shown that rotation-invariant LBP is robust against some common MRI artifacts, including image inhomogeneity even at 40% of intensity variations [112].

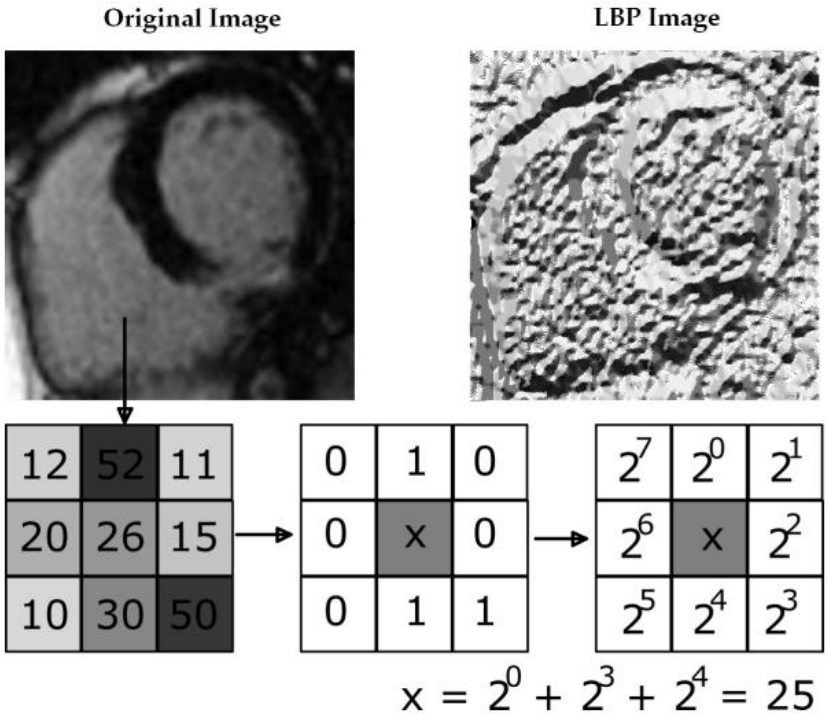

Figure 3.13. *Computation of a basic local binary pattern (LBP) image. For each pixel in the original image, its gray-level is compared to the surrounding pixels. A value of unity is assigned to the pixels with gray-level greater than the central pixel, and a value of zero otherwise. Then a binary number is obtained and this value is allocated in the same position of the original pixel thus obtaining the LBP image.*

3.5.10 Volumetric Texture Feature Extraction

Feature extraction methods have been firstly proposed for 2D texture analysis. The

advantage of the volumetric nature of MRI datasets can be obtained by extending the original methods to 3D. A simple approach to capture volumetric information is to compute 2D features in all MRI slices and then average these values [113], but in this case the gray-level distributions in the third dimension are not taken into account. Nevertheless, it has been shown that even this simple averaging method outperforms the typical 2D approach where only one slice is analyzed [114]. The extension of 2D approaches to 3D is not straightforward as factors such as translation and scaling become more complex. A review of 3D feature extraction methods is presented in [83]. Compared with 2D texture analysis, 3D approaches increase the dimensionality and the information captured from the image, thus improving the discrimination power [78], [115], [116].

Dynamic texture analysis is possible by including the temporal dimension available in some image datasets, i.e. cine CMR. For this purpose, an extension of the LBP texture analysis method was proposed to analyze video sequences [117]. The latter approach can be considered as $2D + t$, so conventional 3D features can also be implemented by using time (t) as the third dimension. Notable results were observed for discrimination of benign and malignant breast lesions [118] and for localization and segmentation of the heart [119], when using 4D spatio-temporal texture analysis in MRI.

3.5.11 Feature Extraction Tools

The widely used software package MaZda (Institute of Electronics, Technical University of Lodz, Lodz, Poland) [120] is freely available and allows computation of texture features based on histogram, GLCM, GLRLM, absolute gradient, autoregressive model and the Haar wavelet transform. MATLAB (MathWorks Inc., Natick, MA) toolboxes can also be found for texture feature extraction, like the one provided by Vallières et al. [85][1] that allows computation of features based on GLCM, GLRLM, GLSZM, and NGTDM; and the implementations of the local binary pattern operator provided by Ojala et al. [110][2].

[1] Available from https://github.com/mvallieres/radiomics

[2] Available from http://www.cse.oulu.fi/CMV/Downloads/LBPSoftware

Chapter 4

Overview on Machine Learning

4.1 Introduction

Machine learning is an interdisciplinary field that include statistics, artificial intelligence, data mining and pattern recognition, by applying algorithms that can learn from data to accurately represent and make use of the underlying information. Machine learning in medical imaging is a promising growing field and its main applications include image segmentation, registration, computer aided detection and diagnosis. Essentially, machine learning algorithms aim to identify patterns that could help radiologists and specialists make accurate decisions when assessing imaging data [10], [121].

Machine learning problems can be *supervised* or *unsupervised*. In supervised learning, each sample in the dataset consists of input observations, or *features,* and output observations, or *responses.* When the responses are categorical variables the problem is known as *classification,* and when the responses are numeric the problem is known as *regression.* The purpose of supervised learning is to build a model that can "learn" from available data in order to predict new samples with unknown *classes*. In contrast, unsupervised learning is more challenging as the output classes are not available. Therefore, it is necessary to seek relationships between features and observations. Cluster analysis is a typical unsupervised tool that aims to find if the observations fall into distinct groups [122].

In the texture analysis process that was illustrated in Figure 3.1, machine learning techniques play a key role. Considering the large number of texture features that can be computed it is necessary to find those that better describe the data, this process is known as *feature selection.* The reduced set of features is then used to differentiate tissues or lesions; this is the *classification* step. This chapter focus on describing the supervised classification methods, and feature selection techniques, that were implemented in the experimental studies of this thesis.

4.2 Feature Selection

The vast variety of feature extraction methods for texture analysis allows us to obtain a myriad of features. This generates a problem, because as much features we have, the more complex the classification model becomes. The computed features have different discrimination power depending on the application. The redundant or irrelevant features hinder the classification performance and can yield issues of dimensionality. This phenomenon arises when dealing with a high-dimensional feature space. The classification performance decreases when more features are added to the model. Feature selection is the process to choose the most relevant features for a specific application. Reducing the number of features speed up the testing of new data and make the classification problem easier to understand, but the main benefit is the increase in classification performance by avoiding the curse of dimensionality. This phenomenon occurs when adding more features degrade the performance of a predictive model [123], [124].

While methods like principal component analysis (PCA) or linear discriminant analysis (LDA) are used for dimensionality reduction, they are not considered as feature selection methods since they still require the computation of all the original features [125]. Apart from *embedded* feature selection methods that are built-in within some predictive models, i.e. classification trees, most approaches for reducing the number of features fall into two main groups: *filter* and *wrapper* methods [126].

4.2.1 Filter Methods

Filter feature selection methods make use of a certain parameter to measure the discriminatory power of each feature. For example, typical statistical methods, such as *t*-test or the Mann–Whitney U test, can be used to find and select features with statistical significance [93]. Filter methods are performed as a preprocessing step that is independent of the choice of the predictive model. Most filter methods consider the individual predictive power of each feature and consequently, redundant features may be selected and important interactions between features will not be quantified. For example, filters may not select a feature that is completely useless by itself but is relevant to provide a significant improvement in performance when used together with other features [123]. Filter methods rank features according to the measuring parameter and usually a predefined number of features is selected, e.g. 5 or 10, as

input to a predictive model.

A popular filter method is based on the *Fisher* coefficient, which is defined as the ratio of between-class variances to within-class variances:

$$F = \frac{D}{V} = \frac{\frac{1}{1-\sum_{k=1}^{K} P_k^2} \sum_{k=1}^{K} \sum_{j=1}^{K} P_k P_j (\mu_k - \mu_j)^2}{\sum_{k=1}^{K} P_k V_k} \tag{4.1}$$

Where D denotes between-class variance, V is the within-class variance, μ_i and V_i are the mean and variance of class i respectively, P_i is the probability of class i (ratio of number of data samples from class i to the number of all data samples) [127]. The Fisher method is computationally efficient but it generates highly correlated features with the same discriminatory power that may not be optimal for classification. In these cases, wrapper methods may be preferred.

4.2.2 Wrapper Methods

The main drawback of the filter methods is that the feature selection is based on the intrinsic information of the training data and does not consider the predictive capability of a certain subset of features. Wrapper methods take advantage of a predictive model in combination with a search algorithm in order to select the subset of features that provides optimal performance. The downside of wrapper methods is that many models have to be trained and tested thus increasing the computation time. There is also an increased risk of over-fitting with wrappers if the evaluation with resampling methods are not performed correctly. Resampling methods to avoid over-fitting will be described in Section 4.4.

The quality of the selected subset of features in wrappers depends fundamentally on the search algorithm. A straightforward approach to find the most discriminative features, or the combination of features that yields the best classification, is to perform an exhaustive search. In the exhaustive search method, all possible combinations of features are tested as input to a classifier and those that yield the best discrimination are selected. The problem with this method is that it becomes tremendously expensive to compute when the feature space is very high. Therefore, an algorithm that uses some type of search strategy has to be chosen.

Forward selection and backward elimination are the most popular search strategies. In *forward* selection, features are progressively added into larger subsets, whereas in *backward* elimination the algorithm starts with the full set of features and

progressively removes the least promising ones. Both strategies generate subsets of features that are evaluated by using an objective function, i.e. the classification accuracy of a predictive model. Forward selection is computationally more efficient than backward elimination. However, it has been argued that weaker subsets are found by forward selection because the importance of features is not assessed in the context of other features not included yet [123].

The recursive feature elimination (RFE) is a backward search algorithm that avoids refitting many models at each step of the search. The full model is created and a measure of feature importance is computed to rank them from most to least important. It ranks the features by recursively training a classifier and removing the feature with the smallest ranking score and selecting the subset of features that yields the best classification. Any classifier can be used in conjunction with the RFE to compute the feature scores [126]. The feature selection technique known as support vector machine-recursive feature elimination (SVM-RFE), first proposed for gene selection in cancer classification [128], has gained major attention for selecting texture features due to its good performance over other methods [129].

The SVM-RFE algorithm returns a ranking of features by recursively training a linear SVM and removing the feature with the smallest ranking score. At each iteration, the coefficients of the weight vector $\boldsymbol{w}$ of a linear SVM are used to compute the feature ranking score. The feature with the smallest ranking score $c_i = {w_i}^2$ is eliminated. The ranking criterion is chosen to remove the feature whose removal least affects the objective function, which is defined as $J = 1/2\|\boldsymbol{w}\|^2$. The choice of the objective function is explained by expanding the function in Taylor series to second order [128]:

$$\Delta J(i) = \frac{\partial J}{\partial w_i}\Delta w_i + \frac{\partial^2 J}{\partial {w_i}^2}(\Delta w_i)^2. \tag{4.2}$$

At the optimum of J, the first-order term can be neglected and with $J = 1/2\|\boldsymbol{w}\|^2$, Equation 4.2 becomes $\Delta J(i) = \Delta {w_i}^2$. The change in weight $\Delta w_i = w_i$ corresponds to removing the i^{th} feature. A modified version of the SVM-RFE method was proposed by Duan et al. [130]. The modified approach, named multiple SVM-RFE, incorporates resampling at each step of the SVM-RFE algorithm to stabilize the feature selection. If t linear SVMs are trained on different subsamples of the original data: let $\boldsymbol{w}_j$ be the weight vector of the j^{th} SVM, $\boldsymbol{w}_{ji}$ the corresponding weight value associated with the

j^{th} feature, and $v_{ji} = {w_{ji}}^2$. The feature ranking score becomes:

$$c_i = \frac{\overline{v_i}}{\sigma_{vi}} \tag{4.3}$$

Where v_i and σ_{vi} are the mean and standard deviation of variable v_i:

$$\overline{v_i} = \frac{1}{t}\sum_{j=1}^{t} v_{ji} \tag{4.4}$$

$$\sigma_{vi} = \sqrt{\frac{\sum_{j=1}^{t}(v_{ji} - \overline{v_i})^2}{t-1}} \tag{4.5}$$

The weight vectors are normalized $\boldsymbol{w}_j = \boldsymbol{w}_j / \|\boldsymbol{w}_j\|$ before computing the ranking score. The algorithm described by Duan et al. [130] is:

1. Start: ranked feature set $R = [\,]$; selected subset $S = [1, \dots, d]$;
2. Repeat until all features are ranked:
    a. Train t linear SVMs on subsamples of the original training data, with features in set S as input variables;
    b. Compute and normalize the weight vectors;
    c. Compute the ranking scores c_i for features in S using Equation 4.3;
    d. Find the feature with the smallest ranking score: $e = \arg min_i c_i$;
    e. Update: $R = [e, R], S = S - [e]$;
3. Output: Ranked feature list R.

4.3 Classification Methods

The main goal in texture analysis applications is the classification of different tissues and/or lesions to automate or aid the diagnosis decision. Simple statistical methods can be used to determine the texture features with statistical significance for discrimination of two or more classes. However, following the feature selection step described in the previous section, we focus on more complex classification algorithms that make use of proper combination of features to achieve the highest discrimination. The feature selection and classification steps are not specific for texture analysis, so instead of providing a full description of the existing methods, we briefly describe the classifiers that were implemented in the experimental studies of this thesis. These classifiers were chosen due to their well-known performance in application to other datasets [131].

4.3.1 K-Nearest Neighbors

K-Nearest Neighbors (KNN) classifies data by using a sample's neighborhood to predict the sample's class. The classification is performed using the K-closest samples from the training set. How close is a sample from its neighbors is determined by a metric like the Euclidean distance. To allow each feature to contribute equally to the distance calculation they need to be centered and scaled, otherwise the distance values will be biased towards features with larger scales [126].

Given a positive integer K and a test observation x_0, the KNN classifier tries to identify the K points in the training data that are closest to x_0, represented by N_0. Then it estimates the conditional probability for class j as the fraction of points in N_0 whose response values equal j. The test observation x_0 is assigned to the class with the largest probability [122]:

$$\Pr(Y = j|X = x_o) = \frac{1}{K}\sum_{i \in N_0} I(y_i = j). \tag{4.6}$$

The choice of the number of neighbors K is important as it affects the classification performance. Figure 4.1 shows an example for two different values of K that results in different classifications. When $K = 1$, the decision boundary is overly flexible and assigns the class of the closest point. At larger values of K, the method becomes less flexible and produces nearly linear decision boundaries [122]. Therefore, the parameter K should be tuned by using resampling techniques in order to choose the value that yields the best performance.

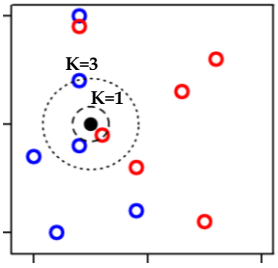

Figure 4.1. *An example of the KNN classifier for a dataset with two classes, illustrated in blue and red. A new sample at which a predicted class label is desired is shown as a solid black point. The closest points to the new observation are identified when using K=1 and K=3 neighbors. The new sample will belong to the most commonly occurring class depending on the number of neighbors. In this case, if K=1 the new sample will be predicted as red and if K=3 it will be predicted as blue.*

4.3.2 Artificial Neural Network

Artificial neural networks (ANNs) simulate the way the human brain processes information by implementing neurons and inter-connections. They function based on the following principles:

- Elements called neurons, or nodes, process the information.
- The nodes transmit signals to other nodes through connections.
- Each connection has an associated weight.
- Each node applies an activation (transfer) function (usually non-linear) to the sum of entries weighted according to the connection weights, thus obtaining an output value that will be transmitted to the rest of the network.

The artificial neuron is the one that receives the signal from the neighboring neurons and then calculates an output value, which is sent to other neurons. Nodes can represent the input and output information as discrete or continuous values. For the sake of classification purposes, the outputs are represented as discrete values. Basically, an ANN consists of three types of layers: input, hidden and output layers; where each layers is composed by several nodes. Figure 4.2 shows a basic ANN with three layers. The input nodes receive the information from the predictors, or features, and these are transmitted through the network. The output nodes receive the weighted information and produces the output value, which in the example correspond to a binary classification [132].

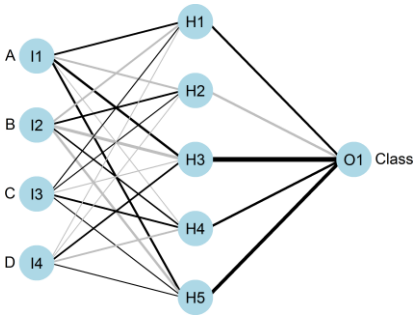

Figure 4.2. *An artificial neural network (ANN) with three layers. The input layer consists of nodes I1 to I4, the hidden layer of nodes H1 to H5, and the output layer consists of node O1. This model has four features (A, B, C, D) that correspond to the input nodes, and one output node that correspond to a binary prediction. The interconnections between nodes are indicated with lines, where thicker lines indicate larger weights and color indicates the sign of the weight value (+ black, - gray).*

The arrangement of nodes together with the activation function used by the neurons make it possible to solve non-linear classification problems. Typical activation functions are: the step function, sigmoid function and Gaussian function, among others. An ANN model can be designed with different number of nodes and/or hidden layers and with different inter-connections between them. The ANN discrimination power depends on the density and complexity of these interconnections [10].

The most popular ANN model is known as *multilayer perceptron*. This type of ANN starts with an input layer in which each node corresponds to a predictor variable. The input nodes are connected with each of the nodes making up a hidden layer. The nodes in the hidden layer connect with the nodes in another hidden layer or to the output layer. The output layer consists of one (binary prediction) or more output nodes. A multilayer perceptron can be made up of different number of hidden layers with a minimum of one. The illustration of Figure 4.2 represents a multilayer perceptron with one hidden layer. In the multilayer perceptron, the information is always transmitted form the input layer towards the output layer [132], [133].

The advantage of ANNs is that as more complex they are, with more nodes and interconnections, they can learn complex patterns in data that are not easily separable. However, the complexity of ANNs makes them computationally expensive [10]. Usually the weights of an ANN are initialized using random values near zero. Hence, the model starts nearly linear and becomes non-linear as the weights increase. Often ANNs have too many weights and will over-fit the data, so they will adapt almost perfectly to the training data but the performance will notably decrease when trying to predict new samples. An explicit penalization method used to regularize the model is by using the parameter *weight decay*. For a three layer multilayer perceptron, the weight decay and the number of nodes in the hidden layer need to be properly chosen in order to obtain the best generalization performance [133].

4.3.3 Random Forest

A classification tree is a predictive model that uses a set of binary rules to calculate a target value. To better understand how a classification tree works we refer to Figure 4.3. In this example, the dataset consists of four input features (*A, B, C, D*), and two output classes. The classification tree algorithm randomly chooses predictor *B* and

makes a split into two nodes according to the values that are above or below a certain threshold. The data is split in this manner several times until a stopping criterion is met. The aim of classification trees is to partition the data into smaller and pure groups, those that contain a larger proportion of one class in each node [126].

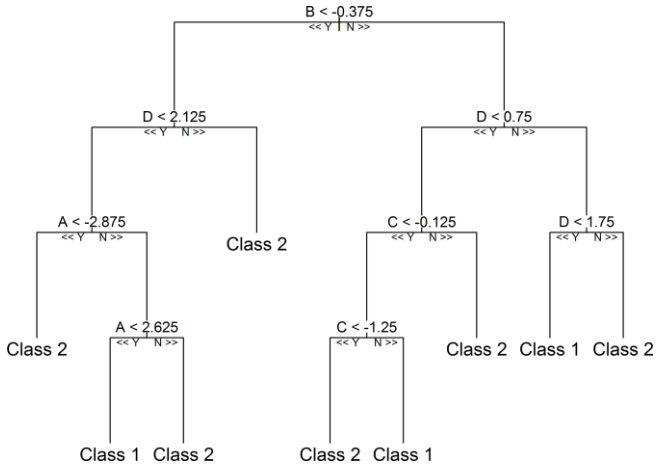

Figure 4.3. *A classification tree that uses four predictors (A, B, C, D) in order to find a decision rule that predicts the data into class 1 or class 2.*

An ensemble model is one that combines the results from different models that can be of similar or different type. The result of an ensemble model is usually better that the individual ones. The *random forest* classifier is an ensemble model of many classification trees. Classification trees are easy to interpret but they are prone to over-fitting and ignorance of a variable in case of small sample size. Random forests generally improve the performance of the model at the expense of lossing interpretability. Each tree in a random forest grow as follows [126], [133]:

- A subset of 2/3 of the training data is selected to train each tree.
- Some features are selected randomly out of all predictor variables and the best candidate is used to split the node.
- For each tree, using the remaining 1/3 of data, the misclassification rate is calculated. Aggregated error from all trees is used to determine the overall classification.
- Each tree provides a classification, or vote for a certain class. The forest classifies according to those having the most votes over all trees.

Two parameters determine the overall performance of a random forest classifier: the strength, how accurate the individual classifiers are; and the correlation, the dependence between the trees. To improve accuracy, it is necessary to minimize the correlation while maintaining the strength. Reducing the number of random variables $mtry$ used in each tree reduces both the strength and correlation. Increasing it increases both. Therefore, the parameter $mtry$ has to be chosen in order to find the best balance that will yield the best performance. Another tuning parameter that affects the performance of the random forest classifier is the number of trees, which also needs to be chosen according to the application [134].

4.3.4 Support Vector Machines

Support vector machines (SVMs) are advanced classifiers that were developed in the early 1990s and have gained popularity in many fields. They have been shown to perform well in a variety of applications and are often considered one of the best classifiers for computer-aided detection in radiology [10], [122].

To understand the SVM, first we define the *hyperplane.* In a p-dimensional space, a hyperplane is a flat subspace of dimension $p - 1$. Therefore, in two dimensions, a hyperplane will be a line and in three dimensions it will be a plane. In higher dimensions is difficult to visualize a hyperplane but the same notion still applies. Considering the classification problem shown in Figure 4.4 where two features are used to predict two classes that are perfectly separable, there are multiple possibilities of a hyperplane that can be used to separate the classes, so it is necessary to use an appropriate metric to decide which hyperplane to use [122], [126].

The *margin* distance is a metric defined by Cortes and Vapnik [135]. The margin is the distance between the classification boundary and the closest data point. For example, Figure 4.4 (right) shows one possible hyperplane as a solid line. The dashed lines on both sides of the boundary are at the maximum distance from the line to the closest training data. In this example, the highlighted data points that are equally closest to the classification boundary are known as *support vectors*. The *maximum margin classifier* is defined by the slope and intercept of the boundary that maximize the buffer between the boundary and the data.

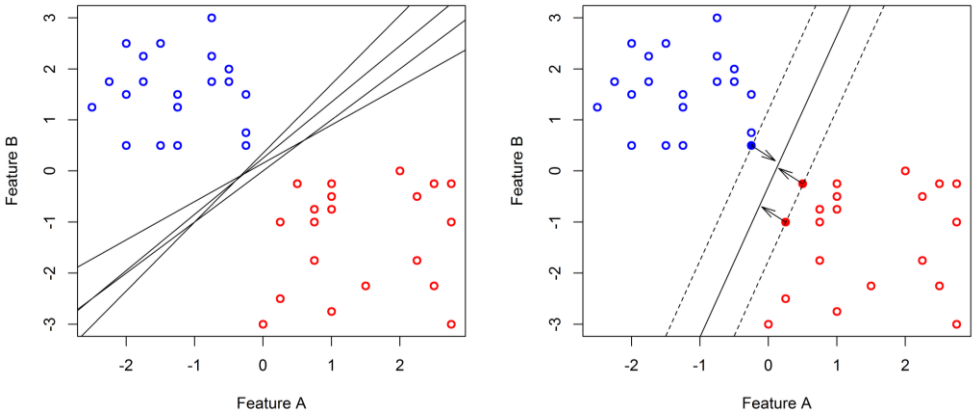

Figure 4.4. *A data set with linearly separable classes, illustrated in blue and red. The left plot shows four separating hyperplanes, out of many possible, that correctly separate both classes. The right plot illustrates the maximal margin hyperplane shown as a solid line. The margin is the distance from the solid line to either of the dashed lines. The solid points that lie on the dashed lines are the support vectors, and the distance from those points to the margin is indicated by arrows.*

In many cases, due to complexity of the data, it is not possible to construct the hyperplane that satisfies the maximal margin classifier. This motivated the concept known as *soft margin*, which also separates both classes but allowing some misclassifications aiming to perform better when grouping the remaining data points. The support vector classifier is based on the soft margin because it allows some data points to be on the incorrect side of the margin. The data point can be on the incorrect side of the margin and also on the incorrect side of the hyperplane.

The use of hyperplanes allows classification of linearly separable data. However, many classification problems are non-linear. For this purpose, the SVM seeks to map the input space to a higher dimension via a *kernel* function, and in the transformed feature space, it tries to find a hyperplane that will result in maximal discrimination. Here, a kernel is a matrix that encodes the similarities between samples that can be used to achieve discrimination between classes that are non-linearly separable [10].

For a two-class classification problem with N samples $\{\boldsymbol{x}_i, y_i\}^N$, where x_i denotes the i^{th} sample vector in the original space and $y_i \in \{+1, -1\}$ represents the class label of $\boldsymbol{x}_i$, the optimization problem to find the discriminant function $f(x) = w.\varphi(\mathbf{x}) + b$ is

formulated as [135]:

$$\min_{\mathbf{w},b} \frac{1}{2}\left|\left|\boldsymbol{w}\right|\right|^2 + C\sum_{i=1}^{N} \xi_i \tag{4.6}$$

Subject to: $y_i(\boldsymbol{w}\cdot\boldsymbol{x}_i + b) \geq 1 - \xi_i$, where $\xi_i \geq 0$ are slack variables that allow a data point to be in the margin error $0 \leq \xi_i \leq 1$ or to be misclassified $\xi_i \geq 1$. The parameter $C > 0$ sets the trade-off between margin and losses. By using Lagrange multipliers, (4.6) can be written into its dual form:

$$\max_{\alpha} \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} y_i y_j \alpha_i \alpha_j (\boldsymbol{x}_i \cdot \boldsymbol{x}_j) \tag{4.7}$$

Subject to: $\sum_{i=1}^{N} y_i\alpha_i = 0\,,\ 0 \leq \alpha_i \leq C.$ The solution of the dual formulation is the weight vector $\boldsymbol{w}$:

$$\boldsymbol{w} = \sum_{i=1}^{N} \alpha_i y_i \boldsymbol{x}_i\,. \tag{4.8}$$

The support vectors are those samples ($\boldsymbol{x}_i$ for which $\alpha_i > 0$) that are on the margin or within the margin when a soft margin SVM is used. A kernel function, defined as $k(\mathbf{x},\mathbf{x}') = \varphi(\mathbf{x})\cdot\varphi(\mathbf{x}')$, can replace the dot products in the SVM formulation (4.7); thereby performing large margin separation in the feature-space of the kernel [135], [136]. The *polynomial* kernel is defined as:

$$k(\mathbf{x},\mathbf{x}') = (\mathbf{x}\cdot\mathbf{x}' + 1)^d \tag{4.9}$$

The feature space for this kernel consists of all monomials up to degree d. The kernel with $d = 1$ is the *linear* kernel, and in that case the additive constant in Equation 4.9 is usually omitted. As the degree of the polynomial kernel is increased, the classifier will have a more flexible decision boundary. Another widely used kernel is the *radial* or Gaussian kernel defined by:

$$k(\mathbf{x},\mathbf{x}') = \exp(-\sigma\|\boldsymbol{x} - \boldsymbol{x}'\|^2) \tag{4.10}$$

Where $\sigma > 0$ is a parameter that controls the width of the Gaussian. It plays a similar role as the degree of the polynomial kernel in controlling the flexibility of the classifier [136]. Examples of non-linearly separable data is shown in Figure 4.5, where a radial and polynomial kernel perfectly define the classification boundaries.

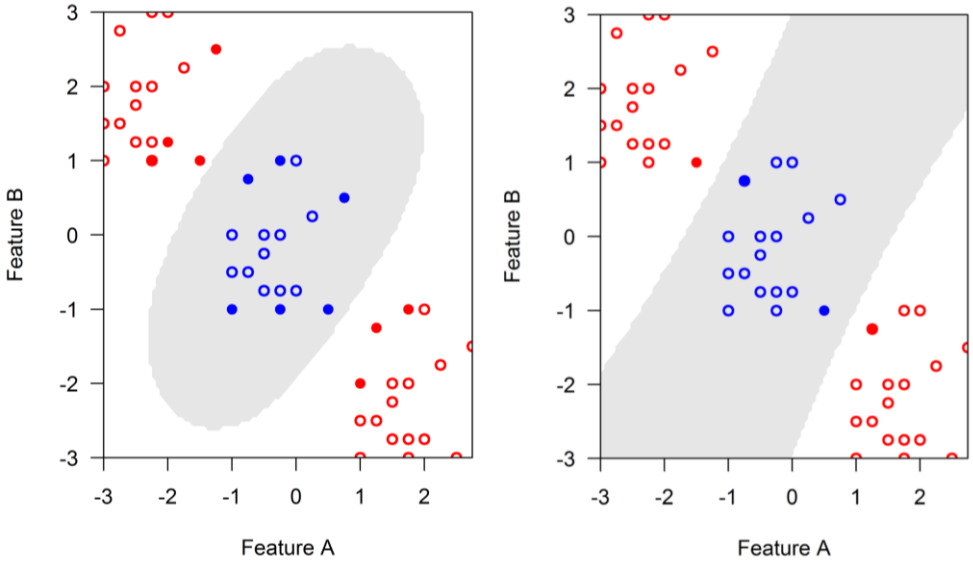

Figure 4.5. *A data set with non-linearly separable classes, illustrated in blue and red. The plots show how a radial kernel (left) and polynomial kernel of degree 3 (right) can separate both classes. The classification boundaries are shown in gray while the support vectors are highlighted with solid points.*

The SVM has been designed for two class problems. To extend SVMs to multiclass problems there are two alternatives: *one-versus-one* and *one-versus-all* classification. In the first approach, if there are $K > 2$ classes then a combination of $\binom{K}{2}$ SVM classifiers is constructed, each comparing a pair of classes. The final classification is performed by assigning the test observation to the class to which it was most frequently assigned in the pairwise classifications. The one-versus-all approach assumes that one class is of interest and the rest of classes are collectively represented by a unique class, and this is repeated for all classes obtaining a total of K SVMs. A test observation is assigned to the class for which the fitted SVM is largest, as this amounts to a high level of confidence that the test observation belongs to the kth class rather than any of the other classes [122].

4.4 Model Validation and Evaluation

4.4.1 Over-fitting and Bias-variance Trade-off

Complex non-linear models are capable of learning data so well that can correctly predict every sample. However, the interest is to accurately predict new samples that were not used when fitting the model. As more flexible is the model it will better fit the data but its generalization to new samples will be very poor. This problem is commonly known as *over-fitting*. To avoid overestimated values, it is always recommended to split the data into training and testing sets so that results on new data can be reported. The testing set will only be used to estimate the classification performance of the model developed with the training set [126].

There is another concern when estimating the model performance on testing set and it is referred as the bias-variance trade-off. *Bias* is the error that is introduced by approximating a real-life problem; the difference between the predicted and true values of performance. *Variance* refers to the amount by which the model would change if it is estimated by using a different training set. Ideally the model performance should not vary too much if different training sets are used. If a model has high variance, then small changes in the training set can result in large changes in the estimated performance. Good test set performance will have low variance and low bias but in general this is not the case. As the model flexibility increases, the variance will increase and the bias will decrease, therefore it is important to find the best bias-variance trade-off [122].

4.4.2 Model Tuning

Many models have parameters that cannot be directly estimated from the data but have to be chosen when building the model. For example, in a KNN classifier, choosing too few neighbors may over-fit the samples of the training data while too many neighbors may not be sensitive enough to yield reasonable performance. Here, the number of neighbors is referred as *tuning parameter* because there is no formula to calculate an appropriate value. The classification models described in the previous section have at least one tuning parameter. These parameters control the complexity of the models, so they are likely to cause over-fitting if values are not suitable.

One approach to choose the best tuning parameters is to define a set of candidate

values and then fit a model for each of these values. Then, each model has to be evaluated on a hold-out or test set and the tuning parameters that yield the best generalization performance will be taken as optimal. The recommendation for choosing the tuning parameters is to repeat this procedure several times and then averaging the results. The latter can be done by using some resampling techniques [126].

4.4.3 Resampling techniques

The ideal procedure to estimate the model performance on new data is to use a separate testing set, but unfortunately data samples are usually sparse. Resampling techniques use a subset of samples to fit the model and the remaining to estimate the efficacy of the model. This process is repeated multiple times and the results are then aggregated.

In the resampling technique known as *k-fold cross-validation*, the samples are randomly partitioned into k groups (or folds) of equal size. First, the model is trained using all samples except those on the first fold. Those samples not used to fit the model are used to estimate performance measures. The same process is repeated with the second fold as held out, and so on. Lastly, the k resampled estimates of performance are summarized with the mean and standard deviation.

A special case where k equals the number of samples is known as *leave-one-out* cross-validation. In this case, only one sample is held-out at a time so the final performance is calculated from the k individual predictions. Repeated k-fold cross-validation is another variant that replicates the original procedure multiple times. For example, if 5-fold cross-validation is repeated ten times, 50 different held-outs will be used to estimate the model performance. There is no formal rule to choose the value of k but it have been shown empirically that using $k = 5$ or $k = 10$ yields an accuracy that suffer neither from excessively high bias nor from very high variance. Additionally, repetitions on the k-fold cross-validation reduces the bias [122], [126].

An important consideration has to be made when performing feature selection and resampling. A typical mistake is that resampling is performed after the features had been selected. Over-fitting occurs as consequence because the feature selection algorithm already "saw" the resampling held-outs. An outer resampling loop is needed to properly resample the feature selection process. The outer loop is used to

estimate the generalization performance of the model with the selected features that are estimated using an inner loop [126]. This resampling method is referred as *nested cross-validation*.

4.4.4 Measures of Classification Performance

Considering that we predict the class labels on a testing set, the *confusion matrix* (Figure 4.6) is the common method for describing the classification performance. In the examples we consider a two-class problem but the same notion applies when constructing a confusion matrix for more than two classes.

		True values	
Predicted values		**Positive**	**Negative**
	Positive	True Positives (TP)	False Positives (FP)
	Negative	False Negatives (FN)	True Negatives (TN)

Figure 4.6. *Diagram showing how to construct a confusion matrix for a classification problem with two classes, namely positive and negative.*

From the first row of the confusion matrix, the values of the positive class that are predicted correctly are the true positives (TP), whereas those that are inaccurately classified are the false positives (FP). Similarly, the second row contains the predicted negatives with false negatives (FN) and true negatives (TN). Several performance measures can be calculated from the confusion matrix:

- **Accuracy:** The overall accuracy reflects the agreement between the observed and predicted classes.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \tag{4.11}$$

- **Sensitivity:** The sensitivity of the model is the rate that the positive class is predicted correctly for all samples having that class. The sensitivity is also known as the *true positive rate*.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{4.12}$$

- **Specificity:** The specificity is the rate that the negative class is predicted correctly for all negative samples. The *false positive rate* is defined as one minus the specificity.

$$\text{Specificity} = \frac{\text{TN}}{\text{FP} + \text{TN}} \tag{4.13}$$

- **Positive Predictive Value:** Indicates the probability that the prediction assigned to the positive class is true.

$$\text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{4.14}$$

- **Negative Predictive Value (NPV):** Indicates the probability that the prediction assigned to the negative class is true.

$$\text{NPV} = \frac{\text{TN}}{\text{FN} + \text{TN}} \tag{4.15}$$

Measures calculated from the confusion matrix are useful to evaluate the classification performance. Nevertheless, it should be noted that the confusion matrix is constructed by considering discrete predictions. Most classification models also generate continuous predictions that are in the form of a probability. For example, the predicted values of class association for any individual sample are between zero and one and sum one. In some applications, the probabilities for each class can be very useful for measuring the confidence of the model about the predicted classification. The class probabilities can be used to set a threshold value that will determine the boundary between classes. Classification models usually take the probability of 0.5 as threshold but in some applications it may be of interest to set a different value [126].

The receiver operating characteristic (ROC) curve is a method used to evaluate the classification performance by considering class probabilities. It is constructed by evaluating the class probabilities across different thresholds. For each threshold, the sensitivity and false positive rate (one minus the specificity) are plotted against each other. Figure 4.7 shows a ROC curve in which two threshold probabilities are indicated on the curve, one at 0.5 and the other at 0.9. It can be noticed how the resulting sensitivity and specificity varies depending on the threshold value. A model that perfectly separates two classes would have unity sensitivity and specificity. On the ROC curve this will be represented as a single step between (0, 0) and (0,1) and

constant from (0, 1) to (1, 1) [126].

The area under the ROC curve (AUC) is a measure used to determine the overall performance of a classifier. A perfect classifier will have AUC = 1 whereas a completely ineffective classifier will have AUC = 0.5, which is plotted as a diagonal line. The ROC curve has the advantage to be insensitive to disparities in the class proportions since it is a function of sensitivity and specificity. Therefore, it is suggested to report results using the AUC instead of the overall accuracy when dealing with imbalanced datasets, as in these cases the accuracy will be mostly influenced by the majority class [10], [126].

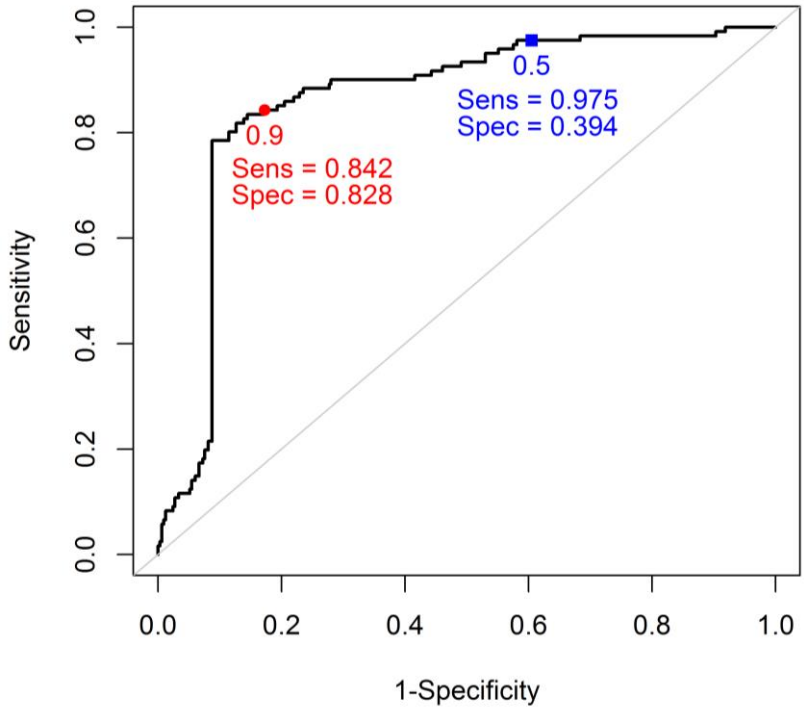

Figure 4.7. *A receiver operating characteristic (ROC) curve for a given data. The blue square indicates the value corresponding to a probability threshold of 50% while the red dot corresponds to a threshold of 90%.*

Chapter 5

Segmentation of Infarcted Myocardium in LGE CMR. A preliminary multicenter evaluation

Part of this chapter was published in:

Larroza A, López-Lereu MP, Monmeneu JV, Bodí V, Moratal D.
Texture analysis for infarcted myocardium detection on delayed enhancement MRI.
IEEE International Symposium on Biomedical Imaging (ISBI 2017), April 18-21, 2017, Melbourne, Australia.

5.1 Introduction

Assessment of myocardial infarct size is relevant for prognosis of patients suffering myocardial infarction. The well-established technique for infarct detection is LGE CMR. Manual segmentation is possible thanks to hyperenhancement of the infarcted area but it is prone to inter and intra-observer variability [46]. State-of-the-art methods for quantification of infarct size rely on setting an intensity threshold (2-5 standard deviations) above remote myocardium, or the FWHM. New techniques have been developed in order to overcome the limitations of intensity based methods [47].

Texture analysis in medical imaging has been applied for segmentation of anatomical structures and detection of lesions, as different lesion areas exhibit different textural patterns that sometimes are beyond human visual perception [6]. A previous study about the application of texture analysis in CMR reported an improvement in detection of infarcted myocardium in LGE CMR [20].

In this first experimental study, we explored the capability of texture features in combination with image histogram features to distinguish infarcted from healthy myocardium. A method for segmentation of the infarcted myocardium in LGE CMR by training a support vector machine (SVM) classifier with previously selected features is proposed. This study is a preliminary evaluation of the application of

texture analysis for segmentation using data from two different centers.

5.2 Material and Methods

The steps followed towards segmentation are summarized in Figure 5.1 and are described in the following sections.

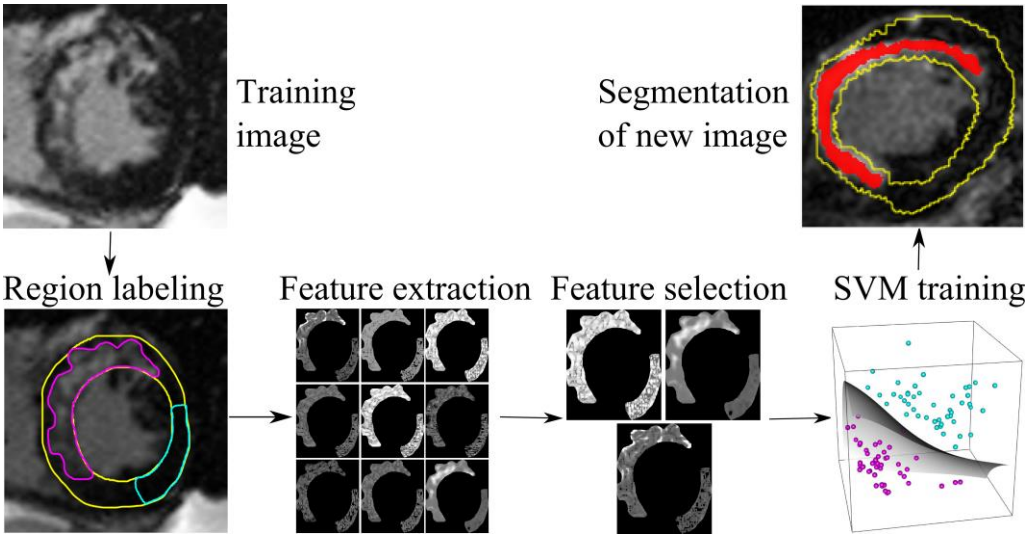

Figure 5.1. *General approach of the proposed infarcted myocardium segmentation method.*

5.2.1 MRI Data

The data consisted of short-axis stacks of LGE CMR covering the left ventricle, derived from 10 men (age range, 58 - 70 years old) suffering chronic myocardial infarction (MRI scan at six months after infarction). All patients gave written informed consent and the study was approved by the Medical Ethical Committee of our hospital (Hospital Clínico Universitario de Valencia, Valencia, Spain). Imaging was performed in breath-hold using a 1.5T MRI scanner (Sonata Magnetom, Siemens, Erlangen, Germany). LGE CMR sequences were acquired after 10 min of gadolinium injection using segmented inversion recovery sequences, cancelling out the signal from the healthy myocardium (repetition time/echo time: 2.5/1.2 ms, flip angle: 45°, voxel size 1.56 × 1.56 × 7 mm).

5.2.2 Region Labeling

The left myocardium and infarct were manually segmented by mutual consent of two expert cardiologists with more than 10 years of experience. A small region of non-infarcted myocardium opposite to the infarct was labeled as remote myocardium to avoid possible overlapping between classes. Remote myocardium was not labeled in some slices due to considerable infarct size. The number of image slices showing a well-defined infarct varied from 4 to 9 depending on the patient. In total, 119 regions were identified: 61 as infarct and 58 as remote myocardium.

5.2.3 Feature Extraction

Feature extraction was performed using the MaZda software, version 4.6 (Institute of Electronics, Technical University of Lodz, Lodz, Poland) [137]. Feature maps were generated for each segmented slice by using a small window (5 × 5 pixels) sliding over the image. Texture features are not pixel independent but using a 1-pixel step for the sliding window was possible to obtain the feature information for each pixel in its surrounding neighborhood. Only pixels with surrounding neighborhood completely inside the infarct or the remote myocardium were used as training vectors, otherwise information corresponding to other regions would be added.

Before computation of texture features and to avoid variability due to intrinsic differences between images, the image intensity was encoded to 8 bits/pixel for all images and pixels inside each window were normalized to the intensity mean ± 3 standard deviations. A total of 122 features derived from the following methods were extracted: histogram (9 features), GLCM (88 features), GLRLM (20 features), autoregressive model (5 features) [99].

5.2.4 Feature Selection

A feature selection step was necessary to keep only the most relevant features and specially to reduce the computation time associated with the generation of feature maps. The multiple SVM-RFE feature selection was implemented [130] using 10-fold cross-validation with 10 repetitions. Thus, at each step of the multiple SVM-RFE $n = 100$ linear SVMs were trained with a fixed value of the C parameter. The parameter $C = 1$ was determined beforehand from the finite set $C \in \{10^{-4}, \dots, 10^{0}, \dots, 10^{4}\}$ on the basis of the highest classification accuracy estimated with the resampling method

when including the full set of features.

Prior to feature selection, the vector data (2976 pixels, 1814 remote myocardium and 1162 infarcts) was split into half training and half testing, so the feature selection process was carried out solely on the training set. The ranked features returned by the multiple SVM-RFE were progressively added one by one from most to least important. Each feature subset was then used to train a linear SVM and to predict the outcome on the test set. The plot of accuracy vs. size of feature subsets on the test set is shown in Figure 5.2. Classification accuracy improved when the feature dimension was incremented but at certain point stabilized around 0.95. Taking the first 17 features (Table 5.1), an accuracy of 0.944 was obtained, so we decided to use these features to train the final classifier model.

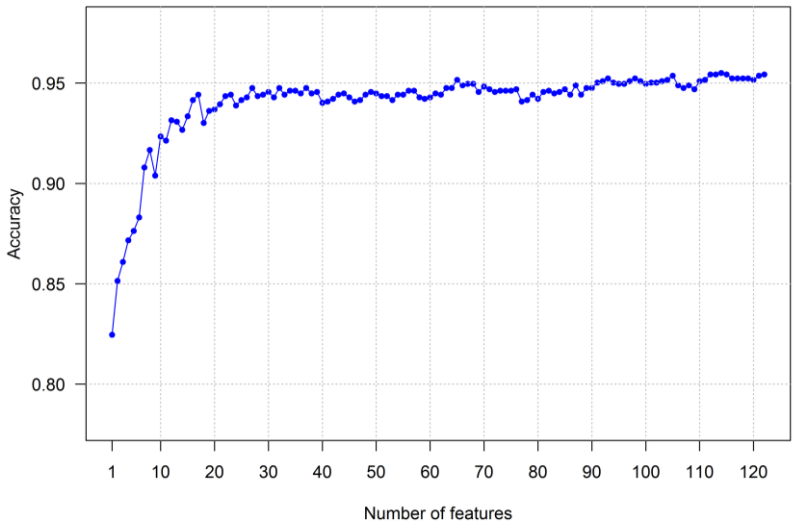

Figure 5.2. *Accuracy values on test set for different subsets of features ranked by the multiple SVM-RFE feature selection technique.*

5.2.5 SVM Training

The entire vector data available along with the selected features were used to train a nonlinear SVM classifier with Gaussian radial basis function (RBF) kernel using the Kernlab package [138] in R language version 3.0.1 (R Development Core Team, Vienna, Austria). The kernel parameter σ was set to a value between the 0.1 and 0.9

quantile of the $\|x - x'\|$ statistics calculated with the sigest function in the Kernlab package. The SVM C parameter was chosen from $C \in \{10^{-4}, \ldots, 10^{0}, \ldots, 10^{4}\}$ according to the highest classification accuracy assessed with 10-fold cross-validation with 10 repetitions. The chosen parameters values were: $\sigma = 0.045, C = 10$.

The final classifier model was implemented in Matlab 2014b (MathWorks Inc., Natick, MA) to segment new images provided with prior segmentation of the entire myocardium as the method limits the search inside this region. Pixels near myocardial boundaries which neighborhood lies outside the myocardium might cause small areas segmented as infarct due to the inclusion of information from non-myocardial regions. To deal with this problem, a refined segmentation was obtained by performing a morphological opening to remove objects within an area smaller than a given threshold of 20 pixels.

5.3 Experiments and Results

To validate the performance of the segmentation method on new images regardless of the acquisition parameters, and for comparison reasons, we decided to use images taken from the dataset provided by the STACOM challenge at MICCAI 2012 [139]. Five human LGE CMR with ground truth segmentations of the left ventricular myocardium and the corresponding infarct were available. These five human cases were used to evaluate the segmentation by comparing the results with the given ground truth. All cases had a history of myocardial infarction of at least three months prior their MRI scan. The images were acquired in breath-hold on a 1.5T MRI scanner (Achieva, Philips, Best, The Netherlands). Image acquisition parameters were: repetition time/echo time: 3.4/2.0 ms, flip angle: 25º, voxel size 1.8 × 1.8 × 8 mm.

The segmentation results were compared with the ground truth by calculating the degree of area overlap measured by the Dice coefficient, which is given by equation 5.1, where A and B are the two segmentations for comparison. Dice values vary between 0 and 1. Higher Dice indicates better agreement between the segmentation and the ground truth. Average Dice coefficients over slices per case are shown in Table 5.2.

$$\text{Dice} = \frac{2 * \|A \cap B\|}{\|A\| + \|B\|} \tag{5.1}$$

Table 5.1. *Ranked features used to train the final SVM classifier.*

Method	Feature	Rank
Histogram	Percentile 1%	1
GLCM	Sum variance (1 pixel, horizontal)	2
Histogram	Skewness	3
GLCM	Sum variance (1 pixel, vertical)	4
GLCM	Sum variance (1 pixel, 45º)	5
GLCM	Correlation (2 pixels, vertical)	6
GLCM	Contrast (1 pixel, vertical)	7
GLCM	Sum variance (2 pixels, horizontal)	8
GLCM	Difference variance (1 pixel, 135º)	9
GLRLM	Run percentage (horizontal)	10
GLRLM	Run length non-uniformity (horizontal)	11
GLCM	Contrast (1 pixel, horizontal)	12
Histogram	Variance	13
GLCM	Sum of squares (2 pixels, 45º)	14
Autoregressive model	θ4	15
GLRLM	Run length non-uniformity (135º)	16
GLRLM	Long run emphasis (135º)	17

Table 5.2. *Dice coefficients shown on test cases.*

Case	1	2	3	4	5	Overall
Mean	0.69	0.59	0.59	0.76	0.72	0.71
SD	(0.23)	(0.06)	(0.11)	(0.04)	(0.08)	(0.12)

Figure 5.3 shows the segmentation on a representative slice for each case with its associated Dice coefficient. The first two images show lower contrast than the rest, in consequence the Dice coefficients were also lower. In general, the infarcted areas segmented by the proposed method were smaller than the ground truth in the images with low contrast, and larger in those with high contrast. When the infarct was located near the myocardial boundaries, the segmented images showed an inclusion of the boundary pixels as can be seen in the third column of Figure 5.3. This situation is a limitation of the method since pixels on the boundaries tend to capture texture information belonging to the adjacent region.

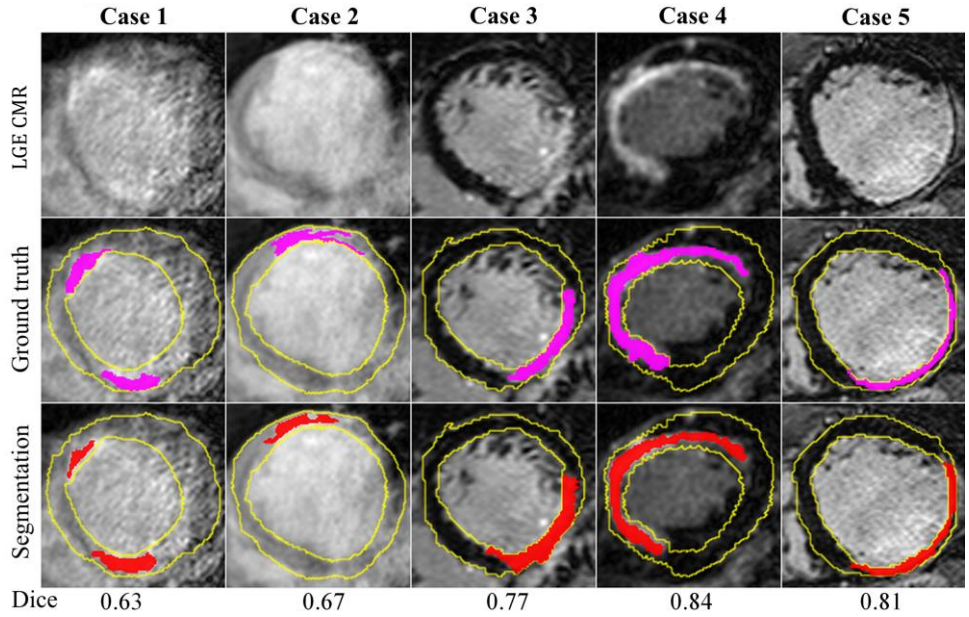

Figure 5.3. *Segmentation results (bottom row) for a representative slice per case.*

As expected, histogram features were important for detection of infarcted myocardium. Three histogram features were selected for the final model including the first ranked feature (percentile 1%). The remaining 14 features included in the final model were derived from texture analysis methods, most of them from GLCM, indicating the important discrimination properties of texture features. The model was trained on a dataset with different acquisition parameters, obtaining an overall good segmentation and therefore good generalization. Therefore, these results indicate the cross-center transferability of texture analysis but it has to be validated on larger datasets.

The results are in accordance with those proposed at the left ventricle LGE CMR segmentation challenge at MICCAI 2012, were obtaining a Dice of 0.8 or above in human scans was challenging for all submissions [139]. A direct comparison with the methods presented in the challenge was not feasible as those were compared using testing cases segmented by three raters and using the STAPLE method to obtain ground truth segmentations.

5.4 Conclusion

This chapter presented a method for detection of infarcted myocardium in LGE CMR, by training an SVM classifier with a subset of features derived from the image histogram and from texture analysis methods. The results obtained on a validation data from the STACOM segmentation challenge at MICCAI 2012 showed good overlap with the ground truth segmentation. Texture features showed important discriminatory properties for detection of infarcted myocardium. This was a preliminary evaluation using small sample data from two different centers, therefore further validation is needed.

Chapter 6

Differentiation between Acute and Chronic Myocardial Infarction

Part of this chapter was published in:

Larroza A, Materka A, López-Lereu MP, Monmeneu JV, Bodí V, Moratal D.
Differentiation between acute and chronic myocardial infarction by means of texture analysis of late gadolinium enhancement and cine cardiac magnetic resonance imaging.
European Journal of Radiology. 2017; 92: 78-83.

6.1 Introduction

The problem of differentiating between acute (AMI) and chronic myocardial infarction (CMI) is of clinical importance for patient treatment and follow-up in cases of preexisting CMI and limited possibility of localizing the acute lesion by means of ECG or coronary angiography [140]. Previous studies reported the capability of other MRI sequences, such as T2-weighted edema CMR [141] and contrast-enhanced cine CMR [142], for differentiation of both entities but technical limitations still exist [143]. Therefore, alternatives to address this problem are open to investigation.

Late gadolinium enhancement (LGE) is the well-established sequence for detection and evaluation of myocardial infarction [4]. However, there is no difference in the hyperenhancement of the infarcted myocardium in LGE CMR regarding the infarct's age [140]. The presence of edema in AMI and fibrosis in CMI characterize the most important corresponding pathologic changes of the heart. They have an effect on the internal structure of the tissue. This suggests there may be an inherent texture difference between images of the tissues affected by AMI and CMI.

Significant differences between AMI and CMI were found in animal models using texture analysis in LGE CMR [25] but, to our knowledge, applications to human studies and other MRI sequences have not been reported. Therefore, in this experimental study we investigated the capability of texture analysis of LGE CMR to

differentiate AMI from CMI. We also studied the possibility to address this problem using standard cine CMR solely. Infarctions are visually imperceptible in most cine CMR images and we believe that necrotic and/or fibrotic areas can be enhanced using texture analysis. Figure 6.1 shows examples of cine and LGE CMR in acute and chronic stage.

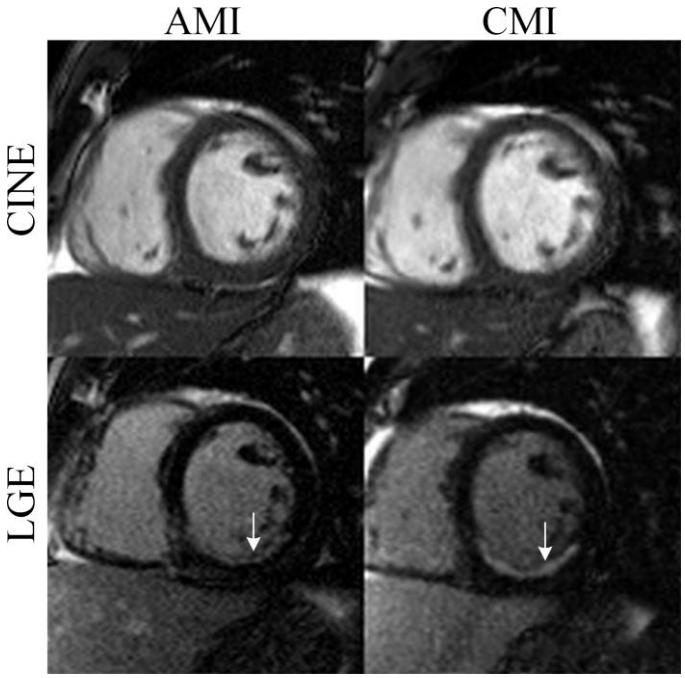

Figure 6.1. *Representative slice of a patient with myocardial infarction in acute (AMI) and chronic (CMI) stage. Infarcted areas are hyperenhanced in LGE CMR (white arrows) but these are imperceptible in the corresponding cine images.*

6.2 Materials and Methods

6.2.1 Study Group and Imaging Protocol

This retrospective study included 44 cases: 22 with AMI and 22 with CMI diagnosis. These cases correspond to 27 patients: 17 of them had MRI scans at one week (AMI) and six months (CMI) after a first ST-segment myocardial infarction, five patients had only AMI scans, and five had only CMI scans. Baseline characteristics of the study group are displayed in Table 6.1. All cases included only one type of

myocardial infarction, AMI or CMI, so the cases of a prior myocardial infarction were discarded on the basis of previous admissions due to cardiovascular diseases and/or electrocardiographic abnormalities.

All MRI examinations were performed using a 1.5T scanner (Sonata Magnetom, Siemens, Erlangen, Germany). Cine images were acquired at rest in 2-, 3-, and 4-chamber views and every 1 cm in short-axis views with steady-state free precession (SSFP) imaging sequences (repetition time/echo time: 3.2/1.6 ms; flip angle: 61°; voxel size: 1.56 × 1.56 × 7 mm). Late gadolinium enhancement (LGE) images were acquired 10 min after the administration of gadolinium at a dose of 0.075 mmol/kg body weight (gadobenato dimeglumine; MultiHance, Bracco, Italy). A segmented, inversion recovery steady-state free precession imaging sequence was used (repetition time/echo time: 2.5/1.2 ms, flip angle: 45°, voxel size: 1.56 × 1.56 × 7 mm), and inversion time was adapted to null the signal from the normal myocardium. Cine and LGE CMR in short-axis views were used for analysis and all slices of interest had identical location and orientation in space.

Table 6.1. *Baseline characteristics of the study group.*

	AMI	**CMI**	**Total**
Number of cases	22	22	44
Baseline characteristics			
Age (years)	61 ± 9	61 ± 10	61 ± 9
Male sex (%)	20 (90)	20 (90)	40 (90)
Diabetes mellitus (%)	5 (22)	7 (31)	12 (27)
Hypertension (%)	9 (40)	8 (36)	17 (38)
Dyslipidemia (%)	8 (36)	7 (31)	15 (34)
Current Smoker (%)	5 (22)	4 (18)	9 (20)
Heart rate (beats per min)	82 ± 12	86 ± 12	84 ± 12
Systolic blood pressure (mm Hg)	144 ± 25	131 ± 24	138 ± 26
Magnetic resonance parameters			
LV Ejection fraction (%)	48 ± 15	51 ± 17	50 ± 16
LV End-diastolic volume index (ml/m^2)	92 ± 28	89 ± 30	91 ± 29
LV End-systolic volume index (ml/m^2)	50 ± 30	47 ± 32	49 ± 31
LV mass (g/m^2)	81 ± 19	77 ± 21	79 ± 20
Infarct size (% of LV mass)	24 ± 11	22 ± 10	23 ± 11

*Continuous variables are expressed as mean ± standard deviation.

6.2.2 Region of Interest Definition

Images were segmented using the freely available software Segment, version 2.0 (Medviso, Lund, Sweden) [144]. The left ventricular myocardium was manually segmented on short-axis cine images at end-diastole. The myocardium segmentation was then exported to the corresponding LGE views. Infarcted myocardium regions were identified in LGE images using the FWHM technique [46].

Texture analysis was performed independently in LGE and cine CMR. Regions of interest (ROIs) for computing texture features were defined as the infarcted myocardium areas in LGE CMR, and the entire myocardium in cine CMR, since delineation of infarcted areas are challenging in this image modality (Figure 6.2). Only slices for which the infarcted area in LGE images was larger than 2 cm^2 were chosen for the analysis, resulting in 92 AMI and 97 CMI slices for each MRI modality.

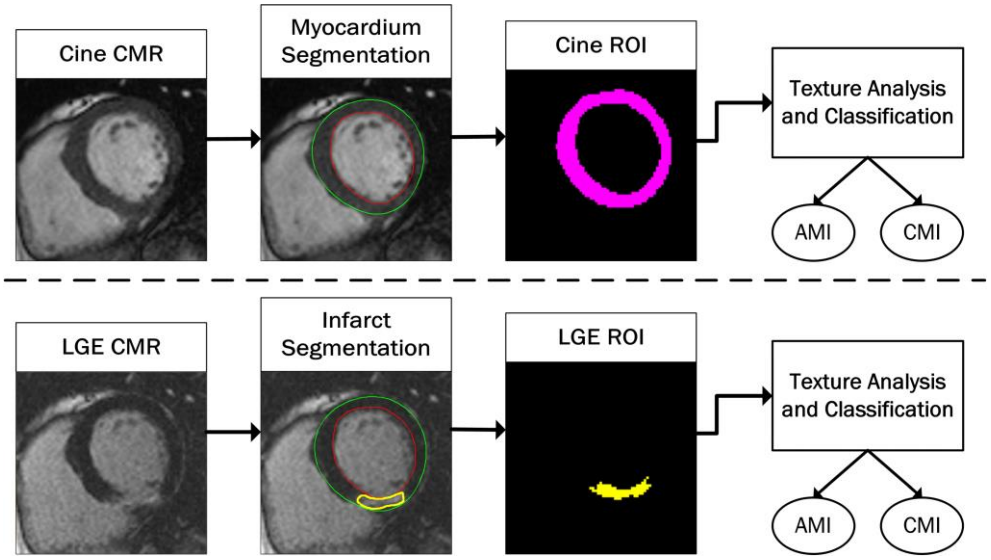

Figure 6.2. *Workflow of the texture-based method to differentiate AMI from CMI. Cine and LGE CMR were analyzed independently. Texture features were extracted from predefined ROIs: the left ventricular myocardium in cine CMR, and the infarcted area in LGE CMR. The supervised classification was evaluated with nested cross-validation.*

6.2.3 Region of Interest Preprocessing

To ensure that the outcome of the analysis truly characterizes the ROI texture, some image preprocessing was implemented. Firstly, the original images were interpolated to a pixel size of 0.5 × 0.5 mm^2 using the cubic B-spline method in order to invoke the heterogeneous information that may have the potential to improve the texture-based classification [82]. Subsequently, inhomogeneity correction was applied throughout the ROI using the additive model described by [88]. The latter procedure is critical especially for large ROIs as of cine CMR. Lastly, ROI normalization was performed to minimize the influence of image contrast and brightness variation, by using a method that remaps the ROI histogram to fit within the intensity mean ± 3 standard deviations [84].

6.2.4 Texture Feature Extraction

Texture features were computed using the MaZda software, version 4.6 (Institute of Electronics, Technical University of Lodz, Lodz, Poland) [120]. A total of 279 features were calculated for each ROI based on the six texture modeling and characterization methods available in MaZda that were defined in Section 3.5 and are summarized in Table 6.2. Matrix-based features were computed using 256 gray levels and in four directions: horizontal, vertical, 45º and 135º.

Table 6.2. *List of texture features used in this study.*

Method	Features	N
Histogram	Mean, variance, skewness, kurtosis, percentiles (1, 10, 50, 90, and 99%).	9
Absolute gradient	Gradient mean, variance, skewness, Kurtosis, non-zeros.	5
Gray-level Co-occurrence matrix	Angular second moment, contrast, correlation, sum of squares, inverse difference moment, sum average, sum variance, sum entropy, entropy, difference variance, difference entropy. Matrices were calculated for one to five inter-pixel distances in four directions: horizontal, vertical, 45º and 135º.	220
Gray-level Run-length matrix	Run-length non-uniformity, gray-level non-uniformity, long run emphasis, short run emphasis, and run percentage; in four directions: horizontal, vertical, 45º and 135º.	20
Autoregressive model	θ_1, θ_2, θ_3, θ_4, σ.	5
Wavelets	Subbands LL, LH, HL, and HH at five scales.	20

6.2.5 Feature Selection and Classification

Two feature selection techniques were implemented: one filter method based on the Fisher coefficient, and one wrapper method; the multiple support vector machine - recursive feature elimination (SVM-RFE). These feature selection techniques were described in Section 4.2. The multiple SVM-RFE was implemented using a repeated five-fold cross-validation to stabilize the feature selection, and the *C* parameter of the linear SVM was set to unity.

Both feature selection techniques return a ranking of features, so the optimal number of features should be chosen by comparing the classification performance of a certain model for different subsets of features. Six predictive models were tested using the *caret* package [145] in R language, version 3.3.0 (R Development Core Team, Vienna, Austria). These models were chosen due to their well-known performance in application to other datasets [131]:

- **K-Nearest Neighbors (KNN):** using the method *knn* in the caret package with 11 neighbors in the range 1:2:21.
- **Artificial Neural Network (ANN):** using the method *nnet* in the caret package, training a multilayer perceptron with a single hidden layer. Tuning parameter *size* (of the hidden layer) with values {1, 3, 5, 7} and tuning parameter weight *decay* with values {0, 0.1, 0.01, 0.0001}, at different random weight initializations.
- **Random Forest (RF):** using the method *rf* in the caret package with *ntree* = 500 and tuning the parameter *mtry* chosen from {2, 4, 8, 10}.
- **SVM-linear:** using the method *svmLinear* in the caret package with linear kernel tuning parameter *C* in the range $\{2^{-2} \ldots 2^{4}\}$.
- **SVM-radial:** the SVM with Gaussian kernel using the method *svmRadial* in the caret package. The tuning parameters *C* and *sigma* were chosen from the finite set $\{2^{-2} \ldots 2^{4}\}$ and $\{10^{-2} \ldots 10^{2}\}$ respectively.
- **SVM-poly:** the SVM using a cubic kernel (*degree* = 3) using the method *svmPoly* in the caret package. The tuning parameters *C* and *scale* were chosen from the finite set $\{2^{-2} \ldots 2^{4}\}$ and $\{10^{-2} \ldots 10^{-2}\}$ respectively.

6.2.6 Model Evaluation

To evaluate their generalization performance, the different models were tested using the nested cross-validation approach illustrated in Figure 6.3. Basically, the dataset was split into training and testing sets $n = 50$ times by using a repeated 5-fold cross-validation. For each training set, a feature ranking was obtained and the classification performance was tested for different number of features that were progressively added one by one according to the ranking. An inner 5-fold cross-validation was used for tuning the parameters of the classifier model. Using this approach, we avoid any possible over-fitting because everything, including the feature selection, was performed in a separate training set. Care was taken to ensure that samples from the same patient were either in the training or in the testing set at each loop, thus avoiding another source of bias. The same data partition was used for each of the six classifiers in order to obtain a fair comparison among them.

Feature values were standardized to zero mean and unit variance across training samples. Area under the curve (AUC) of the receiver operating characteristic (ROC) was used as an index of classification accuracy. The class probabilities estimated by each model were used to plot average ROC curves. Sensitivity and specificity were computed for the cutoff on the ROC curve that maximizes the product of both measures. Student *t*-test was used for comparison among classification models. Results were reported with 95% confidence intervals and statistical significance was set at $p < 0.05$. Bonferroni correction was applied when multiple comparisons were performed.

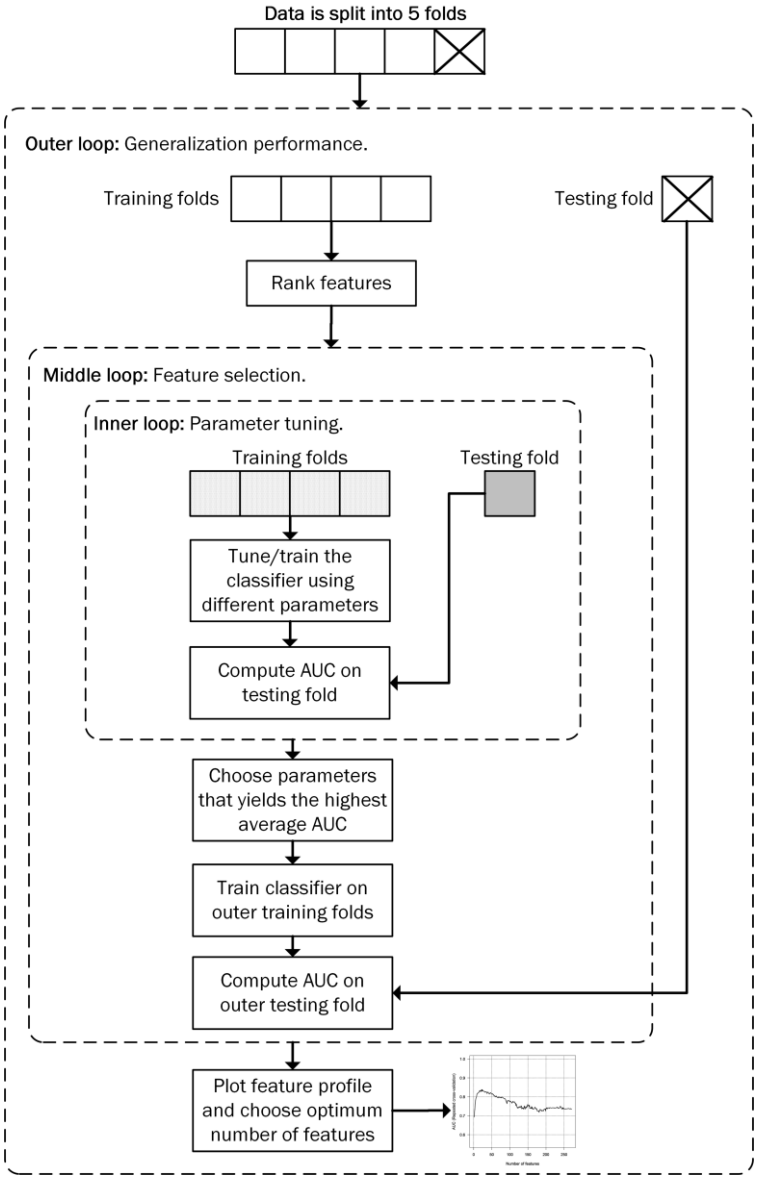

Figure 6.3. *Nested cross-validation procedure. The outer loop is used to estimate the generalization performance on the test fold, the process is repeated 50 times (5 folds x 10 repetitions). The middle loop is used to test different subset of features that are added one by one according to feature ranking. The 5-fold cross-validation inner loop is used to tune the classifier using a predefined set of parameters according to the model.*

6.3 Results

6.3.1 LGE CMR

For each training set in the nested cross-validation, the feature selection algorithm returned feature rankings and these were averaged over resamples to obtain a single ranking. The top ten features for each feature selection technique are shown in Table 6.3. Noticeably, the features ranked by both feature selection techniques were very different. The top ten features ranked by the Fisher technique were derived from the gray-level co-occurrence matrix (GLCM) and most of them correspond to the feature *difference entropy* at different directions and scales. One feature derived from the gray-level run-length matrix (GLRLM) was ranked second by the SVM-RFE but the remaining features were also from the GLCM. Scatter plots for the entire LGE CMR dataset using the best two features are shown in Figure 6.4.

In the Fisher ranking, the average rank value for each feature was very similar to its position in the ranking thus indicating that the Fisher technique tended to select the same feature at each resample. In contrast, the SVM-RFE showed higher average rank values indicating that different features were selected at each resample. This phenomenon can be appreciated by looking at the resampling profiles in Figure 6.5; using only one feature the AUC values in the Fisher profiles are around 0.8 whereas in the SVM-RFE profiles are below 0.7. Nevertheless, the SVM-RFE profiles are smoother and show that after achieving a peak value it stabilizes, so it may give more reliable information when testing the generalization performance.

Optimal classification models were selected according to the number of features that yielded the highest AUC value. An interesting finding was that pairwise comparisons did not show statistical significant differences among models (Table 6.4), thus suggesting that the choice of the classifier is not as relevant as the feature selection is. Paired *t*-test between the reduced set and full set of feature showed that in most cases the reduced set outperformed significantly the results of the full set of features ($p < 0.05$). This improvement was more noticeable for the KNN and SVM-radial models (Table 6.5).

Table 6.3. *Top ten ranked features in LGE CMR.*

Fisher - LGE CMR	**Rank**	**SVMRFE - LGE CMR**	**Rank**
Difference entropy (3 pixels, 0º)	1.8	Difference variance (5 pixels, 135º)	24.0
Difference entropy (5 pixels, 0º)	1.9	Run percentage (45º)	28.2
Difference entropy (2 pixels, 0º)	2.8	Sum of squares (4 pixels, 135º)	32.5
Difference entropy (1 pixel, 0º)	3.8	Sum variance (1 pixel, 135º)	36.7
Difference entropy (3 pixels, 45º)	4.6	Angular 2nd moment (4 pixels, 135º)	37.0
Difference entropy (4 pixels, 0º)	6.0	Angular 2nd moment (3 pixels, 135º)	37.7
Inverse diff. moment (4 pixels, 0º)	8.8	Sum variance (1 pixel, 45º)	38.5
Difference entropy (4 pixels, 45º)	9.3	Angular 2nd moment (2 pixels, 135º)	39.5
Difference entropy (2 pixels, 45º)	9.4	Sum average (3 pixels, 45º)	39.9
Inverse diff. moment (5 pixels, 0º)	10.4	Sum of squares (1 pixel, 45º)	41.5

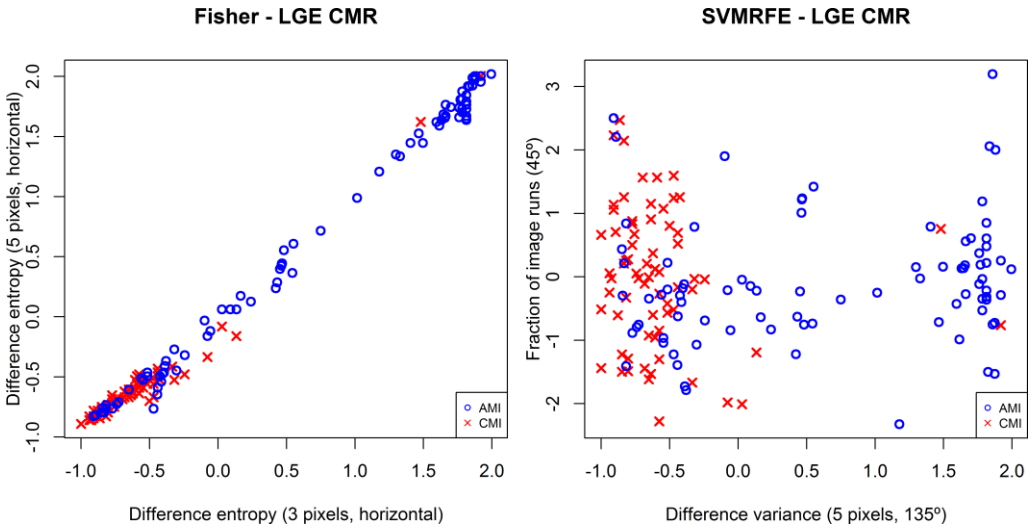

Figure 6.4. *Scatter plots for the entire LGE CMR dataset using the top two features according to the average ranking returned by the Fisher and SVM-RFE feature selection techniques. AMI: acute myocardial infarction, CMI: chronic myocardial infarction.*

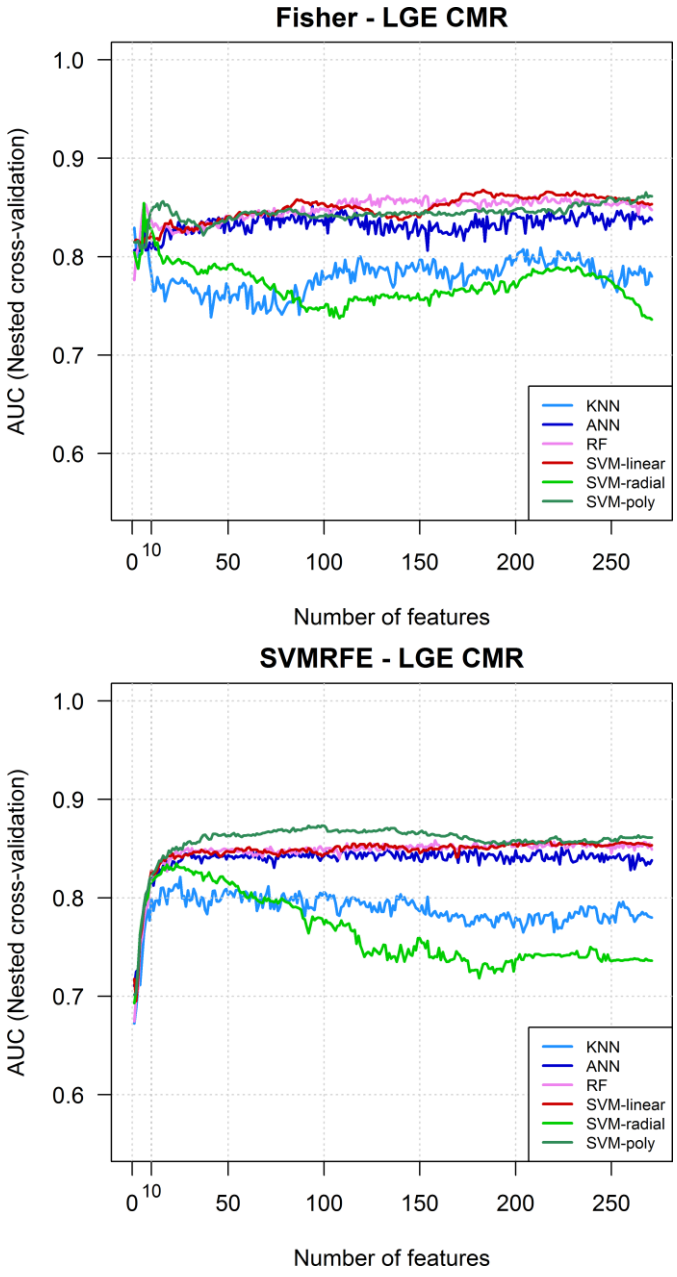

Figure 6.5. *Resampling profiles for different classification models in LGE CMR. Average AUC values over resamples of nested cross-validation for different subset of features added according to the ranking returned by the Fisher and SVM-RFE feature selection techniques.*

Good classification (AUC > 0.8) was obtained for the optimal models driven by the selected features. The best feature subsets were used to compute the classification performance parameters (Table 6.6) and to plot the ROC curves (Figures 6.6 and 6.7). The SVM-poly with 99 features ranked by the SVM-RFE has the best estimate of performance (AUC = 0.87). However, based on the confidence intervals, this value is similar in performance to other models including RF and the SVM variants obtained with the Fisher ranking. All performance values fall within the confidence intervals suggesting that there are not reasons for concern with regards to potential over-fitting.

Sensitivity and specificity were computed according to the optimal operating point on the ROC curve that maximizes the product between both measures. Sensitivity measures the ability of the model to detect AMI, whereas specificity measures the ability to detect CMI. Sensitivity and specificity values were also higher than 0.8 for the best models.

Table 6.4. *Pairwise comparison among classification models in LGE CMR.*

	Fisher – LGE CMR					
	KNN	**ANN**	**RF**	**SVM-linear**	**SVM-radial**	**SVM-poly**
KNN	-	0.1404	0.0232	0.0088	0.0944	0.0143
ANN	0.1404	-	0.4218	0.2475	0.8419	0.3249
RF	0.0232	0.4218	-	0.7233	0.5458	0.8559
SVM-linear	0.0088	0.2475	0.7233	-	0.3383	0.8631
SVM-radial	0.0944	0.8419	0.5458	0.3383	-	0.4322
SVM-poly	0.0143	0.3249	0.8559	0.8631	0.4322	-

	SVMRFE – LGE CMR					
	KNN	**ANN**	**RF**	**SVM-linear**	**SVM-radial**	**SVM-poly**
KNN	-	0.0243	0.0062	0.0075	0.3224	***0.0001***
ANN	0.0243	-	0.6224	0.6689	0.2045	0.1167
RF	0.0062	0.6224	-	0.9483	0.0787	0.2809
SVM-linear	0.0075	0.6689	0.9483	-	0.0902	0.2531
SVM-radial	0.3224	0.2045	0.0787	0.0902	-	0.0047
SVM-poly	***0.0001***	0.1167	0.2809	0.2531	0.0047	-

*Significance was corrected with the Bonferroni method $p = 0.05/15$, or $p = 0.0033$. Statistical significant p values are highlighted in bold.

Table 6.5. *Nested cross-validation results for the full and reduced sets in LGE CMR.*

	Full set	Fisher – LGE CMR			SVMRFE – LGE CMR		
Model	**AUC ± SD**	**Size**	**AUC ± SD**	***p* value**	**Size**	**AUC ± SD**	***p* value**
KNN	0.78 ± 0.07	6	0.83 ± 0.08	**0.000**	25	0.82 ± 0.07	**0.000**
ANN	0.83 ± 0.06	94	0.85 ± 0.05	0.149	157	0.85 ± 0.06	**0.036**
RF	0.84 ± 0.07	124	0.86 ± 0.07	**0.004**	202	0.85 ± 0.07	**0.004**
SVM-linear	0.85 ± 0.05	183	0.86 ± 0.06	0.089	217	0.85 ± 0.05	0.179
SVM-radial	0.73 ± 0.11	6	0.85 ± 0.07	**0.000**	23	0.83 ± 0.07	**0.000**
SVM-poly	0.86 ± 0.06	268	0.86 ± 0.06	0.226	99	0.87 ± 0.05	**0.014**

*The *p value* column corresponds to a *t*-test that evaluates whether the AUC values for the reduced model was larger than the value associated when using the full set of features. Statistical significance was set to $p < 0.05$.

Table 6.6. *Classification performance parameters in LGE CMR.*

Fisher - LGE CMR

Model	**N features**	**AUC**	**CI**	**Sensitivity**	**CI**	**Specificity**	**CI**
KNN	6	0.831	(0.80 - 0.85)	0.732	(0.69 - 0.76)	0.848	(0.81 - 0.88)
ANN	94	0.851	(0.83 - 0.86)	0.819	(0.79 - 0.84)	0.826	(0.80 - 0.85)
RF	124	0.862	(0.84 - 0.88)	0.819	(0.79 -0.84)	0.861	(0.83 - 0.89)
SVM-linear	183	0.867	(0.85 - 0.88)	0.846	(0.81 - 0.87)	0.84	(0.81 - 0.86)
SVM-radial	6	0.854	(0.83 - 0.87)	0.789	(0.75 - 0.82)	0.865	(0.83 - 0.89)
SVM-poly	268	0.865	(0.84 - 0.88)	0.831	(0.80 - 0.85)	0.851	(0.82 - 0.87)

SVMRFE - LGE CMR

Model	**N features**	**AUC**	**CI**	**Sensitivity**	**CI**	**Specificity**	**CI**
KNN	25	0.821	(0.79 - 0.84)	0.73	(0.69 - 0.76)	0.835	(0.80 - 0.86)
ANN	157	0.851	(0.83 - 0.86)	0.789	(0.75 - 0.81)	0.859	(0.83 - 0.88)
RF	202	0.858	(0.83 - 0.87)	0.8	(0.77 - 0.82)	0.865	(0.84 - 0.88)
SVM-linear	217	0.857	(0.84 - 0.87)	0.811	(0.78 - 0.83)	0.829	(0.80 - 0.84)
SVM-radial	23	0.834	(0.81 - 0.85)	0.795	(0.76 - 0.82)	0.866	(0.84 - 0.88)
SVM-poly	99	0.873	(0.85 - 0.88)	0.82	(0.79 - 0.84)	0.848	(0.82 - 0.87)

*Sensitivity and specificity were computed according to the optimal operating point on the ROC curve that maximizes the product between both measures. Sensitivity measures the ability of the model to detect acute myocardial infarction, whereas specificity measures the ability to detect chronic myocardial infarction. The *CI* column corresponds to 95% confidence interval.

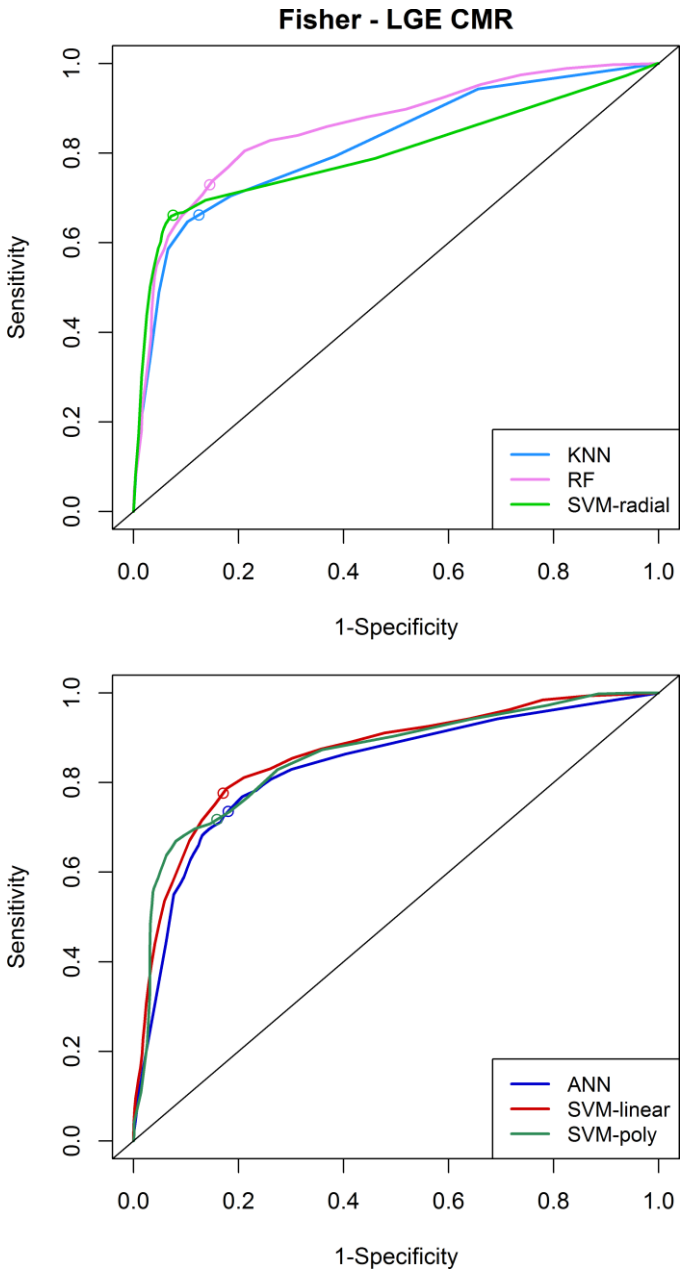

Figure 6.6. *Average ROC curves over test sets in LGE CMR using the features selected by the Fisher technique. The circles on the curves indicate the optimal operating point that maximizes the product between sensitivity and specificity.*

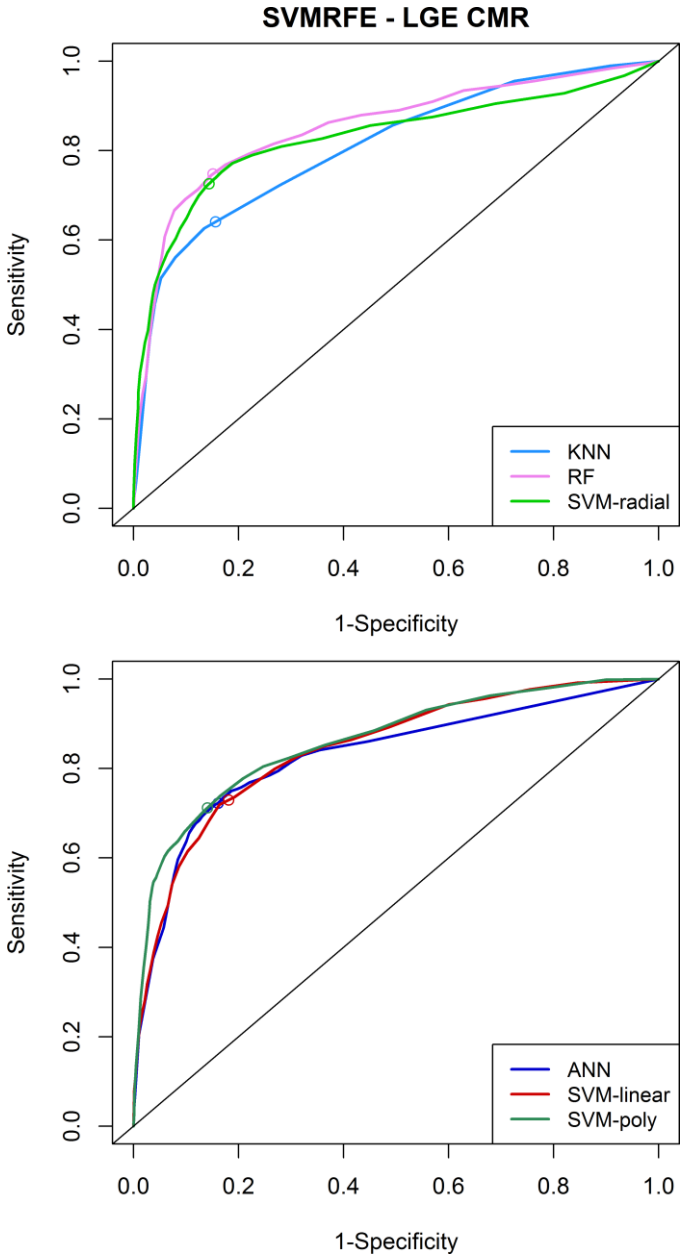

Figure 6.7. *Average ROC curves over test sets in LGE CMR using the features selected by the SVM-RFE technique. The circles on the curves indicate the optimal operating point that maximizes the product between sensitivity and specificity.*

6.3.2 Cine CMR

The averaged feature rankings obtained with the cine CMR dataset are shown in Table 6.7. As in LGE CMR, the features ranked by both feature selection techniques were also very different. In cine CMR, most of the top ten features ranked by Fisher were derived from the Wavelets method whereas the tendency of the SVM-RFE was to rank GLCM features first. Scatter plots for the entire cine CMR dataset using the best two features are shown in Figure 6.8.

Table 6.7. *Top ten ranked features in cine CMR.*

Fisher - cine CMR	**Rank**	**SVMRFE - cine CMR**	**Rank**
Wavelet HL (scale 3)	1.1	Sum average (1 pixel, 90º)	12.8
Wavelet HL (scale 1)	2.2	Inverse diff. moment (4 pixels, 0º)	14.6
Wavelet HL (scale 2)	4.2	Wavelet HL (scale 1)	16.7
Wavelet LL (scale 4)	5.0	Difference variance (5 pixels, 90º)	23.9
Wavelet LL (scale 3)	7.4	Sum average (5 pixels, 45º)	27.7
Wavelet LL (scale 2)	8.1	Difference entropy (1 pixel, 45º)	36.9
Difference variance (1 pixel, 0º)	13.2	Difference entropy (2 pixels, 135º)	38.7
Wavelet HH (scale 4)	17.5	Sum average (3 pixels, 0º)	39.3
Difference variance (2 pixels, 0º)	18.0	Sum entropy (1 pixel, 0º)	39.6
Correlation (1 pixel, 0º)	20.8	Contrast (5 pixels, 90º)	41.6

As in the case of LGE CMR, Fisher average rank values for each feature were very similar to its position in the ranking, whereas the SVM-RFE showed higher average rank values indicating that different features were selected at each resample. However, the resampling profiles show that in general the SVM-RFE achieves a peak AUC value and then stabilizes, which indicates that the feature selection at each resample is more specific.

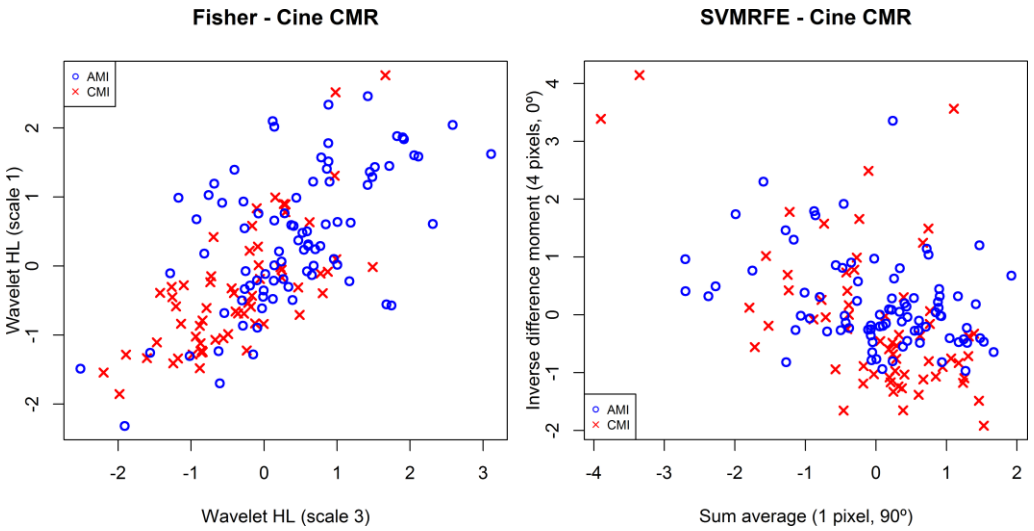

Figure 6.8. *Scatter plots for the entire cine CMR dataset using the top two features according to the average ranking returned by the Fisher and SVM-RFE feature selection techniques. AMI: acute myocardial infarction, CMI: chronic myocardial infarction.*

Paired *t*-test between the reduced and full set of features showed that the SVM-RFE improved significantly the results of the full set of features for most classifiers ($p < 0.05$). This improvement was only noticeable for RF and SVM-radial regarding the Fisher technique (Table 6.8). Pairwise comparisons indicate that KNN had a poor performance in relation to the other models (Table 6.9). The best feature subsets were used to compute the classification performance parameters (Table 6.10) and to plot the ROC curves (Figures 6.10 and 6.11). The SVM-linear with 22 features ranked by the SVM-RFE has the best estimate of performance (AUC = 0.831). Based on the confidence intervals, this value is similar in performance to other models including ANN and SVM-poly. Interestingly, SVM-RFE results outperformed those of Fisher.

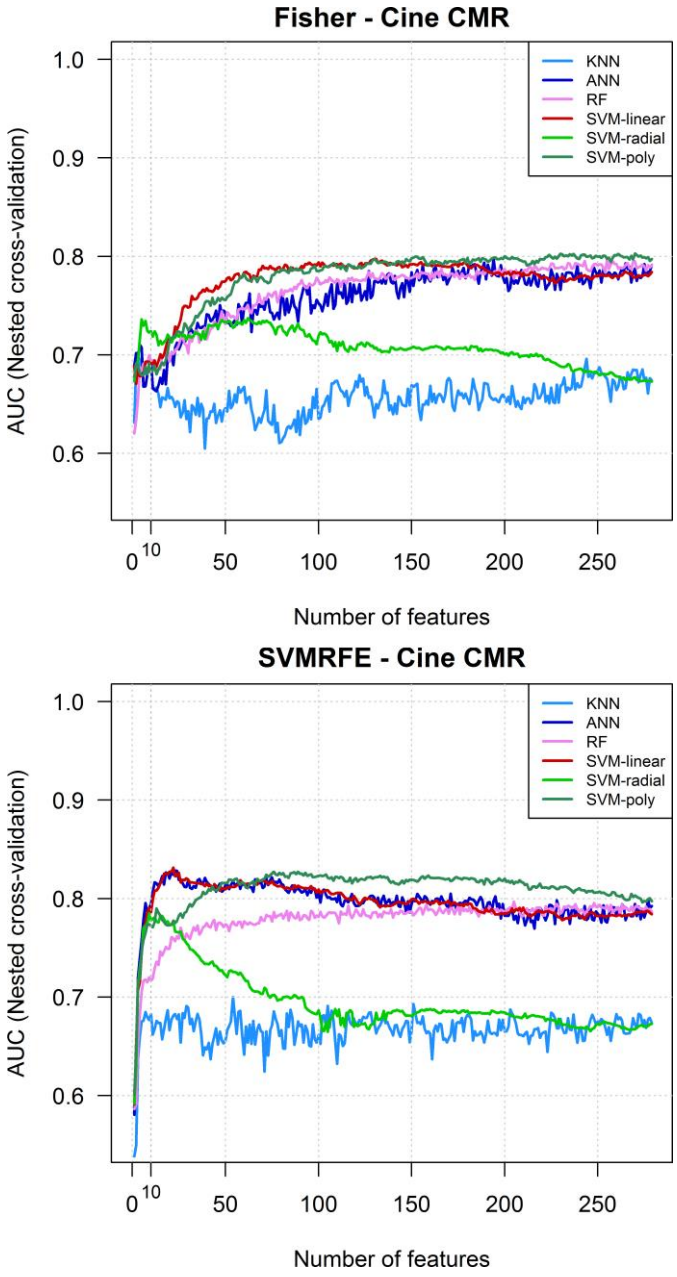

Figure 6.9. *Resampling profiles for different classification models in cine CMR. Average AUC values over resamples of nested cross-validation for different subset of features added according to the ranking returned by the Fisher and SVM-RFE feature selection techniques.*

Table 6.8. *Nested cross-validation results for the full and reduced sets in cine CMR.*

	Full set	**Fisher - Cine CMR**			**SVMRFE - Cine CMR**		
Model	**AUC ± SD**	**Size**	**AUC ± SD**	***p* value**	**Size**	**AUC ± SD**	***p* value**
KNN	0.68 ± 0.08	3	0.71 ± 0.11	0.154	54	0.70 ± 0.09	0.345
ANN	0.79 ± 0.06	194	0.79 ± 0.08	0.535	24	0.82 ± 0.07	**0.000**
RF	0.78 ± 0.07	255	0.79 ± 0.07	**0.008**	205	0.79 ± 0.07	**0.026**
SVM-linear	0.78 ± 0.07	131	0.79 ± 0.07	0.111	22	0.83 ± 0.07	**0.000**
SVM-radial	0.61 ± 0.16	62	0.73 ± 0.08	**0.000**	9	0.78 ± 0.09	**0.000**
SVM-poly	0.79 ± 0.07	270	0.80 ± 0.06	0.165	76	0.82 ± 0.07	**0.003**

*The *p value* column corresponds to a *t*-test that evaluates whether the AUC values for the reduced model was larger than the value associated when using the full set of features. Statistical significance was set to $p < 0.05$.

Table 6.9. *Pairwise comparison among classification models in cine CMR.*

Fisher - Cine CMR

	KNN	**ANN**	**RF**	**SVM-linear**	**SVM-radial**	**SVM-poly**
KNN	-	***0.0000***	***0.0000***	***0.0000***	0.1184	***0.0000***
ANN	***0.0000***	-	0.9684	0.9886	***0.0003***	0.7546
RF	***0.0000***	0.9684	-	0.9797	***0.0003***	0.7247
SVM-linear	***0.0000***	0.9886	0.9797	-	***0.0003***	0.7438
SVM-radial	0.1184	***0.0003***	***0.0003***	***0.0003***	-	***0.0000***
SVM-poly	***0.0000***	0.7546	0.7247	0.7438	***0.0000***	-

SVMRFE - Cine CMR

	KNN	**ANN**	**RF**	**SVM-linear**	**SVM-radial**	**SVM-poly**
KNN	-	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***
ANN	***0.0000***	-	0.0583	0.8412	0.0075	0.9563
RF	***0.0000***	0.0583	-	0.0364	0.4285	0.0659
SVM-linear	***0.0000***	0.8412	0.0364	-	0.0041	0.7986
SVM-radial	***0.0000***	0.0075	0.4285	0.0041	-	0.0088
SVM-poly	***0.0000***	0.9563	0.0659	0.7986	0.0088	-

*Significance was corrected with the Bonferroni method $p = 0.05/15$, or $p = 0.0033$. Statistical significant *p* values are highlighted in bold.

Table 6.10. *Classification performance parameters in cine CMR.*

Fisher - Cine CMR

Model	N features	AUC	CI	Sensitivity	CI	Specificity	CI
KNN	3	0.711	(0.67 - 0.74)	0.697	(0.65 - 0.73)	0.671	(0.63 - 0.70)
ANN	194	0.797	(0.77 - 0.82	0.770	(0.74 - 0.79)	0.804	(0.77 - 0.83)
RF	255	0.797	(0.77 - 0.81)	0.773	(0.75 - 0.79)	0.767	(0.73 - 0.79)
SVM-linear	131	0.797	(0.77 - 0.81)	0.747	(0.72 - 0.77)	0.822	(0.79 - 0.85)
SVM-radial	62	0.737	(0.71 - 0.76)	0.722	(0.68 - 0.75)	0.760	(0.73 - 0.79)
SVM-poly	270	0.802	(0.78 - 0.82)	0.769	(0.74 - 0.79)	0.790	(0.76 - 0.81)

SVMRFE - Cine CMR

Model	N features	AUC	CI	Sensitivity	CI	Specificity	CI
KNN	54	0.700	(0.67 - 0.72)	0.650	(0.61 - 0.68)	0.690	(0.65 - 0.72)
ANN	24	0.828	(0.80 - 0.84)	0.792	(0.76 - 0.82)	0.816	(0.78 - 0.84)
RF	205	0.796	(0.77 - 0.81)	0.749	(0.72 - 0.77)	0.788	(0.75 - 0.81)
SVM-linear	22	0.831	(0.80 - 0.85)	0.812	(0.77 - 0.84)	0.803	(0.77 - 0.83)
SVM-radial	9	0.783	(0.75 - 0.81)	0.762	(0.73 - 0.79)	0.753	(0.72 - 0.78)
SVM-poly	76	0.827	(0.80 - 0.84)	0.790	(0.76 - 0.81)	0.806	(0.77 - 0.83)

*Sensitivity and specificity were computed according to the optimal operating point on the ROC curve that maximizes the product between both measures. Sensitivity measures the ability of the model to detect acute myocardial infarction, whereas specificity measures the ability to detect chronic myocardial infarction. The *CI* column corresponds to 95% confidence interval.

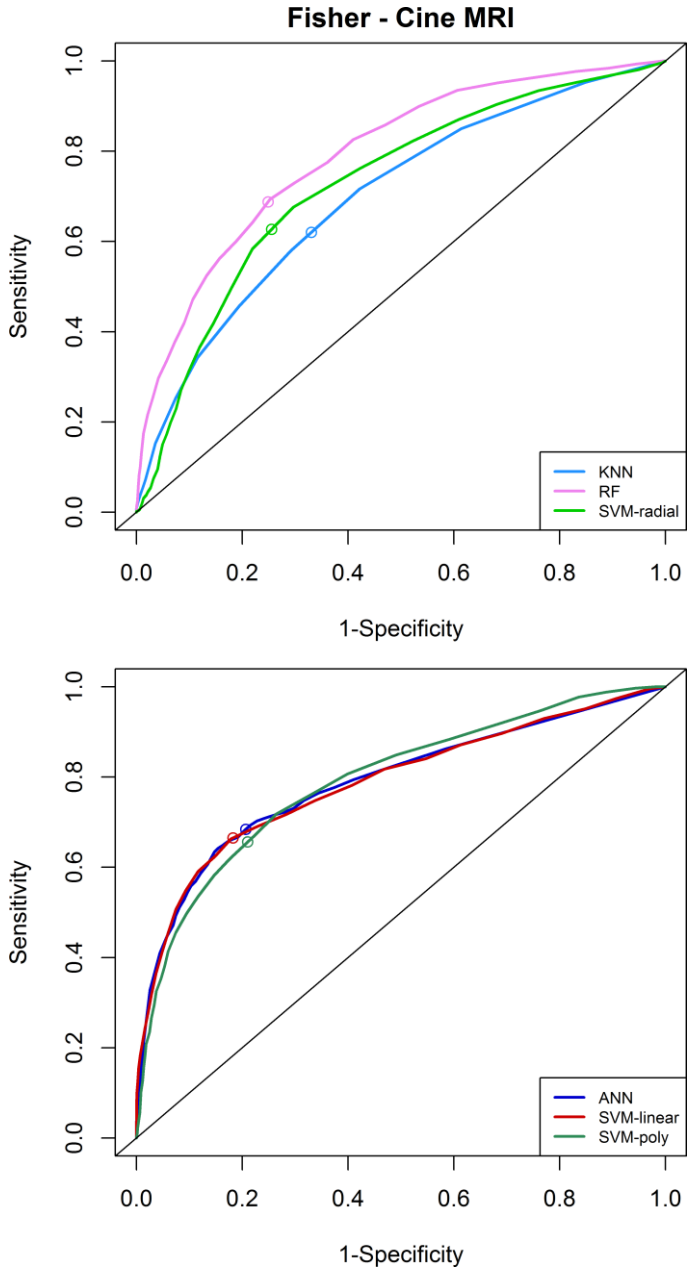

Figure 6.10. *Average ROC curves over test sets in cine CMR using the features selected by the Fisher technique. The circles on the curves indicate the optimal operating point that maximizes the product between sensitivity and specificity.*

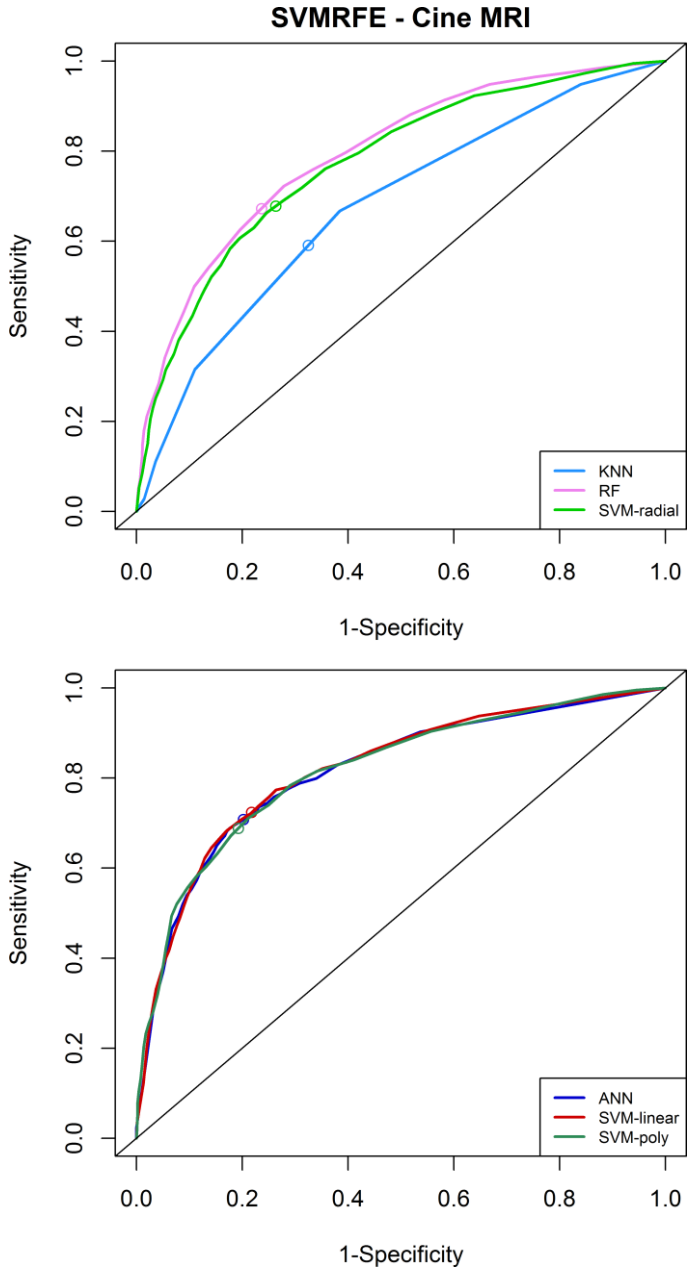

Figure 6.11. *Average ROC curves over test sets in cine CMR using the features selected by the SVM-RFE technique. The circles on the curves indicate the optimal operating point that maximizes the product between sensitivity and specificity.*

6.4 Discussion

We found that texture analysis has the potential to correctly differentiate AMI from CMI regions (AUC > 0.8) contained in LGE and cine CMR. However, the classification was not straightforward as no individual texture feature was capable of accurately differentiate samples of both classes. An appropriate subset of features that were selected and evaluated in combination with some predictive models provided good classification performance. We corroborated that feature selection is relevant to achieve optimal classification performance and by doing so the choice of classification algorithm is not of substantial importance.

The novel finding of this study was the possibility to discriminate slices with AMI and CMI in conventional cine CMR using the delineated myocardium as ROI, since myocardial infarction is not usually detectable in this image modality. It has been reported that regional lipomatous metaplasia are detectable in cine CMR as chemical shift artifacts along the sub-endocardium and are indicative of chronic myocardial infarction [146]. Another study reported the possibility to detect the acute-related edema in cine CMR [147]. These previous studies suggest that the infarction can be detected in cine CMR but the visual appreciation of images is challenging as intensity variations are not always visible in this image modality. Our approach of using the entire myocardium to predict the infarct age may provide more reproducible results and it is indicative that texture analysis can be used to detect infarctions in cine CMR.

Different MRI sequences were used in previous studies aiming at differentiating AMI from CMI. Kim et al. [142] stated that the signal intensity in contrast-enhanced cine CMR is useful for differentiating AMI from CMI. They found that the hyperenhancement in both contrast-enhanced cine CMR and LGE CMR was more accurate in differentiating AMI from CMI than the presence of myocardial thinning. Abdel-Aty et al. [141] claimed that T2-weighted cardiac MRI depicts infarct-related myocardial edema as a marker of AMI when combining LGE with T2-weighted CMR. In fact, T2-weighted CMR has been consensually adopted in clinical practice for detecting edema thus differentiating AMI from CMI [48]. However, there is an ongoing controversial debate about the technical limitations and the pathophysiologic background of conventional T2-weighted edema imaging, as reported by a recent study [143] that stated that using these techniques with simple thresholds did not facilitate the discrimination of AMI and CMI. Additionally, these techniques rely on

the visual assessment of the images.

In the study of [25], texture analysis of high resolution LGE CMR was used to differentiate AMI from CMI in rats. They computed only three texture features derived from the co-occurrence matrix in 3D and all of them were found statistically different. The authors found that the *entropy* texture feature was higher in AMI suggesting its textural complexity due to the presence of microvascular obstruction. Interestingly, our results are somehow in accordance with these previous findings as most of the top ten features ranked by the Fisher feature selection were related to the feature *difference entropy*.

A limitation of this study was that we included orientation dependent texture features, that may influence the results if the short-axis views are not acquired in a standardized position as we did. The inclusion of rotation invariant and 3D texture features may provide a more robust classification [9], [83]. The main limitation of our findings related to cine CMR is that we computed texture features from the entire myocardium rather than from specific segments with known infarcts that should be more clinically relevant. Nevertheless, these findings can be used as a starting point for future research. Differentiating AMI from CMI is clinically important when both infarction entities coexist, which complicates treatment planning and follow-up after treatment [140]. In our study, we included only patients having one type of infarction in order to have a reliable ground-truth for the experiments. Further validation in patients with coexisting AMI and CMI is required for future work.

6.5 Conclusion

Our results show that texture analysis can be used for differentiation of AMI from CMI in cardiac LGE CMR and also in standard cine CMR. Differentiation is not straightforward and requires the application of machine learning techniques in order to take advantage of conventional CMR images. The unexpected findings obtained using cine CMR motivated the study presented in the next chapter in which the objective was to detect visually imperceptible segments in cine CMR.

Chapter 7

Detection of Infarcted Myocardial Segments in Cine CMR

This chapter has been submitted in part as:

Larroza A, López-Lereu MP, Monmeneu JV, Gavara J, Chorro FJ, Bodí V, Moratal D.
Texture analysis on cardiac cine magnetic resonance imaging to detect non-viable segments in patients with chronic myocardial infarction.
Medical Physics [submitted].

7.1 Introduction

Typical MRI protocols for chronic ischemic heart disease and viability includes cine and LGE acquisitions. Cine modalities provide moving images of the heart and are mainly used to assess the contractile function, ventricle volumes and ejection fraction. LGE CMR is the well-established technique to quantify the extent of scar in patients with myocardial infarction but relies on the administration of gadolinium to enhance the scarred area [4]. Hence, LGE CMR is contraindicated in certain patients, i.e. those who have significant renal dysfunction, so gadolinium-free methods to identify diseased myocardium are open to investigation [148].

Histological properties of scarred myocardium differ from normal myocardium and in consequence, textural properties are also different. The fibrosis formation causes distortion of the normal myocardium architecture thus altering the texture properties. In the previous chapter, we have shown that texture analysis of conventional CMR can differentiate acute from chronic myocardial infarction. Moreover, a recent study [27] used texture analysis in standard pre-contrast cine CMR to study different etiologies of left ventricular hypertrophy.

In this study, we part from the hypothesis that infarcted myocardium can be enhanced by using texture analysis in conventional cine CMR, modality in which the affected myocardium is hardly detectable by visual inspection. The purpose was to

determine the capability of texture analysis of cine CMR to detect infarcted non-viable segments in patients with chronic myocardial infarction.

7.2 Materials and Methods

7.2.1 Study Group and Imaging Protocol

This retrospective study included 50 patients (mean age, 61; range, 23 – 80 years old) suffering chronic myocardial infarction examined between April 2012 and July 2015. The inclusion criteria were: 1) first ST-segment elevation myocardial infarction treated with percutaneous intervention within the first six hours after the onset of chest pain; 2) stable clinical course without complications during the first six months; 3) single-vessel disease and a patent (thrombolysis in myocardial infarction (TIMI) flow grade 3 and residual stenosis < 50%) in the infarct-related artery at the end of pre-discharge cardiac catheterization and at the sixth month.

MRI was performed at 179 ± 12 days after a first ST-segment myocardial infarction. Images were acquired in breath-hold using a 1.5T MRI scanner (Sonata Magnetom, Siemens, Erlangen, Germany). Cine images were acquired at rest in 2-, 3-, and 4-chamber views and every 1 cm in short-axis views with steady-state free precession imaging sequences (repetition time/echo time: 3.2/1.6 ms; flip angle: 61°; voxel size: 1.56 × 1.56 × 7 mm). Each short-axis cine sequence consisted of 35 frames. Late gadolinium enhancement (LGE) images were acquired 10 min after the administration of gadolinium at a dose of 0.075 mmol/kg body weight (gadobenate dimeglumine, Multihance, Bracco Diagnostics, Milan, Italy) in identical projections and locations to the cine images. A segmented, inversion recovery steady-state free precession imaging sequence was used (repetition time/echo time: 2.5/1.2 ms, flip angle: 45°, voxel size: 1.56 × 1.56 × 7 mm), and inversion time was adapted to null the signal from normal myocardium.

Left ventricle (LV) ejection fraction (%), LV end-diastolic volume index (ml/m^2), LV end-systolic volume index (ml/m^2) and LV mass (g/m^2) were calculated by manual planimetry of endocardial and epicardial borders in short-axis views cine images. Infarct size (% of LV mass) was assessed as the percentage of LV mass showing LGE. Baseline characteristics and MRI parameters are summarized in Table 7.1.

Table 7.1. *Baseline characteristics of the study group.*

Number of patients	50		
Baseline characteristics		**Magnetic resonance parameters**	
Age (years)	65 ± 12	LV Ejection fraction (%)	51 ± 16
Male sex (%)	45 (90)	LV End-diastolic volume index (ml/m^2)	86 ± 31
Diabetes mellitus (%)	13 (26)	LV End-systolic volume index (ml/m^2)	45 ± 29
Hypertension (%)	17 (34)	LV mass (g/m^2)	73 ± 19
Dyslipidemia (%)	16 (32)	Infarct size (% of LV mass)	21 ± 12
Current Smoker (%)	10 (20)		
Heart rate (beats per min)	82 ± 11		
Systolic blood pressure (mm Hg)	139 ± 26	**Number of segments**	
Infarct location		Non-viable segments (LGE ≥ 50%)	5 ± 3
Anterior (%)	38 (76)	Viable segments (0 < LGE < 50%)	3 ± 2
Inferior (%)	8 (16)	Remote segments	11 ± 5
Other (%)	4 (8)	Abnormal wall motion segments (Wall thickening ≤ 2 mm)	4 ± 3

*Continuous variables are expressed as mean ± standard deviation.

7.2.2 Region of Interest Definition

The left ventricular myocardium was manually segmented on short-axis views of both cine and LGE CMR using the freely available software Segment, version 2.0 (Medviso, Lund, Sweden) [144]. Then, the 17-segment AHA model [38] excluding the apex was used to divide the slices into four or six equal segments depending on the slice level. Various slices were used to create the three thick short-axis sections defined by the 17-segment model. Only slices containing myocardium in all 360° were selected. Infarcted myocardium regions were identified in LGE images using the full width and half maximum (FWHM) technique [46]. The labeled segments were located on the cine slice with the same spatial location and the closest cardiac phase to that of LGE (end-diastole). Segmentations on cine CMR were propagated to all time frames to allow texture analysis of cine sequences including the time dimension (2D + t).

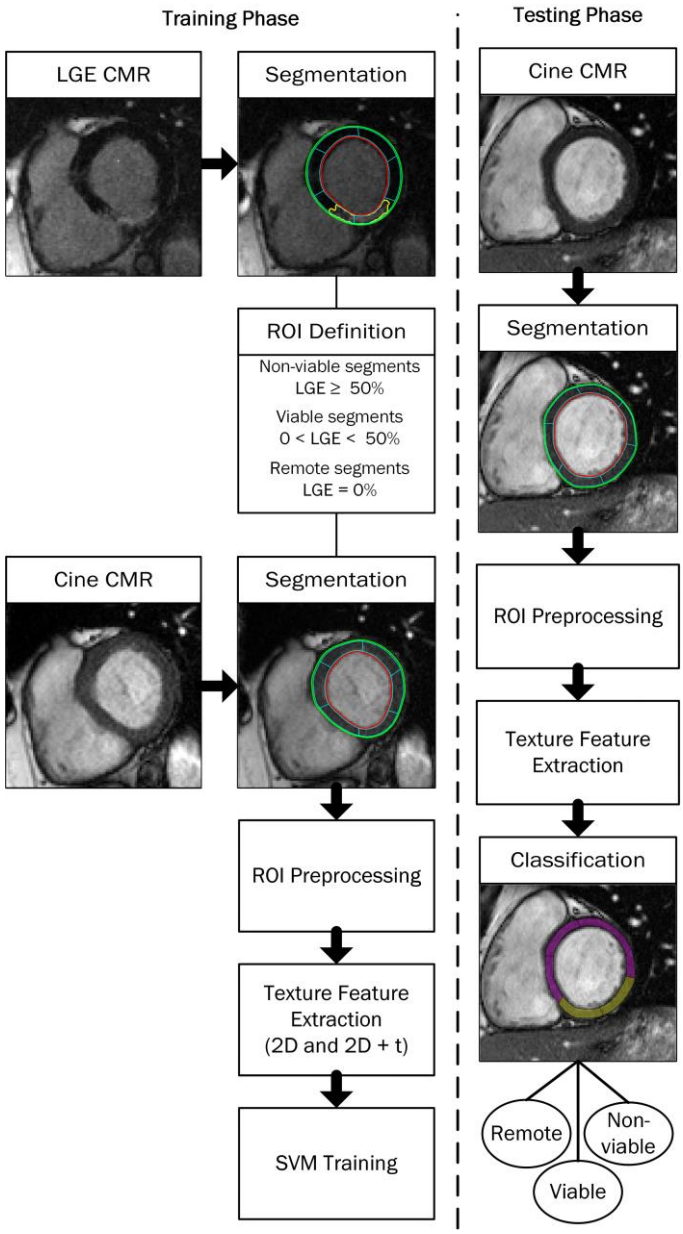

Figure 7.1. *General diagram of the texture-based approach to detect non-viable segments. In the training phase, LGE CMR was used as reference to define ROIs in the corresponding cine CMR slice. Texture features were extracted from cine CMR and were used to train an SVM classifier.*

Non-viable segments were identified as those showing LGE ≥ 50% transmural extension, segments showing 0 < LGE < 50% transmural extension were labeled as infarcted but viable and those not showing enhancement were labeled as remote myocardium [40]. In total, 1400 cine CMR segments were defined as ROIs for texture feature extraction: 340 non-viable (24%), 224 viable (16%), and 836 remote segments (60%). Figure 7.1 shows a general diagram of the proposed methodology.

7.2.3 Region of Interest Preprocessing

Preprocessing was implemented to ensure that the analysis truly characterizes the ROI texture. The original images were interpolated to a pixel size of 0.5 × 0.5 mm^2 using the cubic B-spline method [86]. Inhomogeneity correction was applied throughout the entire myocardium using the additive model described by [92]. Normalization was performed using the "$\pm 3\sigma$" normalization technique [88].

7.2.4 Texture Feature Extraction

Texture features were computed using the toolboxes provided by Vallières et al. [89] and Ojala et al. [114] in Matlab 2015b (MathWorks Inc., Natick, MA). Features were calculated for each ROI segment by means of four matrix-based texture analysis methods and LBPs, which are summarized in Table 7.2.

2D Analysis

Two-dimensional analysis was performed on the end-diastolic cine CMR slice. Texture features were calculated in their rotation invariance form to evade image rotation as a possible source of bias. In matrix-based methods, only one matrix was computed for each method by simultaneously taking into account the neighboring points at one-pixel distance in four directions: horizontal, vertical, 45º and 135º. Image quantization was set to 16 gray-levels based on a preliminary comparison among various levels. The $LBP_{P,R}$ operator, where P is the total number of neighborhood pixels and R is the radius in pixel was set to $LBP_{8,1}$. Rotation invariance values were used thus the LBP histogram was reduced to 36 bins instead of the original 256.

2D + *t* Analysis

To take advantage of the temporal dimension available in cine CMR, texture

analysis was also performed in 2D + *t*. It was implemented as volumetric 3D texture analysis but using time as the third dimension for each slice. The *z* dimension was not included because of the very large slice thickness in comparison to pixel size in the available images.

Similar to 2D analysis, only one GLCM, GLRLM, GLSZM and NGTDM was computed per slice by simultaneously taking into account the neighboring properties of voxels in 13 directions of 3D space [85]. A preliminary evaluation showed that 8 gray-levels yielded the best results in our data, so this value was used for matrix-based features in 2D + *t*. Isotropic image resolution is a requirement for volumetric texture computation, therefore time-spacing was indicated as having the same value as the in-plane resolution.

The spatio-temporal LBP known as LBP-TOP is an extension of the LBP operator that was proposed to analyze video sequences [117]. The LBP-TOP analyze the video sequence by concatenating the LBP histogram on three orthogonal planes: *xy*, *xt* and *yt*. In our analysis, the same $LBP_{P,R}$ operator was used for the three planes and were set to $LBP_{4,1}$, thus obtaining a concatenated histogram of 48 bins.

Table 7.2. *List of texture features used in this study.*

Method	Features	2D	2D+t
GLCM	Angular second moment, contrast, correlation, inverse difference moment, sum of squares, sum average, entropy, dissimilarity.	8	8
GLRLM	Short run emphasis, long run emphasis, gray-level non-uniformity, run-length non-uniformity, run percentage, low gray-level run emphasis, high gray-level run emphasis, short run low gray level emphasis, short run high gray-level emphasis, long run low gray-level emphasis, long run high gray-level emphasis, gray-level variance, run-length variance.	13	13
GLSZM	Small zone emphasis, large zone emphasis, gray-level non-uniformity, zone-size non-uniformity, zone percentage, low gray-level zone emphasis, high gray-level zone emphasis, small zone low gray level emphasis, small zone high gray-level emphasis, large zone low gray-level emphasis, large zone high gray-level emphasis, gray-level variance, zone-size variance.	13	13
NGTDM	Coarseness, contrast, busyness, complexity, strength.	5	5
LBP	2D: *P*=8 and *R*=1, 3D: *P*=4 and *R*=1 in three orthogonal planes.	36	48

*Values under the 2D and 2D + *t* columns indicate the number of features for each method.

7.2.5 Classification

Classification was performed using an SVM with radial kernel implemented in the *caret* package [146] in R language, version 3.3.0 (R Development Core Team, Vienna, Austria). Tuning parameters *C* and *sigma* were chosen from the finite set $\{2^{-2}…2^{4}\}$ and $\{10^{-2}…10^{2}\}$ respectively.

Analysis was performed independently in 2D and 2D + *t* using the following subsets of features as input to the SVM classifier:

- **Matrix features:** All texture features derived from matrix-based methods: GLCM, GLRLM, GLSZM, and NGTDM.
- **LBP features:** The LBP histogram where each bin represent one feature.
- **Matrix + LBP:** The full set of features.
- **SVM-RFE:** The multiple SVM-RFE [130] feature selection was implemented to the full set of features. Five-fold cross-validation was used to stabilize the feature selection and the *C* parameter of the SVM was set to unity. This method returns a ranking of features that was used to choose the optimal subset of features according to the performance of the radial SVM classifier.

Furthermore, the wall thickening (*Wth*), defined as the difference in myocardial thickness between end-diastole and end-systole, was used as input to the classifier. The performance using *Wth* solely was taken as reference for comparison with texture-based models.

7.2.6 Model Evaluation

The study group was randomly split into training (30 patients) and testing sets (20 patients). The training set was used for model building and the testing set to evaluate the generalization performance. Feature selection and model tuning were performed within the nested cross-validation approach that was illustrated in the previous chapter (Figure 6.3). The training set was split $n = 50$ times by using a repeated 5-fold cross-validation. An inner 5-fold cross-validation was used for model tuning.

Feature values were standardized to zero mean and unit variance across training samples. Near-zero values were found in some bins of the LBP histograms and these were excluded from analysis. All classification models were built within a multiclass approach using the one-versus-all technique. Performance parameters were reported on the testing set by considering each class individually: non-viable, viable and

remote myocardium.

Area under the curve (AUC) of the receiver operating characteristic (ROC) was used as an index of classification accuracy. The class probabilities estimated by each model were used to plot ROC curves. Sensitivity and specificity were computed for the cutoff on the ROC curve that maximizes the product of both measures. Class probabilities of the best model were also used to analyze the agreement between the infarct percentage measured on LGE CMR and the SVM probability on cine CMR using Bland Altman analysis [149]. The McNemar's test [150] was used to compare the performance between models trained with 2D and 2D + *t* features.

7.3 Results

The top 10 features selected by the SVM-RFE for 2D and 2D + t subsets are shown in Table 7.3. The top 2D feature was derived from the GLSZM texture analysis method and the top 2D + t feature was from the GLRLM method. The top 10 features included features from all texture analysis methods. However, the average ranks are relatively high thus indicating that at each fold of the nested cross-validation different features were selected. The feature profiles shown in Figure 7.2 indicate that almost all 2D features (67 out of 68) were necessary to achieve optimal classification performance. For the 2D + t subset, 57 out of 87 features were optimal. These top number of features were used to train the final classifier and to test its performance on the testing set. It can be seen that 2D + t features performed better that 2D features.

Table 7.3. Top ten ranked features according to the SVM-RFE algorithm.

2D features	**Rank**	**2D + t features**	**Rank**
GLSZM – GLN	6.52	GLRLM – GLN	19.98
NGTDM – Coarseness	15.78	GLSZM – GLN	20.62
LBP – bin 13	16.90	GLRLM – LGRE	24.04
GLSZM – ZSN	16.98	NGTDM – Busyness	24.64
NGTDM – Strength	17.12	GLCM – Entropy	25.04
GLSZM – LZLGE	18.50	GLRLM – HGRE	25.54
NGTDM – Busyness	18.50	LBP – bin 18	25.56
GLRLM – LGRE	18.60	GLSZM – SZHGE	26.52
LBP – bin 27	20.68	GLSZM – ZP	27.06
LBP – bin 35	20.88	GLSZM – ZSV	27.06

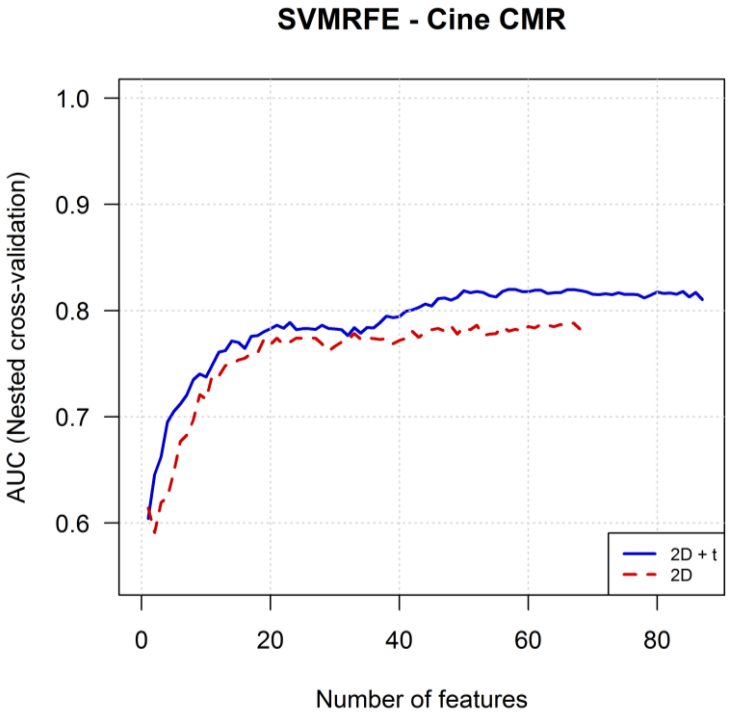

Figure 7.2. *Resampling profiles for 2D and 2D + t texture features. Average AUC values over resamples of nested cross-validation for different subset of features added according to the ranking returned by the SVM-RFE feature selection technique.*

Feature profiles for each class are presented in Figures 7.3 and 7.4 for 2D and 2D + t respectively. Classification performance was better for non-viable than remote segments for both cases. Notably, the classification for viable segments was very poor, which was expected considering that these are the "intermediate" segments that have small percentage of infarcted mass.

Classification performance parameters on testing set are reported in Table 7.4. The AUC values obtained with the SVM-RFE subset are very similar to that of the nested cross-validation profiles. The best overall performance was achieved combining the SVM-RFE subset with the wall thickening feature (AUC = 0.756) when using the 2D approach. The best overall performance for the 2D + t approach was achieved using the LBP features (AUC = 0.849).

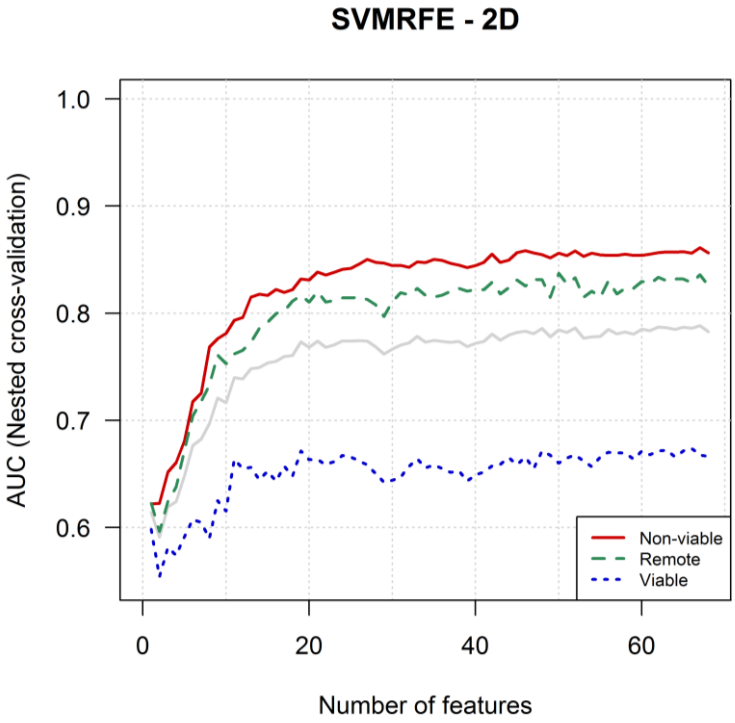

Figure 7.3. *Resampling profiles for 2D texture features according to each class. Average AUC values over resamples of nested cross-validation for different subset of features added according to the ranking returned by the SVM-RFE feature selection technique. The average profile is shown in light gray.*

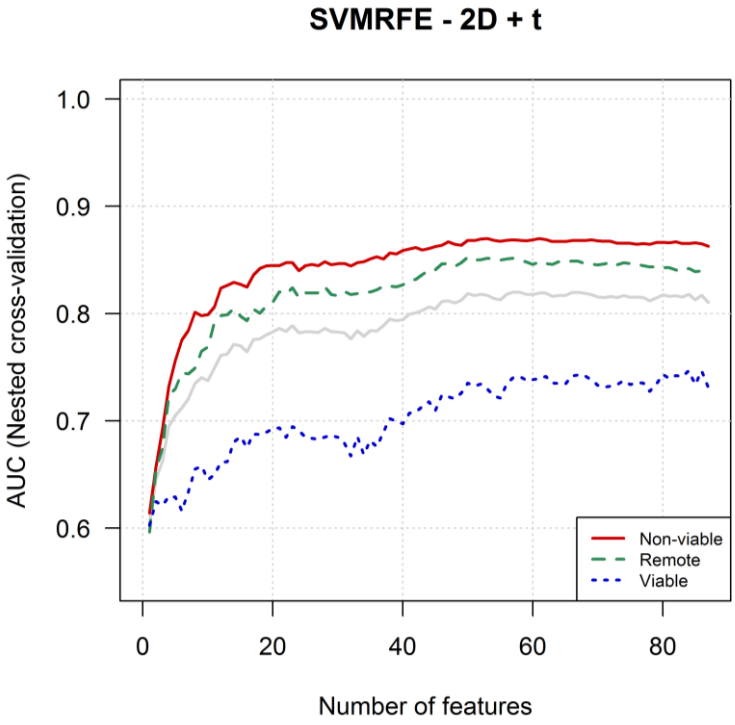

Figure 7.4. *Resampling profiles for 2D + t texture features according to each class. Average AUC values over resamples of nested cross-validation for different subset of features added according to the ranking returned by the SVM-RFE feature selection technique. The average profile is shown in light gray.*

Table 7.4. Classification performance parameters on testing set.

				Non-viable			Viable			Remote		
	Feature subset	**# Features**	**Overall AUC**	**AUC**	**Sens**	**Spec**	**AUC**	**Sens**	**Spec**	**AUC**	**Sens**	**Spec**
	Wall thickening	1	0.561	0.601	0.676	0.476	0.565	0.611	0.572	0.518	0.476	0.601
2D	Matrix	39	0.753	0.733	0.507	0.976	0.715	0.900	0.612	0.813	0.706	0.730
	LBP	29	0.727	0.889	0.977	0.636	0.563	0.377	0.814	0.730	0.787	0.650
	Matrix + LBP	68	0.737	0.810	0.580	0.981	0.635	0.788	0.459	0.765	0.694	0.650
	SVM-RFE	67	0.755	0.815	0.580	0.976	0.656	0.755	0.536	0.795	0.736	0.681
	Wth + SVM-RFE	68	0.756	0.811	0.602	0.948	0.651	0.800	0.485	0.805	0.754	0.716
2D + t	Matrix	39	0.811	0.810	0.647	0.797	0.822	0.855	0.748	0.802	0.631	0.911
	LBP	48	0.849	0.935	0.926	0.799	0.819	0.722	0.789	0.794	0.850	0.584
	Matrix + LBP	87	0.822	0.906	0.757	0.877	0.784	0.922	0.621	0.776	0.655	0.752
	SVM-RFE	57	0.807	0.911	0.794	0.875	0.752	0.877	0.587	0.759	0.631	0.761
	Wth + SVM-RFE	58	0.820	0.939	0.970	0.757	0.738	0.866	0.561	0.783	0.613	0.827

*Sensitivity and specificity were computed according to the optimal operating point on the ROC curve that maximizes the product between both measures.

Chronic scar formation is associated with thinned myocardium; therefore, the wall thickening feature was also used individually to train the SVM classifier. However, the performance using only this feature was very poor (AUC = 0.561) in comparison to those obtained with texture features. Figure 7.5 shows a comparison of 2D and 2D + t features that clearly shows how the latter outperforms the 2D approach. This difference was statistically significant for all cases according to the McNemar's test.

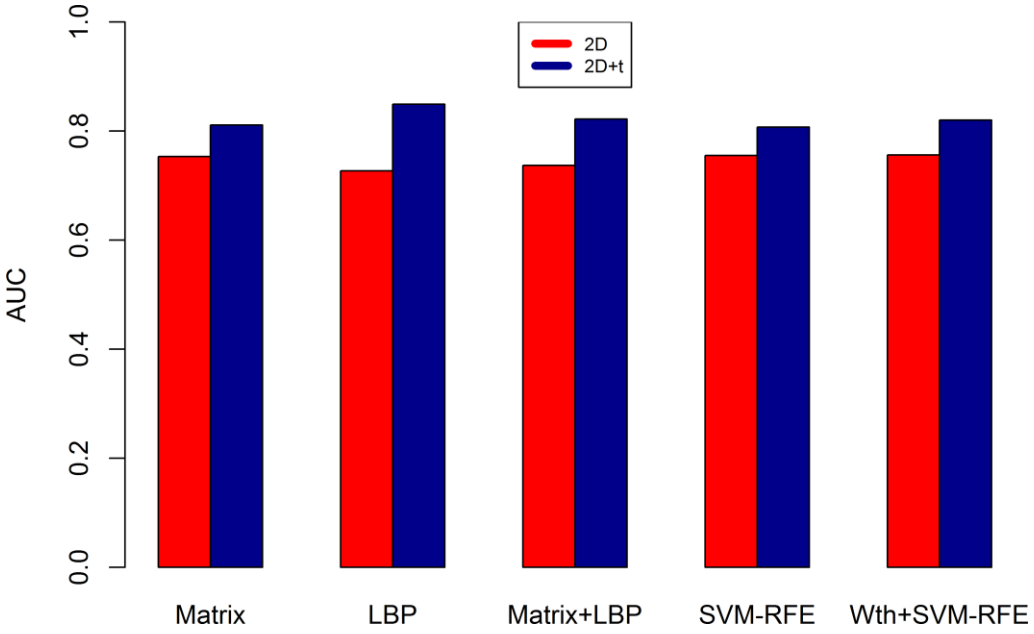

Figure 7.5. *Barplot comparing the AUC values for 2D and 2D + t features on testing set for each subset of features. 2D + t features provided larger values in all cases with statistical significance ($p < 0.01$) according to the McNemar's test.*

ROC curves for the best models in 2D and 2D + t are shown in Figures 7.6 and 7.7 respectively. The best model was achieved with 2D + t LBP features. AUC values were 0.935, 0.819, and 0.794 for non-viable, viable and remote segments respectively. Notably, classification of viable segments performed better than remote segments on testing set, contrary to the expected feature profiles previously shown in Figure 7.4.

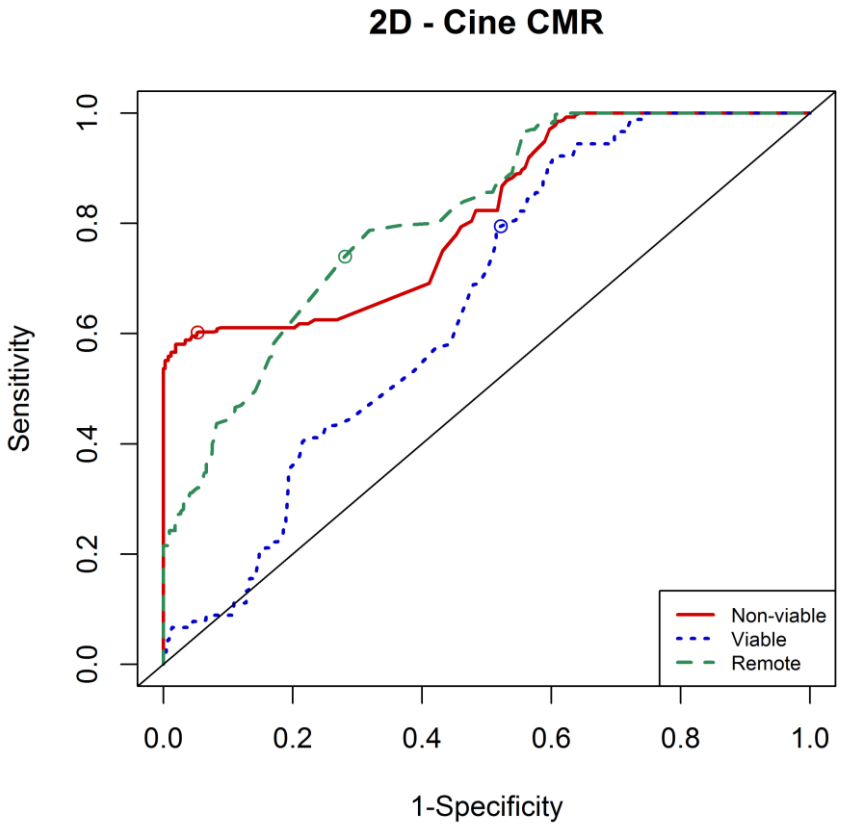

Figure 7.6. *ROC curves for the best model using 2D features (Wth + SVM-RFE). One ROC curve for each class is shown according to the one-versus-all technique. The circles on the curves indicate the optimal operating point that maximizes the product between sensitivity and specificity.*

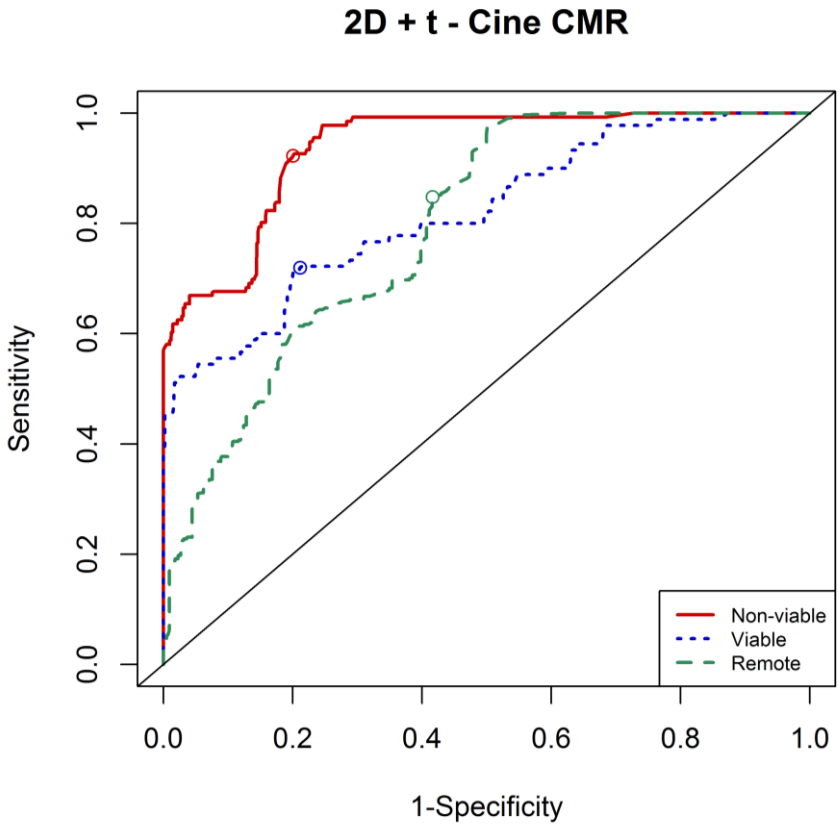

Figure 7.7. *ROC curves for the best model using 2D + t features (LBP). One ROC curve for each class is shown according to the one-versus-all technique. The circles on the curves indicate the optimal operating point that maximizes the product between sensitivity and specificity.*

To evaluate how the probability of the SVM classifier correlates with the mass percentage in each segment, the Bland Altman plot of Figure 7.8 was reproduced. It can be seen that most of the segments fall within the 95% that represents the likelihood of a real difference between both methods.

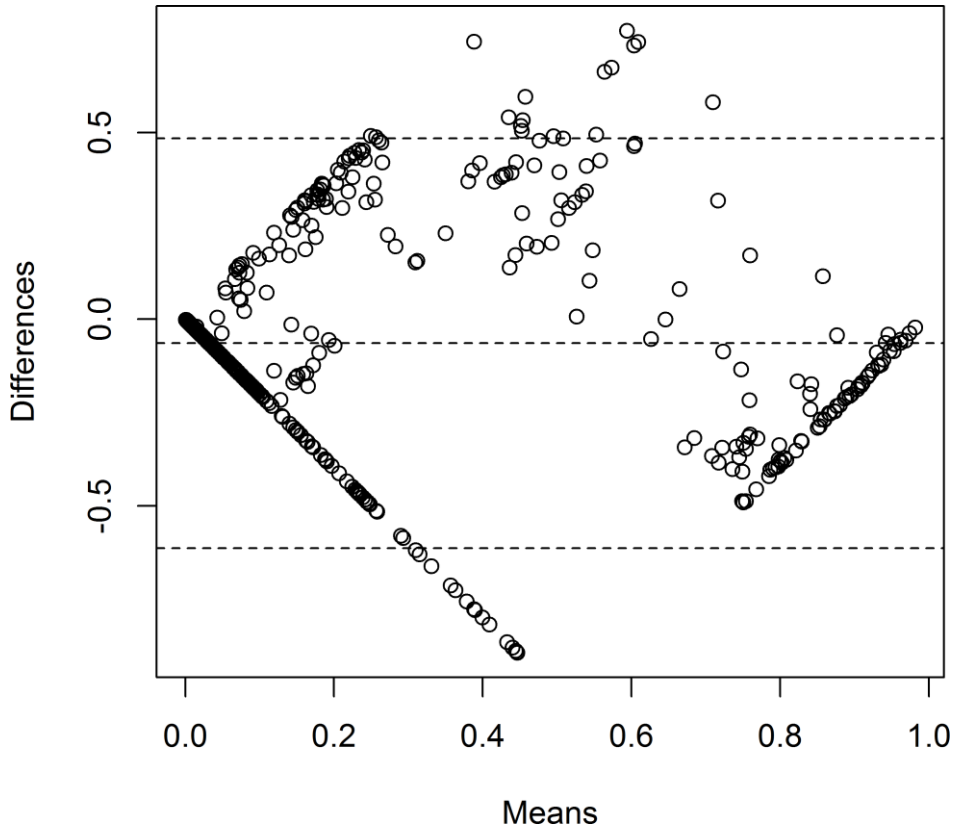

Figure 7.8. *Bland Altman plot of differences between the mass percentage measured in LGE CMR and the probability of the SVM classifier in cine CMR. Most of the segments fall within the upper and lower limits that represent the 95% likelihood of representing a real difference between both methods.*

Figure 7.9 shows examples of testing cine CMR and the predicted segments compared to the LGE ground truth.

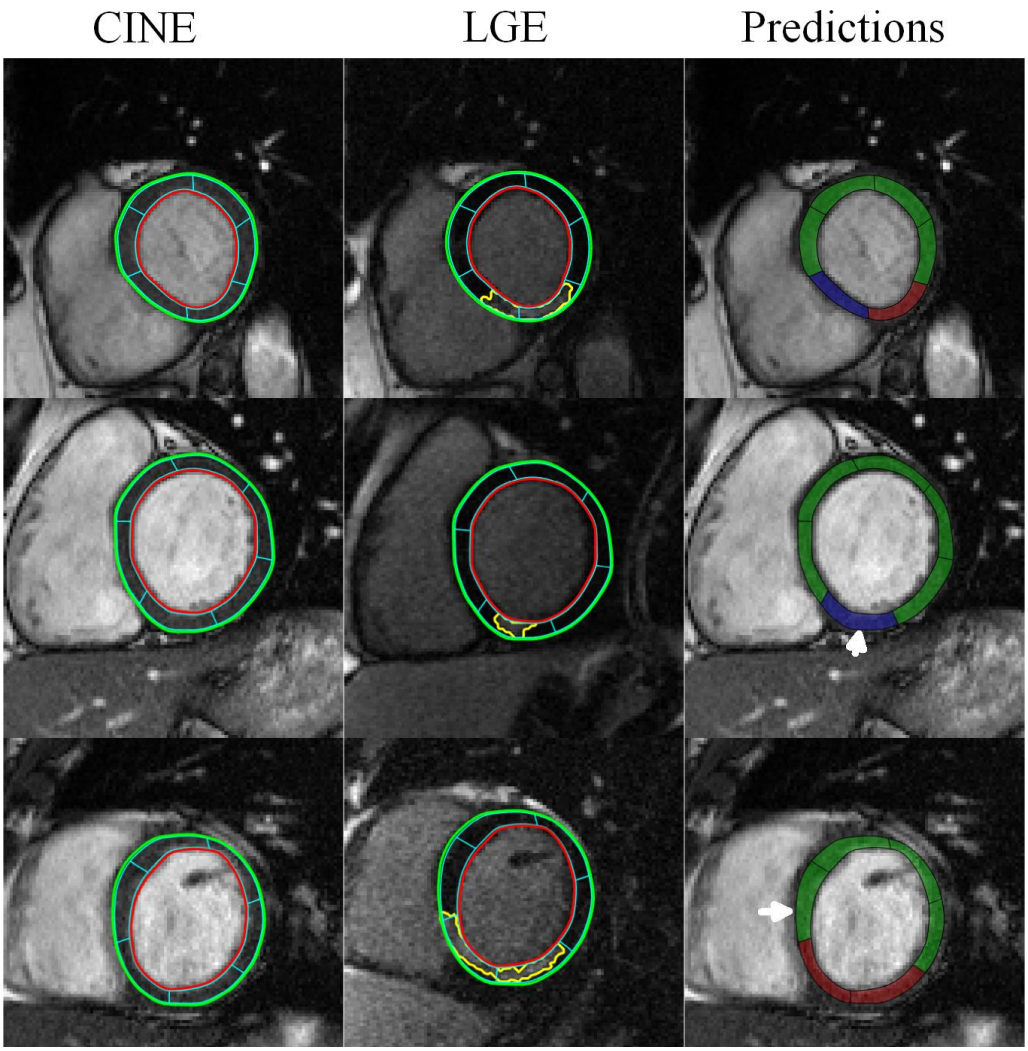

Figure 7.9. *Examples of images on testing set and the predictions using the LBP features in 2D + t. Scar (yellow) is represented on the ground truth LGE images. Segments predicted as non-viable are colored in red, those predicted as viable in blue and those predicted as remote in green. White arrows indicate the incorrectly predicted segments. Top row shows a correctly classified image, the inferoseptal segment is covered by less than 50% of scar and it was predicted as viable. The middle row shows a small scar that covers slightly more than 50% of the segment and it was incorrectly classified as viable. The two non-viable segments of the bottom row were correctly classified but the viable segment was incorrectly predicted as remote.*

7.4 Discussion

This study demonstrated that texture analysis has the potential to discriminate between non-viable, viable and remote segments in cine CMR. High discrimination (AUC > 0.8) was achieved using 2D + t texture features extracted from cine CMR, modality in which the scarred myocardium is visually imperceptible in most cases. However, classification was not straightforward as it was necessary the combination of several texture features and an advanced classifier such as the SVM to achieve high discrimination.

In recent years, CMR has become the gold standard non-invasive imaging technique for a comprehensive evaluation of the structural consequences of myocardial infarction. Beyond the well-established value of cine CMR to accurately quantify relevant parameters in post-infarction patients such as ejection fraction or left ventricular volumes, in this scenario LGE CMR has emerged as a unique tool for calculating the extent of the scar. As derived from LGE CMR, this variable has been demonstrated to be decisive for predicting left ventricular remodeling, late systolic recovery and patient's outcome [151], [152]. However, the use of LGE CMR involves a prolongation of studies and the administration of contrast. In turn, this implies certain limitations for selected subsets of patients and for cardiac MR labs: i) A significant number of post-infarction patients are still clinically unstable at the time of the study and, as a consequence, they cannot tolerate prolonged studies. ii) The use of gadolinium can induce side effects, especially it can worsen renal function in patients with a certain degree of renal insufficiency. iii) The number of studies per shift in cardiac labs has to be necessarily reduced due to the prolongation of cases in whom LGE sequences have to be used.

Thus, if the methodology proposed in this manuscript can be further validated and extended to detect the scar precisely, relevant benefits for post-infarction patients (less side effects, less time of scanning, better tolerance for unstable patients) and logistic advantages for labs (more studies per unit of time) could be achieved.

The presented approach could detect non-viable segments on cine CMR. The main objective would be to precisely delineate the scar using cine CMR solely. To achieve that goal, further multicenter studies including larger study groups will be needed in order to precisely identify the scar extent, so the presented hypothesis can be validated and included in the routine practice of MRI labs.

A myriad of features that might improve classification accuracy could be incorporated in the analysis, technique known as Radiomics [9], including features derived from other texture analysis methods, that we did not analyze. We presented a 2D + t approach without considering the *z* dimension because of the large slice thickness of the available images. Research on data with better image resolution is recommended for future work, so an approach of 3D + t texture analysis can be performed and this may improve the results.

7.5 Conclusion

This study reinforced the hypothesis that implicit differences between non-viable, viable and remote segments are present in cine CMR and can be detected by the application of texture analysis. The results could serve as starting point for the development of several applications, including the delineation of scar in cine CMR, application that would have great impact allowing the quantification of scar extension on cine sequences without the need of further LGE acquisitions.

Chapter 8

Conclusions

This thesis provides three main conclusions according to the experimental studies (Chapters 5 to 7), and are summarized here:

1. Texture and intensity features can be used in combination with an SVM classifier for segmentation of the infarcted myocardium in LGE CMR. The transferability of this application to images acquired with a different scanner was tested on a small sample data. Larger multicenter studies are needed for validation of the proposed method.
2. Differentiation between acute and chronic myocardial infarction is possible using texture features and machine learning techniques in LGE CMR and in standard pre-contrast cine CMR. The choice of the classifier is not relevant as long as a properly feature selection method is implemented. This quantitative approach is an alternative to the most useful existing method that relies on the visual assessment of edema CMR images.
3. Implicit differences between non-viable, viable and remote segments are present in cine CMR and these can be detected by the application of texture analysis. These results could serve as starting point for the development of several applications, including the delineation of scar in cine CMR, application that would have great impact allowing the quantification of scar extension on cine sequences without the need of further LGE acquisitions.

The experimental studies showed that texture analysis can be successfully applied to CMR images to assess patients with myocardial infarction. The main contribution was the possibility to classify infarcted non-viable segments using conventional cine CMR solely. This finding opens a research line aiming to accurately delineate the scarred myocardium as a gadolinium-free alternative, which will have potential advantages if it is successfully validated and applied in clinical routine. Finally, it can be said that texture analysis can enhance the “invisible” details of CMR images.

Publications

[P01] Larroza A, Moratal D, Paredes-Sánchez A, Soria-Olivas E, Chust ML, Arribas LA, Arana E, "Support vector machine classification of brain metastasis and radiation necrosis based on texture analysis in MRI," *J. Magn. Reson. Imaging*, vol. 42, no. 5, pp. 1362–1368, 2015.

[P02] Larroza A, Bodí V, Moratal D. Texture Analysis in Magnetic Resonance Imaging: Review and considerations for future applications. In: Assessment of Cellular and Organ Function and Dysfunction using Direct and Derived MRI Methodologies, Christakis Constantinides (Ed.) InTech, Rijeka, Croatia, 2016. ISBN 978-953-51-2723-9, Print ISBN 978-953-51-2722-2.

[P03] Larroza A, López-Lereu MP, Monmeneu JV, Bodí V, Moratal D. Texture analysis for infarcted myocardium detection on delayed enhancement MRI. In: IEEE International Symposium on Biomedical Imaging (ISBI 2017), April 18-21, 2017, Melbourne, Australia.

[P04] Larroza A, Materka A, López-Lereu MP, Monmeneu JV, Bodí V, Moratal D. Differentiation between acute and chronic myocardial infarction by means of texture analysis of late gadolinium enhancement and cine cardiac magnetic resonance imaging. European Journal of Radiology. 2017; 92: 78-83.

[P05] Larroza A, López-Lereu MP, Monmeneu JV, Gavara J, Chorro FJ, Bodí V, Moratal D. Texture analysis on cardiac cine magnetic resonance imaging to detect non-viable segments in patients with chronic myocardial infarction. Medical Physics [submitted].

References

[1] D. Mozaffarian, E. J. Benjamin, A. S. Go, D. K. Arnett, M. J. Blaha, M. Cushman, *et al.*, "Heart disease and stroke statistics-2016 update a report from the American Heart Association," *Circulation*, vol. 133, no. 4, pp. e38–e48, 2016.

[2] S. Boateng and T. Sanborn, "Acute myocardial infarction," *Lancet*, vol. 389, pp. 197–210, 2017.

[3] F. Von Knobelsdorff-Brenkenhoff and J. Schulz-Menger, "Cardiovascular magnetic resonance imaging in ischemic heart disease," *J. Magn. Reson. Imaging*, vol. 36, no. 1, pp. 20–38, 2012.

[4] M. P. Pfeiffer and R. W. W. Biederman, "Cardiac MRI: a general overview with emphasis on current use and indications," *Med. Clin. North Am.*, vol. 99, no. 4, pp. 849–861, 2015.

[5] C. M. Kramer and J. Narula, "CMR imaging: Creating contrast without cosmetics," *JACC Cardiovasc. Imaging*, vol. 4, no. 12, pp. 1326–1327, 2011.

[6] G. Castellano, L. Bonilha, L. M. Li, and F. Cendes, "Texture analysis of medical images.," *Clin. Radiol.*, vol. 59, no. 12, pp. 1061–1069, 2004.

[7] M. Hájek, D. M, A. Materka, and R. Lerski, *Texture analysis for Magnetic Resonance Imaging*. Prague, Czech Republic: med4 Publishing, 2006.

[8] S. Drabycz, R. G. Stockwell, and J. R. Mitchell, "Image texture characterization using the discrete orthonormal S-transform.," *J. Digit. Imaging*, vol. 22, no. 6, pp. 696–708, 2009.

[9] R. J. Gillies, P. E. Kinahan, and H. Hricak, "Radiomics: images are more than pictures, they are data," *Radiology*, vol. 278, no. 2, pp. 563–577, 2016.

[10] S. Wang and R. M. Summers, "Machine learning and radiology," *Med. Image Anal.*, vol. 16, no. 5, pp. 933–951, 2012.

[11] J. Juntu, J. Sijbers, S. De Backer, J. Rajan, and D. Van Dyck, "Machine learning study of several classifiers trained with texture analysis features to differentiate benign from malignant soft-tissue tumors in T1-MRI images," *J. Magn. Reson. Imaging*, vol. 31, no. 3, pp. 680–689, 2010.

[12] E. I. Zacharaki, S. Wang, S. Chawla, D. Soo Yoo, R. Wolf, E. R. Melhem, *et al.*, "Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme," *Magn. Reson. Med.*, vol. 62, no. 6, pp. 1609–1618, 2009.

[13] M. S. De Oliveira, M. L. F. Balthazar, A. D'Abreu, C. L. Yasuda, B. P. Damasceno, F. Cendes, *et al.*, "MR imaging texture analysis of the corpus callosum and thalamus in amnestic mild cognitive impairment and mild

alzheimer disease," *Am. J. Neuroradiol.*, vol. 32, no. 1, pp. 60–66, 2011.

[14] T. Santos, C. Maistro, C. Silva, M. Oliveira, M. Franca, and C. G, "MRI Texture analysis reveals bulbar abnormalities in Friedreich Ataxia," *AJNR. Am. J. Neuroradiol.*, vol. 36, no. 12, pp. 2214–2218, 2015.

[15] Z. Karimaghaloo, H. Rivaz, D. L. Arnold, D. L. Collins, and T. Arbel, "Temporal hierarchical adaptive texture CRF for automatic detection of gadolinium-enhancing multiple sclerosis lesions in brain MRI," *IEEE Trans. Med. Imaging*, vol. 34, no. 6, pp. 1227–1241, 2015.

[16] K. M. Iftekharuddin, S. Ahmed, and J. Hossen, "Multiresolution texture models for brain tumor segmentation in MRI.," *Conf. Proc. IEEE Eng. Med. Biol. Soc.*, vol. 2011, no. 2, pp. 6985–8, 2011.

[17] J. Yao, J. Chen, and C. Chow, "Breast Tumor Analysis in Dynamic Contrast Enhanced MRI Using Texture Features and Wavelet Transform," *IEEE J. Sel. Top. Signal Process.*, vol. 3, no. 1, pp. 94–100, 2009.

[18] D. Jirák, M. Dezortová, P. Taimr, and M. Hájek, "Texture analysis of human liver," *J. Magn. Reson. Imaging*, vol. 15, no. 1, pp. 68–74, 2002.

[19] G. Soumya, A. Oliver, R. Martí, X. Lladó, J. Freixenet, J. Vilanova, *et al.*, "Prostate segmentation with Local Binary Patterns guided active appearance models," in *Proc. of SPIE7962, Medical Imaging 2011: Image Processing, 796218*, 2011.

[20] L. P. Kotu, K. Engan, T. Eftestøl, S. Ørn, and L. Woie, "Segmentation of scarred and non-scarred myocardium in LG enhanced CMR images using intensity-based textural analysis.," in *33rd Annual International Conference of the IEEE EMBS*, 2011, pp. 5698–5701.

[21] T. Eftestøl, L. Woie, K. Engan, J. T. Kvaløy, D. W. T. Nilsen, and S. Ørn, "Texture analysis to assess risk of serious arrhythmias after myocardial infarction," *Comput. Cardiol. (2010).*, vol. 39, pp. 365–368, 2012.

[22] L. Kotu, K. Engan, T. Eftestol, L. Woie, S. Orn, and K. K. Aggelos, "Local binary patterns used on cardiac MRI to classify high and low risk patient groups," in *20th European Signal Processing Conference (EUSIPCO 2012)*, 2012, pp. 2586–2590.

[23] L. P. Kotu, K. Engan, K. Skretting, F. Måløy, S. Orn, L. Woie, *et al.*, "Probability mapping of scarred myocardium using texture and intensity features in CMR images.," *Biomed. Eng. Online*, vol. 12, no. 91, pp. 1–19, 2013.

[24] P. Beliveau, F. Cheriet, S. A. Anderson, J. L. Taylor, A. E. Arai, and L. Y. Hsu, "Quantitative assessment of myocardial fibrosis in an age-related rat model by ex vivo late gadolinium enhancement magnetic resonance imaging with histopathological correlation," *Comput. Biol. Med.*, vol. 65, pp. 103–113, 2015.

[25] P. Beliveau, F. Cheriet, S. A. Anderson, J. Taylor, A. E. Arai, and L.-Y. Hsu, "Textural analysis of late gadolinium enhanced magnetic resonance images can discriminate acute from chronic myocardial infarction," *J. Cardiovasc. Magn. Reson.*, vol. 16, no. Suppl 1, p. P182, 2014.

[26] R. E. Thornhill, M. Cocker, G. Dwivedi, C. Dennie, L. Fuller, A. Dick, *et al.*, "Quantitative texture features as objective metrics of enhancement heterogeneity in hypertrophic cardiomyopathy," *J. Cardiovasc. Magn. Reson.*, vol. 16, no. Suppl 1, p. P351, 2014.

[27] R. Schofield, B. Ganeshan, R. Kozor, A. Nasis, R. Endozo, and M. Ashley, "CMR myocardial texture analysis tracks different etiologies of left ventricular hypertrophy," *J. Cardiovasc. Magn. Reson.*, vol. 18, no. Suppl 1, pp. 1–2, 2016.

[28] E. N. Marieb and K. N. Hoehn, *Human Anatomy & Physiology*, 9th ed. Boston, Massachusetss USA: Pearson Education USA, 2013.

[29] A. Guyton and J. Hall, *Textbook of medical physiology*, 11th ed. Philadelphia, PA, USA: Saunders Elsevier, 2006.

[30] P. J. Lynch, "'Blausen 0463 HeartAttack' by Blausen Medical Communications, Inc.; 'Heart ant wall infarction' by Patrick J. Lynch, medical illustrator." .

[31] A. E. Stillman, M. Oudkerk, D. Bluemke, J. Bremerich, F. P. Esteves, E. V Garcia, *et al.*, "Assessment of acute myocardial infarction: current status and recommendations from the North American society for Cardiovascular Imaging and the European Society of Cardiac Radiology.," *Int. J. Cardiovasc. Imaging*, vol. 27, no. 1, pp. 7–24, 2011.

[32] H. Mahrholdt, A. Wagner, R. Judd, and U. Sechtem, "Assessment of myocardial viability by cardiovascular magnetic resonance imaging," *Eur. Heart J.*, vol. 23, no. 8, pp. 602–619, Apr. 2002.

[33] M. G. Friedrich, H. Abdel-Aty, A. Taylor, J. Schulz-Menger, D. Messroghli, and R. Dietz, "The salvaged area at risk in reperfused acute myocardial infarction as visualized by cardiovascular magnetic resonance," *J. Am. Coll. Cardiol.*, vol. 51, no. 16, pp. 1581–1587, 2008.

[34] F. Van De Werf, J. Bax, A. Betriu, C. Blomstrom-Lundqvist, F. Crea, V. Falk, *et al.*, "Management of acute myocardial infarction in patients presenting with persistent ST-segment elevation," *Eur. Heart J.*, vol. 29, no. 23, pp. 2909–2945, 2008.

[35] M. Saeed, T. A. Van, R. Krug, S. W. Hetts, and M. W. Wilson, "Cardiac MR imaging: current status and future direction," *Cardiovasc Diagn Ther*, vol. 5, no. 4, pp. 290–310, 2015.

[36] D. T. Ginat, M. W. Fong, D. J. Tuttle, S. K. Hobbs, and R. C. Vyas, "Cardiac imaging: Part 1, MR pulse sequences, imaging planes, and basic anatomy," *Am.*

J. Roentgenol., vol. 197, no. 4, pp. 808–815, 2011.

[37] S. M. Forbat, M. A. Sakrana, K. H. Darasz, F. El-Demerdash, and S. R. Underwood, "Rapid assessment of left ventricular volume by short axis cine MRI," *Br. J. Radiol.*, vol. 69, no. 819, pp. 221–225, 1996.

[38] M. D. Cerqueira, N. J. Weissman, V. Dilsizian, A. K. Jacobs, S. Kaul, W. K. Laskey, *et al.*, "Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart," *Circulation*, vol. 105, no. 4, pp. 539–542, 2002.

[39] R. D. Mosteller, "Simplified calculation of body-surface area," *N. Engl. J. Med.*, vol. 317, no. 17, pp. 1098–1098, Oct. 1987.

[40] V. Bodí, J. Sanchis, M. P. López-lereu, A. Losada, J. Núñez, M. Pellicer, *et al.*, "Usefulness of a comprehensive cardiovascular magnetic resonance imaging assessment for predicting recovery of left ventricular wall motion in the setting of myocardial stunning," *J. Am. Coll. Cardiol.*, vol. 46, no. 9, pp. 1747–1752, 2005.

[41] J. Sandstede, C. Lipke, M. Beer, S. Hofmann, T. Pabst, W. Kenn, *et al.*, "Age- and gender-specific differences in left and right ventricular cardiac function and mass determined by cine magnetic resonance imaging.," *Eur. Radiol.*, vol. 10, pp. 438–442, 2000.

[42] D. Moratal, M. E. Brummer, L. Martí-Bonmatí, and A. Vallés-Lluch, "NMR imaging," in *Wiley Enciclopedia of Biomedical Engineering*, Hoboken, New Jersey, USA: John Wiley & Sons, Inc., 2006, pp. 2590–2606.

[43] D. W. McRobbie, E. A. Moore, M. J. Graves, and M. R. Prince, *MRI: From picture to proton*, 2nd ed. New York, NY, USA: Cambridge University Press, 2003.

[44] M. G. Friedrich, C. Bucciarelli-Ducci, J. A. White, S. Plein, J. C. Moon, A. G. Almeida, *et al.*, "Simplifying cardiovascular magnetic resonance pulse sequence terminology.," *J. Cardiovasc. Magn. Reson.*, vol. 16, no. 1, p. 3960, 2014.

[45] M. Ishida, S. Kato, and H. Sakuma, "Cardiac MRI in ischemic heart disease.," *Circ. J.*, vol. 73, no. 9, pp. 1577–88, 2009.

[46] A. S. Flett, J. Hasleton, C. Cook, D. Hausenloy, G. Quarta, C. Ariti, *et al.*, "Evaluation of techniques for the quantification of myocardial scar of differing etiology using cardiac magnetic resonance," *JACC Cardiovasc. Imaging*, vol. 4, no. 2, pp. 150–156, 2011.

[47] R. Karim, P. Bhagirath, P. Claus, R. James Housden, Z. Chen, Z. Karimaghaloo, *et al.*, "Evaluation of state-of-the-art segmentation algorithms for left ventricle infarct from late Gadolinium enhancement MR images," *Med. Image Anal.*, vol. 30, pp. 95–107, 2016.

[48] J. V. Monmeneu, V. Bodí, J. Sanchis, M. P. López-Lereu, L. Mainar, J. Núñez, *et al.*, "Cardiac magnetic resonance evaluation of edema after ST-elevation acute

myocardial infarction," *Rev. Esp. Cardiol.*, vol. 62, no. 8, pp. 858–866, 2009.

[49] V. Bodi, J. Sanchis, M. P. Lopez-Lereu, J. Nunez, R. Sanz, P. Palau, *et al.*, "Microvascular perfusion 1 week and 6 months after myocardial infarction by first-pass perfusion cardiovascular magnetic resonance imaging," *Heart*, vol. 92, no. 12, pp. 1801–1807, 2006.

[50] M. E. Mayerhoefer, M. J. Breitenseher, J. Kramer, N. Aigner, S. Hofmann, and A. Materka, "Texture analysis for tissue discrimination on T1-weighted MR images of the knee joint in a multicenter study: Transferability of texture features and comparison of feature selection methods and classifiers," *J. Magn. Reson. Imaging*, vol. 22, no. 5, pp. 674–680, 2005.

[51] S. A. Waugh, R. A. Lerski, L. Bidaut, and A. M. Thompson, "The influence of field strength and different clinical breast MRI protocols on the outcome of texture analysis using foam phantoms.," *Med. Phys.*, vol. 38, no. 9, pp. 5058–5066, 2011.

[52] L. Bonilha, E. Kobayashi, G. Castellano, G. Coelho, E. Tinois, F. Cendes, *et al.*, "Texture Analysis of Hippocampal Sclerosis," *Epilepsia*, vol. 44, no. 12, pp. 1546–1550, 2003.

[53] T. Pallavi, P. Prateek, R. Lisa, W. Leo, B. Chaitra, S. Andrew, *et al.*, "Texture descriptors to distinguish radiation necrosis from recurrent brain tumors on multi-parametric MRI.," *Proc. SPIE--the Int. Soc. Opt. Eng.*, vol. 9035, p. 90352B, 2014.

[54] K. Oppedal, T. Eftestøl, K. Engan, M. K. Beyer, and D. Aarsland, "Classifying dementia using local binary patterns from different regions in magnetic resonance images," *Int. J. Biomed. Imaging*, vol. 2015, pp. 1–14, 2015.

[55] S. Herlidou-Même, J. M. Constans, B. Carsin, D. Olivie, P. A. Eliat, L. Nadal-Desbarats, *et al.*, "MRI texture analysis on texture test objects, normal brain and intracranial tumors," *Magn. Reson. Imaging*, vol. 21, no. 9, pp. 989–993, 2003.

[56] J. Fruehwald-Pallamar, J. Hesselink, M. Mafee, L. Holzer-Fruehwald, C. Czerny, and M. Mayerhoefer, "Texture-based analysis of 100 MR examinations of head and neck tumors – is it possible to discriminate between benign and malignant masses in a multicenter trial?," *Fortschr Röntgenstr*, vol. 188, no. 2, pp. 195–202, 2016.

[57] E. Orphanidou-Vlachou, N. Vlachos, N. P. Davies, T. N. Arvanitis, R. G. Grundy, and A. C. Peet, "Texture analysis of T1- and T2-weighted MR images and use of probabilistic neural network to discriminate posterior fossa tumours in children," *NMR Biomed.*, vol. 27, no. 6, pp. 632–639, 2014.

[58] P. Brynolfsson, D. Nilsson, R. Henriksson, J. Hauksson, M. Karlsson, A. Garpebring, *et al.*, "ADC texture-an imaging biomarker for high-grade

glioma?," *Med. Phys.*, vol. 41, no. 10, p. 101903, 2014.

[59] P. Foroutan, J. M. Kreahling, D. L. Morse, O. Grove, M. C. Lloyd, D. Reed, *et al.*, "Diffusion MRI and novel texture analysis in osteosarcoma xenotransplants predicts response to anti-checkpoint therapy," *PLoS One*, vol. 8, no. 12, pp. 1–10, 2013.

[60] F. Khalvati, A. Wong, and M. A. Haider, "Automated prostate cancer detection via comprehensive multi-parametric magnetic resonance imaging texture feature models.," *BMC Med. Imaging*, vol. 15, no. 1, p. 27, 2015.

[61] D. Jirák, M. Dezortová, and M. Hájek, "Phantoms for texture analysis of MR images. Long-term and multi-center study.," *Med. Phys.*, vol. 31, no. 3, pp. 616–22, 2004.

[62] M. E. Mayerhoefer, P. Szomolanyi, D. Jirak, A. Materka, and S. Trattnig, "Effects of MRI acquisition parameter variations and protocol heterogeneity on the results of texture analysis and pattern discrimination: an application-oriented study.," *Med. Phys.*, vol. 36, no. 2009, pp. 1236–1243, 2009.

[63] L. R. Schad, "Problems in texture analysis with magnetic resonance imaging.," *Dialogues Clin. Neurosci.*, vol. 6, no. 2, pp. 235–42, 2004.

[64] H. Z. Tameem, L. E. Selva, and U. S. Sinha, "Texture measure from low resolution MR images to determine trabecular bone integrity in osteoporosis," *Annu. Int. Conf. IEEE Eng. Med. Biol. - Proc.*, pp. 2027–2030, 2007.

[65] S. J. Savio, L. C. V Harrison, T. Luukkaala, T. Heinonen, P. Dastidar, S. Soimakallio, *et al.*, "Effect of slice thickness on brain magnetic resonance image texture analysis.," *Biomed. Eng. Online*, vol. 9, no. 1, p. 60, 2010.

[66] W. G. J. Bradley, "Pros and cons of 3 tesla MRI.," *J. Am. Coll. Radiol.*, vol. 5, no. 8, pp. 871–878, Aug. 2008.

[67] X. Zhang, X. Gao, B. J. Liu, K. Ma, W. Yan, L. Liling, *et al.*, "Effective staging of fibrosis by the selected texture features of liver: Which one is better, CT or MR imaging?," *Comput. Med. Imaging Graph.*, vol. 46, no. September, pp. 227–236, 2015.

[68] M. Giger, H. Li, L. Lan, H. Abe, and G. Newstead, "Quantitative MRI phenotyping of breast cancer across molecular classification subtypes," *Breast Imaging*, vol. 8539, pp. 195–200, 2014.

[69] M. Sanz-Cortes, F. Figueras, E. Bonet-Carne, N. Padilla, V. Tenorio, N. Bargall??, *et al.*, "Fetal brain MRI texture analysis identifies different microstructural patterns in adequate and small for gestational age fetuses at term," *Fetal Diagn. Ther.*, vol. 33, no. 2, pp. 122–129, 2013.

[70] L. C. V Harrison, R. Nikander, M. Sikiö, T. Luukkaala, M. T. Helminen, P. Ryymin, *et al.*, "MRI texture analysis of femoral neck: Detection of exercise

load-associated differences in trabecular bone," *J. Magn. Reson. Imaging*, vol. 34, no. 6, pp. 1359–1366, 2011.

[71] Z. Shi, Z. Yang, G. Zhang, G. Cui, X. Xiong, Z. Liang, *et al.*, "Characterization of texture features of bladder carcinoma and the bladder wall on MRI: Initial experience," *Acad. Radiol.*, vol. 20, no. 8, pp. 930–938, 2013.

[72] C. P. Loizou, S. Petroudi, I. Seimenis, M. Pantziaris, and C. S. Pattichis, "Quantitative texture analysis of brain white matter lesions derived from T2-weighted MR images in MS patients with clinically isolated syndrome," *J. Neuroradiol.*, vol. 42, no. 2, pp. 99–114, 2015.

[73] H. Liu, Y. Shao, D. Guo, Y. Zheng, Z. Zhao, and T. Qiu, "Cirrhosis classification based on texture classification of random features," *Comput. Math. Methods Med.*, vol. 2014, 2014.

[74] K. K. Holli, L. Harrison, P. Dastidar, M. Wäljas, S. Liimatainen, T. Luukkaala, *et al.*, "Texture analysis of MR images of patients with mild traumatic brain injury.," *BMC Med. Imaging*, vol. 10, p. 8, 2010.

[75] Z. Karimaghaloo, H. Rivaz, D. L. Arnold, D. L. Collins, and T. Arbel, "Adaptive voxel, texture and temporal conditional random fields for detection of Gad-enhancing multiple sclerosis lesions in brain MRI," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 8151 LNCS, no. PART 3, pp. 543–550, 2013.

[76] A. Vignati, S. Mazzetti, V. Giannini, F. Russo, E. Bollito, F. Porpiglia, *et al.*, "Texture features on T2-weighted magnetic resonance imaging: new potential biomarkers for prostate cancer aggressiveness.," *Phys. Med. Biol.*, vol. 60, no. 7, pp. 2685–701, 2015.

[77] M. E. Mayerhoefer, D. Stelzeneder, W. Bachbauer, G. H. Welsch, T. C. Mamisch, P. Szczypinski, *et al.*, "Quantitative analysis of lumbar intervertebral disc abnormalities at 3.0 Tesla: value of T(2) texture features and geometric parameters.," *NMR Biomed.*, vol. 25, no. September 2011, pp. 866–72, 2012.

[78] W. Chen, M. L. Giger, H. Li, U. Bick, and G. M. Newstead, "Volumetric texture analysis of breast lesions on contrast-enhanced magnetic resonance images," *Magn. Reson. Med.*, vol. 58, no. 3, pp. 562–571, 2007.

[79] M. Sikiö, K. Holli-Helenius, R. Pertti, L. Harrison, and H. Eskola, "The effect of region of interest size on textural parameters. A study with clinical magnetic resonance images and artificial noise images," in *9th International Symposium on Image and Signal Processing and Analysis (ISPA), IEEE*, 2015, pp. 149–153.

[80] S. B. Antel, D. L. Collins, N. Bernasconi, F. Andermann, R. Shinghal, R. E. Kearney, *et al.*, "Automated detection of focal cortical dysplasia lesions using computational models of their MRI characteristics and texture analysis,"

Neuroimage, vol. 19, no. 4, pp. 1748–1759, 2003.

[81] L. Kjaer, P. Ring, C. Thomsen, and O. Henriksen, "Texture analysis in quantitative MR imaging. Tissue characterisation of normal brain and intracranial tumours at 1.5 T," *Acta radiol.*, vol. 36, no. 2, pp. 127–135, 1995.

[82] M. E. Mayerhoefer, P. Szomolanyi, D. Jirak, A. Berg, A. Materka, A. Dirisamer, *et al.*, "Effects of magnetic resonance image interpolation on the results of texture-based pattern classification: a phantom study.," *Invest. Radiol.*, vol. 44, no. 7, pp. 405–411, Jul. 2009.

[83] A. Depeursinge, A. Foncubierta-Rodriguez, D. Van De Ville, and H. Müller, "Three-dimensional solid texture analysis in biomedical imaging: review and opportunities," *Med. Image Anal.*, vol. 18, no. 1, pp. 176–196, 2014.

[84] G. Collewet, M. Strzelecki, and F. Mariette, "Influence of MRI acquisition protocols and image intensity normalization methods on texture classification," *Magn. Reson. Imaging*, vol. 22, no. 1, pp. 81–91, 2004.

[85] M. Vallières, C. R. Freeman, S. R. Skamene, and I. El Naqa, "A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities," *Phys. Med. Biol.*, vol. 60, no. 14, pp. 5471–5496, 2015.

[86] R. E. Thornhill, M. Golfam, A. Sheikh, G. O. Cron, E. A. White, J. Werier, *et al.*, "Differentiation of lipoma from liposarcoma on MRI using texture and shape analysis," *Acad. Radiol.*, vol. 21, no. 9, pp. 1185–1194, 2014.

[87] P. Kölhi, J. Järnstedt, M. Sikiö, J. Viik, P. Dastidar, T. Peltomäki, *et al.*, "A texture analysis method for MR images of airway dilator muscles: A feasibility study," *Dentomaxillofacial Radiol.*, vol. 43, no. 5, pp. 1–7, 2014.

[88] A. Materka and M. Strzelecki, "On the importance of MRI nonuniformity correction for texture analysis," in *Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA)*, 2013, pp. 118–123.

[89] B. Belaroussi, J. Milles, S. Carme, Y. M. Zhu, and H. Benoit-Cattin, "Intensity non-uniformity correction in MRI: Existing methods and their validation," *Med. Image Anal.*, vol. 10, no. 2, pp. 234–246, 2006.

[90] P. Prasanna, P. Tiwari, and A. Madabhushi, "Co-occurrence of local anisotropic gradient orientations (CoLlAGe): Distinguishing tumor confounders and molecular subtypes on MRI," in *Medical Image Computing and Computer-Assisted Intervention- MICCAI 2014*, 2014, pp. 73–80.

[91] Y. Zhang, A. Traboulsee, Y. Zhao, L. M. Metz, and D. K. Li, "Texture analysis differentiates persistent and transient T1 black holes at acute onset in multiple sclerosis: a preliminary study.," *Mult. Scler.*, vol. 17, no. 5, pp. 532–540, 2011.

[92] D. Yang, G. Rao, J. Martinez, A. Veeraraghavan, and A. Rao, "Evaluation of

tumor-derived MRI-texture features for discrimination of molecular subtypes and prediction of 12-month survival status in glioblastoma," *Med Phys*, vol. 42, no. 11, p. 6725, 2015.

[93] T. K. Chuah, E. Van Reeth, K. Sheah, and C. L. Poh, "Texture analysis of bone marrow in knee MRI for classification of subjects with bone marrow lesion - Data from the Osteoarthritis Initiative," *Magn. Reson. Imaging*, vol. 31, no. 6, pp. 930–938, 2013.

[94] J. G. Sled, A. P. Zijdenbos, and A. C. Evans, "A nonparametric method for automatic correction of intensity nonuniformity in MRI data," *IEEE Trans. Med. Imaging*, vol. 17, no. 1, pp. 87–97, 1998.

[95] P. Gibbs and L. W. Turnbull, "Textural analysis of contrast-enhanced MR images of the breast," *Magn. Reson. Med.*, vol. 50, no. 1, pp. 92–98, 2003.

[96] A. Ahmed, P. Gibbs, M. Pickles, and L. Turnbull, "Texture analysis in assessment and prediction of chemotherapy response in breast cancer," *J. Magn. Reson. Imaging*, vol. 38, no. 1, pp. 89–101, 2013.

[97] D. Mahmoud-Ghoneim, M. K. Alkaabi, J. D. de Certaines, and F.-M. Goettsche, "The impact of image dynamic range on texture classification of brain white matter," *BMC Med. Imaging*, vol. 8, no. 1, p. 18, 2008.

[98] W. H. Nailon, "Texture analysis methods for medical image characterisation," in *Biomedical Imaging*, Rijeka, Croatia: InTech, 2010, pp. 75–100.

[99] A. Materka and M. Strzelecki, "Texture analysis methods – a review," *Tech. Univ. Lodz, Inst. Electron. COST B11 Rep.*, pp. 1–33, 1998.

[100] A. Materka, "Texture analysis methodologies for magnetic resonance imaging," *Dialogues Clin. Neurosci.*, vol. 6, no. 2, pp. 243–250, 2004.

[101] R. M. Haralick, K. Shanmugam, I. I. Dinstein, M. Haralick, Robert, K. Shanmugam, I. I. Dinstein, *et al.*, "Textural features for image classification," *IEEE Transactions on Systems, Man, and Cybernetics SMC-3*, vol. 6, no. 6. pp. 610–621, 1973.

[102] M. J. House, S. J. Bangma, M. Thomas, E. K. Gan, O. T. Ayonrinde, L. a. Adams, *et al.*, "Texture-based classification of liver fibrosis using MRI," *J. Magn. Reson. Imaging*, vol. 328, no. 2, pp. 322–328, 2013.

[103] A. Wibmer, H. Hricak, T. Gondo, K. Matsumoto, H. Veeraraghavan, D. Fehr, *et al.*, "Haralick texture analysis of prostate MRI: utility for differentiating non-cancerous prostate from prostate cancer and differentiating prostate cancers with different Gleason scores," *Eur. Radiol.*, vol. 25, no. 10, pp. 2840–2850, 2015.

[104] M. M. Galloway, "Texture analysis using gray level run lengths," *Comput. Graph. image Process.*, vol. 4, no. 2, pp. 172–179, 1975.

[105] A. Chu, C. M. Sehgal, and J. F. Greenleaf, "Use of gray value distribution of

run lengths for texture analysis," *Pattern Recognit. Lett.*, vol. 11, no. 6, pp. 415–419, 1990.

[106] B. V. Dasarathy and E. B. Holder, "Image characterizations based on joint gray level-run length distributions," *Pattern Recognit. Lett.*, vol. 12, no. 8, pp. 497–502, 1991.

[107] G. Thibault, B. Fertil, C. Navarro, S. Pereira, P. Cau, N. Levy, *et al.*, "Texture Indexes and Gray Level Size Zone Matrix Application to Cell Nuclei Classification," *Pattern Recognit. Inf. Process.*, pp. 140–145, 2009.

[108] M. Amadasun and R. King, "Textural features corresponding to textural properties," *IEEE Trans. Syst. Man Cybern.*, vol. 19, no. 5, pp. 1264–1273, 1989.

[109] A. Materka, "MaZda user's manual," 2002. [Online]. Available: http://www.eletel.p.lodz.pl/mazda/download/mazda_manual.pdf. [Accessed: 28-Feb-2017].

[110] T. Ojala, M. Pietikäinen, and T. Mäenpää, "A generalized local binary pattern operator for multiresolution gray scale and rotation invariant texture classification," in *Second International Conference on Advances in Pattern Recognition*, 2001, pp. 397–406.

[111] M. S. Sheethal, D. B. Kannanm, A. Varghese, and S. T, "Intelligent classification technique of human brain MRI with efficient wavelet based feature extraction using local binary pattern," in *International Conference on Control Communication and Computing (ICCC)*, 2013, pp. 368–372.

[112] D. Unay, A. Ekin, M. Cetin, R. Jasinschi, and A. Ercil, "Robustness of local binary patterns in brain MR image analysis," in *Proceedings of the 29th Annual International Conference of the IEEE EMBS*, 2007, pp. 2098–2101.

[113] D. Assefa, H. Keller, C. Ménard, N. Laperriere, R. J. Ferrari, and I. Yeung, "Robust texture features for response monitoring of glioblastoma multiforme on T1-weighted and T2-FLAIR MR images: a preliminary investigation in terms of identification and segmentation.," *Med. Phys.*, vol. 37, no. 4, pp. 1722–36, 2010.

[114] F. Wagner, A. Gryanik, R. Schulz-Wendtland, P. A. Fasching, and T. Wittenberg, "3D characterization of texture: evaluation for the potential application in mammographic mass diagnosis," *Biomed. Tech.*, vol. 57, no. Suppl. 1, pp. 490–493, 2012.

[115] D. Mahmoud-Ghoneim, G. Toussaint, J. M. Constans, and J. D. De Certaines, "Three dimensional texture analysis in MRI: A preliminary evaluation in gliomas," *Magn. Reson. Imaging*, vol. 21, no. 9, pp. 983–987, 2003.

[116] P. Georgiadis, D. Cavouras, I. Kalatzis, D. Glotsos, E. Athanasiadis, S. Kostopoulos, *et al.*, "Enhancing the discrimination accuracy between

metastases, gliomas and meningiomas on brain MRI by volumetric textural features and ensemble pattern recognition methods," *Magn. Reson. Imaging*, vol. 27, no. 1, pp. 120–130, 2009.

[117] G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 915–928, 2007.

[118] B. J. Woods, B. D. Clymer, T. Kurc, J. T. Heverhagen, R. Stevens, A. Orsdemir, *et al.*, "Malignant-lesion segmentation using 4D co-occurrence texture analysis applied to dynamic contrast-enhanced magnetic resonance breast image data," *J. Magn. Reson. Imaging*, vol. 25, no. 3, pp. 495–501, 2007.

[119] J. Huang, X. Huang, D. Metaxas, and L. Axel, "Dynamic texture based heart localization and segmentation in 4-D cardiac images," in *4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 2007, pp. 852–855.

[120] P. M. Szczypinski, M. Strzelecki, A. Materka, and A. Klepaczko, "MaZda-A software package for image texture analysis.," *Comput. Methods Programs Biomed.*, vol. 94, no. 1, pp. 66–76, 2009.

[121] L. Markides and D. F. Gillies, "Machine Learning in Medical Imaging," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7588, pp. 136–143, 2012.

[122] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An introduction to statistical learning with applications in R*. New York, NY, USA: Springer, 2006.

[123] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.*, vol. 3, pp. 1157–1182, 2003.

[124] C. Chu, A.-L. Hsu, K.-H. Chou, P. Bandettini, and C. Lin, "Does feature selection improve classification accuracy? Impact of sample size and feature selection on classification using anatomical magnetic resonance images," *Neuroimage*, vol. 60, no. 1, pp. 59–70, 2012.

[125] C. C. Reyes-Aldasoro and A. Bhalerao, "Volumetric texture description and discriminant feature selection for MRI.," *Inf. Process. Med. Imaging*, vol. 18, pp. 282–293, 2003.

[126] M. Kuhn and K. Johnson, *Applied predictive modeling*. New York: Springer New York, 2013.

[127] J. Schürmann, *Pattern classification: A unified view of statistical and neural approaches*. New York, NY, USA: John Wiley & Sons, Inc., 1996.

[128] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Mach. Learn.*, vol. 46, pp. 389–422, 2002.

[129] C. Fernandez-Lozano, J. A. Seoane, M. Gestal, T. R. Gaunt, J. Dorado, and C.

Campbell, "Texture classification using feature selection and kernel-based techniques," *Soft Comput.*, vol. 19, no. 9, pp. 2469–2480, 2015.

[130] K. B. Duan, J. C. Rajapakse, H. Wang, and F. Azuaje, "Multiple SVM-RFE for gene selection in cancer classification with expression data," *IEEE Trans. Nanobioscience*, vol. 4, no. 3, pp. 228–233, 2005.

[131] M. Fernández-Delgado, E. Cernadas, S. Barro, and D. Amorim, "Do we need hundreds of classifiers to solve real world classification problems?," *J. Mach. Learn. Res.*, vol. 15, pp. 3133–3181, 2014.

[132] A. Palmer, R. Jiménez, and E. Gervilla, "Data mining: Machine learning and statistical techniques," in *knowledge-oriented applications in data mining*, Rijeka, Croatia: InTech, 2006, pp. 373–396.

[133] T. Hastie, R. Tibshirani, and J. J. J. Friedman, *The elements of statistical learning*, 2nd ed. Springer, 2008.

[134] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[135] C. Cortes and V. Vapnik, "Support-Vector networks," *Mach. Learn.*, vol. 20, pp. 273–297, 1995.

[136] A. Ben-Hur and J. Weston, "A user's guide to support vector machines," *Data Min. Tech. life Sci.*, vol. 609, pp. 223–239, 2010.

[137] P. M. Szczypinski, M. Strzelecki, and A. Materka, "Mazda - a software for texture analysis," *2007 Int. Symp. Inf. Technol. Converg. (ISITC 2007)*, pp. 245–249, 2007.

[138] A. Karatzoglou, A. Smola, K. Hornik, and A. Zeileis, "Kernlab - An S4 package for Kernel methods in R," *J. Stat. Softw.*, vol. 11, no. 9, 2004.

[139] R. Karim, P. Claus, Z. Chen, R. J. Housden, S. Obom, H. Gill, *et al.*, "Infarct segmentation challenge on delayed enhancement MRI of the left ventricle," in *Lecture Notes in Computer Science (LNCS)*, 2013, vol. 7746, pp. 97–104.

[140] B. W. Choi, "Differentiation of acute myocardial infarction from chronic myocardial scar with MRI," *Korean J. Radiol.*, vol. 7, no. 1, pp. 1–3, 2006.

[141] H. Abdel-Aty, A. Zagrosek, J. Schulz-Menger, A. J. Taylor, D. Messroghli, A. Kumar, *et al.*, "Delayed enhancement and T2-weighted cardiovascular magnetic resonance imaging differentiate acute from chronic myocardial infarction," *Circulation*, vol. 109, no. 20, pp. 2411–2416, 2004.

[142] K. A. Kim, J. B. Seo, K. H. Do, J. N. Heo, Y. K. Lee, J. W. Song, *et al.*, "Differentiation of recently infarcted myocardium from chronic myocardial scar: The value of contrast-enhanced SSFP-based cine MR imaging," *Korean J. Radiol.*, vol. 7, no. 1, pp. 14–19, 2006.

[143] F. von Knobelsdorff-Brenkenhoff, M. Prothmann, M. A. Dieringer, R. Wassmuth, A. Rudolph, W. Utz, *et al.*, "Current T1 and T2 mapping techniques

applied with simple thresholds cannot discriminate acute from chronic myocadial infarction on an individual patient basis: a pilot study.," *BMC Med. Imaging*, vol. 16, p. 35, 2016.

[144] E. Heiberg, J. Sjögren, M. Ugander, M. Carlsson, H. Engblom, and H. Arheden, "Design and validation of Segment--freely available software for cardiovascular image analysis.," *BMC Med. Imaging*, vol. 10, pp. 1–13, 2010.

[145] M. Kuhn, "Building predictive models in R using the caret package," *J. Stat. Softw.*, vol. 28, pp. 1–26, 2008.

[146] J. E. Shriki, K. S. Surti, A. F. Farvid, C. C. Lee, S. Samadi, J. Hirschbeinv, *et al.*, "Chemical shift artifact on steady-state free precession cardiac magnetic resonance sequences as a result of lipomatous metaplasis: a novel finding in chronic myocardial infarctions," *Can. J. Cardiol.*, vol. 27, no. 5, p. 664.e17-664.e23, 2011.

[147] J. W. Goldfarb, J. Mclaughlin, C. A. Gray, and J. Han, "Cyclic cine-balanced steady-state free precession image intensity variations: implications for the detection of myocardial edema," vol. 33, no. 3, pp. 573–581, 2011.

[148] T. A. Stromp, S. W. Leung, K. N. Andres, L. Jing, B. K. Fornwalt, R. J. Charnigo, *et al.*, "Gadolinium free cardiovascular magnetic resonance with 2-point Cine balanced steady state free precession," *J. Cardiovasc. Magn. Reson.*, pp. 1–11, 2015.

[149] D. Giavarina, "Understanding Bland Altman analysis," *Biochem. Medica*, vol. 25, no. 2, pp. 141–151, 2015.

[150] S. L. Salzberg, "On comparing classifiers: pitfalls to avoid and a recommended approach," *Data Min. Knowl. Discov.*, vol. 1, pp. 317–328, 1997.

[151] V. Bodí, O. Husser, J. Sanchis, J. Núñez, M. P. López-Lereu, J. V Monmeneu, *et al.*, "Contractile reserve and extent of transmural necrosis in the setting of myocardial stunning: comparison at cardiac MR imaging.," *Radiology*, vol. 255, no. 3, pp. 755–763, 2010.

[152] V. Bodi, J. V Monmeneu, J. T. Ortiz-Perez, M. P. Lopez-Lereu, C. Bonanad, O. Husser, *et al.*, "Prediction of reverse remodeling at cardiac MR imaging soon after first ST-segment-elevation myocardial infarction: results of a large prospective registry.," *Radiology*, vol. 278, no. 1, pp. 54–63, 2016.