Claudia Ximena Mazo Vargas

AUTOMATIC CLASSIFICATION OF HISTOLOGICAL IMAGES

AND HISTOLOGICAL KNOWLEDGE MODELLING OF THE HUMAN CARDIOVASCULAR SYSTEM

University of Valle | University of León

DOCTORAL DISSERTATION, 2016

universidad de león

UNIVERSITY OF VALLE

FACULTY OF ENGINEERING, SYSTEMS ENGINEERING AND COMPUTER SCHOOL

UNIVERSITY OF LEÓN

DEPARTMENT OF ELECTRICAL, SYSTEMS AND AUTOMATIC ENGINEERING

AUTOMATIC CLASSIFICATION OF HISTOLOGICAL IMAGES AND HISTOLOGICAL KNOWLEDGE MODELLING OF THE HUMAN CARDIOVASCULAR SYSTEM

Thesis supervised by

PROF. DR. MARIA TRUJILLO,

AND PROF. DR. ENRIQUE ALEGRE,

advised by

PROF. MSC. LILIANA SALAZAR

and submitted by

CLAUDIA XIMENA MAZO VARGAS

in fulfillment of the requirements for the Degree of

ENGINEERING DOCTOR (ENGD)

León, September 2016

universidad de león

UNIVERSIDAD DEL VALLE

Facultad de Ingeniería, Escuela de Ingeniería de Sistemas y Computación

UNIVERSIDAD DE LEÓN

Departamento de Ingeniería Eléctrica y de Sistemas y Automática

Clasificación Automática de Imágenes Histológicas y Modelado del Conocimiento Histológico del Sistema Cardiovascular Humano

Tesis doctoral dirigida por

Prof. Dr. Maria Trujillo,

y Prof. Dr. Enrique Alegre

Asesorada por

Prof. MSc. Liliana Salazar

y desarrollada por

Claudia Ximena Mazo Vargas

a fin de optar al grado de

Doctor en Ingeniería por la Universidad del Valle y la Universidad de León

León, Septiembre de 2016

Abstract

This thesis proposes firstly, a method to recognise automatically fundamental tissues of the human cardiovascular system using image processing techniques with morphological information. Secondly, a method to classify automatically fundamental tissues and organs using machine learning algorithms based on texture features. Thirdly, a histological ontology to model histological and expert knowledge. Fourthly, two methods to improve the previous classification using the information provided by the histological ontology.

The human body studied through histology is composed of four fundamental tissues: epithelial, connective, muscle and nervous. An organ can be identified using fundamental tissues and analysing some particular criteria and functions which define spatial structures among fundamental tissue relations. Identifying fundamental tissues and organs out of histology images is still an open problem in Computer Vision. Two main challenges are faced during the process: information loss and knowledge representation. On one side, 2D samples of a 3D structure are taken on the acquisition process. Thus, inferring information from a 2D representation of a 3D organ is an inverse problem. Furthermore, the lack of information regarding the physiological functions of an organ results in unstable solutions with respect to measurement errors and, hence, the problem becomes ill-posed. On the other side, histological knowledge formalisation and its processing within machines in a machine-interpretable form to solve complex tasks such as support teaching, medical practices or having natural language interactions are some of the main challenges.

In this dissertation, firstly we present an approach for the automatic recognition of fundamental tissues of the cardiovascular system — epithelial, connective and muscle —, using morphological information. Loose connective tissue, light regions and cell nuclei are recognised on $40\times$ images. In a similar way, light regions, loose connective and muscle tissues are recognised on $10\times$ images. Finally, the tissue's function and composition are used to refine muscle tissue recognition. Experimental validation is then carried out by histologists using manually annotated images that are used as a ground-truth. The proposed automatic recognition approach provides for epithelial tissues a sensitivity of 0.79 for cubic, 0.85 for cylindrical and 0.91 for flat. Furthermore, the experts gave our method an average score of 4.85 out of 5 in the recognition of loose connective tissue and 4.82 out of 5 for muscle tissue recog-

nition.

In the same line of work, later we introduce an approach for the automatic classification of cardiovascular tissues using texture information and a cascade Support Vector Machine (SVM) classifier. Additionally, we used the same pipeline to recognise some cardiovascular organs. The best results were achieved using a concatenation of LBP and LBPri with a cascade SVM yielding a hit rate of 90%. The obtained results yielded 0.9385 F-Score using the cascade SVM, outperforming other classifiers, including Random Forest (RF) that obtained 0.870 F-Score and Linear Discriminant Analysis (LDA) that achieved 0.823 F-Score. We also evaluated the classification of tissues and organs in a histological image using the block-based recognition method. The accuracy obtained with a complete histological image is over 70% compared with 90% obtained with the block proposal, this result is significant considering that the latter contain only one type of tissue per block.

On the other hand, we created a human histological ontology using OWL and Protégé in order to formalise human knowledge and its processing within machines in a machine-interpretable form to solve complex tasks. We propose a four-fold approach to validate our ontology, as follows: firstly, pitfalls on the ontology were detected using the web tool *OOPS!*. Secondly, an assessment by experts is performed using the Conceptual Models (CMs) through a survey of two different sets of experts. The positive results we obtained makes it possible to made publicly available the ontology. Furthermore, to validate the ontology, we conducted surveys with two sets of experts. The first survey was taken by 20 students of Medicine and Surgery in semester 5 at the University of Valle. The second survey was taken by 51 experts in Latin America from different specialties and 32 of them have over 10 years of experience. Thirdly, we used the competency questions (CQ) to verify its answers using SPARQL queries and according to the answers obtained the task was completed successfully. Fourthly, we could verify that the ontology is a heavyweight one considering it was enriched with axioms used to fix the semantic interpretation of concepts and relations.

Furthermore, we propose a process to improve the classification based on the information provided by the histological ontology. This proposal improves the results of automatic classification and recognises the epithelial tissue on $10\times$ images. The obtained results are two-fold to improve the process and classification of epithelial tissue as follows: (i) the organ classification improvement process reclassified correctly between 1 and 24 blocks per image. Additionally, the improvement process increases hit rates classification in all cases, between 0.333% and 23.188%, regarding to the histological image and its automatic classification. (ii) During the improvement process the recognition of epithelial tissue reclassified correctly between 0 and 7 blocks per image increasing the classification hit rates, from 0% and 2.333%, with respect to the tissue area. This improvement is expected since the case of 0% occurs when the image does not contain the tissue and the percentage is low since epithelial

tissue area is commonly small. We concluded that the improvement process enabled us to infer which type of epithelium is present in a sample.

Resumen

En esta tesis se propone en primer lugar, un método que permite reconocer automáticamente los tejidos fundamentales del sistema cardiovascular humano utilizando técnicas de procesamiento de imágenes con base a la información morfológica. Además, se presenta un método para clasificar automáticamente los tejidos y órganos usando algoritmos de aprendizaje automático basados en características de textura. En tercer lugar, se ha creado una ontología histológica para modelar el conocimiento histológico y de los expertos. Finalmente, se presentan dos métodos para mejorar el proceso de clasificación que permiten incrementar los resultados obtenidos utilizando la información proporcionada por la ontología histológica.

El cuerpo humano estudiado a través de la histología se compone de cuatro tejidos fundamentales: epitelial, conectivo, muscular y nervioso. A su vez, un órgano puede identificarse utilizando los tejidos fundamentales y el análisis de criterios y funciones particulares que definen las estructuras espaciales entre ellos. La identificación de los tejidos fundamentales y órganos a traves de imágenes histológicas es todavía un problema abierto en visión por computador, enfrentando principalmente dos retos durante el proceso: la pérdida de información y la representación del conocimiento. Por un lado, en el proceso de adquisición se capturan muestras 2D de una estructura en 3D; por lo tanto inferir información a partir de una representación 2D de un órgano 3D es un problema inverso. Además, la falta de información con respecto a las funciones fisiológicas de un órgano da como resultado soluciones inestables con respecto a los errores de medición y, por tanto, nos enfrentamos a un problema mal planteado. Por otro lado, algunos de los principales desafíos son la formalización del conocimiento histológico y su procesamiento dentro de las máquinas para resolver tareas complejas como el apoyo a la enseñanza, las prácticas médicas y la interacción con lenguaje natural.

En este trabajo presentamos un enfoque para el reconocimiento y clasificación automática de los tejidos fundamentales del sistema cardiovascular — epitelial, conectivo y muscular —, utilizando la información morfológica. En las imágenes a $40\times$ se reconocen el tejido conectivo laxo, las regiones de luz y los núcleos celulares. De manera similar, en las imágenes a $10\times$ se reconocen las regiones de luz y los tejidos conectivo laxo y muscular. Por último, la función y la composición del tejido muscular se utilizan para mejorar su reconocimiento. La validación experimental se llevó a cabo mediante imágenes etiquetadas de forma manual que se utilizaron como

ground-truth. El enfoque del reconocimiento automático propuesto clasifica los tipos de tejidos epiteliales con una sensibilidad de $0{,}79$ para cúbico, $0{,}85$ para cilíndrica y $0{,}91$ para plano. Por otra parte, los expertos evaluaron el método propuesto con una puntuación media de $4{,}85$ de 5 en el reconocimiento de tejido conectivo laxo y $4,82$ de 5 para el reconocimiento de tejido muscular.

En la misma línea de trabajo, presentamos un enfoque para la clasificación automática de los tejidos cardiovasculares utilizando la información de textura y un clasificador con máquinas de soporte vectorial (SVM) en cascada, así mismo utilizado este método se logran reconocer algunos órganos. Los mejores resultados se obtuvieron usando una concatenación de LBP y LBPri con un SVM en cascada, obteniendo una tasa de acierto de $90\,\%$. Los resultados obtenidos arrojaron un F-Score de $0{,}939$ utilizando SVM en cascada, superando a otros clasificadores como *Random Forest* (RF) que obtuvo un F-Score de $0{,}870$ y Análisis Discriminante Lineal (LDA) que alcanzó un F-Score de $0{,}823$. También se evaluó la clasificación de los tejidos y órganos en una imagen histológica completa usando el método de reconocimiento basado en bloques, descrito anteriormente. La medida de la precisión obtenida es superior al $70\,\%$ en comparación con el $90\,\%$ arrojado en las pruebas de la propuesta de bloques, lo cual es significativo teniendo en cuenta que estos últimos contienen solo un tipo de tejido por bloque.

Por otro parte, hemos creado una ontología histológica humana utilizando OWL y Protégé. Se propone un enfoque de cuatro etapas para validar la ontología propuesta, de la siguiente manera: primero, se realiza una detección de errores y fallos utilizando la herramienta web *OOPS!*. Segundo, se realiza una evaluación por parte de expertos utilizando los Modelos Conceptuales (CM) por medio de una encuesta a dos grupos diferentes de expertos. Los resultados positivos obtenidos hacen que sea posible publicar la ontología. Las encuestas se realizaron con dos grupos de expertos: 20 estudiantes de Medicina y Cirugía de 5 semestre de la Universidad del Valle y 51 expertos de América Latina con diferentes especialidades de los cuales 32 tienen más de 10 años de experiencia. En tercer lugar, hemos utilizado las cuestiones de competencias (CQ) para verificar sus respuestas usando consultas SPARQL y de acuerdo con los resultados obtenidos la tarea se completó con éxito. Finalmente, nosotros verificamos que la ontología es *heavyweight* teniendo en cuenta que contiene axiomas utilizados para la interpretación semántica de los conceptos y las relaciones.

Además, en esta tesis, se propone un proceso de mejora de la clasificación basado en la información obtenida de la ontología histológica. Está propuesta mejora los resultados de la clasificación automática y reconoce el tejido epitelial en imágenes a $10\times$. Los resultados obtenidos para mejorar el proceso de clasificación y lograr el reconocimiento del tejido epitelial son: (i) se reclasifican correctamente entre 1 y 24 bloques por imagen; además el proceso de mejora de la clasificación aumenta la tasa de éxito en todos los casos, entre $0{,}333\,\%$ y $23,188\,\%$, de acuerdo con la imagen

histológica y su clasificación automática. (ii) Durante el proceso de mejora de la clasificación permite el reconocimiento del tejido epitelial reclasifica correctamente entre 0 y 7 bloques por imagen incrementando las tasas de acierto de la clasificación, entre $0\,\%$ y $2,333\,\%$, dependiendo de la zona que contiene el tejido epitelial. Los casos con una mejora del $0\,\%$ se producen cuando la imagen no contiene el tejido epitelial y por otra parte el incremento de la tasa de éxito es bajo debido a que el área de este tejido es normalmente reducida. Llegamos a la conclusión de que el proceso de mejora de la clasificación utilizando la ontología histológica permite inferir el tipo de epitelio que está presente en una muestra.

Contents

List of Figures

List of Tables

Índice general

Anexo D: Resumen de la Tesis en Castellano

Acknowledgements

A doctoral thesis is sometimes portrayed as a solitary endeavour; however the long list that follows absolutely proves the opposite. Many are the people and institutions that have collaborated to allow a successful end to these studies.

I am thankful to my supervisors Maria Trujillo, from University of Valle, and Enrique Alegre, from University of León, and my advises Liliana Salazar, from University of Valle, for giving me the opportunity to carry out my doctoral studies under their supervision and guidance. I would like to thank you for encouraging my research and for allowing me to grow as a research scientist. Your advice on both research as well as on my career have been priceless. I am deeply grateful for your continuous support, insight and patience. Your invaluable guidance has made possible this thesis work.

I am grateful to Oscar Corcho for his supervision and guidance about ontologies and to the committee members for your thorough review of this thesis. A sincere thanks to the administrative stuff for easing my life at University. The agreement between both Universities would have never been possible without your help and hard work.

This work would not materialised without COLCIENCIAS and AUIP for letting me enjoy a grant destined to fund EngD students and international academic mobility.

Thanks to David Vescio for helping me review my use of English in some parts of my essay and to Frédéric Commandeur, for helping me with his recent experience in your EngD.

I feel fortunate and honoured to have had the opportunity to share my academic environment surrounded by great scientists and teachers in an enjoyable atmosphere. Many thanks to Martha Millan, Ivan Cabezas and many more.

I am very grateful to my lab colleagues in Colombia for the stimulating discussions, for the sleepless nights we were working together before deadlines, and for all the fun we have had in the last five years have helped me during good and hard times of this thesis: Deisy, Jose, MariaC, Jhonma, Cristina, and et al. I feel very fortunate and happy that the great people of work colleagues have become my lovely

family in León: Maite, Laura and Óscar, thanks for teaching me to be more Spanish using regionalisms. I would like to express my special appreciation to each member of the research group Teblami from the University of Valle. A special thanks to Jaime Fuente, Neco and Jhon Osorio who always have words of encouragement for me and the goals I set for myself, growing my confidence. To my step-family in León: Marta, Claudia, Adriana, José, Isidoro and Carla for making me feel the warmth of family. To my dear friends Ynn, Vile, Cami, Jean, Ricardo and everyone from Colombia and Spain, because in many occasions this work has drift me apart from you. Thanks for your support and friendship.

Words cannot express how grateful I am to my parents Fanny Vargas and Yesid Mazo, for their love, dedication and education they gave me which helped reach my goal. I dedicate this paper to you. *Mil gracias por todo el amor, la confianza, la motivación y por la educación que me han proporcionado. Este logro es de y para ustedes.* I thank my sister Estefania and my brothers Yesid, Christian and Victor for all their support and belief that I was capable of taking this and many more challenges. I also want to thank my nephews Sebastian and Santiago, who taught me to feel a matchless love, to express my feelings in every moment and to better myself every day for them; and in general, for their enduring support and patience, in all nights, weekends, and special dates that I spent with my laptop instead of with them.

Finally, to numerous questions about my future academic endeavours from family and friends, I shall answer in the words of Sir Winston Churchill: *"Now, this is not the end. It is not even the beginning of the end. But it is, perhaps, the end of the beginning"*.

Claudia Ximena Mazo
León
14th September 2016

Chapter 1

Introduction

When you are finished changing, you are finished.
Benjamin Franklin

1.1. Motivation

A computer vision problem, with histology images, consists of identifying the fundamental tissues and recognising distinctive patterns formed by spatial structures between them to infer an organ. It has been a challenge for researchers and still remains as an open problem (Zhao et al., 2005). Since an organ is a 3D structure and an image is a 2D representation of an organ sample, inferring information from a 2D representation of a 3D organ is an inverse problem (Hansen, 2010). Considering that information about the physiological functions of an organ is not contained in an image; this lack of information makes the solution unstable and makes the problem an ill-posed one.

Moreover, a histological knowledge representation and its processing within machines in a machine-interpretable form to solve complex tasks such as support teaching, medical practices or having natural language interactions are some challenges that must be confronted in this context.

This dissertation presents automatic classification of histological images and histological knowledge modelling through four different proposals:

- Recognising fundamental tissues — epithelial, connective and muscle — using image processing techniques.
- Classifying cardiovascular tissues — epithelial, connective and muscle — and organs — the heart, the muscular artery, the elastic artery and the large vein — using Local Binary Pattern (LBP) based descriptors and a cascade Support Vector Machine (SVM).
- Building a histological ontology.
- Improving the automatic classification using the ontology of the human cardiovascular system.

These proposals may be helpful to reinforce the learning process of histologists, biologists, pathologists, and those in related disciplines (Izet, 2008). Automated tissue recognition may increase the number of cases that a student may analyse, promoting self-learning to on-campus students and also facilitating on-line learning to external or remote students through E-Learning systems (Ruiz et al., 2006). This all translates into better-formed professionals without any great social or economic investment required. Another benefit of such systems is that they could allow the automatic annotation of large repositories or sets of images obtained by digital technology — a digital camera is connected to a microscope to capture images — available in hospitals or distributed through the storage devices of histologists. Automatic labelling solves some problems present in manual annotation, such as subjectivity, time costs, difficult, and impracticality (Hernandez et al., 1990).

Finally, the two aforementioned applications are highly relevant tasks that pose challenging current computer vision problems. The motivation of each application is presented in the next sections.

1.1.1. Recognition and Classification of Fundamental Tissues and Organs

In histology, an organ has complex spatial structures. Figure 1.1 shows the spatial structures of four different organs: the carotid artery, the coronary artery, the heart and the large vein. It can be observed how the fundamental tissues interact to form specific and distinctive patterns in each organ. Additionally, it is important to note that the differences observed are achieved with the same organ by changing the capture zone or sample. Figures 1.1(a) and 1.1(b) show some specific patterns of the arteries when separated into three very definite layers: a thin layer near the light, a compact and succinct layer which is thicker in the middle than the other two layers and finally a porous layer with more separate structures. Figure 1.1(c) shows a homogeneous and compact structure. Figure 1.1(d) show some specific patterns of the large vein when separated into three very definite layers similar to artery show in cases (a) and (b) with two particular differences: the layer in the middle is thinner than arteries and the porous layer is thicker than arteries.

This challenge has been studied for histological images using techniques based on texture features to approximate the distribution of texture patterns Zhao et al. (2005), supervised learning strategies combining colour and texture features Herve et al. (2011), Markov model to to automatically learn and characterize the semantic context of histological images Yu et al. (2008), among others. Additionally, cell, tissue and organ recognition has been presented in many works. However, the main difference, to the best of our knowledge, is that healthy fundamental tissues have not been segmented or classified and there are not works classifying healthy tissues of the cardiovascular system. The computer-aided recognition of fundamental

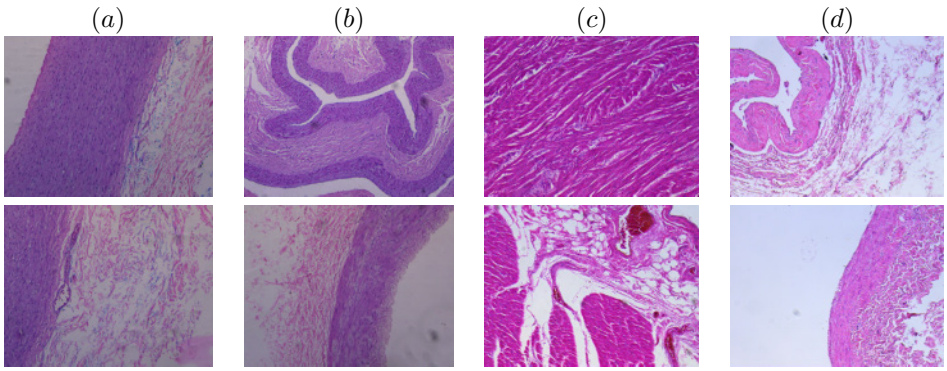


Figure 1.1: Patterns in different organs of the cardiovascular system. (a) The carotid artery. (b) The corony artery. (c) The heart. (d) The large vein.

tissues and organs encounters several problems due to the hard boundary among fundamental tissues, and the low definition in image areas such as at the edges of the tissues and organs. In this work, we present two different approaches where the recognition and classification of fundamental tissues are carried out using different techniques based on image processing, texture features, machine learning algorithms for pattern recognition, and computational learning.

1.1.2. Improving the Automatic Classification of Histological Images using an Ontology of the Human Cardiovascular System

Knowledge representation makes possible to describe information about the real world in a form such that a computer system can use it to solve complex tasks. In this case, it is necessary to incorporate how humans solve problems and represent knowledge in order to incorporate formalisms. One way to achieve this goal is to create a histological ontology. A histological ontology would allow: common understanding of the structure of information among people or systems, histological and expert knowledge reuse and analysis — assumptions, deduction and reasoning. Ontologies enable the reuse of domain knowledge, thus making domain assumptions explicit while clarifying any ambiguities. Histological ontology construction is a complex task which requires the collaboration of both engineers and domain experts. The complexity of the medical domain and the formal description languages makes this collaboration necessary.

On the other hand, heterogeneous data sources produces different types or representations of data that cannot operate or be treated in the same way. In this work, we present a histological ontology and an automatic classification of tissues and organs. The main objective is to use the histological ontology to improve the auto-

matic classification and achieve lower margins of error or uncertainties than those obtained when using a single information source. This work constitutes a new way to perform multiple heterogeneous information integration in the histological context in order to identify patterns and infer the organisation of the fundamental tissues. This would also allow the identification of specific organs.

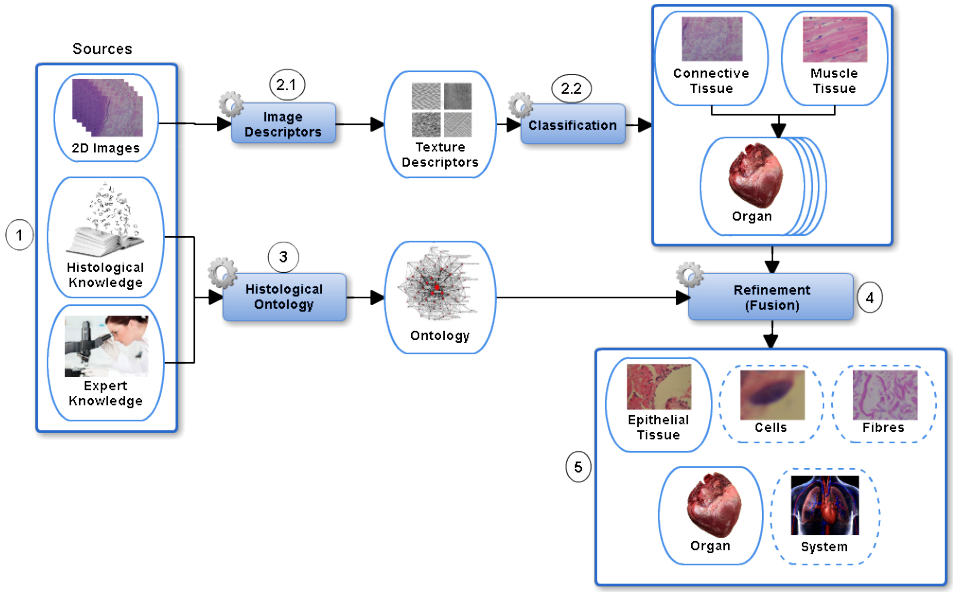


Figure 1.2: Improving the automatic classification of histological images using an ontology of the human cardiovascular system process diagram. (1) Input data source. (2.1) Image description process using texture features. (2.2) Automatic classification of the fundamental tissues and organs process. (3) Process to build a histological ontology. (4) Improvement and classification of epithelial tissue processes. (5) Final classification.

1.2. Objectives

The main goal of this work is to propose and evaluate an automatic method to classify fundamental tissues and modelling the histological and experts' knowledge upon the human cardiovascular system.

Given the previous general goal, we defined the following particular objectives:

1. To propose an automatic method to recognise fundamental tissues using image processing techniques and tissue morphological information.

2. To propose an automatic block-based classification method of cardiovascular tissues and organs using texture descriptors and machine learning algorithms.

3. To propose a representation of histological and expert knowledge using models that capture essential features of relations and their usefulness.

4. To combine the automatic classification based on texture descriptor with the histological ontology to improve the obtained results.

1.3. Main Contributions

The main contributions of this dissertation are summarised as follows:

1. *An approach to automatic recognition and classification of fundamental tissues, using morphological information*. Its output is epithelial — flat, cubic and cylindrical —, loose connective and muscle tissues recognition.

2. *A method to classify fundamental tissues based on its texture information*. Its output is cardiac muscle of heart, loose connective tissue — vein, arteries and heart — and smooth muscle of muscular artery, large vein and elastic artery with success rates greater than 90%

3. *A histological ontology of the human cardiovascular system*.

4. *A method that improves the previous automatic classification using the knowledge contained in the human histological ontology created*. This solution increase hits rates and decrease misses of the automatic classification process.

1.4. Thesis Organisation

This introductory chapter has been focused on motivating the work presented, its main objectives and its original contributions. The remaining chapters of this thesis are organised as follows:

In Chapter 2, the state-of-the-art about recognition of fundamental tissues and organs as well as histological ontologies is reviewed. Firstly, some published methods that deal with recognition of fundamental tissues and cell nuclei are presented. Secondly, the state-of-the-art that evaluate organ identification are discussed. Thirdly, it presents published histological ontologies in order to assess if there are parts of them that may be reusable.

Chapter 3 presents a method to recognise cardiovascular tissues using morphological information. Epithelial (flat, cubic and cylindrical), loose connective and muscle tissues are recognised. The K-means algorithm along with information coming from the structural tensor, red and green colour channels, Feret's diameter and spatial projection are considered within the proposed method which is shown in

this chapter. The proposed method was later on evaluated by histologist using an experimental validation.

Chapter 4 presents an approach to automatically classify cardiovascular tissues using texture information and a cascade SVM classifier. Additionally, some cardiovascular organs are recognised through the same process. The concatenation of LBP and LBPri was selected to describe the tissues after evaluating different texture features. Moreover, SVM was chosen for the classification after assessing different classifiers, such as RF and LDA. The best classification results were achieved using a cascade SVM, firstly with a linear kernel and after with a polynomial kernel.

Chapter 5 contains firstly, a human histological ontology of the cardiovascular system to represent histological and expert knowledge. Secondly, a classification improvement process using the histological ontology.

Chapter 6 presents the conclusions of this thesis and gives an outlook of future working lines to extend this work.

Regulations regarding doctoral studies at the University of León claim that if a thesis is not written in Spanish, at least the table of contents, conclusions, and an abstract of each chapter must be written in Spanish. In order to comply with these regulations, we include a Spanish version of the conclusions in Chapter 7, and a summary of all chapters in Part II.

Chapter 2

State-of-the-Art

Research is seeing what everybody has seen and thinking what nobody has thought yet.
Albert Szent-Györgyi

In previous decades, there has been substantial work in the computer vision field that tackles the problem of recognition of fundamental tissues and organs in histological images. On the other hand, knowledge representation research has allowed to obtain more complete solutions in the medical area.

This chapter presents a review of the state-of-the-art of the following research lines: (i) feature selection and machine learning algorithms in histological images; (ii) recognition of fundamental tissues including recognition of cell nuclei necessary for epithelial tissue identification; (iii) organ identification; (iv) ontology of human histology and taxonomies to model the histological knowledge; and (v) improving the classification based on ontology.

2.1. Feature Selection and Machine Learning Algorithms applied to Histological Images

Two important steps in image processing techniques are feature extraction and classification being these ways for understanding human perception. Humans have innate abilities to process and understand imagery, though they do not tend to explain how they reach their decisions. There are many works dedicated to object recognition for Content Based Information Retrieval (CBIR) on histological images. Some of them using techniques based on: (i) texture, many works have compared several feature types on healthy and pathological histological images demonstrating a clear improvement of the algorithms performance and classification process (Markkongkeaw et al., 2013; Peyret et al., 2015; Lai et al., 2011); (ii) shape-based features may capture the particular characteristics to recognise cells or some pathologies — tumours or cancer —, some results suggest that these shapes are not accurate diagnostic features but it can be used for specific aims (Kothari et al., 2013; Melnyk, 2015; Sharma et al., 2012); (iii) colour representation may be used for diagnosis diseases and as an sign to recognise cells or tissues. However, the colour of histological images may vary according to the stain, the preparation procedure

of the tissue and image capture (Claridge et al., 2002; Kothari et al., 2011; Nedzved and Starovoitov, 2010); (iv) edge detection is useful to obtain nuclear morphometric features such as shape of nuclei and diagnosing of some pathologies. However, this technique encounters several problems due to the low definition in some edges of the tissues and organs (Alén et al., 2006; Fujii et al., 2013; Fakhrzadeh et al., 2013). On the other hand, classification and recognition process are based on: (i) clustering methods such as K-means, mean shift, K-nn, among others are employed for histological image segmentation (Wu et al., 2015; He et al., 2011); (ii) SVM, Bayesian networks and another machine learning algorithms are used to characterise and classify healthy and pathology cells, tissues and organs (d. A. Zampirolli et al., 2010; Krishnan et al., 2010; Veillard et al., 2012); (iii) neural networks are used to the classification task producing good results on automatically learned features (Bevilacqua et al., 2015; Jitaree et al., 2013; Kashif et al., 2016); (iv) learning based approach using bag-of-words model, this method is attractive as it offers good classification accuracy at low computation cost than the texture-feature-based methods (Galaro et al., 2011; Nguyen et al., 2015; Cheng et al., 2012). However, some of these works are focus in histopathologic images which have different characteristics to healthy tissues.

In Table 2.1, some features that have been used in histopathology images Gurcan et al. (2009) are presented. Large feature sets are generated in the hope that some subset of features incorporates the information used by the human expert for analysis. Therefore, many of the generated features could be redundant or irrelevant. A feature selection method may require an exhaustive search, which is not practical for a large set of features generated from a large dataset. Therefore, several heuristic algorithms have been developed which use classification accuracy as the optimization criterion. Sequential Forward Selection (SFS) (Cateni and Colla, 2015) and Sequential Backward Selection (SBS) (Pudil et al., 1994) are widely used. SFS works by sequentially adding the feature that most improves classification; similarly, SBS begins with the entire feature set and sequentially removes the feature that most improves classification. Both SFS and SBS suffer from the "nesting effect" whereby features that are selected (SFS) or discarded (SBS) cannot be revisited in a later step and are thus sub-optimal. Other algorithms are also used as genetic algorithms such as simulated annealing, boosting and grafting (Gurcan et al., 2009). After feature selection, a dimensionality reduction is applied taking into account three well-known and commonly used methods: Principal Component Analysis (PCA), Independent Component Analysis (ICA), and Linear Discriminant Analysis (LDA). Dimensionality reduction can bring an improved understanding of the data apart from a computational advantage.

In Caicedo (2011) a short summary of the main algorithms and strategies for feature extraction in histology images is presented. This literature review covers histological images such as cervical tissue, skin tissue, gastrointestinal tract tissue,

Table 2.1: Summary of object-level features used in histopathology image analysis (Gurcan et al., 2009)

Category	Features
Size and shape	Area
	Elliptical features: major and minor axis length, eccentricity, orientation, elliptical deviation
	Convex hull features: convex area, convex deficiency, solidity
	Filled image features: filled area, euler number
	Bounding features: perimeter, radii, perimeter Fourier energies, perimeter curvature, bending energy, perimeter fractal dimension
	Other shape features: equivalent diameter, sphericity, compactness, inertia shape
	Center of mass
	Reflection symmetry
Radiometric and densitometric	Image bands, intensity
	Optical density, integrated optical density, and mean optical
	Hue
Texture	Co-occurrence matrix features: inertia, energy, entropy, homogeneity, maximum probability, cluster, cluster shade
	Fractal dimension
	Run-length features: short runs emphasis, long runs emphasis, gray-level non-uniformity, run-length non-uniformity, runs percentage, low gray-level run emphasis, high gray-level runs
	Wavelet features: energies of detail and low resolution images
	Entropy
Chromatin-specific	Area, integrated optical density, mean optical density, number of regions, compactness, distance, center of mass

neural tissue and prostate tissue. This paper is divided into six parts for analysis and identification of histological images:

1. *Image Segmentation:* the first step in the automatic analysis of histology images. Some of the algorithms used are: block based (split histology images) (Diamond et al., 2004a; Yu and Ip, 2008) and region based (colour analysis) (Doyle et al., 2007).

2. *Feature Extraction:* presents the most used colour (Kong et al., 2009a) and texture features, such as Haralick texture features (Canada et al., 2008), Gabor

features (Yu and Ip, 2008), Tamura texture (Orlov et al., 2008) and grey statistics (Caicedo et al., 2009).

3. *Architecture:* the architectural features for histology images have been designed to measure the spatial arrangement of all the objects in the picture. These algorithms are the Voronoi Diagram (Doyle et al., 2007), the Delaunay Triangulation (**?**), fractal dimension (Tambasco et al., 2009) and topology features (Sertel et al., 2008).

4. *Morphology:* pathologists usually describe the characteristics many gland or cell characteristics in terms of morphology and, hence, this information is essential for obtaining good results (Diamond et al., 2004a; Doyle et al., 2007; Sertel et al., 2008).

5. *Image Transforms:* in this part basic image transforms are used such as the Fourier, Chebyshev and Wavelet transforms to represent image contents for different classification tasks obtaining a variety of image content descriptors (Orlov et al., 2008).

6. *Image Representation:* after the previous five steps it is necessary to apply decision rules, learning algorithms or similarity measures in the histology images for its representation (Diamond et al., 2004a; Sertel et al., 2008; Yu and Ip, 2008).

However, this work does not cover the cardiovascular system which is one of the more complex systems to be treated because of its differences with other systems and the peculiarities of its organs — due to the differences between spatial relations of fundamental tissues on each tubular and visceral organ. In addition to this, it is not as explicit as the morphological information which is used as part of the image processing.

In Cruz et al. (2011) an histological image repository was developed, accessible by RENATA (National Academic Network of Advanced Technology available at http://www.renata.edu.co/). This system allows us to study these images in regard to the four fundamental tissues: epithelial, connective, muscular and nervous. This project contains a module for image retrieval by content, which is composed of feature extraction as colour, borders and textures, and the similarity calculation used to sort the search results. This work was evaluated with 11284 histological images, obtaining an accuracy between 67% and 80%. However, it is also important to note that the use of colour as one of the descriptors of the images may limit the application of the approach, since the colour of images of histology may vary according to the stain used.

The objective of Orlov et al. (2009) was to compare effectiveness of features derived from several image transforms to those derived from several filters in classi-

fying a benchmark set of biological images as well as two sets of images related to cancer diagnosis using H&E-stained biopsies. A two-stage method was employed in which types of derived images were used as inputs for a bank of feature extraction algorithms. The approach takes into account descriptors based on polynomial coefficients (they used Chebyshev, Chebyshev-Fourier, and Zernike polynomials), textures (Haralick, Gabor, and Tamura families were computed), and several multi-purpose families, including Radon, singular values, multi-scale histograms, moments calculated from a four-directional comb filter, edge and blob statistics. For each feature a ranking scheme is employed based on the Fisher score. The classification of a malignant tissue was made using two datasets, the first set is a representative collection of lymph node biopsies from three types of malignancies and the second dataset consisted of benign, primary and five secondary melanoma tumours. The highest classification rate achieved in the first set was 0.97 and in the second set 0.93. This work shows that image transforms perform with a consistently higher accuracy than image filters, possibly because transform-derived features represent more varied image content compared to filter-derived features. However, assessment those algorithms cannot be carried out since no details are given about the set on which the tests were performed.

In Meng et al. (2010) a framework is proposed based on the Collateral Representative Subspace Projection Modeling (CRSPM) supervised classification model for general histology image classification. In the proposed framework, a cell image is first divided into 25 blocks, nine overlapping blocks. The reason for the overlapping blocks are included is to recover the information in the boundary area near the dividing lines between neighbouring non-overlapping blocks. The information in these areas may be missing if only non-overlapping blocks are used. To reduce the spatial complexity of computation, a C-RSPM model is built on each block set which contains blocks in the same location from different images. This model is constructed with a set of 505 features, including colour and texture features which are extracted from every block – colour dominant (16 features), colour histogram (51 features), colour moment (108 features), edge histogram (47 features), texture Co-occurrence (36 features), texture Wavelet (219 features), texture Tamura (3 features), texture Gabor (24 features), and LBP (1 feature). For each testing image, the proposed framework first classifies each of its blocks using the C-RSPM classification model built for that block set, and then applies a multimodal late fusion algorithm with a weighted majority voting strategy to decide the final class label of the whole image — chronic lymphocytic leukemia, follicular lymphoma, and mantle lymphoma. Experimenting using three-fold cross validation with three benchmark histology data sets shows that the proposed framework outperforms other well-known classifiers with 92.7% in the comparison with the highest accuracy reported with 85%. However, the considered datasets have similarities between them, the technique generates good results for these matches between histological images.

In Caicedo et al. (2011) a system of annotation and retrieval of histological images for query by example or by semantic concepts is proposed. The problem is limited to images used for the diagnosis of one particular type of cancer known as basal cell carcinoma. The semantic concepts used are 18 histological terms. The main steps of the proposal are extraction of multiple features, combined with features based on kernel methods theory, and the automatic image annotation. For the extraction of different visual characteristics, seven features spaces have been selected: grey scale histogram, invariant feature histogram, LBP, RGB colour histogram, bag SIFT features histogram and sobel Tamura texture histogram. The combination of features for kernel-based methods was performed using SVM. A weight is assigned to each histological concept regarding its discriminatory power in the seven characteristics with different kernels. According to their tests, the histogram intersection kernel yields better results than other kernels used. The features with higher discriminatory power were LBP, SIFT and Tamura texture histogram. The percentages of accuracy obtained in automatic annotation of histopathology images ranging between 44% and 77% depending on the evaluated term. However, the results obtained in this study does not contain annotations to differentiate between normal and abnormal images.

In Canada et al. (2011) a CBVIR system is proposed for automatic annotation of images with histological abnormalities in the eye of zebrafish larvae. The proposed method covers several stages, initially performing a pre-processing of the images that are to be used before performing a removal of the organ of interest, namely the eye of the larvae. Considering the image of the organ of interest, the image is divided into a series of 64×64 pixel blocks in order to identify anomalies. For each image block, a total of 54 features are extracted, consisting of a combination of: grey-level co-occurrence features, Lacunarity, grey-level morphology feature, Markov Random Field model parameters and Daubechies wavelet packed feature. For training and testing purposes a total of 176 eye images were manually extracted, 100 to training and 76 to testing. The results of the application are given using the correct, incorrect or acceptable labels. In different types of test accuracy results from a 62% to 98% acceptable counting were obtained. Although, it is important to note that the number of images may not be sufficient to be useful as a completely automated — the training process can be sensitive to the number of images used —, this work will require improvements in accuracy, precision, and speed.

2.2. Recognition of Fundamental Tissues

Automatic recognition of healthy and pathological fundamental tissues has been addressed using techniques based on the spatial organization inter and intra regions using a Bag of Words model (Garnier et al., 2014), segmentation applying iterative

edge labeling and using Voronoi diagram (Wang et al., 2014), adaboost, support vector machine, random forest and convolutional neural networks (Carneiro et al., 2015). In Chen et al. (2011) pixel intensity neighbourhoods are used to assign each pixel of a test image its correct class through a supervised learning strategy. Nuclei segmentation in fluorescence microscopy images to recognised bone, cartilage, fat, and background in magnetic resonance and histopathology microscopy images are proposed in this approach. First, data are normalised to extract image neighbourhoods using windows of $N \times N$. Second, images patches with different scales and rotations are including in the set of image. Third, the K-means algorithm is used for select representative pixel neighbourhoods. Fourth, the support vector machine is used to classify each pixel with Gaussian radial basis function kernel. Finally, the predicted results were obtained with majority voting and confidence-based voting — each model containing different scale information. Recognition of teratoma tumour images obtain an accuracy between 59% and 91%. Colour represents useful information for classification, although it is clear that a method relying purely on colour information cannot perform well in this context since most tissues are heterogeneous in terms of the colour of the content and histological images use different stains.

Some papers also deal with objects segmentation. In Tosun et al. (2009) the segmentation of cancerous and normal regions based on a homogeneity measure of objects — cells and crypts — on colon biopsy images is performed. The proposal is based on a new object-oriented segmentation algorithm focused on colon biopsy image in which tissue components are organised to form glandular structures. A K-means algorithm taking into account the colour intensities of pixels, homogeneity texture features and object spatial distribution with uniformity measures is used. This algorithm performs segmentation of cancerous and normal regions with 94.89% of accuracy on average. However, this proposal is useful in histopathologic images of colon biopsy which have specific characteristics, and could work with similar colour distribution images.

Other approaches use texture features to identify fundamental tissues. For instance, Diamond et al. (2004b) use a Haralick texture features to identify tissue composition in prostatic neoplasia and classified as normal, stroma, or prostatic adenocarcinoma. This method identified tissue abnormalities in prostate histology images with $40\times$ of magnification, image processing techniques are applied on subregions of 100×100 pixels. This work was evaluated with 12 images, four to training and eight to testing, obtaining an average of 79.3% of subregions correctly classified. However, the number of cases is reduced and this proposal is for a specific pathology images. In Simsek et al. (2012), a proposal based on co-occurrence frequency for unsupervised segmentation regions of cancer in colon tissue images is presented. First, a new set of high-level texture features to represent the prior knowledge of a particular spatial relation of the tissue components is introduced. Second, multiple

segmentations are obtained with a multilevel partitioning of a graph constructed on the tissue objects and combine by an ensemble function, they consider the object segmentation as a graph partitioning problem. The K-means clustering to defines the tissue objects and co-occurrences matrices over these objects were used. The experiments were performed on 200 colon tissue images obtaining an accuracy and F-score results greater than 90%. This work is applied on a specific cancer in colon tissue images with particular characteristics, such as a magnification of $40\times$. However, the dataset is not publicly available and the results obtained using Haralick features to identify normal tissues in $10\times$ magnification in cardiovascular images does not have the same performance according to our tests.

In general, normal fundamental tissues have not been segmented and the cardiovascular system has not been taken into account, to the best of our knowledge. Additionally, some image processing works in histopathology recognise fundamental tissues indirectly. Texture descriptors (Chen et al., 2012), the Support Vector Machine (SVM) (Chen et al., 2012), and segmentation process for gland detection using colour-threshold, position of the cell, and contour (Ficsor and Molnar, 2009), are some examples. However, the indirect results may not be accurate since authors are interested in obtaining other information — glands or cancerous tissue recognition — from processed images.

Moreover, automatic recognition of epithelial tissue led to the analysis of the cell nuclei since according to shape and position of cell nuclei the type of epithelial tissue and some pathologies are determined. In the state-of-the-art, much effort is dedicated to recognise the cell nucleus since it is a key part for determining biological structures. Previously, the computer-aided recognition of cell nuclei has been addressed using different techniques such as clustering (Gharipour and Liew, 2015; Chankong et al., 2014; Kothari et al., 2009), segmentation (Mohammed et al., 2013; Palacios and Beltran, 2007; Rogojanu et al., 2010) adaptive active contour (Kang et al., 2015; Hafiane et al., 2008; Zeng et al., 2013), and machine learning (Kancherla and Mukkamala, 2013; Han et al., 2012b; Shir et al., 2007).

Clustering techniques to recognise cell nuclei purposes are used in medical images. In Tonkin et al. (2011) a segmentation of epithelial regions of images using an algorithm based on binary graph cuts taking into account the probabilities obtained from colour histogram models is proposed. The proposal was trained using 38 images of four types of odontogenic cyst and tested using a separate data set of 35 images of the same four cyst types. The results using training set are sensitivity of $91.5 \pm 17\%$ and overall mean specificity of $85.1 \pm 18.6\%$. Dentigerous and odontogenic keratocysts results were sensitivities of $91.9 \pm 6.15\%$ and $96.1 \pm 1.98\%$ and specificities of $97.4 \pm 2.15\%$ and $98.7 \pm 3.16\%$, respectively. However, this method is applicable in pathological conditions with similar tissues, such as skin and mucous membranes, where there is a clear microscopic distinction between epithelium and connective tissues to obtain a similar performance. In Kong et al. (2009b) a

computer-aided prognosis system for neuroblastoma, a cancer of the nervous system, is classified into favourable or unfavourable based on the tissue morphology is proposed. The proposed approach uses texture features extracted through co-occurrence statistics, local binary patterns and a modified K-nearest neighbour classifier. They used 43 image tissue samples collected from Nationwide Children's Hospital, 32 images samples are associated with stroma-poor and the rest are associated with stroma-rich and provided an overall classification accuracy of 88.4%. However, this method is applied in a specific pathological condition which presented differences with features selected.

In Song et al. (2013) a method for automated cell nucleus segmentation is proposed. This method is composed of three steps: (i) an initial segmentation based on the maximally stable extremal regions algorithm; (ii) inter-region feature discrimination using texture features based on local binary patterns and feature distance are used; and (iii) refinement of cell nuclei boundary using intra-region contrast information with K-means clustering. The proposed approach was evaluated on U2OS and NIH3T3 datasets of fluorescence microscopic images with 4009 cells publicly available, and achieved superior performance compared to popular state-of-the-art methods with 0.94 and 0.87 Dice Similarity Coefficient (DSC) in each dataset respectively. In Lou et al. (2012) a method to segment multiple cell nuclei from GFP or Hoechst stained microscope images with a shape prior is presented. The method is composed of shape prior extension for multiple nuclei and predicted by Random Forest parametrised via structure learning. Achieving a rand index of 8.2% and 4.3 decrease on Hausdorff distance with respect to the second best method on a public hand-labelled 2D benchmark. However, these methods are useful in fluorescence microscope images which are acquired by a different process to identify specific properties.

Another way of recognising cell nuclei is using adaptive active contour. In Zeng et al. (2015) a method based on an adaptive active contour modelling to segment the cell nuclei from cervical smear images is presented. The method is composed of three steps: (i) the cervical smear image after coarse segmentation using morphological opening and closing operation with circular structuring elements are detected; (ii) active contour with adaptive local region fitting energy modelling using a Gaussian kernel function is applied; and (iii) the split Bregman method to obtain a robust numerical solution and to generate the final segmentation results is used. The spatial precision of the segmentation results are evaluated by using true positive rates and DSC obtaining a mean of 0.87 and 0.85. However, this method was tested only with cervical smear images and does not provide information about shape of cell nuclei.

Machine learning techniques are used in cell nuclei recognition. For instance, Han et al. (2012a) investigated the use of SVM classification based on Laplace edge features for detection of cell nuclei. The edge value is calculated using the Laplace

operator and other feature information such as image moments, histograms and their combination using raw pixel values. This proposal identifies NIH/3T3 fibroblasts cultures grown on glass and stained with Hematoxylin or Hematoxylin and Eosin in a data set of 75 images and 25 images contained NIH/3T3 fibroblast cultures stained with Hematoxylin only. The average detection rate was above 90%. However, they do not specify the proportion of data used for training and testing.

In general these studies used different staining and type of images, their objectives are focused on nuclei segmentation — area, contours, and counts.

2.3. Organ Identification

An organ may be identified using fundamental tissues and analysis criteria such as location, type, interaction, and unique characteristics. Additionally, an organ has functions which defines spatial structures among fundamental tissue relations. The study and the automatic recognition of patterns that make every organ unique is an open problem since an organ has complex structures.

Some papers deal with organ identification using texture features. In Zhao et al. (2005) a statistical model for human categorization of histological images was proposed. The study was performed with texture features of Multi-channel Gabor. The probabilistic distribution of texture patterns in each category is approximated by a finite mixture models of Gaussian. This proposal identifies ten different organs — adrenal, heart, kidney, liver, lung, pancreas, spleen, testis, thyroid, and uterus. Method validation was performed with 778 histological images. The approach yields an accuracy rate between 44% and 93% which varies depending on the organ being identified. Nevertheless, this work has varied percentages of accuracy according to the organ, and the heart is the only organ of the cardiovascular system with obtained results 80% of accuracy.

Similar proposal are used in gastrointestinal tract, as in Yu et al. (2008) a novel 2D stochastic method for semantic analysis of the content of histological images called Spatial-Hidden Markov Model (SHMM) is presented. This proposal aims to identify five organs of the gastrointestinal tract — oesophagus, stomach, small intestine, large intestine, and anus. They propose a block-based, 64×64 pixels, classification approach. For each block a 25-dimensional feature vector is obtained concatenating the total Gabor energy and the mean grey value. The approach yields to an accuracy rate between 59% and 82% depending on the organ that is being identified. However, this work has high computational complexity and the image collection consists of 200 histological images, 40 images for each region, which is a limited dataset.

Other approaches use colour and texture features to identify organs. In Herve et al. (2011) a comparison of different combinations of colour and texture features for

particular cases of histological images is presented. They proposed the Softly Quantised Colour Local Binary Pattern (SQCLBP) searching highlight some tissue structures that will be first isolated by some quantised colours before being described by texture features. Co-occurence matrices are also often used to extract Haralick features. These approaches were tested on three different datasets: LG6MAL, LYMPH and GLOMDB. When colour and texture are considered jointly, the best overall performances are reached on all three datasets. The Mean Average Precision (MAP) measure is used to evaluate the performances. The accuracy of the tests was 0.973 for LG6MAL — male class — 0.974 for GLOMDB — glomerulus class — and 0.633 for LYMPH — three class of malignant lymphoma. However, the results obtained are different according to each set, this would indicate that no method could outperforms the other when there is no prior information about the dataset.

2.4. Ontology of Human Histology

Many ontologies and taxonomies are available in electronic form with Open Source licenses. Some medical taxonomies best known are: GALEN (basic clinical concepts — fracture, bone, and so on — controlling the combination relations concepts — bones fractures — and concepts complex — clavicle fracture), UMLS (Unified Medical Language System), MeSH (Medical Subject Heading), Kingsbury Center for Cancer Care Glossary, MedicineNet Medical Dictionary, Multilingual Glossary of Technical, and Popular Medical Terms in nine European Languages, ICD (International Classification of Diseases) among others (Vasquez et al., 2010). Some ontologies are used in web retrieval systems (Paslaru Bontas et al., 2004), identification of relations between diseases (Schofield et al., 2013), diagnosis (Colantonio et al., 2008), among others.

Ontology research and analysis were performed using different approaches, histological and anatomical in order to assess if there are parts that may be reusable (Rubin et al., 2008). BioPortal (2005) contains some histological terms. However, this ontology has different kind of guidance to our research due to its organisation does not contain a specific order, some terms are randomly located, for this reason it cannot be reused. BioPortal (2008b) and BioPortal (2008a) have similar terms to those required in our research, for instance terms related to the epithelial tissue. Nevertheless, these concepts are linked by a different route, tissues blood vessels. These ontologies contain many concepts without giving much detail, leaving some ways inconclusive for instance with muscle tissue. Concepts are linked in one-way allowing to connect from a large to a small structure but not reverse. Some methods to reach for a concept are not intuitive or logical so the user should guess taking more effort and needed knowledge to found possible routes for these terms. BioPortal (2014c) contains the cardiovascular system and its organs. It is a complete

ontology and close to what is sought in our research. However, some terms are not in this ontology such as the type of epithelium, connective and muscle tissues which has another classification — cutaneous, corneal and lymphatic. Moreover, it is a fairly complete cardiovascular system and organs ontology. It has large shortcomings regarding the fundamental tissues — epithelial tissue and muscle tissue can be referenced as indivual terms. BioPortal (2014b) contains concepts similar to those required in our research such as some tissues and cells. Nevertheless, this is a human histopathological ontology which contain an abnormal cell type which can occur in either disease states or disease models, then this ontology cannot be applied in this research. Additionally, this ontology does not contain the organs of the cardiovascular system and tissues classification consider other specific approaches — retinal, mammary, urethral, and so on. Finally, some terms can be referenced as individual concepts. BioPortal (2014a) is a mouse ontology which an adult gross anatomy focus, for this reason does not contain microscopic terms such as cells, fibres, and tissue with histological information. However, this ontology contains some similar organ and system terms which can be referenced how individual concepts in our ontology.

2.5. Classification based on Ontology

Ontologies and taxonomies contain relevant knowledge represented with rich structural and semantic information. Approaches that use this tools into automatic classification process are dividing in two: (i) model the relation between visual and semantic information (Wu et al., 2010; Yang et al., 2007) and (ii) use these ontologies and taxonomies in the classification algorithm (Othmani et al., 2010; Smith et al., 2015; Abdollahpour et al., 2015; Paulson et al., 2006; Breen et al., 2002). We will focus in the second group to perform the classification process using images and ontologies in the same way.

On the one hand, in histological context Othmani et al. (2010) proposes to leverage the high-level reasoning and knowledge formalization ability of ontology-based software to make annotation of high-content images more efficient and interactive. The low-level image processing aims at outlining and describing general biological objects — the nuclei, the lumina and the invasive areas — in the histopathological images. In this step thresholding, morphological closure and snake-based method are used. They use an anatomical ontology to improve the specificity and sensibility rate using SPARQL query language. The results show that the proposed method detected all mitoses but the detection has many false positive. Additionally, the algorithm with geometric constraints is more specific but that it decreases sensibility. However, this work is focused in histopathological images and anatomical ontology. Smith et al. (2015) provides a survey of the biomedical imaging ontologies that have

been developed thus far. It outlines the challenges, particularly faced by ontologies in the fields of histopathological imaging and image analysis, and suggests a strategy for addressing these challenges. This review presented the use of ontologies in annotation or tagging, investigations and biobanking. Furthermore, a critical survey of major existing contributions to ontologies and ontology-related standards in the imaging domain is provided. They proposed a image ontology initiative taking into account the following elements: image acquisition, specimen protocol, image processing parameters and organizational levels. Finally, they presented a example case of quantitative histopathology image ontology. However, this review is focused in histopathological imaging.

On the other hand, other application domains have used ontologies to improve the classification process using different strategies. Abdollahpour et al. (2015) introduces a new visual word generation and feature representation method for multiclass image classification based on semantic taxonomies. They used visual feature of all the sub-concepts modelled in the taxonomy to represent the images as a histogram of visual words' occurrences. Finally, they used the semantic relation between classes based on WordNet hierarchy and then assign a set of linear SVM to each semantic node. The CIFAR-10 data set was used for testing given better accuracy results than the baseline methods of one vs. one, one vs. all and another reference method. Experimental results show that the proposed method has improved the accuracy of classification results. However, they use a taxonomy which is not able to do inferences about its content. In Paulson et al. (2006) a methodology for integrating images and text for object identification research in data fusion and information retrieval is proposed. This paper presents an application to compare images and description of vehicles. The description is a brief text similar to police logs reported in many local newspapers. This proposal is composed of three steps: first, the system extracts fundamental features such as colour, wheel base, and so on. Second, the system makes an inference about information. Abstract knowledge signatures are created in steps one and two. Third, knowledge signature are improved with an ontology. The ontology contains knowledge, this knowledge allows automated reasoning to improve observed knowledge signatures in order to find hidden similarity. The system choices matched 58.7% of the subjects' selections, which bettered the average subject's score of 55.3%. However, it is a very explicit application to vehicles context. Breen et al. (2002) proposed a scalable system capable of examining images and accurately classifying the image based on its visual content combining ontologies and neural networks. Neural networks are used as object identifiers to do the automatic classification of an image based in its content — colour distribution. Ontologies are used to represent relations that reveal information useful in classifying the entire image. After the network identifies a set of objects from an input image, these objects may be used to select concepts from ontologies. They used a rank to determine if an images belongs to a concept or not, then the ontologies allow the

child or parent concept to be discarded improving the automatic classification. For instance it is possible an image be wrongly classified as both an NBA basketball game and a college basketball game at the same time, this can be corrected using the ontologies. The combined system was tested with 15 sample images from the sport domain obtaining between 33% and 87% of the images are associated with relevant concepts and 13% and 77% images are associated with at least one irrelevant concept according the threshold.

Chapter 3

Recognition of Fundamental Tissues

You will never know if you never try.
Anonymous

3.1. Dataset

Tissue samples from organs were stained with Masson's trichrome and Hematoxylin and Eosin using a laboratory protocol to control the process. The image capture protocol was defined taking into account microscope configuration, software configuration, sample manipulation and image capture to reduce errors in the automatic recognition. Finally, 400 images samples belonging to different organs and persons were obtained, 300 acquired at $40\times$ — 100 for each type of epithelial tissue — and the other 100 at $10\times$ objective. The images were acquired with a microscope Leica *DM750-M* with 2048×1536 pixels of resolution and were stored in *PNG* format. The microscope has an eyepiece with a magnification factor of $10\times$ and a field of view of 20 obtaining 400 and 100 end magnifications for a $40\times$ and $10\times$ objectives, respectively.

The group of histology experts consisted of six members of the research group Teblami, from the University of Valle. In this thesis, we used the images belonging to the project *Desarrollo del Banco de Imágenes Histológicas sobre el Sistema Cardiovascular* (BISCAR), CI-2714 Vice-rectorate for Research at the University of Valle. We left this dataset publicly available at `http://biscar.univalle.edu.co/?page_id=1003`. A desktop application has been implemented and this is available at `http://biscar.univalle.edu.co/?page_id=1049`. Algorithms were implemented in *C++*, using the *CImg* library in a computer of *4-cores* and *4Gb* of *RAM*.

Figure 3.1 shows examples of histological images taken with $40\times$ and $10\times$ objectives.

3.2. Motivation

A fundamental tissue has unique patterns that allow us to identify and differentiate it from others. For instance, the coating epithelial has two locations, the

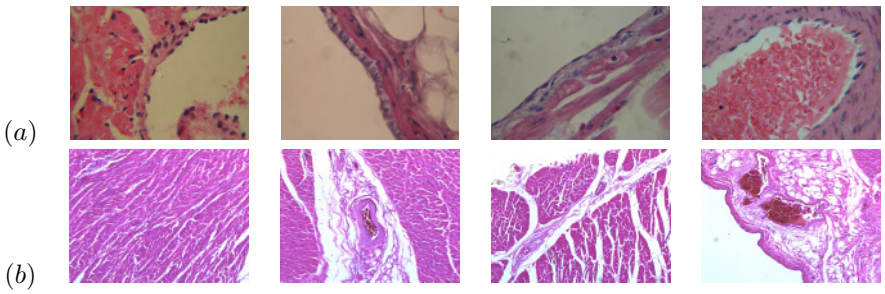


Figure 3.1: Histological images. (a) $40\times$ objective. (b) $10\times$ objective.

epidermis and the lumen — the inner region — of hollow internal organs and the coated external body surfaces. Microscopically, these locations are always close to light areas and have specific features that enable identification. Flat, cubic and cylindrical epithelial tissue are some of epithelium's type according to the shape and position of cell nuclei. Cubic epithelial cells have sphere shape, flat and cylindrical cells have ellipse shapes. However, flat cells are parallel to light regions whilst cylindrical cells are perpendicular to light regions. Muscle tissue has a homogeneous and compact structure with variable direction and organisation according to the region or cutting sample. Loose connective tissue has more separate or scattered structures. Furthermore, the appearance of a tissue may be vary in the same organ, that can be observed by changing regarding the captured zone, direction of the cut or sample. Regarding the visual characteristics, colour is not a reliable feature in histological images due to changes in the muscular staining that may invalidate the results. Figure 3.2 contains examples of epithelial, loose connective and muscle tissues.

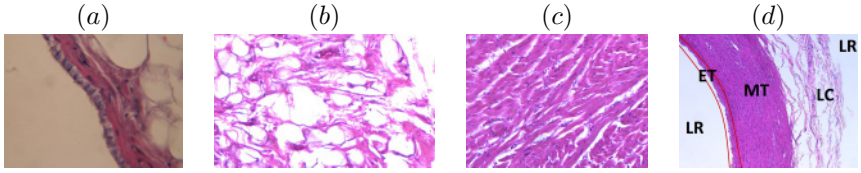


Figure 3.2: Examples of fundamental tissues. (a) Epithelial. (b) Loose connective. (c) Muscle. (d) Histological image. LR represents light regions; LC represents loose connective tissue region; MT represents muscle tissue region; ET represents epithelial tissue region between red lines marked.

Thermal drift and colour balance affect the analysis introducing small variations in the intensities of the colours and lighting. Other variations are introduced during the preparation of samples (cut and stain processes) under environmental conditions or inducted by settings (microscope and software) during the acquisition of

the image. This situation implies that the proposed method has to be robust to small variations in colour balance and thermal drift.

3.3. Segmentation of Fundamental Tissues

In this section we proposed an approach for automatic recognition of fundamental tissues using image processing techniques to extract information about tissue's morphology — composition, location and spatial relations. Automatic recognition of fundamental tissues is a n-fold process according to each tissue. A brief summary of the proposed method is given below and a detailed explanation is presented in the following subsections. A general outline of our proposal is illustrated in Figure 3.3. (1) The process for images taken with the $40\times$ and $10\times$ objectives are presented on the left and right side, respectively. (2) In both cases, three images are used from a histological image by extracting information from the Structure Tensor (Lu et al., 2010) and the red and the green channels. (3) A pixel position is represented by a 3-element feature vector with values indicated above. A K-means clustering (Kanungo et al., 2002) is performed using as input the set of these vectors, obtaining three different groups. (4) The black regions correspond to segmented areas for both magnifications and characters a and b are used to reference $40\times$ and $10\times$ images, respectively. The obtained results are the following: (4.1.a) and (4.1.b), correspond to loose connective tissue that is recognised just after performing the K-means clustering and therefore does not need further processing; (4.2.a) and (4.2.b) correspond to light regions; (4.3.a) is cell nuclei and (4.3.b) is muscle tissue. To recognise epithelial tissue, only with $40\times$ images, three additional steps are necessary: (i) Epithelial cell recognition using the Flood-fill algorithm and the size of the regions on (4.2.a) obtaining the result (6.a). Pixels belonging to epithelial tissue are determined based on the distance between light regions in (4.3.a) and cells obtained in the previous results. (ii) Epithelial cell classification according to shape and position of cell nuclei are determined (8.a) using proportion of circularity and projection ratio; (iii) Epithelial tissue classification based on the cell nuclei frequency. On the other hand, recognition of muscle tissue is performed in two additional steps: (i) Removing irrelevant details or small areas using erosion and thresholding applied to (4.3.b) to obtain (6.b) which represents muscle tissue with red blood cells; (ii) Removing red blood cells to obtain the final result (8.b).

Small variations in colour balance and thermal drift are due to the flexibility of the K-means algorithm and the definition of our protocols. The details of the complete process are presented in this section.

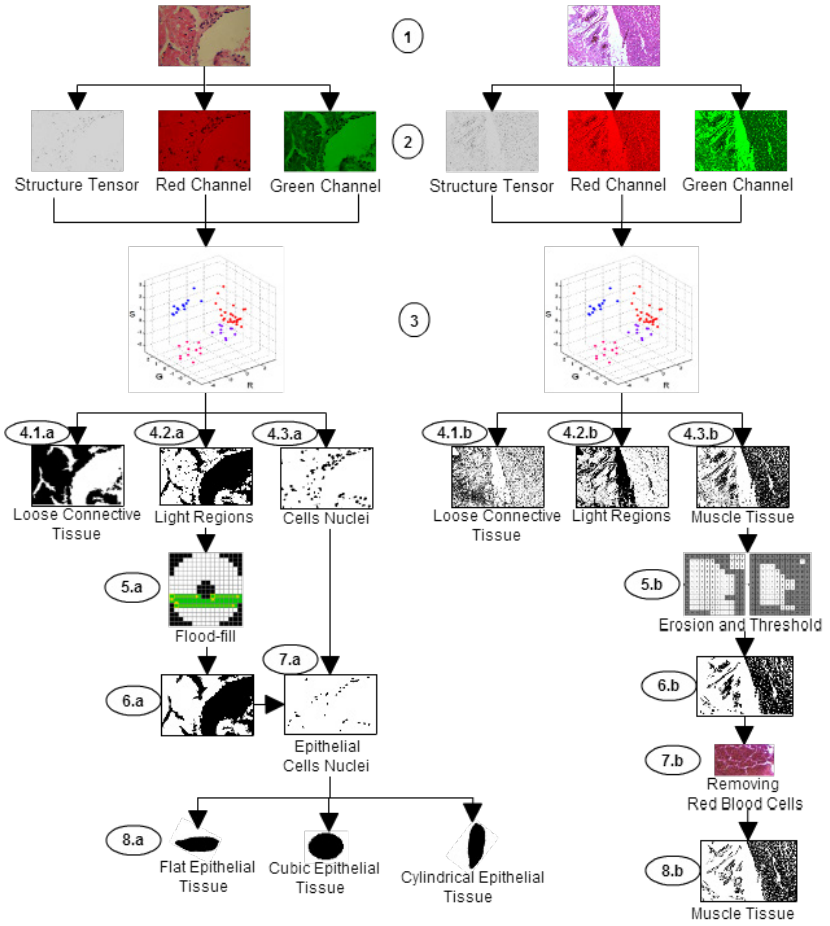


Figure 3.3: Illustration of the proposed automatic recognition of fundamental tissues. The process for images taken with the $40\times$ and $10\times$ objectives are presented on the left and right side, respectively.

3.3.1. Method

In our approach, we segment each image obtaining three classes — according to magnification in a similar way to the conventional method employed by histologist — using K-means (Kanungo et al., 2002). Loose connective tissue and light regions are segmented regardless of magnification level. Epithelial tissue is recognised through cell nuclei using $40\times$ images, due to small size of cell nuclei to identify some characteristics and avoid confusion with cell nuclei which belong to another tissues. Muscle tissue is recognised in $10\times$ images. This phase is illustrated in Figure 3.3 (3), (4.1), (4.2) and (4.3).

Clustering using K-means algorithm

Some experiments were carried out in order to recognise the importance of each channel in *RGB*. We observed that the amount of information provided by the blue channel was not as useful as the other two channels for performing the clustering. However, the results achieved using only *R* and *G* had some space for improvement. Under this perspective, we consider to add edge information to delimit different areas in histological images and the results after this approach were more promising. We evaluated different edge detection algorithms in order to include this information. Figure 3.4 shows some results of epithelial tissue, using the edge segmentation by computing a gradient-magnitude image and combining with non-maximum gradient suppress (Canny, 1986; Sonka et al., 1998), the Hessian tensor using the maximum and the minimum eigenvalues (Sato et al., 1997; Frangi et al., 1999; Rohr, 2001), and the Structure tensor using the maximum and the minimum eigenvalues (Rao and Schunck, 1991; Weickert and Scharr, 2002). The largest eigenvalues of structure tensor was selected based on its capability to identify the cell nuclei and highlight edges clearly, eliminating most of remaining information.

The segmentation of fundamental tissues is conducted using a feature vector of dimension three. Let $I : \mathbb{I} \times \mathbb{I} \rightarrow \mathbb{R}^3$ be a histological image of the cardiovascular system in RGB colour space; $Q : \mathbb{I} \times \mathbb{I} \rightarrow \mathbb{R}^3$ be a matrix in which each element is formed by three values (x, y, z) where x represents the intensity of the red channel of I, y corresponds to the intensity of the green channel of I, and z is the largest eigenvalue of Structure Tensor (Lu et al., 2010) of I; $H_k(t)$ is a cluster represented by a set of vectors in $\mathbb{R}^3$ in the $t - th$ iteration; and $C_k(t) \in \mathbb{R}^3$ be a centroid k of the cluster $H_k(t)$. We use the red and green pixel values since they contain relevant information about cell nuclei, tissues and light regions according to the stains used — Hematoxylin and Eosin (H&E) (Fischer et al., 2008). The third dimension is obtained as a result of the distribution of gradient directions within the neighbourhood of a point defined by a window (Mazo et al., 2012).

The initial parameters of the K-means algorithm are set: $t = 0$, $C_1(0) = \{70, 30, 150\}$, $C_2(0) = \{160, 70, 30\}$, and $C_3(0) = \{110, 154, 0\}$. These values are set up using a heuristic.

$$Q_{ij} \in H_k(t) \rightarrow k = \underset{k \in \{1,2,3\}}{arg\ min} | \sum_{u=0}^{2} (Q_{ij}[u] - C_k(t)[u])|^2, \tag{3.1}$$

$$C_k(t+1) = \frac{1}{|H_k(t)|} \sum_{i=1}^{M} \sum_{j=1}^{N} a_{ij}.Q_{ij}, \tag{3.2}$$

where a_{ij} is 1 if $Q_{ij} \in H_k(t)$ and 0 in other case. Increase $t = t + 1$ and repeat (1) and (2) until $H_k(t + 1) = H_k(t)$. Let $O = \{O_1, O_2, O_3\}$ be a binary image such that:

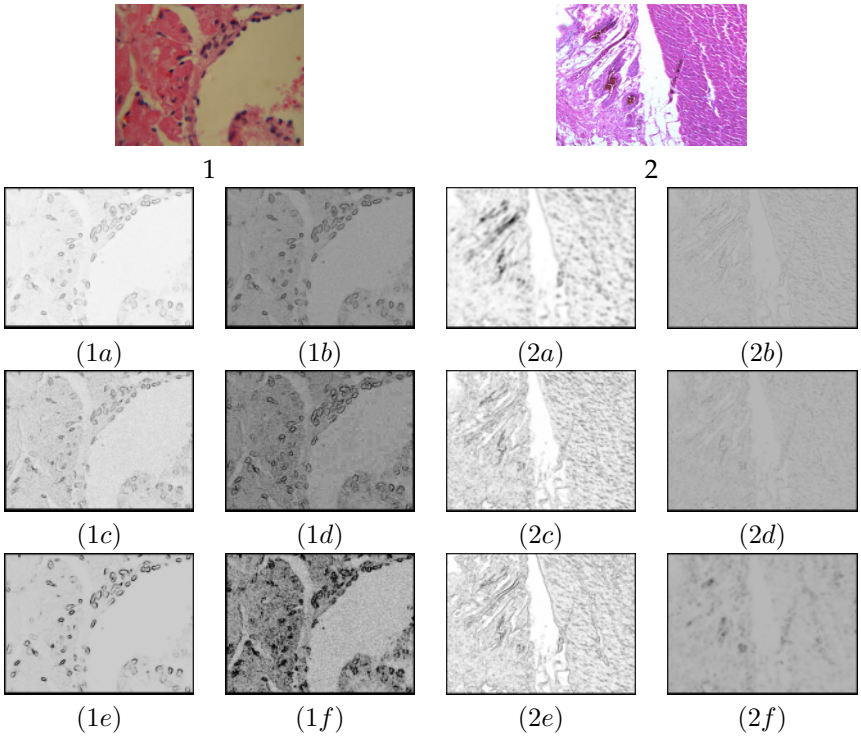


Figure 3.4: Results obtained with edge detection algorithms of the original image. (1) $40\times$ Histological images. (2) $10\times$ Histological images (a) Using the Gradient-magnitude. (b) Using the gradient-magnitude combining with non-maximum gradient suppress. (c) Using the Hessian tensor maximum eigenvalues. d) Using the Hessian tensor minimum eigenvalues. (e) Using the Structure tensor maximum eigenvalues. (f) Using the Structure tensor minimum eigenvalues.

$$(O_k)_{ij} = \begin{cases} 1 & Q_{ij} \in H_k(t) \\ 0 & \text{else,} \end{cases} \tag{3.3}$$

where a value of k represents a different tissue in I, such that $k = 1$ corresponds to pixel positions associated with connective tissue, $k = 2$ corresponds to light regions in the image, and $k = 3$ corresponds to cell nucleus or muscle tissue, according to magnification. The groups' connective tissue and light regions do not require further processing. However, the segmentation obtained with $k = 3$, in images taken with the $10\times$ objective, contains cell nuclei which does not belong to epithelial cells. Also, the segmentation obtained with $k = 3$, in images taken with the $40\times$ objective, contains regions of red blood cells or dense connective tissue which does not belong to muscle tissues. Thus, a further processing is required for the group $k = 3$.

3.4. Epithelial Tissue Classification

In this subsection, we explain the further processing required for epithelial tissue segmentation and classification. Histologists, biologists, and pathologists commonly use light regions as information that helps them to recognise epithelial tissue given that epithelial cells are always found close to these regions. Additionally, images taken with the $40\times$ objective are the ones usually employed by specialists to analyse and classify this type of tissue. In our approach, epithelial tissue classification is performed in three steps: (i) epithelial cell recognition; (ii) epithelial cell classification; and (iii) epithelial tissue classification. In the next subsections, the complete process is presented.

3.4.1. Method

Epithelial Cells Recognition

After the previous segmentation, light region sets are represented in O_2. However, while small and large white regions are recognised as light, only the large ones are of interest for epithelial cell recognition.

The flood-fill algorithm (Nosal, 2008) is used to fill small blank spaces within the segmented areas and, since the algorithm returns the size of regions, it is also used to remove small and isolated areas. This algorithm is applied to image regions associated with O_2. The result is a binary image where pixels equal to 0 represent the background and pixels equal to 1 are the regions of interest — light regions.

Definition 1: Let a region R be a set of pixels with the similar intensity values excluding the background.

According to the Definition 1, let δ be a region that represents a cell nuclei and γ be a region that represents a light region. The histologists consider cell nuclei at a distance of no more than $640\mu m$ to light regions as epithelial. This value was provided by histology experts of the research group Teblami from the University of Valle, Colombia, which collaborated in this work.

A tuning parameters procedure is used in different parts. Tuning of parameters is performed by selecting the value t such that the absolute error Ae between a set of ground truth images ρ_g and a set of automatic results ρ is minimised. Epithelial cells are characterised by being close to a light region. Distances between cell nuclei and light regions depend on cell nuclei sizes, which are determined by the objective used in a microscope.

$$\underset{t\in\mathbb{R}}{\operatorname{argmin}}\, Ae(\rho) = \sum_{\rho_i\in\rho} |\rho_g - \rho_i(t)|, \tag{3.4}$$

where $\rho_i(t)$ is the result obtained automatically when the the value t is used as input

of the algorithm. Therefore, the set of epithelial cells, Ec, is defined as follows:

$$\exists\, p_i \in \delta_i \ \wedge \exists\, p_j \in \gamma_j \ : \|p_i - p_j\|_2 \leq 200 \rightarrow \delta_i \in Ec, \tag{3.5}$$

where $\delta_i \in O_3$ is a cell nucleus, $\gamma_j \in O_2$ is a light region and p_i and p_j are a pair of pixels with coordinates (x, y).

Epithelial Cell Classification

Once a cell is recognised as epithelial, we perform a classification as either flat, cubic or cylindrical according to its morphological information; namely its circularity and the cell's projection into the nearest light region.

First, the circularity of the cell is measured using Feret's diameter (Merkus, 2009) and cells with a circular shape are classified as cubic epithelial cells. Then, for the rest of the cells, two points of the widest diameter are calculated, naming the line segment between the points the baseline. Finally, the cell's projection into the nearest light region is calculated to obtain a projection ratio. This ratio is used to determine whether the cell is flat or cylindrical. This phase is illustrated in Figure 3.3 (5.1) and the complete details of the procedure is below.

Cubic Epithelial Cell Classification: circularity is the ratio between the minimum and maximum Feret's diameter (Merkus, 2009). The Feret's maximum diameter is the supremum of the distance between points belonging to edges of an epithelial cell, and the Feret's minimum diameter is the minimum existing distance between points belonging to edges of an epithelial cell. Hence, circularity is:

$$Circularity = \frac{FeretMinimum}{FeretMaximum}, \tag{3.6}$$

for this case, when the circularity is greater than or equal to 0.7, a value calculated using equation (3.4), a cell is classified as cubic epithelium. Where circularity is less than 0.7, epithelial cells will be flat or cylindrical and they will be classified in the next step.

Flat and Cylindrical Epithelial Cell Classification based on Orientation of Cells with Respect to the Nearest Light Region: we calculate the angle formed between line segment obtained by two endpoints of a cell nuclei and the projection line obtained from the nearest light region. For this case two methods were considered: the equation of the slope (Eves, 1971) and a technique based on the Law of Cosine (Heath and Euclid, 1956).

The first method is based on the slope (Eves, 1971) and is defined as follows:

$$m_0 = \frac{((y_1 - y_0) - (x_1 - x_0)) - ((y_2 - y_0) - (x_2 - x_0))}{(1 + ((y_1 - y_0) - (x_1 - x_0)) * ((y_2 - y_0) - (x_2 - x_0)))}, \tag{3.7}$$

Given the coordinates of two pixels (x_0, y_0) and (x_1, y_1) located inside the cell and the coordinates of a pixel (x_2, y_2) placed in a light region, the slope of those

pixels, m, is calculated using equation (3.7) and the inclination of the cell is obtained as the angle:

$$\theta = \arctan(m_0). \tag{3.8}$$

A drawback of the slope method is that it does not work with cells of vertical baselines whose tangent tends to infinity.

The second method used the Law of Cosines (Heath and Euclid, 1956):

$$\theta = \arcsin \frac{a + b^2 - c^2}{2ab}, \tag{3.9}$$

where θ is the angle between a and b, and c is the side opposite to this angle. For our case, a is the segment between q_1 and r_1, b is the segment between q_2 and r_1, and c is the segment between q_1 and q_2. θ is the opposite angle of side c, reflecting the inclination of the cell.

However, when the cell is located on the edge of the light region segments a or b may be 0 and it is not possible to obtain the orientation of the cell.

Since the two previous methods are not able to obtain the cell orientation, we used another procedure. We calculate the cell's projection over the closest light region and we obtain a ratio between the lengths of the baseline and the projection line as is illustrated in Figure 3.5. The baseline is the length of the line defined by the coordinates of points (q_1, q_2) of the maximum Feret's diameter of a cell nucleus. The $ProjectionRatio$ is the length of the projection line of a cell over the closest light region $(r_1^* r_2^*)$ and it is obtained as explained below.

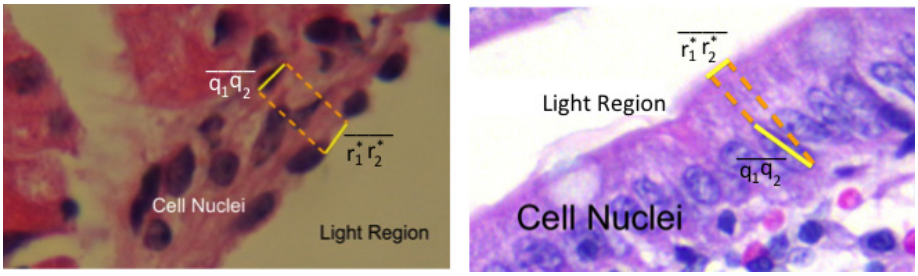


Figure 3.5: Illustration of cell projection in histological images taken with the $40\times$ objective. The image on the left presents an example with the obtained projection for a flat epithelial cell. The image on the right presents an example with the obtained projection for a cylindrical epithelial cell.

Let NLR be the set of the Nearest Light Region coordinates, q_1 and q_2 are the extremes of the segment of the maximum Feret's diameter, the aim is to identify the corresponding pair of points r_1 and r_2 lying on the edge of the NLR such that the distance between the two lines is minimised.

$$r_1^* = \text{argmin}_{r_1 \in NLR} \sqrt{(q_{1(x)} - r_{1x})^2 + (q_{1(y)} - r_{1y})^2}, \tag{3.10}$$

$$r_2^* = \text{argmin}_{r_2 \in NLR} \sqrt{(q_{2(x)} - r_{2x})^2 + (q_{2(y)} - r_{2y})^2}. \tag{3.11}$$

The $ProjectionRatio$ is calculated as:

$$ProjectionRatio = \frac{\ell(\overline{r_1^* r_2^*})}{\ell(\overline{q_1 q_2})}, \tag{3.12}$$

where $\ell(\overline{r_1^* r_2^*})$ is the length of the line formed by r_1^* and r_2^* and $\ell(\overline{q_1 q_2})$ the length of the line formed by q_1 and q_2. Note that the vector between r_1 and q_1 should be orthogonal with respect to the edge of the NLR. The same situation happens in the case of r_2 and q_2.

An epithelial cell with a $ProjectionRatio$ larger or equal to 0.8, a value calculated using equation (3.4), is classified as a flat epithelium. This means that the baseline of cell nucleus is parallel to the closest light region. Otherwise, the epithelial cell is classified as cylindrical epithelium, this indicating that the baseline is perpendicular to the closest light region.

Epithelial Tissue Classification

Once the type of epithelial cells is recognised, epithelial tissue is classified in one of three categories — flat, cubic or cylindrical — based on the plurality rule; the winner is the cell type that obtains more votes than any other cell type.

3.4.2. Experiments and Results

In this section, we discuss the results obtained with the proposed method and we compare them with the ones obtained by the experts. We present experiments and results obtained in classification of epithelial tissue using sensitivity and specificity as performance measures as following equations 3.13 and 3.14.

$$Sensitivity = \frac{TruePositive}{TruePositive + FalseNegative}, \tag{3.13}$$

$$Specificity = \frac{TrueNegative}{TrueNegative + FalsePositive}. \tag{3.14}$$

Our approach to automatically segment nuclei of epithelial cells was qualitatively evaluated by a group of six experts. Sensitivity and specificity were obtained comparing the results of our approach with the ground-truth provided by them (Mazo et al., 2012). The sensitivity indicates how many elements, of all mem-

bers of a class, are classified correctly, while the specificity expresses the percentage of elements, which do not belong to that class, that were assigned correctly as belonging to another class. The results obtained with manual and automatic classification are included in Figure 3.6. In Figure 3.6 can be observed that in the first and the second row our approach for automatic classification produces the same result as the manual one. In the last row of Figure 3.6, we present some cases where cubic epithelial tissue is misclassified as cylindrical tissue. This happens when cell nuclei are very close to each other and they are evaluated as a single cell. A quantitative evaluation of classifying epithelial tissues is presented in Table 3.1.

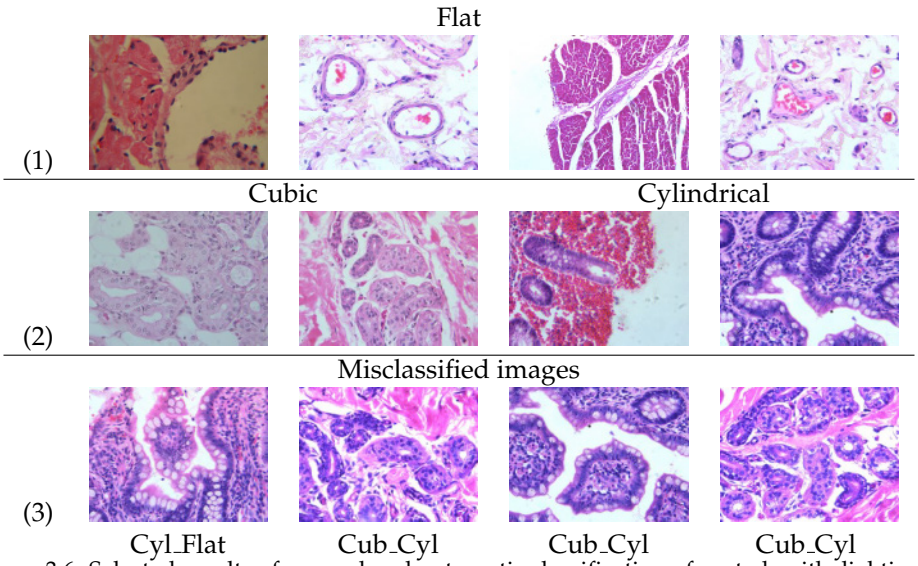


Figure 3.6: Selected results of manual and automatic classification of coated epithelial tissue. In the first row (1), some samples of flat epithelial tissue correctly classified by the automatic method proposed. In the second row (2), The first two images on the left correspond to cubic epithelial tissue and the last two images correspond to cylindrical epithelial tissue, all of them correctly classified. In the third row (3), some misclassified images are presented, below the images we presented, by the accompanying text, the ground truth assigned by experts — first part of the label — followed by the class assigned by automatic classification — second part of label. The results show that the cubic epithelial tissue have the most errors.

3.5. Connective and Muscle Tissues Recognition

The first step to recognise muscle tissue is carried out by the K-means algorithm represented in O_3. In O_3, small areas of loose connective tissue and red blood cells were observed. In this section we will explain the method used to finally recognise

Table 3.1: Performance evaluation of tissues classification.

Confusion Matrix	Flat	Cubic	Cylindrical	Total
True Positive	31	26	28	85
False Positive	12	8	10	30
False Negative	3	7	5	15
True Negative	54	59	57	170
Sensitivity	0.91	0.79	0.85	0.85
Specificity	0.81	0.88	0.85	0.85

muscle tissue first getting rid of irrelevant details and later removing red blood cells. In the next subsections, the process carried out using images taken with the $10\times$ objective is presented.

3.5.1. Method

Removing Irrelevant Details

Small areas of loose connective tissue are sometimes mistakenly classified as muscle. This may be caused by creases, cutting effects or subtraction operation between values. The segmented muscle image obtained previously is eroded to eliminate small regions. Later, regions under $3840\mu m$ are removed, selecting this value heuristically considering different samples and the specific magnification using equation (3.4). This phase is illustrated in Figure 3.7.

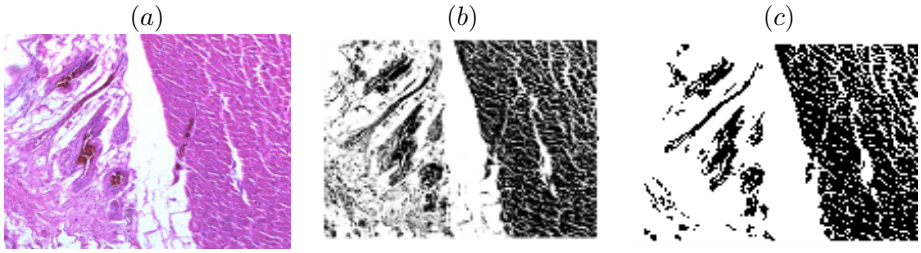


Figure 3.7: Illustration of removal irrelevant details in the segmented muscle image taken with the $40\times$ objective. (a) Original Image. (b) Segmented muscle image. (c) Result after removed irrelevant details.

Removing Red Blood Cells

In the images O_3, red blood cells appear in some samples wrongly labelled as muscle tissue. We removed red blood cells by the following procedure: (i) A segmentation of the regions recognised as red blood cells is performed using threshold-

ing on the red channel of the original image. The threshold is a range of values between $\{146-255\}$, values calculated using equation (3.4). Although small changes in colour could impact the process, this effect is reduced with the use of sample preparation and image capture protocols. (ii) Erosion is applied to regions recognised as red blood cells to remove irrelevant details. (iii) Regions under $480\mu m$ are removed to avoid segmented muscle or connective cells nuclei. This value was set by heuristics considering different samples and magnification using equation (3.4). (iv) The segmented red blood cells image is subtracted to O_3, the result is:

$$O_3 - (O_3 \cap B), \tag{3.15}$$

where B is the segmented red blood cells image. The regions under $320\mu m$ are removed to eliminate irrelevant regions resulting from the subtraction, a value calculated using equation (3.4).

3.5.2. Experiments and Results

In this section, we discusses the performance of the loose connective and the muscle tissues classification. This part is assessed using expert criteria based on a scale from 1 to 5 due to the lack of a ground truth dataset for conducting an objective assessment.

Results of the automatic recognition of loose connective and muscle tissues were evaluated by a group of six experts and using a scale from 1 to 5 to represent poor, average, good, very good, and excellent. A selected set of original images and automatic segmentation of loose connective and muscle tissues is included in Figure 3.8. It can be observed that loose connective tissue is recognised even when immersed in muscle tissue and muscle tissue is recognised without red blood cells.

Figure 3.9 contains a graphical representation of the median of the expert judgments to evaluate four different characteristics about the segmentation of the tissues in the set of test images. In the automatic segmentation of loose connective tissue (Figure 3.8) two aspects of the proposed approach were evaluated: (i) the ability to recognise loose connective tissue and; (ii) the ability to differentiate loose connective tissue from muscle tissue. This evaluation was carried out taking into account issues such as the correct recognition of loose connective tissue and the close relation with muscle tissue —loose connective tissue is in thin layers surrounding the muscle tissue that is sometimes difficult to recognise, even manually. In automatic segmentation of muscle tissue (Figure 3.8) two aspects of the proposed approach were evaluated: (i) the ability to recognise muscle tissue and; (ii) the ability to differentiate muscle tissue from connective tissue. The correct recognition of muscle tissue, the close relation with loose connective tissue and high similitude with dense connective tissue that are sometimes difficult to differentiate even manually, were

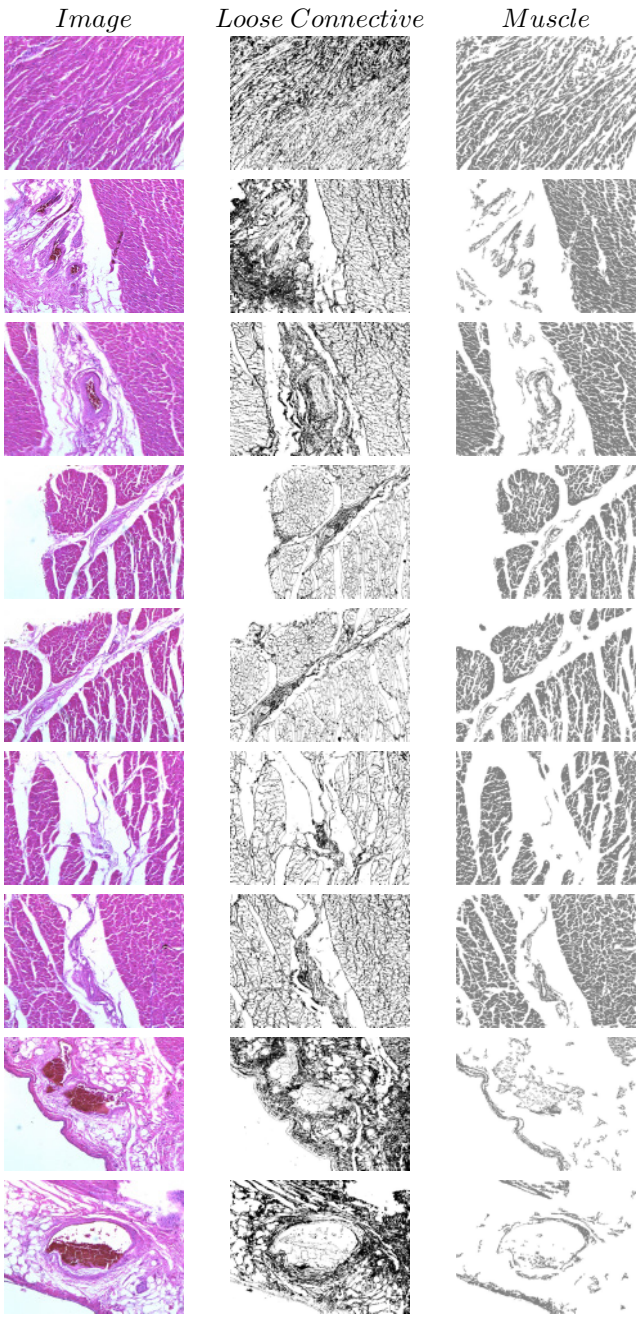


Figure 3.8: Selected results of loose connective tissue and muscle tissue recognition. In the first column, original images are presented. In the second column, automatic segmentation of loose connective tissue. In the third column, automatic segmentation of muscle tissue.

some of the issues considered during the evaluation.

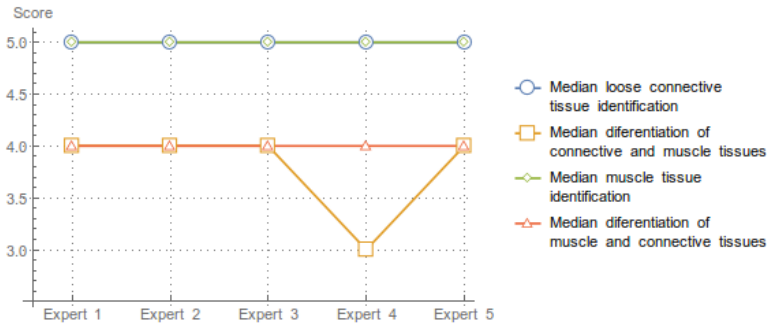


Figure 3.9: Results obtained to loose connective tissue and muscle tissue recognition.

The ability of the proposed approach to recognise loose connective tissue was given an average score of 4.85 by the experts while its ability to differentiate loose connective tissue from muscle tissue was given an average score of 3.96 out of 5. For its ability to recognise muscle tissue, experts gave the proposed approach an average score of 4.82, but gave its ability to differentiate muscle tissue from connective tissue an average score of 3.67 out of 5. This lower score is due to the potential confusion between dense connective and muscle tissues.

3.6. Conclusions

The contributions of the work presented in this chapter are: first of all, we proposed an automatic method to recognise fundamental tissues — epithelial, loose connective and muscle — on histology images of the human cardiovascular system. Secondly, we classified epithelial cells into flat, cubic and cylindrical even when the conditions were challenging, such as variation in cells size, the same shape of flat and cylindrical cells, and where some cells may lose their characteristic shape due to tissue cutting. Variations in cell size are overcome by measuring circularity, a feature conserved regardless of cell size. Thirdly, we propose a procedure for determining the inclination of cell nuclei, called Projection Ratio. Fourthly, we recognise muscle tissue and loose connective tissue.

In the future, we are planning to continue working on this proposal in the following way: (i) to separate particles in order to differentiate cubic epithelial tissue; (ii) to use additional image features in order to overcome the current potential misclassification between muscle and dense connective tissues; (iii) to carry out a comparison with available slide scanners used in pathology and; (iv) to integrate the proposed method as an additional functionality in BISCAR (*Banco de Imágenes Histológicas del Sistema Cardiovascular*) available at `http://biscar.univalle.edu.co`.

Chapter 4

Classification of Cardiovascular Tissues Using LBP Based Descriptors and a Cascade SVM

4.1. Dataset

Tissue samples and image capture protocol were described in Section 3.1. The dataset used was composed by 3000 blocks belonging to a subset of images acquired at $10\times$ objective from tissue samples of different organs and persons, 600 per class. We have made the dataset publicly available at `http://biscar.univalle.edu.co/?page_id=1003`. Algorithms were implemented in *C++* and *MATLAB*, using the *CImg* and *libSVM* libraries in a computer of *8-cores* and *8Gb RAM*.

4.2. Motivation

An organ can be identified by knowing the tissues that are present in a histological sample of it. Thus, recognising the tissue at hand is a primary task to carry out. For instance, muscle tissue is similar in different organs showing a homogeneous and compact structure with variable direction and organisation according to the region or cutting sample; and loose connective tissue has more separate or scattered structures. However, the appearance of a tissue may vary in the same organ, that can be observed by changes regarding the capture zone, cut or sample. Nonetheless, observed spatial patterns are indicated that texture descriptors may provide relevant information for tissue recognition. Figure 4.1 shows examples of blocks containing a unique tissue of size 100×100 pixeles. Note that, as explained in the Subsection 4.4.1.1 a block is a fixed non-overlapping $m \times m$ partition of a histology image. Each row shows specific patterns for each kind of tissue: (a) cardiac muscle tissue of the heart, (b) smooth muscle tissue of muscular artery, (c) smooth muscle tissue of the elastic artery, (d) smooth muscle tissue the large vein, and (e) loose connective tissue. It can be observed in Figure 4.1 that there exist similarity intra-class among the blocks of different tissues and difference inter-class among the blocks of the same tissues according to colour, texture and the effect of fibre orientation.

Moreover, the classification of fundamental tissues is an inverse and an ill-posed

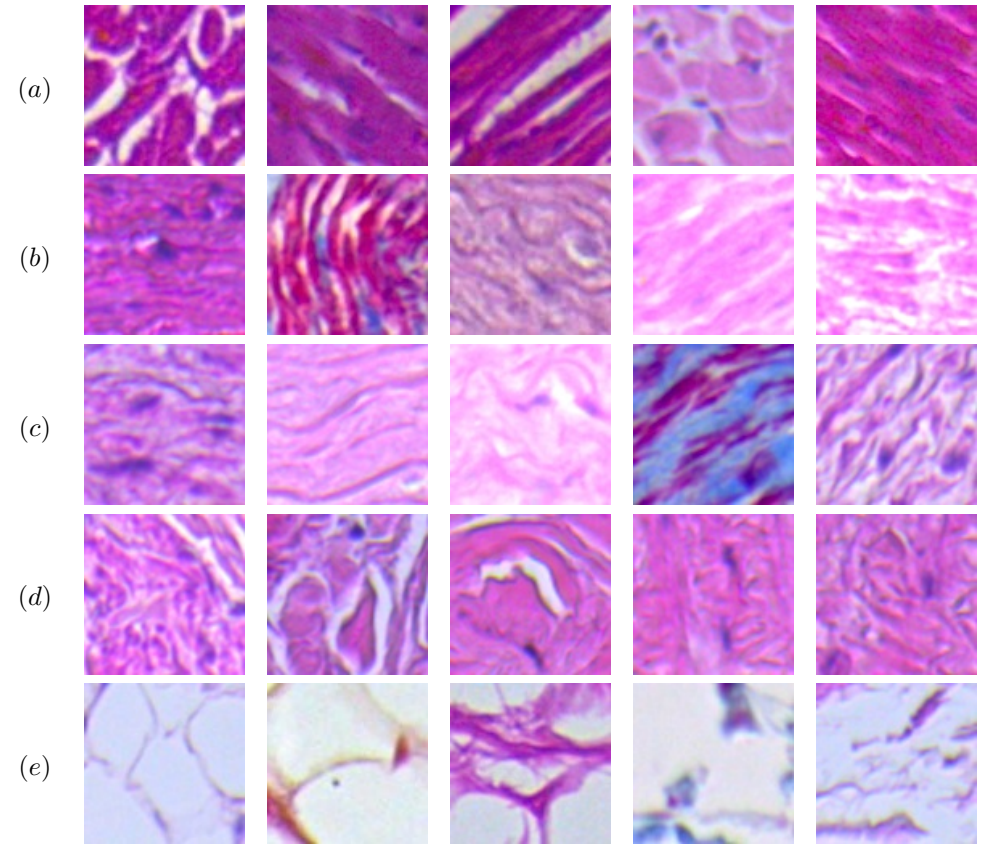


Figure 4.1: Examples of blocks. (a) Cardiac muscle tissue of the heart. (b) Smooth muscle tissue of the muscular artery. (c) Smooth muscle tissue of the elastic artery. (d) Smooth muscle tissue of the large vein. (e) Loose connective tissue

problem. Thus, the classification of fundamental tissues is heavily affected by the existing variability among blocks of the same tissue, making it an open problem.

4.3. Automatic Classification of the Fundamental Tissues of the Cardiovascular System

In this chapter we propose an approach to automatically classify the fundamental tissues and, in some cases, the organs of the cardiovascular system. The proposed classification is three-fold: (i) initially, an image is divided into blocks, (ii) then, information is extracted from blocks using the LBP-LBPri texture descriptor

and (iii) finally, LBP-LBPri texture descriptors are used to classify blocks using a cascade SVM. Figure 4.2 shows a general outline of our proposal. 1.) Blocks of 100×100 pixels are obtained from histological images. A block has the following two characteristics: (i) it contains only one type of tissue and (ii) it has discriminant information so that tissue recognition is possible. 2.) A feature extraction process is performed to obtain relevant information from a block. After evaluating different texture descriptors and combinations among them. We propose to use LBP and LBPri to represent efficiently local micro-patterns, which is possible due to the robustness of LBP and its variations. The feature extraction process returns a feature vector of 292 values obtained by concatenating LBP $(256\ values)$ and LBPri $(36\ values)$. 3.) A cascade SVM is used as a classifier, since it outperforms in comparison to the RF (Bader-El-Den, 2014) and the LDA (Ghassabeh et al., 2015). A SVM with a linear kernel is used to classify blocks into one of four classes: 3.1) the first class corresponds to the group that contains the smooth muscle of the large vein and the elastic artery; 3.2) the second class is smooth muscle of muscular artery; 3.3) the third class is cardiac muscle of the heart; and 3.4) the fourth class is loose connective tissue. 3.1.1) Organs in the first class are separated, using SVM with a polynomial kernel, into: 3.1.1.1) smooth muscle of the elastic artery and 3.1.1.2) smooth muscle of the large vein. The details of the complete process are presented in the following sections.

4.4. Tissues Description

In this section we propose an approach using texture descriptors to extract information about fundamental tissue morphology — composition and characteristics — for block-based recognition. Thereupon, tissue recognition in a histology image is achievable based on the obtained features.

4.4.1. Method

Our proposal is a block-based strategy that will be used to extract the content of an image. The recognition of each part in a image is an advantages that, we believe, will allow us to recognise organs in a complete image according to its blocks.

Partitioning Images into Blocks

One of the major issues in this work is the selection of the block size. If the block size is too large, then a block may contain more than one fundamental tissue and, thus may not provide much discriminating information. On the other hand, if the block size is too small, the fundamental tissues may be easily misclassified. Hence, a trade-off between block size and classification accuracy has to be achieved.

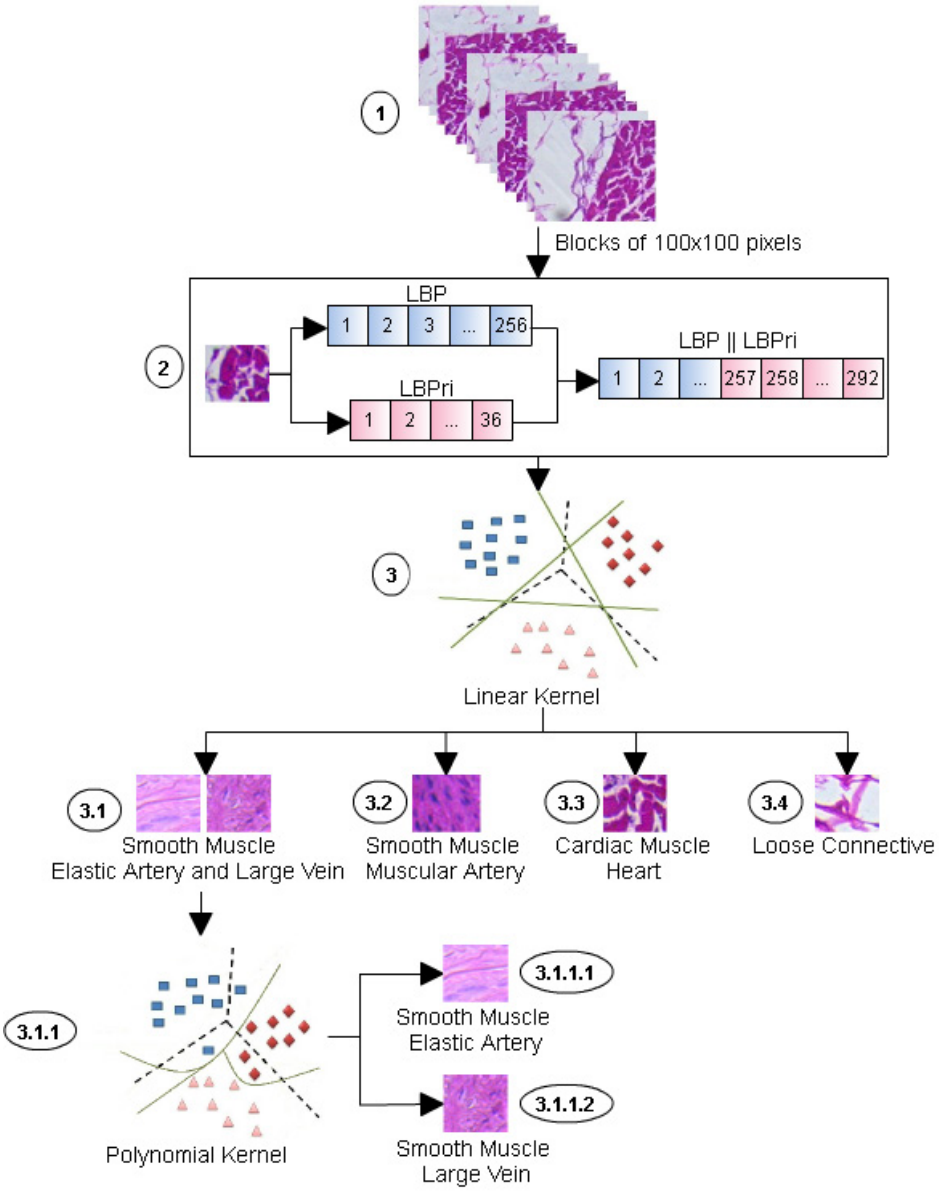


Figure 4.2: Proposed approach for automatic classification of fundamental tissues associated with an organ: (1) image block of size 100×100 pixels. (2) LBP and LBPri histograms from blocks. Feature vector is composed by concatenating histograms. (3) Classification using SVM with a linear kernel. (3.1), (3.2), (3.3) and (3.4) blocks classified. (3.1.1) SVM with a polynomial kernel is used to separate the first group into (3.1.1.1) and (3.1.1.2).

We considered four block sizes, $\{25 \times 25, 50 \times 50, 100 \times 100, 200 \times 200\}$ pixeles, in order to find a suitable size for obtaining a single fundamental tissue and being able to recognise a specific tissue automatically. An analysis of the partition size is performed using 25 histological images of different organs and stains, adding variability even in the same tissue. Examples of different block sizes are shown in Figure 4.3.

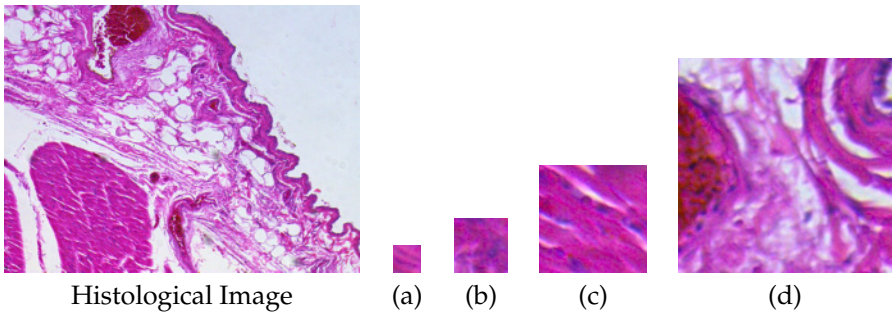


Figure 4.3: Examples of block sizes in a histological image. block sizes: (a) 25×25; (b) 50×50; (c) 100×100; and (d) 200×200.

We observe that blocks of size 25×25 and 50×50 are too small for tissues recognition. Blocks of size 100×100 present the following characteristics: (i) fundamental tissues can be recognised, (ii) a large number of blocks were classified with only one type of tissue (see Figure 4.4a); (iii) cases with two tissues are a consequence of the close relation between muscle and loose connective; and (iv) dense connective tissue is mistaken with muscle tissue, because both are highly similar and sometimes it is difficult to differentiate them even manually. Blocks of size 200×200 have some advantages and disadvantages, for instance: (i) it is easy to recognise epithelial tissue; (ii) the number of blocks with two or more tissues was increased (see Figure 4.4b); (iii) some blocks with presence of epithelial tissue were easily mistaken with connective tissue or light areas, according to the proportion of epithelial tissue in the block; and (iv) dense connective and muscle tissues may be wrongly classified. As a result of this analysis, we decided to use block size of 100×100.

Finally, the classes were modified to recognise the organ according to the muscle tissue — the smooth muscle of the muscular artery, the cardiac muscle of the heart, the smooth muscle of the elastic artery and the smooth muscle of the large vein.

Feature Selection

LBP features have become commonly used as texture descriptors in recent years (Kylberg and Sintorn, 2013). LBP (Ojala et al., 2002), LBPri (Pietikainen

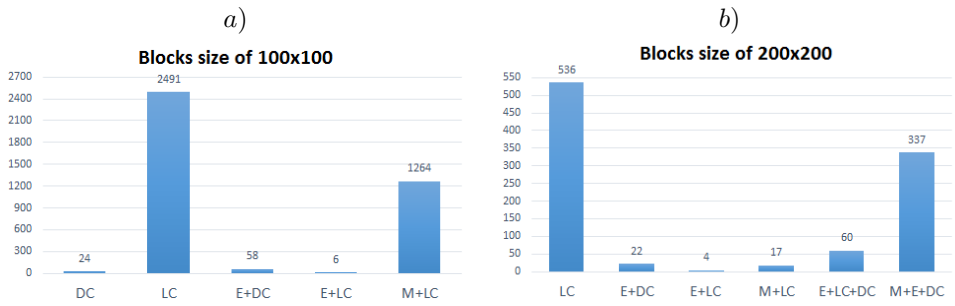


Figure 4.4: Classes by blocks identified for 13 histological images. E represents the epithelial tissue, M represent the muscle tissue, LC represents the loose connective tissue and DC represents the dense connective tissue.

et al., 2000) and Haralick features (Suganya and Rajaram, 2013) — contrast, angular second moment, energy, correlation, entropy, and first and second correlation measures — were tested individually and, concatenations of them were considered. In order to describe block content, LBP, LBPri, Haralick features, LBP+LBPri and LBP+LBPri+Haralick were tested. We use LBP with a radius equal to 1 and with 8 neighbours, LBP_8^{ri36} (for simplicity, this term will be referred to as LBPri henceforth) and Haralick features with distance equal to 3 pixels and 4 directions. As a result of the analysis, we decided to use LBP+LBPri as texture features in accordance with the obtained results in Figure 4.5 presented in Subsection 4.4.2.

Given a block from a histological image, the LBP ($lbp = [l1, l2, ..., l256]$) and the LBPri ($lbpri = [lr1, lr2, ..., lr36]$), the texture descriptor is created by concatenating histograms as follows:

$$F_{td} = [lbp||lbpri] = [l_1, l_2, ..., l_256, l_{r_1}, l_{r_2}, ..., l_{r_{36}}], \tag{4.1}$$

where F_{td} is the texture descriptor and $||$ symbolised concatenation.

Hence, the concatenation between LBP and LBPri increases the dimension of the feature vector but also improves the description of the image. A vector of size 256 is obtained by the LBP and vector of size 36 is obtained by the LBPri, which means that the total length of the F_{td} descriptor is $256 + 36 = 292$.

4.4.2. Experiments and Results

In this section we discuss the results obtained using the concatenation of LBP and LBPri, and comparisons with other texture descriptors. We present Receiver Operating Characteristic (ROC) curves for each test and Area Under the ROC Curve (AUC) is the measurement used to assess the response of the proposed approach (Robin

et al., 2011).

We use a metaheuristic procedure to generate a heuristic, in feature selection, that may provide a sufficiently good image description. We evaluate relevance features which are able to differentiate the classes by considering the selection of a set of features as a search problem, where different combinations are prepared, evaluated and compared with other combinations. We present the obtained results in Figure 4.5, for each set of features by class.

LBP+LBPri+Haralick was the best descriptor with an AUC of 0.9995 for smooth muscle of the muscular artery. LBP+LBPri has the best AUCs of 0.9981, 0.9818 and 0.9871 for cardiac muscle of the heart, smooth muscle of the elastic artery and smooth muscle of the large vein, respectively. LBP has a AUC of 0.9821 for the loose connective tissue class which is the best value for this class. However, the LBP+LBPri combination has high True Positive Rates which indicates good classification results globally. The worst F-Scores are achieved by Haralick features in all cases.

4.5. Tissues and Organs Classification

In this section we present the proposed approach based on machine learning algorithms to recognise fundamental tissues and, in some cases, the organs is obtained. The classification is done using the following five classes: (i) cardiac muscle of the heart, (ii) loose connective tissue — vein, arteries and the heart —, (iii) smooth muscle of the muscular artery, (iv) smooth muscle of the large vein and (v) smooth muscle of the elastic artery. A process was conducted in order to select the classifier, SVM, RF with different parameters and a LDA were evaluated and compared. SVM performs classification by finding the hyperplane that maximizes the margin between the two classes (Yang et al., 2012). RF is an ensemble of decision trees, where the number of trees will have a significant effect on the resulting model's accuracy (Bader-El-Den, 2014). LDA uses a linear combination of features that compute the directions, which will represent the axes that maximise the separation of two or more classes (Ghassabeh et al., 2015). Finally, we decided to use SVM based on the analysis of results, shown in subsection 4.5.2.

A brief summary of the proposed automatic classification is given below and a detailed explanation is presented in the following subsections. Given the texture descriptors for each block we propose a cascade classification using SVM with two different kernels. The first step consists of classifying each block into the following four classes using SVM with a linear kernel: (i) cardiac muscle of the heart, (ii) loose connective tissue — vein, arteries and the heart —, (iii) smooth muscle of the muscular artery, and (iv) smooth muscle of the large vein and the elastic artery. The second step is a cascade process using SVM with a polynomial kernel to separate

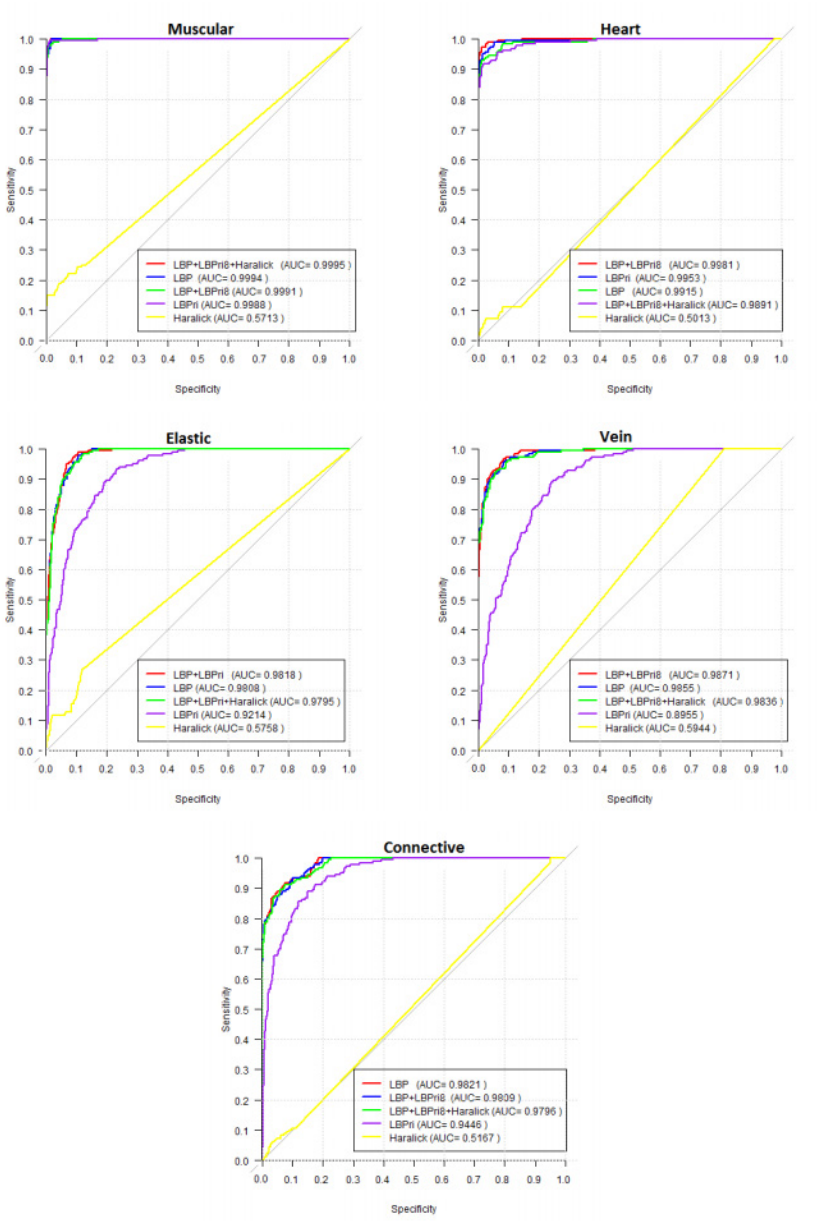


Figure 4.5: ROC curves with texture descriptors per class.

the smooth muscle of the large vein and the elastic artery, which were classified to be in the same class (as is illustrated in Figure 4.2 (3) to (3.1.1.2)).

4.5.1. Method

Given a set of blocks of 100×100 pixels manually labelled with a tissue — cardiac muscle of the heart, loose connective tissue, smooth muscle of the muscular artery, smooth muscle of the large vein, and smooth muscle of the elastic artery — and a texture descriptor concatenating LBP and LBPri, detailed in 4.4. We propose a classification method building a cascade SVM. This proposal is based on the obtained results from assessment SVM and Cascade SVM in subsection 4.5.2.1 in which the overall results show that SVM was outperformed by Cascade SVM with higher success rates for the classes of loose connective tissue, smooth muscle of the large vein and smooth muscle of the elastic artery. Finally, an evaluation of different kernels in the cascade proposal was conducted to select the appropriate kernel at each stage.

Kernel Selection

The kernel is defined as follows: let $x_1, ..., x_l$ be vectors in the input space, R^n, and $\phi(x)$ be a nonlinear mapping. The kernel-trick is used to calculate a kernel function $k : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$, $k(x_i, x_j) = \phi(x_i)^T \phi(x_i)$, avoiding explicit mappings (evaluations of) $\phi()$ (Yang et al., 2012). In this new space, the data are considered linearly separable and allow for an optimal separation (see Figure 4.6). Normally, a kernel is a function that domain experts provide to the SVM algorithm. We tested linear, polynomial, Radial Basis Functions (Gaussian) and sigmoid kernels. Finally, based on the results obtained in subsection 4.5.2.4, we selected to use a linear kernel in the first SVM and a polynomial kernel in the second one.

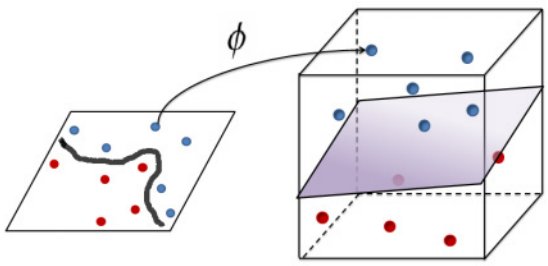


Figure 4.6: Example polynomial kernel function.

4.5.2. Experiments and Results

In this subsection we discuss the results obtained using SVM versus RF and LDA classifiers with the whole training data. We divided the evaluation into the following four subsections: (i) a comparison between SVM and Cascade SVM, (ii) a parameter selection optimisation for RF, (iii) a comparison between Cascade SVM, RF

and LDA, and (iv) kernel selection. We present ROC curves for each test and AUC is the measurement used to assess the response of the proposed approach (Robin et al., 2011). Muscular represents the smooth muscle of the muscular artery, Heart represent the cardiac muscle of the heart, Elastic represents the smooth muscle of the elastic artery, Vein represent the smooth muscle of the large vein, and Connective represent the loose connective tissue — veins, arteries and the heart.

SVM vs Cascade SVM

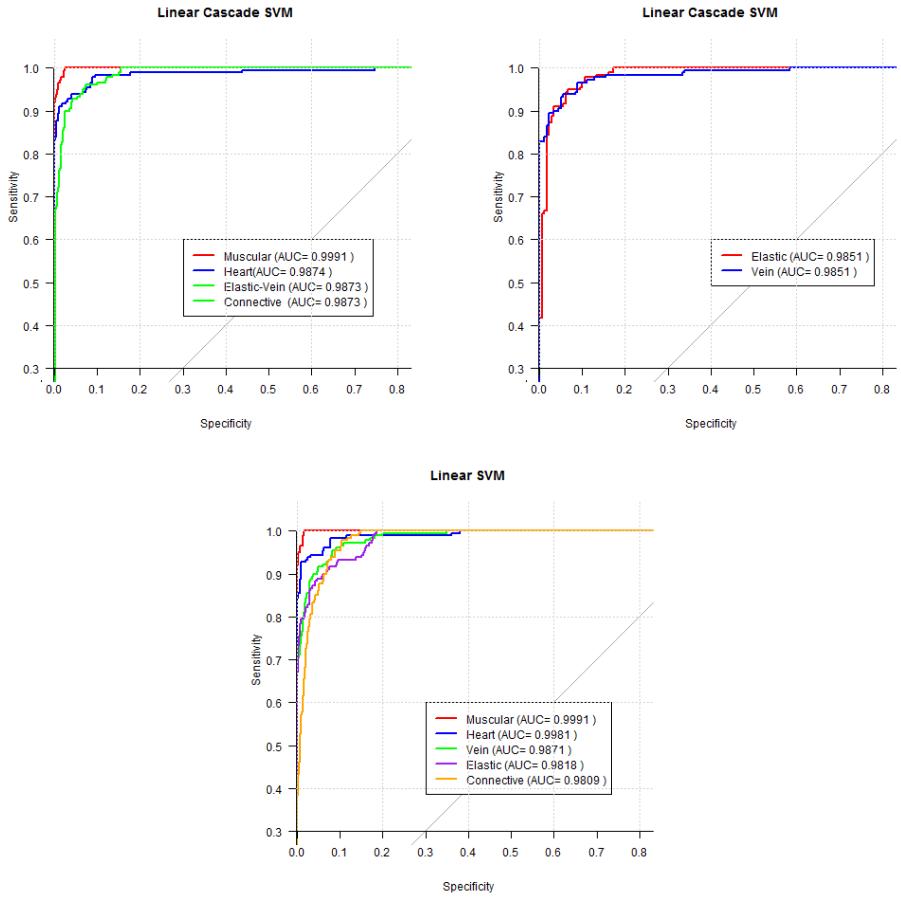


Figure 4.7: ROC curves for comparing linear SVM with linear Cascade SVM. Top to bottom, linear Cascade SVM and linear SVM. The axes are not displayed from the origin to improve visualisation.

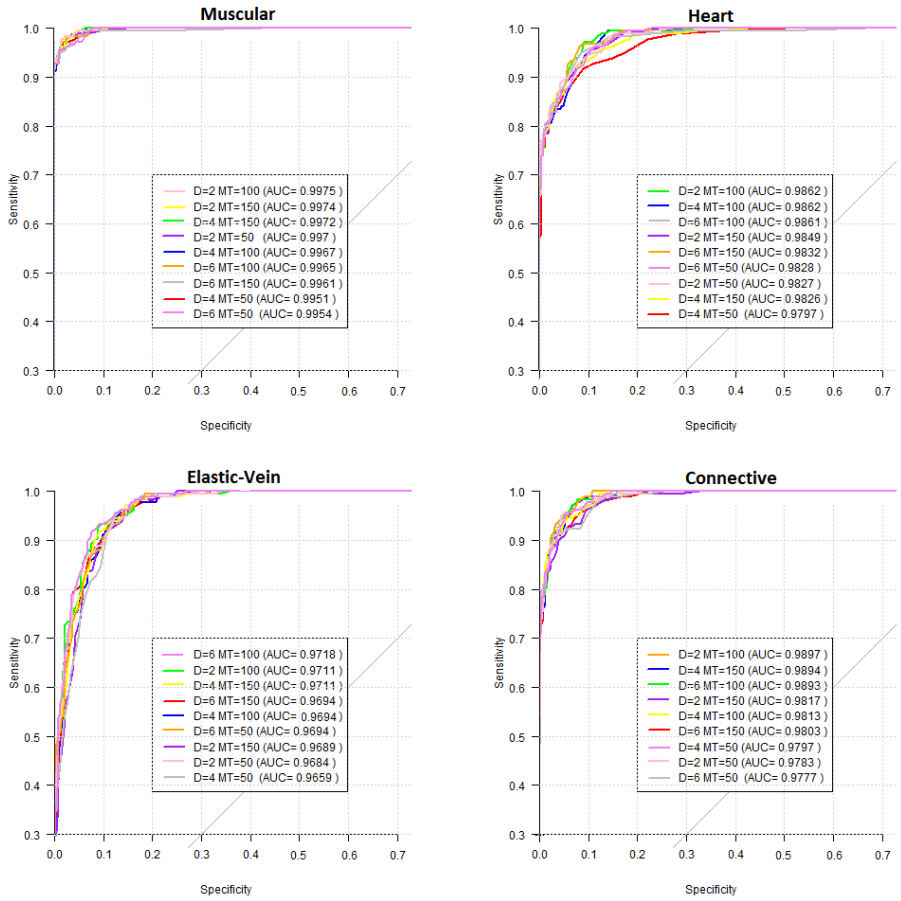


Figure 4.8: ROC curves with parameters selection optimisation for RF per class. The axes are not displayed from the origin to improve visualisation.

A linear SVM versus a cascade linear SVM is used in this tests and the comparison is summarised in Figure 4.7.

The best results are obtained using Cascade SVM for the loose connective tissue, smooth muscle of the large vein and smooth muscle of the elastic artery classes with an AUC of 0.9873. Smooth muscle of the muscular artery has the same results with both strategies, an AUC of 0.9991. Cardiac muscle of the heart is better classified using SVM with an AUC of 0.9981 compared with Cascade SVM with an AUC of 0.9876. Therefore, Cascade SVM outperforms SVM in most of the cases. The smooth muscle of the large vein and artery and smooth muscle of the elastic artery have low hit rates when a SVM is used. In our cascade proposal, these classes are classified

together in the first SVM and reclassified in the second SVM, increasing the hit rates.

Parameters Selection for RF Classification

Tree Depth and Max Trees are adjustable parameters to which RF is somewhat sensitive. We evaluated Tree Depth= $\{2, 4, 6\}$ and Max Trees= $\{50, 100, 150\}$ in order to obtain the most accurate, useful, and generalisable model. Figure 4.8 presents the results obtained, separated by class.

The ROC curve raises quickly towards high True Positive Rates, which indicates good classification results using RF with different parameters. Also, there is great variation among the different parameters per class. Nevertheless, True Positive Rates tend to decrease as Max Trees values decrease. The best global scores are achieved with Deep=2 and Max Trees=100.

Cascade SVM vs RF vs LDA

We evaluated the performance of each classification method through experiments using the testing data and analysed the values returned by the tests. Figure 4.9 depicts ROC curves and AUC values of testing performance comparisons.

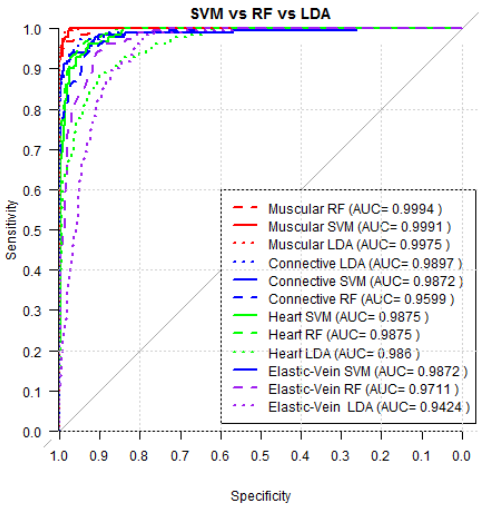


Figure 4.9: Comparative evaluation using ROC curves classification of SVM, RF and LDA for each class.

As shown in Figure 4.9, a direct comparison of the classification performances shows that the SVM yielded the biggest difference, while RF and LDA performed

similar although in some cases lower values are obtained with LDA. Cascade SVM yielded the best results for the cardiac muscle of the heart and the smooth muscle of the large vein and the elastic artery classes with AUC of 0.9875 and 0.9872 respectively. RF performed the best in classifying smooth muscle of the muscular artery with an AUC of 0.9994 and LDA had the best result classifying loose connective tissue with an AUC of 0.9897, but it is notable that SVM had the second best result in each case with AUCs of 0.9991 and 0.9872, correspondingly.

Kernel Selection

A good kernel selection in SVM provides more versatility to the obtained model. We tested different kernels in order to select the best kernels. In this case we use a combination of a linear kernel in the first SVM, because it is much faster and can yield good results in many cases, and a polynomial kernel to separate smooth muscle of the large vein and the elastic artery, which were recognised in the same class by the first SVM. Figure 4.10 illustrates the results obtained for each class.

Moreover, analysis indicated that varying performances of SVM were connected to the choice of different kernels for each SVM, during model selection. The usage of a linear kernel in the first SVM and a polynomial kernel for the second SVM yields high True Positive Rates, which indicates the best classification results in most of the cases. The Radial Basis kernel achieves the best classification for the cardiac muscle of the heart class with an AUC of 0.9954, and the Linear kernel is the second best with an AUC of 0.9875. The worst AUCs are achieved by Sigmoid kernel in every case except for the smooth muscle of the muscular artery.

Figure 4.10 illustrates that the performance increase becomes visible when a linear kernel is used in the first SVM and a polynomial kernel is used in the second one.

4.6. Classification of a Histological Image Using Block-based Recognition

In this section we present an approach to classify tissues and an organ in a histological image using the block-based recognition method described in the previous Section. In contrast to the block-based method, this approach include blocks which may contain one or more type of tissues. A brief summary of the proposed method is given below and a detailed explanation is presented in the following subsections. A general outline of our proposal is illustrated in Figure 4.11. (1) A histological image acquired at $10\times$ objective. (2) Blocks of 100×100 pixels are obtained from the histological image. (3) Block-based recognition method is used to classify blocks. (4) A histological image classification by blocks.

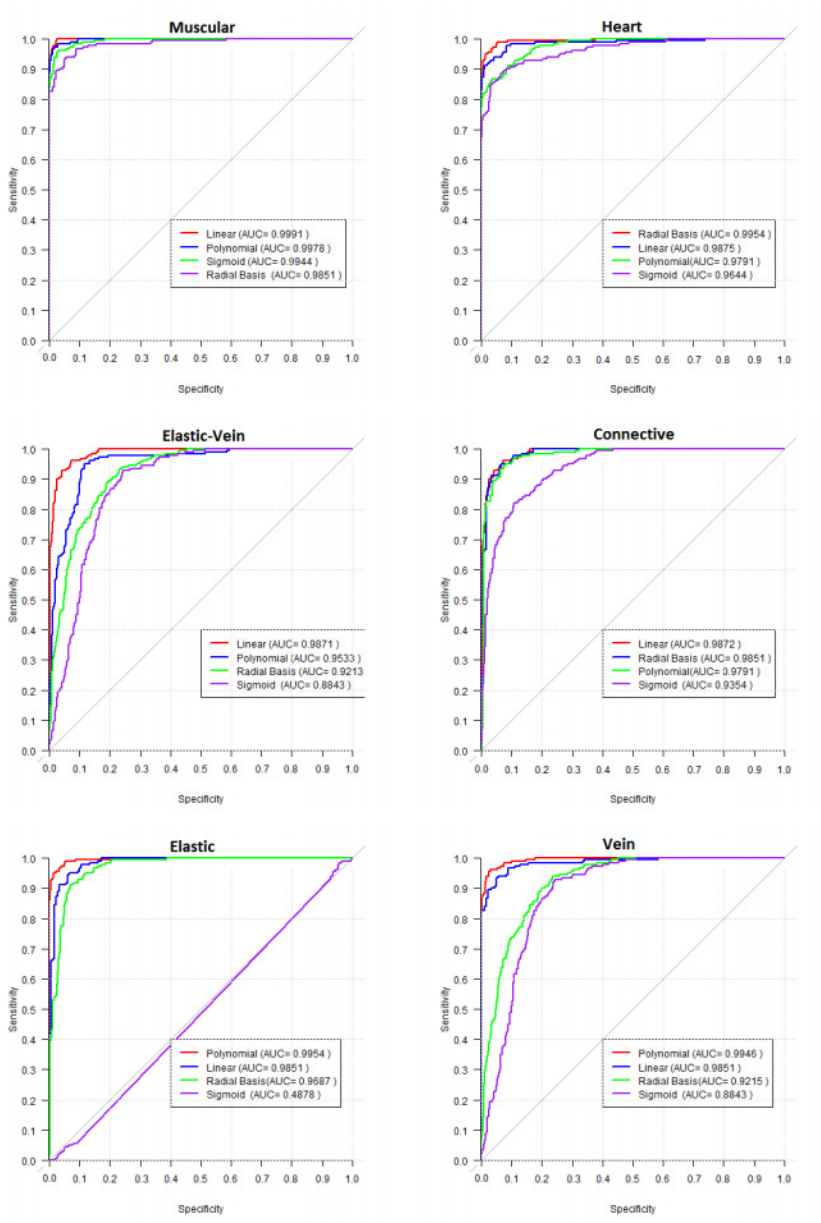


Figure 4.10: Comparative evaluation using ROC curves classification of SVM with different kernels for each class.

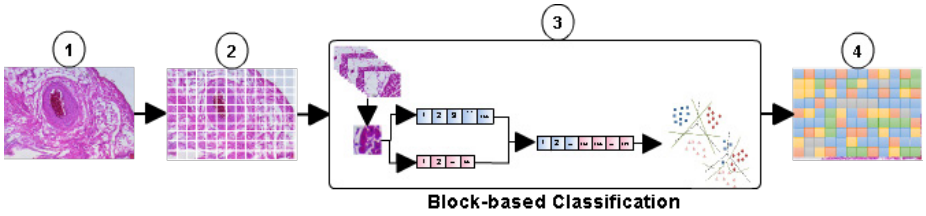


Figure 4.11: Proposed approach for automatic classification of fundamental tissues and an organ in a histological image: (1) Histological image taken with the $10\times$ objective. (2) Blocks of 100×100 pixels. (3) Block-based recognition method. (4) Histological image classification.

4.6.1. Dataset

Tissue samples and image capture protocol were described in Section 3.1. 1500 blocks, belonging to five histological images of different organs and persons, which were acquired at $10\times$ objective, were manually labelled and used as a ground-truth.

4.6.2. Method

In this subsection, the classification of a histological image process is presented. A classification of a histological image using block-based recognition process is defined as:

Let $I : \mathbb{I} \times \mathbb{I} \rightarrow \mathbb{R}^3$ be a histological image of the cardiovascular system in RGB colour space; $I_s = \{I_0, I_1, ..., I_K\}$ be a set of histological images; B be a matrix of blocks in which each B_{ij} represents the j-th block of the image i, note that blocks in a histological image may contain more than one tissue; $M_{rbc}(B_{ij})$ be the block-based recognition method described in Section 4.3 which classified blocks into one of the six classes: (i) cardiac muscle of heart, (ii) loose connective tissue — vein, arteries and heart —, (iii) smooth muscle of muscular artery, (iv) smooth muscle of large vein, (v) smooth muscle of elastic artery, and (vi) light regions; R_I be a $m \times n$ matrix of labels where $m = widthImage/blockSize$ and $n = heightImage/blockSize$, in our case 20×15. Then, classification of a histological image using block-based recognition is:

$$R_I = \begin{pmatrix} M_{rbc}(B_{11}) & M_{rbc}(B_{12}) & \cdots & M_{rbc}(B_{1n}) \\ M_{rbc}(B_{21}) & M_{rbc}(B_{22}) & \cdots & M_{rbc}(B_{2n}) \\ \vdots & \vdots & \cdots & \vdots \\ M_{rbc}(B_{n1}) & M_{rbc}(B_{n2}) & \cdots & M_{rbc}(B_{nn}) \end{pmatrix} \tag{4.2}$$

4.6.3. Experiments and Results

In this section, we discuss the classification of a histological image performance. Results of the classification process were evaluated taking into account hits and misses — the number of blocks that were corrected and incorrect classified, respectively — with their accuracy.

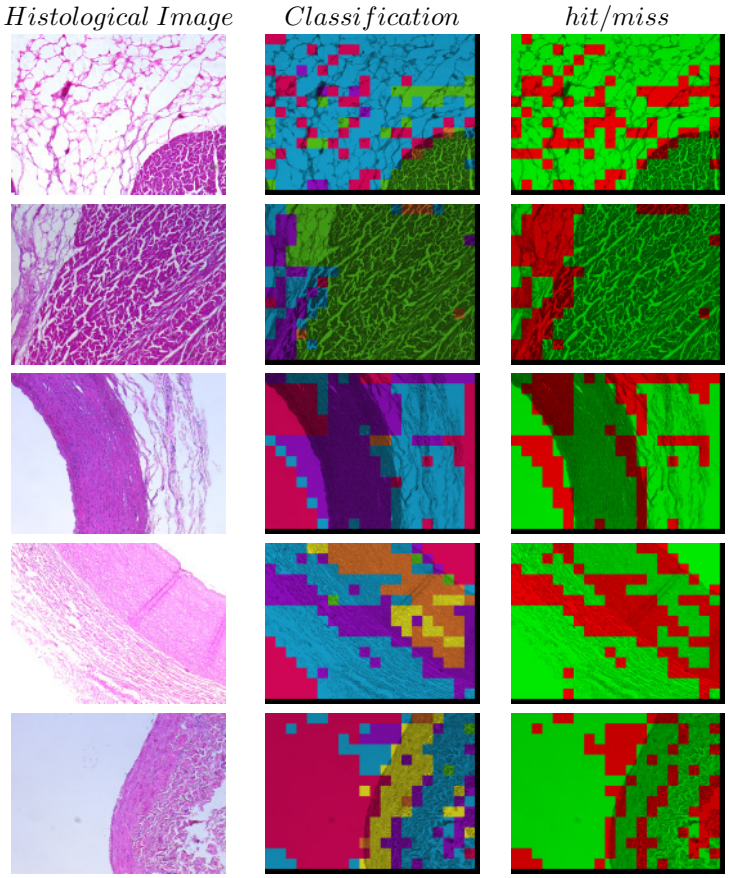


Figure 4.12: Results of automatic classification of a histological image. In the first column, histological images. In each row from top to bottom: *Img-He* and *Img-He1* represent the heart images; *Img-MA* represents the muscular artery image; *Img-EA* represents the elastic artery images; and *Img-LV* represents the large vein image. In the second column, automatic classification. In the third column, hits and misses of automatic classification.

A selected set of histological images, the automatic classification results R_I and their hits and misses are included in Figure 4.12. In Figure 4.12 the automatic classification process results represent each class of R_I with a distinctive colour as follows:

(i) cardiac muscle of the heart with green, (ii) loose connective tissue with blue, (iii) smooth muscle of the muscular artery with violet, (iv) smooth muscle of the large vein with yellow, (v) smooth muscle of the elastic artery with orange, and (vi) light regions with fuchsia. On the other hand, hits and misses results are represented in green and red colours, respectively.

Figure 4.13 contains a graphical representation of the hits and misses by blocks obtained with automatic classification process in the set of test images. 211 to 228 out of 300 blocks per image correctly classified were obtained. The accuracy is between 70.333% and 76.000% according to the histological image. The highest accuracy is obtained on the second step classification in the cascade SVM process with 37 and 53 blocks correctly classified and accuracy of 77.083% and 76.812% per image. This occur because this phase considers only two possible classes and less number of blocks. The accuracy measure obtained is over 70% compared with 90% obtained using the blocks that contains only a single kind of tissue.

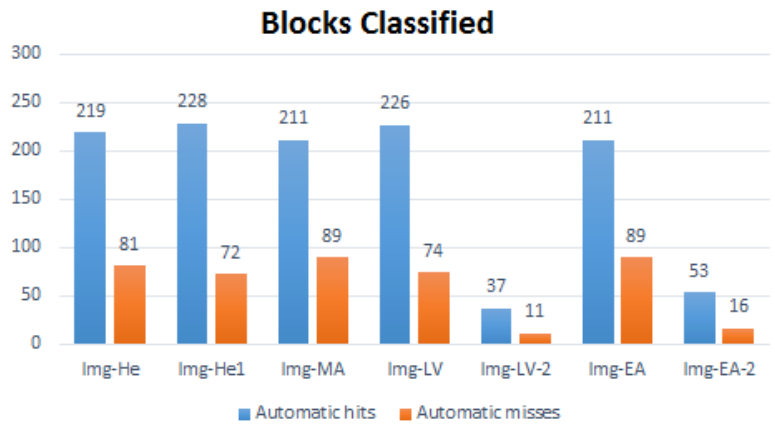


Figure 4.13: Hits and misses blocks with automatic classification process. *Img-He* and *Img-He1* represent the heart images; *Img-MA* represents the muscular artery image; *Img-EA* represents the elastic artery images; and *Img-LV* represents the large vein image. The name of each images with -2 at end correspond to the second classification in the cascade SVM process.

4.7. Conclusions

In this chapter we presented: First, a method that allows the recognition of the cardiac muscle of the heart, loose connective tissue — vein, arteries and the heart — and smooth muscle of the muscular artery, the large vein and the elastic artery with success rates greater than 90%. Second, we have determined that, among those tested, the best texture feature is concatenated LBP with LBPri and among the eval-

uated classifiers, a cascade SVM with linear and polynomial kernel since it yields to hit rates above 90%. Third, using the recognised tissues from a block, we were able to identify some of the organs associated to these tissues. For instance, when cardiac muscle tissue is recognised, we know without any doubt that the specific image contains at least a partial view of tissue from the heart. In a similar way, the correct classification of smooth muscle tissue allows for the recognition of the presence in the image of muscular arteries, veins and large elastic arteries. However, the correct classification of loose connective tissues does not provide information for organ identification since such tissues could be present in all the previously mentioned organs. Forth, we used the recognition of small blocks to recognise every part of a complete histological image to classify tissues and an organ in a histological image with accuracy greater than 70%. Additionally, we have created and made publicly available two dataset consisting of: (i) 3000 blocks — 600 per class — and (ii) 1500 blocks — 300 per image — that can be used to validate the results obtained in our work or to improve upon the proposed method.

Chapter 5

Improving the Automatic Classification of Histological Images using an Ontology of the Human Cardiovascular System

Knowledge is power, if you know it about the right person.
Ethel Watts Mumford

Expert knowledge contains all the knowledge granted by the experience. Knowledge representation and reasoning is necessary to formalise human knowledge and its processing within machines in a machine-interpretable form to solve complex tasks such as support teaching, medical practices or having natural language interactions. We propose to create a histological ontology to represent histological and expert knowledge. We left the ontology publicly available at `https://w3id.org/def/System`, this ontology used `https://w3id.org/def/Tissue` and `https://w3id.org/def/Organ`.

A research problem, in the histological context, consists of the recognition of patterns and infer the organisation of the fundamental tissues which enable the identification of specific organs. Histological knowledge and expert knowledge representation are described in Section 5.1. In Section 5.2 a refinement process using the histological ontology that improves the automatic classification of histological images is presented.

5.1. Building a Histological Ontology

Histological and expert knowledge are important sources in histology studies and practices. This sources have drawbacks — vocabulary heterogeneity, unambiguous language, semantic differences, subjectivity, language as well as procedural and ways of structuring — which affect research, analysis and information retrieval processes.

Two challenges are identified: (i) communicate specifically, clearly and precisely complex medical concepts and (ii) represent or model the knowledge of these data sources, interact and process with both automatically. These challenges require a profound analysis of the structure and the concepts of histological terminologies.

This can be achieved by constructing histological domain ontologies. Recently, the use of ontologies for representing the prior knowledge of a domain has been proposed in medicine focusing mainly on the representation and organisation of medical terminologies. The union between ontologies and medical information is presented as the necessary alternative to solve the main problems regarding these sources of information (Wang and Tansel, 2013; Guefack et al., 2012; Bertaud-Gounot et al., 2012).

The term "ontology" has many definitions depending on the author and the way it can be built and used for computer systems. One of the most widespread definition is the following: "Ontology is an explicit and formal specification of a shared conceptualisation" (Studer et al., 1998). Ontologies create models to formalise knowledge in a same way to be used in each area. From a histology perspective, an ontology would consist of concepts defined by histological knowledge. Additionally, relations, attributes, rules and axioms are among concepts that enrich and contribute to expand vocabulary of domain or the work area. Rules and relations between histology knowledge and expert knowledge will be represented. Finally, ontologies would contribute to define intimate relations between different components to study — cells, tissues, organs and systems.

We proposed to represent histological and expert knowledge creating a histological ontology of the human cardiovascular system. To the best of our knowledge, there is no histology ontology in the literature, so this would be a contribution to the research community.

5.1.1. Method

NeOn methodology is one of the most useful for ontology engineering (Suarez-Figueroa and Gomez-Perez, 2009). This methodology does not prescribe a rigid workflow, but instead it suggests nine scenarios for developing ontologies. However, the methodology cover commonly occurring situations which mostly focus on reusing, merging, restructuring and re-engineering ontological resources. Taking into account that we will create a histological ontology without reusing ontological resources according to the State-of-the-Art, we decided use (Castro et al., 2006) methodology. This methodology consists of the following steps: (i) identification of purpose, scope, competency questions and scenarios, (ii) identification of those ontologies we could reuse, (iii) domain analysis and knowledge acquisition, (iv) iterative building of informal ontology models, (v) formalisation and (vi) evaluation. We modify minimally (Castro et al., 2006) methodology in i, iii and vi steps, Figure 5.1 presents the methodology to develop ontologies of (Castro et al., 2006) with our modifications. Firstly, we merge step (i) and (iii) which will be our first step called capturing expert and histological knowledge. Secondly, we use four evaluation criteria — detecting pitfalls, expert evaluation, answering competency questions and

heavyweight ontology — and (Castro et al., 2006) uses two evaluation criteria — CMs and the protégé axiom language plug-in provided by Protégé.

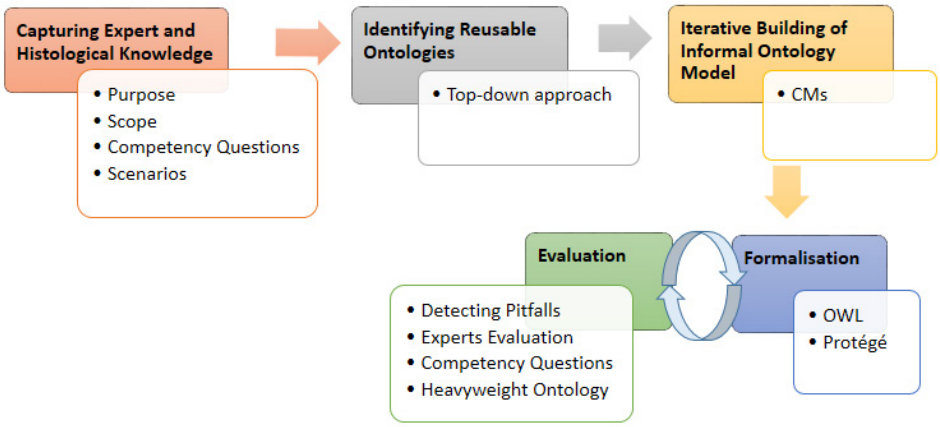


Figure 5.1: Methodology to develop ontologies.

Capturing Expert and Histological Knowledges

In this step, the aim is to capture the domain knowledge using activities to extract the information. We planned a series of activities with the experts through which the principle bases of our ontology were built: purpose, scope, competency questions and scenarios. We hosted a series of meetings with the group of histology experts conformed by members of the research group Teblami[1], belonging to the University of Valle[2], during which the domain experts discussed the terminology and structure used to describe the processes to analyse a histological sample.

Some of the questions to answer at this stage are the following: (i) what is the ontology going to be used for?, (ii) what do we want the ontology to be aware of?, (iii) what is the scope of the knowledge we want to have in the ontology?, and (iv) how is the ontology going to be used?. The answers are represented in the next elements:

Purpose, Scope and Scenarios

Commonly, ontology development is not the final goal of the process. Instead, an ontology becomes a tool to be used by other systems. Under this perspective, the purpose is defined by taking into account the main reasons that can lead to create an ontology (Noy and Mcguinness, 2001). Our Ontology will be constructed for:

[1] https://sites.google.com/a/correounivalle.edu.co/grupo-de-tejidos-blandos-y-mineralizados/

[2] www.univalle.edu.co

- Sharing a common understanding of histology knowledge between people and machines in processes such as automatic recognition and identification of cells, tissues and organs.
- Allowing reuse of domain knowledge.
- Allowing change specifications of histology knowledge, if changes occur in it. In addition, explicit specifications of histology knowledge are useful for new users who must learn the meaning of histological terms.

However, this work covers only the human cardiovascular system which is one of the more complex systems to be treated because of its differences with other systems and the peculiarities of its organs.

Two scenarios are described in order to illustrate and motivate the need for the histological ontology. These scenarios are later used to develop a set of competency questions and indicate how the ontology would be used in these cases.

Medical: a histology expert works in a hospital analysing samples in a cardiovascular system context. When receiving a sample, they analyse, labels and validates different characteristics of the sample from perspectives — cells, tissues and organs.

Professor: a histology expert works as a professor in a university teaching histology of the cardiovascular system. The expert teaches different group levels, covering histology of cells, tissues, organs and systems. The professor needs to cover each topic considering its components, relations and organisations. Additionally, they could also promote self-learning to on-campus students and facilitate on-line learning to external or remote students.

Having defined the purpose, scope, and scenarios of the ontology, we discussed the competency questions with our histology experts, henceforth experts. These competency questions were used at a later stage in order to help evaluate our model.

Competency Questions

Competency questions are the kind of questions for which we want the ontology to be able to provide support for reasoning and inferring processes. Additionally, those questions are essential for evaluating ontologies (Grüninger and Fox, 1995). Experts should express the competency questions in natural language without any constraint. Based on the above scenarios, we have identified four categories of competency questions: classification, properties, constraints and inferences. Examples of those competency questions are presented in Table 5.1, and Annex 7 contains the complete document.

Classes and Properties

In this step, we illustrate the construction of our ontology and explain its primitive classes and properties. The core of histological ontology are the following

Table 5.1: Examples of competency questions.

Classification
What are the organs of the cardiovascular system?
What is the composition of the myocardium?
What are the muscular arteries?
Properties
What are the tunics in veins?
Which is the constitution of a media tunic?
What are the structures present in the large veins?
Constraints
A simple epithelial tissue cannot be stratified
A capillary is only composed of endothelium
An organ can have three tunics maximum
Inferences
If a set of cells is close to a light region, then the tissue is probably an epithelial tissue
If an organ has a thin media tunic as well as a thick adventitia tunic and a wide light region, it is probably a vein
If an organ has a thick media tunic and a small light region, it is probably an artery

classes: cells, tissues, organs and systems. These are the main structures to represent. Some examples of leading properties are: layers, cell morphology, ducts, specialisation, mechanism of secretion, nature of secretion, valves and nodes. Some examples of object properties of histology ontology are included in Table 5.2.

Table 5.2: Object properties in histology ontology.

Property	Domain Class	Range Class	Inverse Property
isOrganOf	Organ	System	hasOrgan
isTypeOf	TypeOrgan	Organ	hasType
isCellOf	Cell	Tissue	hasCell
isMorphologyOf	Cell Morphology	Epihelial Tissue	hasMorphology
hasNumberLayer	Epithellial Tissue	Number Layer	isNumberLayerOf

A modular implementation taking into account tissues, organs and systems was used in our ontology to facilitate integration and/or reuse of histological data.

Two tasks were developed in this stage: (i) build the glossary of terms with definition and synonyms, and (ii) build the taxonomy of concepts. Figure 5.2 shows the complete glossary of terms obtained for the human cardiovascular system. Figures 5.3 to 5.9 show the taxonomies to cells, tissues and organs; these taxonomies are divided to show in more detail the different components and relations.

Órgano	Región	Capa	Sector	Estructura	Célula	Tejido
Corazón	Ventrículo izquierdo	Endocardio	Valvula atrioventrícular derecha o tricúspide	Fibra elástica	Cardiomiocito	Epitelio plano simple
Arteria de gran calibre o arteria elastica	Ventriculo derecho	Miocardio	Valvula atrioventrícular izquierda o mitral	Lámina elástica interna	Miocito liso	Músculo estriado cardiaco
Vena	Atrio izquierdo	Pericardio	Válvula aórtica	Lámina elástica externa	Adipocito	Musculo liso
Vena de pequeño calibre	Atrio derecho	Túnica intima	Válvula pulmonar	Vasa vasorum	Fibroblasto	Conectivo laxo
Vena de mediano calibre		Túnica media	Cuerdas tendinosas		Endotelial	Tejido adiposo
Vena de gran calibre		Túnica adventicia	Músculos papilares		Fibra de Purkinje	Conectivo denso irregular
Capilar Sinusoide		Mesotelio	Músculos pectinados		Mesotelial	
Capilar Fenestrado		Endotelio	Trabécula cárneas			
Capilar continuo		Subendotelio	Nódo atroventricular			
Capilar Linfatico		Subendocardio	Nodo Sinusal			
Vénula		Pericardio Visceral	Sistema conducente de corazón			
Arteriola		Pericardio Parietal	Haz de his			
Arteria Muscular			Fibra de Purkinje			
Arteria coronaria						
Arteria Carótida						
Arteria Femoral						
Arteria Renal						
Arteria Pulmonar						
Arteria Cerebral						
Vena cava						

Figure 5.2: Glossary of human cardiovascular system.

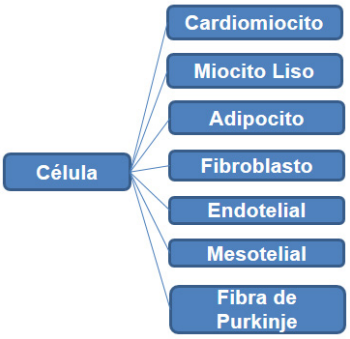


Figure 5.3: Taxonomy of main cells observed in a sample of the circulatory system.

Identifying Reusable Ontologies

We followed a 'top-down' approach (Gandon, 2002) in which histology experts work together to identify requirements and create the Conceptual Models (CM). Ontology search and analysis were performed using different approaches, histological and anatomical, in order to assess if some parts were reusable (Rubin et al., 2008). However, we did not find in the State of the Art an ontology according to our proposal. Finally, we did not reuse any available ontology, nevertheless, terms which are related to existing ontologies will be linked using *rdfs:seeAlso*.

Figure 5.4: Taxonomy of the fundamental tissues. The epithelial tissue is not completely displayed here to improve visualisation.

Iterative Building of Informal Ontology Models

We use CMs in each step of our methodology. CMs are graphs comprised of nodes connected by arcs representing concepts and relations between them (Canas et al., 1999). CMs were useful to share and capture knowledge, to facilitate communication with experts as well was to formalise use cases, and evaluation.

Histology and expert knowledge are represented very specific with instances and relations with as much detail as possible in CMs. Subject-predicate structures are easily identified with this knowledge modelling. Subjects are entities that perform or receive an action, whereas the predicate is everything may be said about a subject. The subjects, predicates and objects are extracted since histological knowledge manually.

Classes and subclasses were identified using the CMs representation; for example, epithelial tissue is_a fundamental tissue and simple flat epithelium *is_an* epithelial tissue. Similarly, attributes were obtained. For instance, *has_attribute or is_attribute_of*. An iterative process was carried out to represent histological and expert knowledge by providing through a full narration of the instances, specific properties, and relations. Experts did a validation process after obtaining our representation of the knowledge.

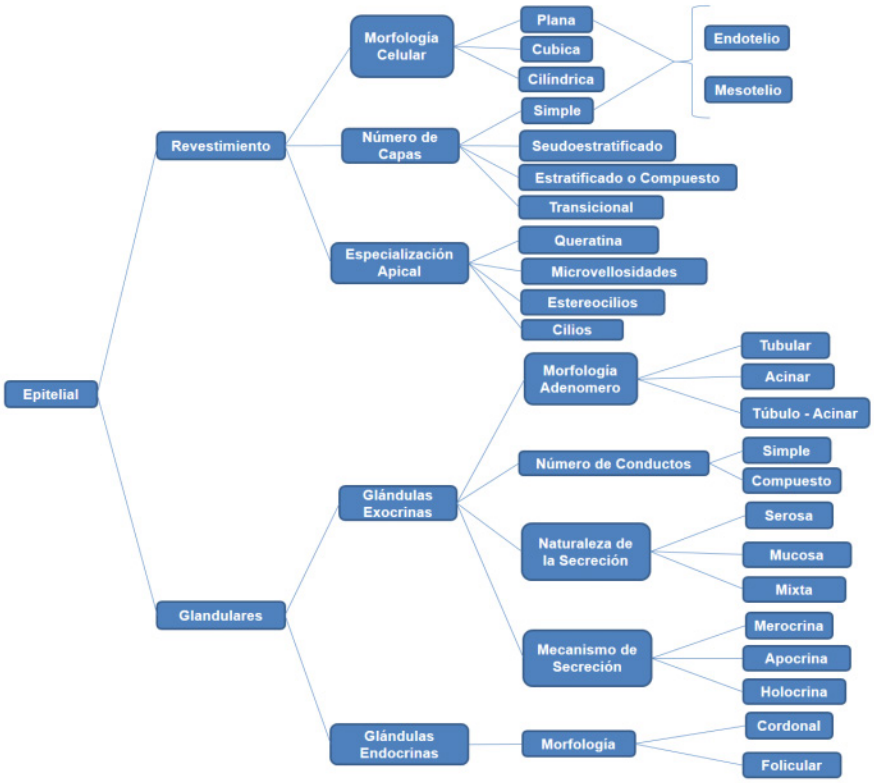


Figure 5.5: Taxonomy of the epithelial tissue.

Formalisation

Informal models obtained in the last step with CMs are converted to formal models which are valid computationally. This will be done using OWL and Protégé. Formal languages enable the encoding of knowledge and often include reasoning rules. Our histological ontology is expressed in OWL (OWL Working Group, 2009; Hitzler et al., 2009) and this was implemented using Protégé (Horridge et al., 2004).

Web Ontology Language (OWL): is characterised by formal semantics. The purpose of OWL is to develop ontologies that are compatible with the World Wide Web. OWL was developed to overcome the weak expressive power of RDF(S). The expressivity of RDF(S) is enhanced by OWL with tools for: describing relations between classes, defining property characteristics, cardinality and value restrictions on properties, and among others (Kapoor and Sharma, 2010; Sabou, 2006).

Protégé: is an open-source platform that provides a growing user community with a suite of tools to construct domain models and knowledge-based applications

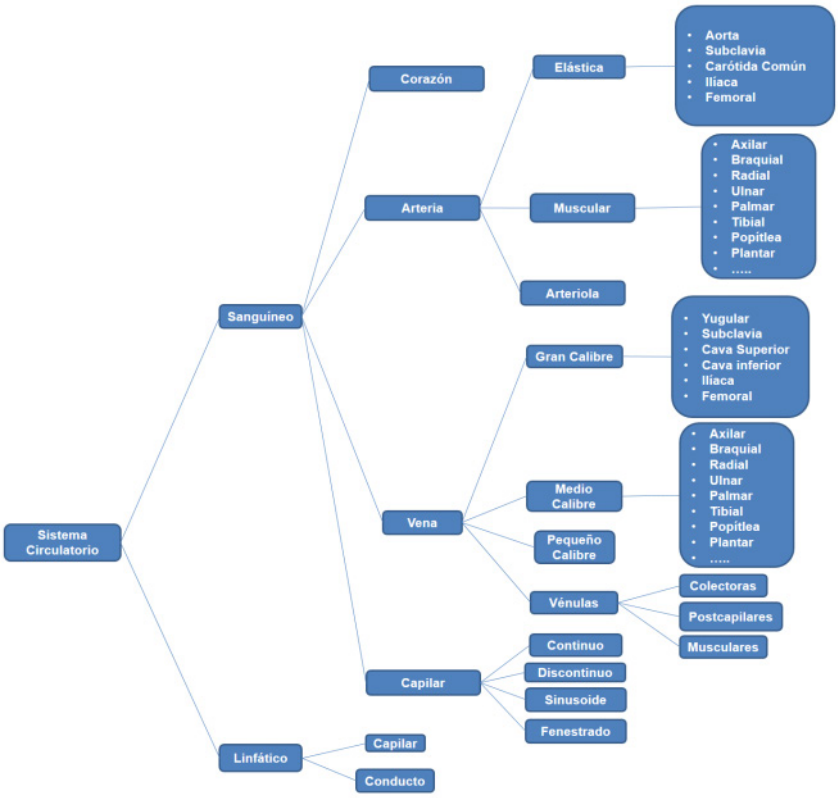


Figure 5.6: Taxonomy of histological classification of the circulatory system.

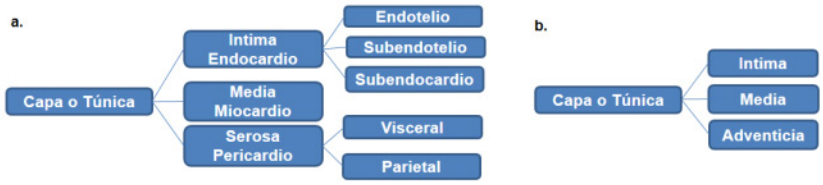


Figure 5.7: : (a) layers of the heart. (b) layers of blood vessels.

with ontologies (Horridge et al., 2004). Protégé was developed at Stanford Medical Informatics. Protégé can be currently used to load, edit and save ontologies in different formats including XML, RDF, UML, and OWL.

The transformation from CMs models to an OWL model requires interdisciplinary work. Domain experts develop part of the ontology, modelling their knowledge, with the assistance of knowledge engineers. Experts defined with as much detail as possible classes, properties and relations to obtain a consistent OWL model.

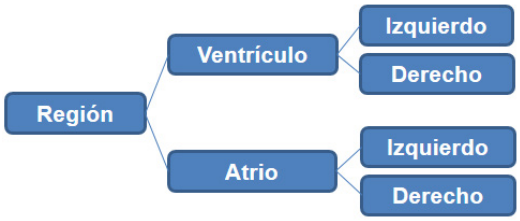


Figure 5.8: Taxonomy of classification of anatomical regions present in the heart.

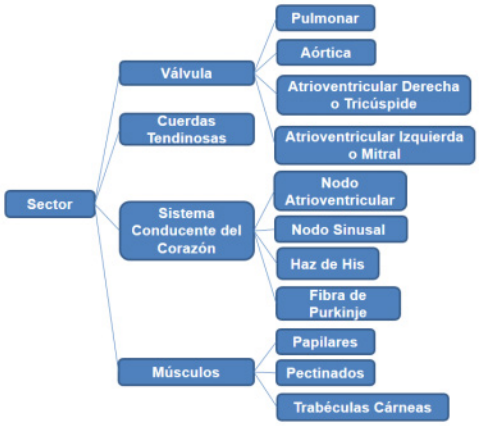


Figure 5.9: Taxonomy of classification of anatomical sectors present in the heart.

Evaluation

We propose a four-fold approach to validate our ontology before putting this into use. First of all, detect some of the most common pitfalls using OOPS. Secondly, expert evaluation using CMs. Thirdly, we will evaluate how accurately the ontology could answer our competency questions. Fourthly, we will evaluate if our ontology is a heavyweight or lightweight ontology.

Detecting Pitfalls: we use *Ontology Pitfall Scanner! (OOPS!)*, a web tool to detect the most common pitfalls in ontologies. *OOPS!* detects warnings in cases such as: reasoning problems, naming conventions, unconnected elements, modelling as well as reasoning problems and many others described in the catalogue (Poveda-Villalón et al., 2014). This evaluation enables to improve the maintainability, the accessibility and the clarity of the ontology.

OOPS! shows the results for each pitfall in four different ways depending on the kind of pitfall — critical, important, minor and suggestion. After executing *OOPS!* with the histological ontology, we obtained a summary of the pitfalls encountered as presented in Figure 5.10 and Figure 5.11. Figures show two pitfalls being detected

as well as one suggestion and one warning in each case.

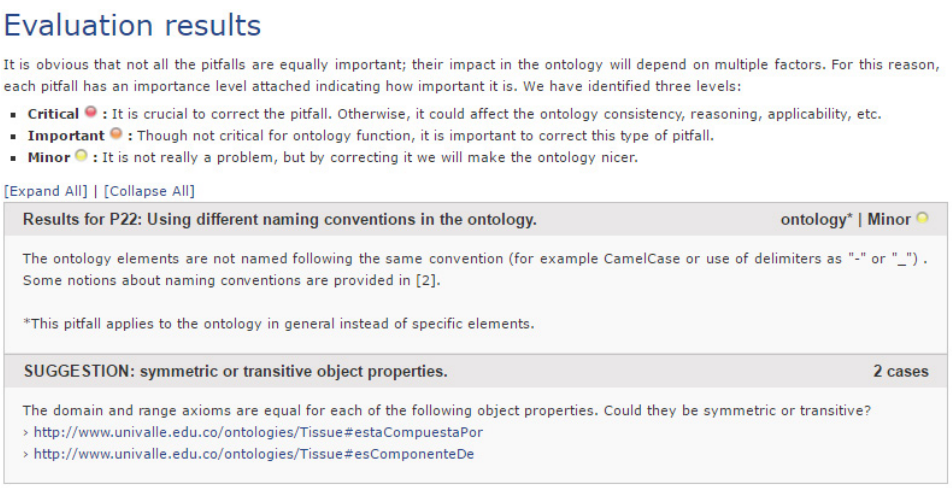


Figure 5.10: Evaluation results for tissues.

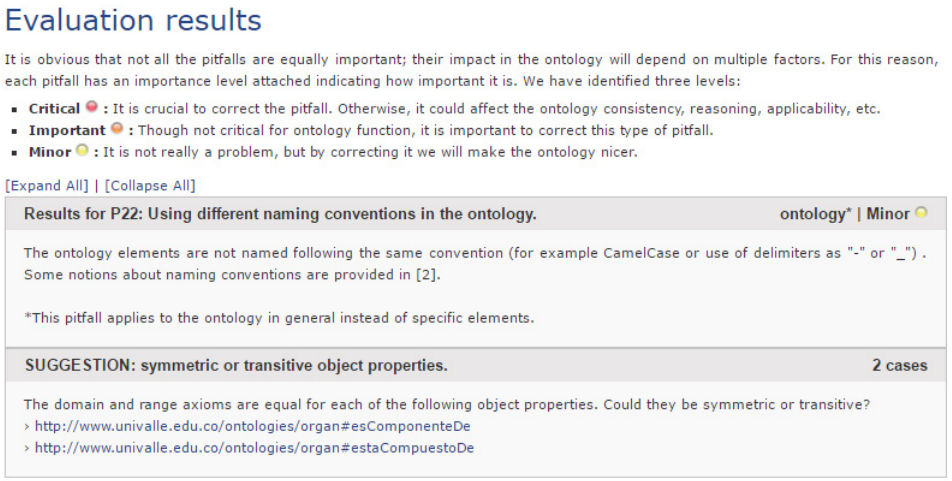


Figure 5.11: Evaluation results for organs and system.

The results suggest that "the domain and range axioms are equal in an object properties" and that a warning refers to the convention used. However, those are not pitfalls in our case and do not affect its correctness. Note that *OOPS!* gives us suggestions but according to modelling it could fit, or not, to *OOPS!*'s results.

Expert Evaluation: we use our CMs to evaluate the ontology taking into account that CMs represent the conceptual scaffold of the knowledge we are representing.

Although several criteria are used to validate ontologies, we are interested in the formal correctness of the ontology such as described in Gómez-Pérez (2004):

- Completeness: the concepts presented cover all terms related to the the cardiovascular system.
- Duplication errors: some elements of the ontology are redundant.
- Disjunction errors: defining a class as a conjunction of distinct classes.
- Consistency and coherence: check if the current definitions have been accurately represented — syntactically and semantically.

Abacha and Zweigenbaum (2015) propose a validation of medical ontologies through simple questions with only two possible answers (Yes/No) and a textual feedback. This method makes the evaluation easier for the medical experts and interpret their feedback. We used this method through the construction of a survey.

The elaboration of this survey was addressed with four basic objectives: (i) identify elements need to be validated, (ii) organise the elements to be validated, (iii) identify the characteristics to be validated in these elements, and (iv) interpret the feedback and make the necessary updates.

The first step consists in generating appropriate, natural language questions according to our purpose, which was focused on evaluating the histological ontology — possible answers Yes/No and a textual feedback. We have made the complete survey publicly available at the following URL: `http://survey-megaspace.rhcloud.com/index.php/741917` and Annex 1. The second step consists in applying the survey to our group of experts. The third step consists in interpreting expert's feedback to validate or modify the ontology.

We applied two different surveys. The first survey was applied in order to do an initial evaluation on the first version of our ontology, which was enhanced following the expert recommendations. This survey was taken by 20 students of Medicine and Surgery in semester 5 of the University of Valle. The second survey was taken by 51 experts from Latin America (See Figure 5.14) with different specialties presented in Figure 5.12, of which 32 have over 10 years of experience and the rest have less than 10 years of experience. Additionally, their action fields are presented in Figure 5.13. The results of the surveys are summarised in Figures 5.15, 5.16 and 5.17.

Taking into account our criteria to be evaluated, the experts' evaluation tackles issues concerning concepts and logical relations. Where possible, the first version of the ontology was enhanced by following the experts' recommendations. However, one of the drawbacks of the first survey was the lack of experience of the participants. For this reason, their answers were previously revalidated with an expert in order to take them into account.

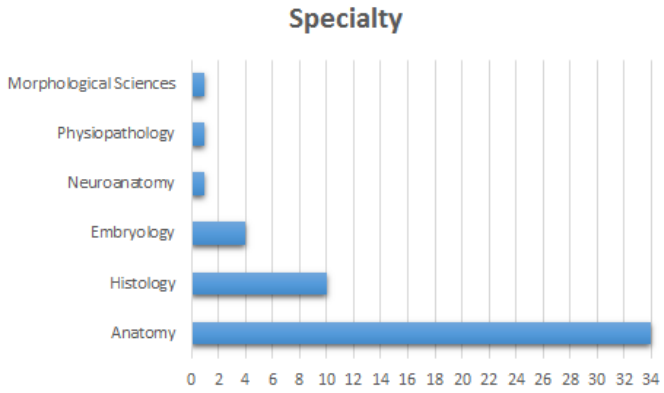


Figure 5.12: Experts by specialty of the second survey.

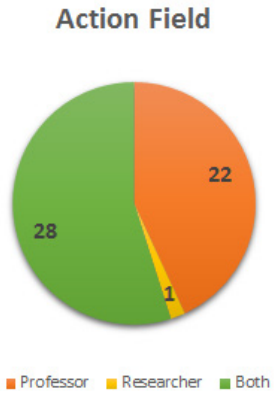


Figure 5.13: Experts by action field of the second survey.

The results show that the new version of the ontology is better than the first one, according to the results obtained between the surveys. This confirms that the former version had been improved thanks to the experts' suggestions. Additionally, we make sure that our ontology is designed in a way that enables the definition of new terms or modifications without needing to redefine existing terms. This characteristic is called extendibility.

The final results were as follows: completeness was tackled by the first question in each CM; some relevant concepts were added to the ontology after the first evaluation. Duplication and disjunction were evaluated by the second question in each CM; we have also ensured that there was neither duplication nor conflict in concepts. Consistency and coherence were covered by the third question in each CM; the results showing that the experts agreed with this aspect of our ontology.

The results obtained in the last survey are crucial for us as they were provided

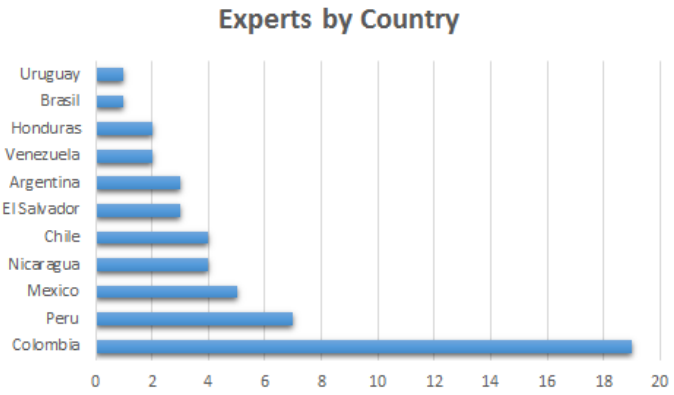
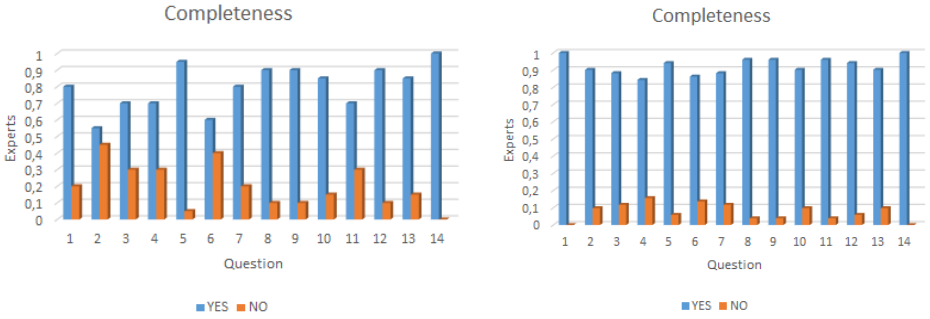


Figure 5.14: Experts by country of the second survey.



(a) (b)

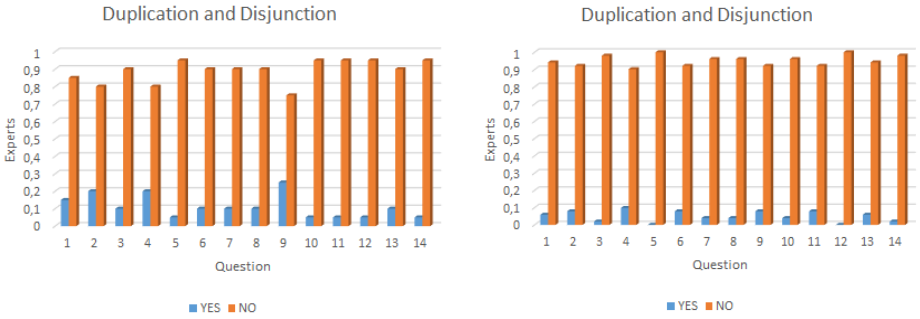
Figure 5.15: Completeness results:(a) First survey. (b) Second survey.

by experts with a lot of experience in histology. This means that their feedback was really valuable for our research and the fact that we obtained positive results makes it possible to put the ontology into use.

Answering Competency Questions: we evaluate the capability of the ontology to answer the competency questions of Section 5.1.1, using SPARQL (Prud'hommeaux and Seaborne, 2008). SPARQL was used to represent the competency questions to retrieve the data from the ontology according to the query. SPARQL queries were created to verify if the ontology gives the correct answer for each Competency Question (CQ). CQ, SPARQL query and a figure with the result obtained are presented in the following examples:

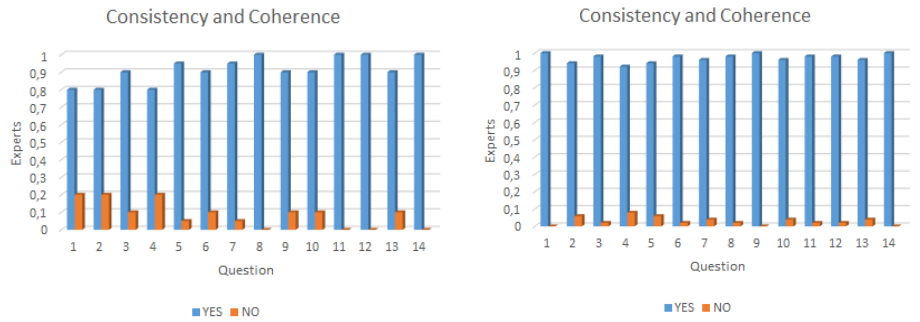
CQ-0: What are the fundamental tissues? Figure 5.18 shows the obtained results.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
```



(a) (b)

Figure 5.16: Duplication and disjunction results:(a) First survey. (b) Second survey.



(a) (b)

Figure 5.17: Consistency and coherence results:(a) First survey. (b) Second survey.

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX tissue: <https://w3id.org/def/Tissue#>
SELECT ?s ?name
WHERE { ?s rdfs:subClassOf tissue:Tejido;
                  rdfs:label ?name .}
```

CQ-1: What are the types of connective proper tissue? Figure 5.19 shows the obtained results.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX tissue: <https://w3id.org/def/Tissue#>
SELECT ?s ?name
WHERE { ?s rdfs:subClassOf
tissue:TejidoConectivoAdultoPropiamenteDicho;
                rdfs:label ?name}
```

s	name
TejidoNervioso	"Tejido Nervioso"@es
TejidoNervioso	"Nervous Tissue"@en
TejidoEpitelial	"Epithelium Tissue"@en
TejidoEpitelial	"Tejido Epitelial"@es
TejidoMuscular	"Tejido Muscular"@es
TejidoMuscular	"Muscle Tissue"@en
TejidoConectivo	"Connective Tissue"@en
TejidoConectivo	"Tejido Conjuntivo"@es
TejidoConectivo	"Tejido Conectivo"@es

Figure 5.18: Obtained results for CQ-0.

s	name
ConectivoEspecializado	"Specialised Connective"@en
ConectivoEspecializado	"Conectivo Especializado"@es
ConectivoDenso	"Tejido Conectivo Denso"@es
ConectivoDenso	"Dense Connective Tissue"@en
ConectivoLaxo	"Tejido Conectivo Laxo"@es
ConectivoLaxo	"Loose Connective Tissue"@en

Figure 5.19: Obtained results for CQ-1.

CQ-2: What are the layers present in the heart? Figure 5.20 shows the obtained results.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX organ: <https://w3id.org/def/Organ#>
SELECT ?s ?name
WHERE { ?s rdfs:subClassOf
organ:TunicaCorazón;
                  rdfs:label ?name .}
```

s	name
TunicaCorazónSubendotelio	"Subendotelio"@en
TunicaCorazónSubendotelio	"Subendotelio"@es
TunicaCorazónMiocardio	"Miocardio"@es
TunicaCorazónMiocardio	"Myocardial"@en
TunicaCorazónEndocardio	"Endocardium"@en
TunicaCorazónEndocardio	"Endocardio"@es
TunicaCorazónPericardio	"Pericardium"@en
TunicaCorazónPericardio	"Pericardio"@es
TunicaCorazónSubendocardio	"Subendocardio"@es
TunicaCorazónSubendocardio	"Subendocardial"@en

Figure 5.20: Obtained results for CQ-2.

CQ-3: Which are the elastic arteries? Figure 5.21 shows the obtained results.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX organ: <https://w3id.org/def/Organ#>
SELECT ?s ?name
WHERE { ?s rdfs:subClassOf
organ:OrganoArteriaElastica;
                rdfs:label ?name}
```

s	name
OrganoArteriaElasticaAorta	"Arteria Aorta"@es
OrganoArteriaElasticaAorta	"Aorta Artery"@en
OrganoArteriaElasticaIliaca	"Arteria Iliaca"@es
OrganoArteriaElasticaIliaca	"Iliac Artery"@en
OrganoArteriaElasticaCarotidaComun	"Common Carotid Artery"@en
OrganoArteriaElasticaCarotidaComun	"Arteria Carotida Común"@es
OrganoArteriaElasticaFemoral	"Femoral Artery"@en
OrganoArteriaElasticaFemoral	"Arteria Femoral"@es
OrganoArteriaElasticaSubclavia	"Arteria Subclavia"@es
OrganoArteriaElasticaSubclavia	"Subclavian Artery"@en

Figure 5.21: Obtained results for CQ-3.

Heavyweight Ontology: An lightweight ontology is an ontology simply based on a hierarchy of concepts and a hierarchy of relations. On the other hand, a heavyweight ontology is a lightweight ontology enriched with axioms used to fix the semantic interpretation of concepts and relations. We use the restrictions to evaluate if our ontology have axioms which make possible reasoning. Axioms are used to fix the semantic interpretation of concepts and relations, some examples are presented:

```
Arteria tieneTunica some
    (TunicaVasosSanguineosMedia
      and (grosor value "Gruesa"))


Arteria tieneRegionDeLuz some
    (EstructuraRegionLuz
      and (amplitud value "Angosta"))


Arteria tieneTunica some
    (TunicaVasosSanguineosMedia
      and (grosor value "Delgada"))


    Vena tieneTunica some
     (TunicaVasosSanguineosAdventicia
      and (grosor value "Gruesa"))


    Vena tieneRegionDeLuz some
    (EstructuraRegionLuz
      and (amplitud value "Amplia"))
```

5.2. Improving the Classification based on an Ontology of Human Histology

The histological ontology has rules and inferences which may be used to improve the automatic classification of histological images. A improvement process of classification based on an ontology represent an alternative for obtaining information of higher consistency which may not be accomplished by automatic classification or histological ontology individually. This proposal can help to extract knowledge in a unified manner such that users do not perceive the heterogeneity of data sources but get the benefits out of them.

In this section we propose a classification improvement process based on a classification method and a histological ontology. A brief summary of the proposed process is given below and a detailed explanation is presented in the following subsections. Figure 5.22 shows a general outline of our proposal which consists of five steps: 1) A histological image is obtained from a histological sample. (2) Classification using image processing techniques. In this case, we use a classification method using LBP based descriptors and a cascade SVM. (3) Classified blocks are obtained. (4) Occurrences for each discriminating class. (5) RDF triples are built. (6) A histological ontology is used to improve the classification. (7) Blocks reclassified. The details of the complete process are presented in this section.

5.2.1. Dataset

Tissue samples from organs were stained with Hematoxylin and Eosin, the capture protocol were described in Section 3.1. 1500 blocks belonging to five histological images of different organs and persons acquired at $10x$ objective were manually labelled. The histological ontology expressed in OWL language and using Protégé obtained previously was used. We left this dataset publicly available at `http://biscar.univalle.edu.co/datasets`. Algorithms were implemented in *C++*, using the *CImg* library in a computer of *8-cores* and *8Gb* of *RAM*.

5.2.2. Method

In this research, the classification method proposed in Chapter 4 is used in the image classification step (see Figure 5.22 (2)). On the other hand, the histological ontology presented in Section 5.1 is used (see Figure 5.22 (6)). Finally, a improvement method between sources is applied to identify more complete knowledge. We propose two different improvement methods as follows:

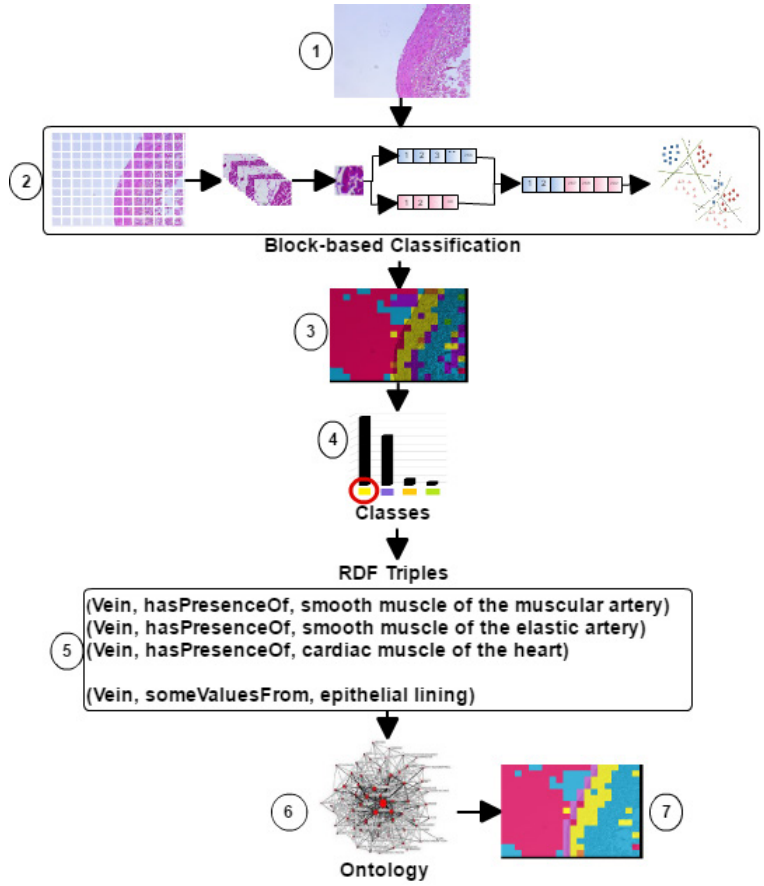


Figure 5.22: Proposed approach for improvement process of classification using an example image. (1) Histological image. (2) Classification based on texture features and a cascade SVM. (3) Classified blocks. (4) Occurrences for each discriminating class. Yellow represent the smooth muscle of the large vein, violet represents the smooth muscle of the muscular artery, orange represents the smooth muscle of the elastic artery, and green represent the cardiac muscle of the heart. (5) RDF triples. (6) Histological ontology. (7) Blocks reclassified.

Improvement Process of Organ Classification

Initially, occurrences of each class, taking into account the results of the classification method, are obtained for the histological image. A part from this, two set are defined: discriminating and non-discriminating classes. Discriminating classes are tissue associated to an organ, such as cardiac muscle of the heart, smooth muscle of the muscular artery, smooth muscle of the large vein, and smooth muscle of the elastic artery. Non-discriminating classes are not directly associated to an organ, such as loose connective tissue and light regions. In this step, the organ of discrim-

inant class with higher occurrences is selected as a subject and the other organs as objects to build a RDF triples for each object in the form of subject, predicate, and object. We worked with the higher occurrences taking into account the accuracy of our classification method which describe the image content. For example, in a case in which the cardiac muscle of the heart is the tissue with higher occurrences, then three RDFs are constructed using the heart as subject and the muscular artery, the large vein, and the elastic artery as objects, respectively.

Rules in RDF form are built taking into account subject and objects obtained in the last step from discriminating classes and the predicate *"tienePresenciaDe"* or *"hasPresenceOf"*. In our example the RDFs are:

- (the heart, hasPresenceOf, the muscular artery)
- (the heart, hasPresenceOf, the large vein)
- (the heart, hasPresenceOf, the elastic artery)

These rules are used to make a query with the histological ontology's reasoner, using SPARQL queries, to obtain a result. If the obtained result is empty, then blocks which were classified with the organ used as object in the RDF triple should be reclassified. The new labelled will be decided according to the behaviour of false positives in the classification process. In other case, the classification is confirmed and it is not modified.

Recognition of Epithelial Tissue Using the Histological Ontology

Epithelial tissue is recognised by histologist, biologist and pathologists using commonly light regions as key information, given that epithelial cells are always found close to these regions. These specialists usually employed images taken with the $40\times$ objective to classify this type of tissue according to the shape of cells — flat, cubic or cylindrical. However, we propose a approach to recognise the epithelial tissue on images taken with the $10\times$ objective using the discriminating classes obtained in the preview step, size of light regions, and the histological ontology presented in Section 5.1 (see Figure 5.22 (6)).

Firstly, the size of light regions are evaluated to decide if it is possible the presence of epithelial tissue in a image. Taking into account the organs which we are recognised — the heart, the muscular artery, the elastic artery and the large vein —, then the presence of epithelial tissue is more probably if the light regions are ten or more consecutive blocks. This value was selected heuristically considering different samples and the specific magnification using Equation (3.4). Secondly, the light regions' neighbourhood should contain smooth muscle of the muscular artery, smooth muscle of the large vein or smooth muscle of the elastic artery. This restriction eliminates false positives cases when exist light regions between loose connective tissue

or close to the adventitia tunic (see Figure 5.23). Thirdly, rules in RDF form are built taking into account the subject obtained of the discriminant class with higher occurrences; the object *"TejidoEpitelialRevestimiento"* or *"EpithelialLining"* which correspond to the epithelial tissue; and the predicates *"someValuesFrom"* and *"subClassOf"*. These rules are used to make a query with the histological ontology's reasoner, using SPARQL queries, to obtain a result. The possible results are two cases as follows: (i) the type of epithelial tissue present in the subject, and (ii) a empty result which mean that the histological images is highly probably to doesn't have presence of epithelial tissue. Finally, blocks over the limit between light region and muscle region are identified as the type of epithelial tissue resulting of the SPARQL query.

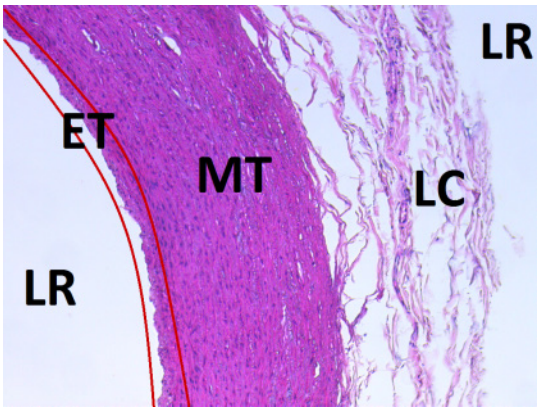


Figure 5.23: Fundamental tissues in a histological image. LR represents light regions; LC represents loose connective tissue region; MT represents muscle tissue region; ET represents epithelial tissue region between red lines marked.

5.2.3. Experiments and Analysis of Results

In this section, we discuss the performance of improvement processes of classification: (i) organ improvement, and (ii) recognition of epithelial tissue.

Improvement Process of Organ Classification

Four specific rules are used to illustrate and evaluate our improvement process of organ classification — the examples of empty queries results in our case — are:

- The heart does not have presence of elastic arteries.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
```

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX organ: <https://w3id.org/def/Organ#>
SELECT ?s ?name
WHERE { ?s rdfs:subClassOf organ:Corazon;
               rdfs:tienePrecensiaDe organ:ArteriaElastica}
```

- The heart does not have presence of the large vein.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX organ: <https://w3id.org/def/Organ#>
SELECT ?s ?name
WHERE { ?s rdfs:subClassOf organ:Corazon;
               rdfs:tienePrecensiaDe organ:VenaGranCalibre}
```

- The large vein does not have presence of elastic arteries.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX organ: <https://w3id.org/def/Organ#>
SELECT ?s ?name
WHERE { ?s rdfs:subClassOf organ:VenaGranCalibre ;
               rdfs:tienePrecensiaDe organ:ArteriaElastica}
```

- The large vein does not have presence of the muscular artery

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX organ: <https://w3id.org/def/Organ#>
SELECT ?s ?name
WHERE { ?s rdfs:subClassOf organ:VenaGranCalibre;
               rdfs:tienePrecensiaDe organ:ArteriaMuscular}
```

Although, arteries and veins are anatomically close, each one accompanied by their concomitant, are evaluated separately when histological samples are obtained. This part is assessed using the ground-truth classification. According to our experiments, the new labelled for the improvement process will be the loose connective tissue taking into account the analysis of positive false in classification process.

Results of the classification improvement process were evaluated taking into account the increasing in the total hit rate of blocks in an image and the number of blocks that were corrected classified due to the applied rules. The increasing of the total hit rate is defined as follows:

$$I_{th} = \mid hitRateAutomaticClassification - hitRateImprovementProcess \mid . \quad (5.1)$$

A selected set of histological images are presented in Figure 5.24 and their automatic and classification improvement results with their hits and misses are included in Figure 5.25. In Figure 5.25 the automatic classification and the improvement process results represent each class with a distinctive colour as follows: (i) cardiac muscle of the heart with green, (ii) loose connective tissue with blue, (iii) smooth muscle of the muscular artery with violet, (iv) smooth muscle of the large vein with yellow, (v) smooth muscle of the elastic artery with orange, and (vi) light regions with fuchsia. On the other hand, hits and misses results are represented in green and red colours, respectively. It can be observed that hits are increase and misses are decrease using the improvement method.

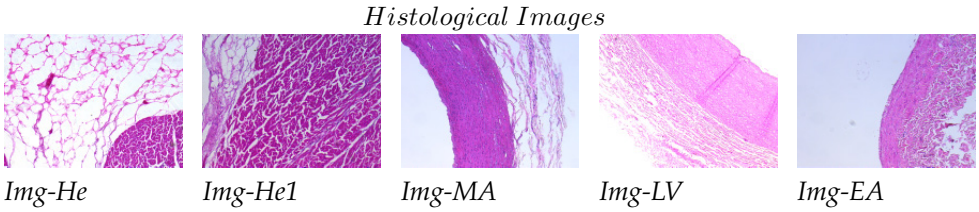


Figure 5.24: Histological images. *Img-He* and *Img-He1* are the heart images; *Img-MA* is a the muscular artery image; *Img-EA* is an the elastic artery images; and *Img-LV* is a the large vein image.

Figure 5.26 contains a graphical representation of the hits and misses by blocks obtained with automatic classification and improvement process in the set of test images. Figure 5.27 contains a graphical representation of the rate increase with improvement process. The reclassified blocks correctly classified between 1 and 24 per image and the improvement process increases hits classification between 0.333% and 23.188% according to the histological image and its automatic classification. The highest increasing in the total hit rate of images' blocks are obtained on the second step classification in the cascade SVM process. This occur because this phase considers only two possible classes. Although rates of improvement are highly variable between images, the behaviour after the improvement process corresponds to increasing hits and decreasing misses in all cases.

The restrictions modelled in the ontology corrects blocks misclassification. The reclassification of blocks appears in the followings two cases:

- The confidence of the classification is not high enough to give a verdict about the presence of a tissue, this case takes place when more than one tissue in the same block.

- The classifier find a suitable class, meaning that this tissue can be perceived in the block, but it is discarded in the ontology process.

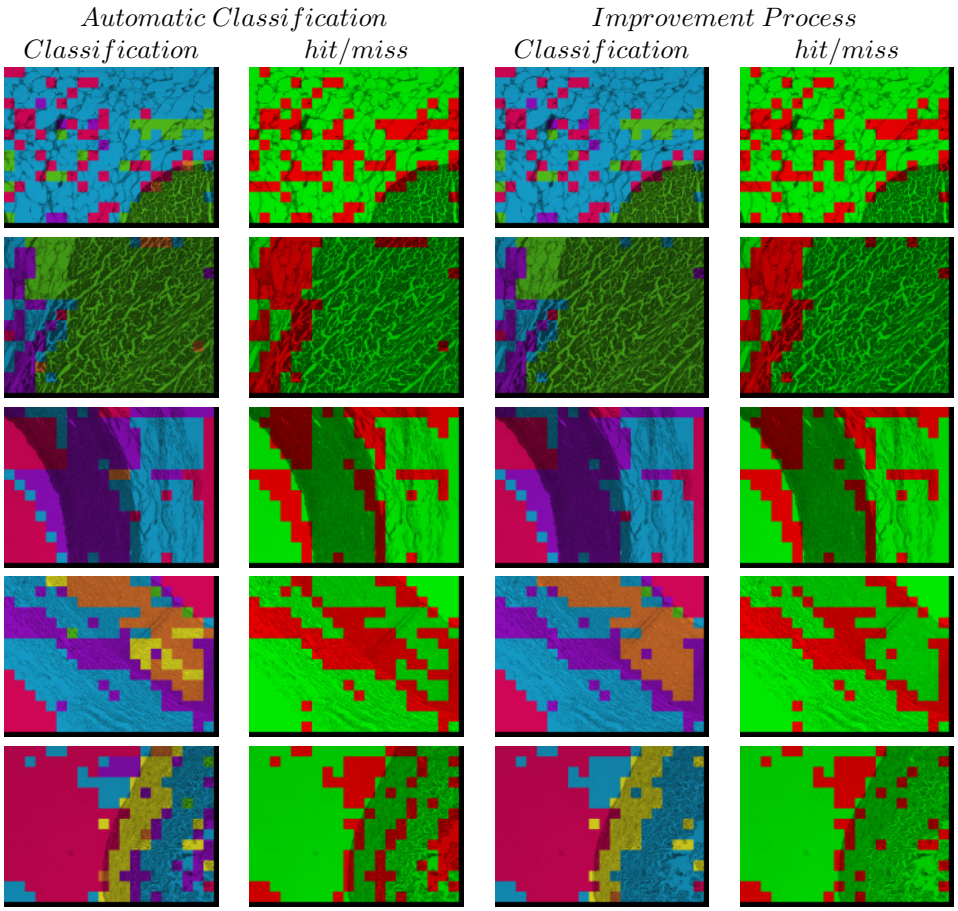


Figure 5.25: Results of classification improvement process. In the first column, automatic classification. In the second column, hits and misses of automatic classification. In the third column, classification using improvement process. In the fourth column, hits and misses of improvement process. In each row a histological image is represented, from top to bottom: *Img-He*, *Img-He1*, *Img-MA*, *Img-EA*, and *Img-LV*.

Recognition of Epithelial Tissue

Four specific rule cases are used to illustrate and evaluate our method to recognise epithelial tissue improving the classification:

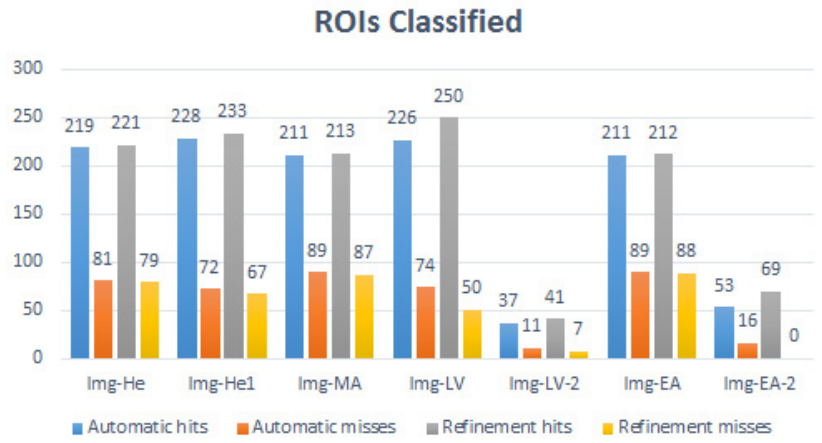


Figure 5.26: Hits and misses blocks with automatic classification and improvement process. The name of each image with -2 at end correspond to the second classification in the cascade SVM process.

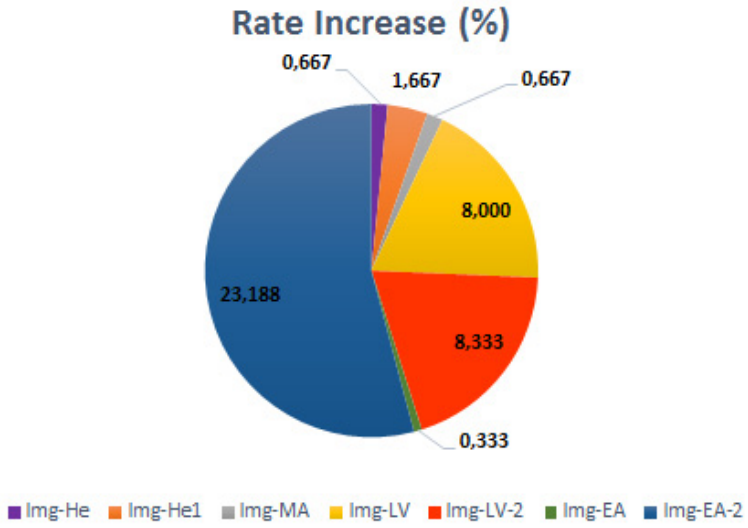


Figure 5.27: Rate increase with improvement process. The name of each image with -2 at end correspond to the second classification in the cascade SVM process.

- Subject= the heart.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
```

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX organ: <https://w3id.org/def/Organ#>
PREFIX tissue: <https://w3id.org/def/Tissue#>
SELECT ?name
WHERE {organ:OrganoCorazon ?p ?o .
?o ?p1 ?o2 .
?o2 ?p2 organ:TunicaCorazónEndocardio .
?o3 owl:someValuesFrom ?o4 .
?o4 rdfs:subClassOf tissue:TejidoEpitelialRevestimiento ;
rdfs:label ?name}
```

- Subject= the elastic artery.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX organ: <https://w3id.org/def/Organ#>
PREFIX tissue: <https://w3id.org/def/Tissue#>
SELECT ?name
WHERE {organ:OrganoArteriaElastica ?p ?o.
?o ?p1 ?o2.
?o2 ?p2 organ:TunicaVasosSanguineosIntima .
?o3 owl:someValuesFrom ?o4.
?o4 rdfs:subClassOf tissue:TejidoEpitelialRevestimiento ;
rdfs:label ?name}
```

- Subject= the muscular artery.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX organ: <https://w3id.org/def/Organ#>
PREFIX tissue: <https://w3id.org/def/Tissue#>
SELECT ?name
WHERE {organ:OrganoArteriaMuscular ?p ?o.
?o ?p1 ?o2.
?o2 ?p2 organ:TunicaVasosSanguineosIntima .
?o3 owl:someValuesFrom ?o4.
?o4 rdfs:subClassOf tissue:TejidoEpitelialRevestimiento ;
rdfs:label ?name}
```

- Subject= the large vein.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX organ: <https://w3id.org/def/Organ#>
PREFIX tissue: <https://w3id.org/def/Tissue#>
SELECT ?name
WHERE {organ:OrganoVenaGranCalibre ?p ?o.
```

```
?o ?p2 organ:TunicaVasosSanguineosIntima .
?o2 owl:someValuesFrom ?o3.
?o3 rdfs:subClassOf tissue:TejidoEpitelialRevestimiento ;
rdfs:label ?name}
```

A selected set of histological images, recognition of epithelial tissue results and their hits and misses are included in Figure 5.28. In Figure 5.28 the recognition image results represent each class with a distinctive colour as follows: (i) cardiac muscle of the heart with green, (ii) loose connective tissue with blue, (iii) smooth muscle of the muscular artery with violet, (iv) smooth muscle of the large vein with yellow, (v) smooth muscle of the elastic artery with orange, (vi) light regions with fuchsia, and (vii) flat simple epithelial tissue with pink. On the other hand, hits and misses results are represented in green and red colours, respectively.

Figure 5.29 contains a graphical representation of the hits and misses by blocks obtained with improvement process and epithelial tissue recognition in the set of test images. Figure 5.30 contains a graphical representation of the rate increase with improvement process and epithelial tissue recognition proposal. The reclassified blocks correctly classified between 0 to 7 per image and the recognition of epithelial tissue process increases hits classification between 0% and 2.333% according to the area which contain epithelial tissue. It is important to highlight that epithelial tissue regions occupy a smaller proportion in histological images, for this reason rates of improvement are highly variable between images and less than 3%. However, this proposal does give an additional step to pattern recognition in which was not recognised epithelial tissue. Additionally, the behaviour after the epithelial recognition process corresponds to increasing hits and decreasing misses in images with epithelial tissue.

5.3. Conclusions

The contributions of the work presented in this chapter are two-fold. First of all, building ontologies gives rise to a significant improvement in representation of histological information such as recovery systems, searches, queries, analysis, inference, reducing the time spent, maximising the use of information and reliability, among other criteria. The ontology enables the integration of different types of knowledge — expert knowledge and histological knowledge — with the purpose of processing, inferring and obtaining new, and more complete, knowledge. The histological ontology was built from histological analysis perspective, potentiating its use in teaching and medical practices. The second contribution is an improvement process on the automatic classification of the fundamental tissues and organs, based on an histological ontology, thus achieving a more complete and accurate knowledge. Our refinement proposal enables us to obtain more consistent information and to

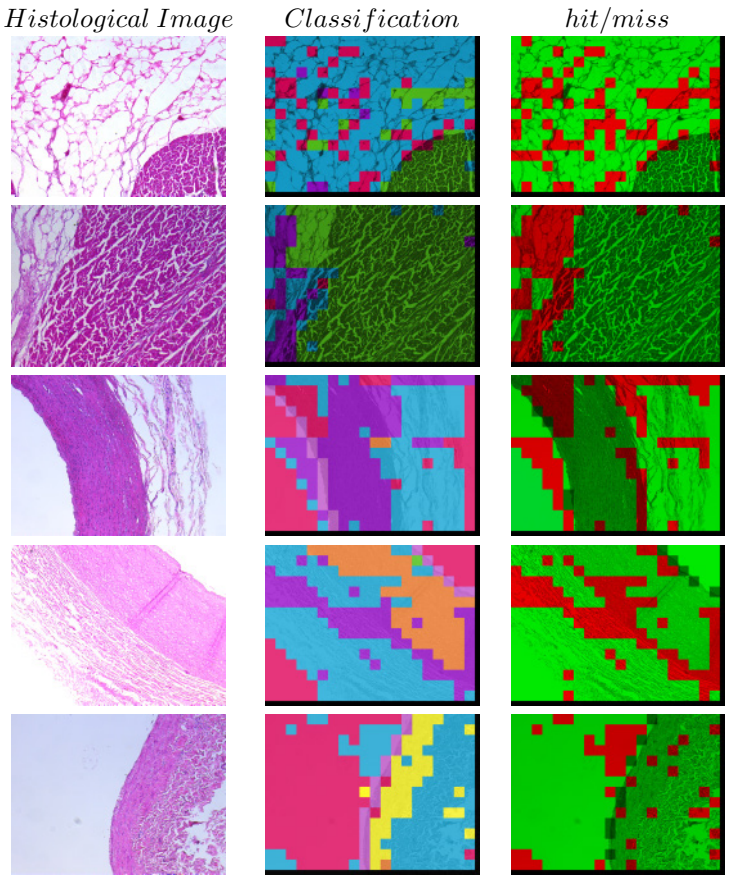


Figure 5.28: Results of epithelial tissue recognition in a histological image. In the first column, histological images. In each row from top to bottom: *Img-He* and *Img-He1* represent the heart images; *Img-MA* represents the muscular artery image; *Img-EA* represents the elastic artery images; and *Img-LV* represents the large vein image. In the second column, automatic classification. In the third column, hits and misses of automatic classification.

reduce margins of error and uncertainty through corroboration and verification, by comparing every analysis of the different data sources separately. The refinement process increases the classification's hits between 0.333% and 23.188%, according to the histological image and its automatic classification. In addition, the ontology enables us to infer which type of epithelium is present in a sample, providing us with an additional step to pattern recognition. Thus, epithelial tissue is identified in images with $10\times$ magnification whereas images with $40\times$ magnification are not always available to recognise areas of epithelial tissue. Besides, more information of

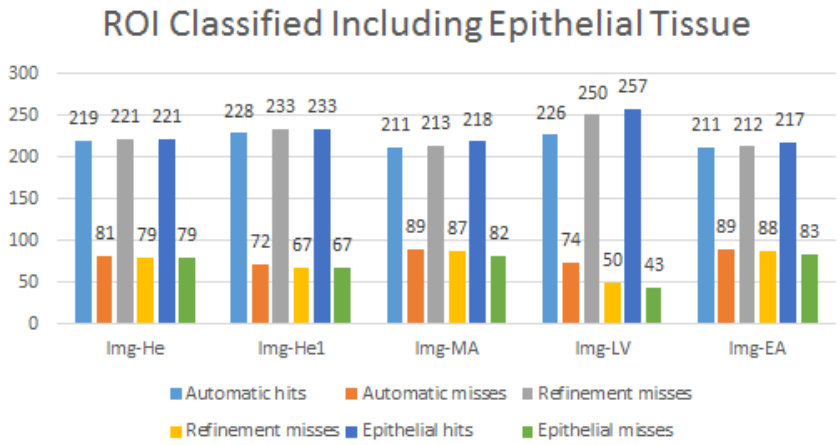


Figure 5.29: Hits and misses blocks with automatic classification and improvement process. The name of each image with -2 at end correspond to the second classification in the cascade SVM process.

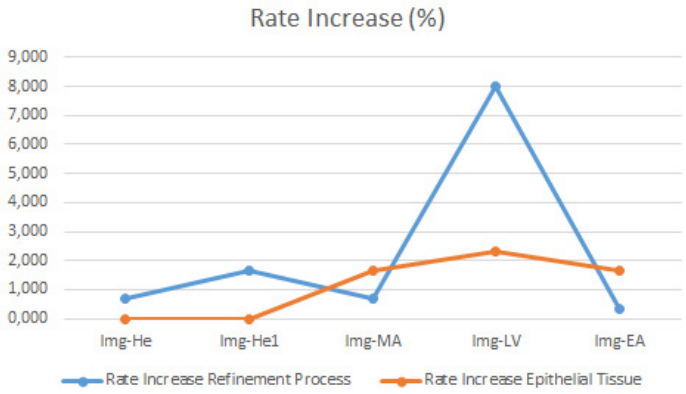


Figure 5.30: Rate increase taking into account improvement process and epithelial tissue recognition. *Img-He* and *Img-He1* represent the heart images; *Img-MA* represents the muscular artery image; *Img-EA* represents the elastic artery images; and *Img-LV* represents the large vein image.

histological knowledge as a system, a composition, but also as structures, relations, regions, layers, sectors, tissues and cells is obtained or inferred from an image.

As future work, automatic identification of micro-circulation organs using macro-circulation identified in this work could be proposed.

Chapter 6

Conclusions and outlook

6.1. Work Summary

Four main proposals have guided the work presented in this dissertation: the recognition of fundamental tissues using morphological information; the classification of fundamental tissues and organs; histological and expert knowledge representation; and the improvement process of histological images classification using a histological ontology. Image classification techniques and knowledge representation, which have seen huge activity in the last years in the computer vision and artificial intelligent fields, are required to provide a solution for these applications.

The recognition and classification of fundamental tissues and organs done in daily medical practice and learning processes is completed manually. Experts prepare histological samples and analyse them through a microscope or on a computer using images obtained by digital technology — a digital camera is connected to a microscope to capture images. Additionally, the number of cases that a student may analyse is limited. Thus, the current process faces many drawbacks such as subjectivity, time costs, difficulty, impracticality and does not promote self-learning. In this work, we recognised and classified the fundamental tissues using morphological information and images processing techniques for the first time. Additionally, we have proposed a method to classify automatically cardiovascular tissues and, in some cases, the organs are identified based on image processing techniques and machine learning algorithms. Using the proposed classification, we outperform our initial results by using more general and robust computer vision and machine learning algorithms. This allows us to recognise organs through the same process.

Histological and expert knowledge are important tools in histology studies and practices, which are transmitted verbally or written. Thus, with the current process, it is not possible to always have expert presence, interaction with machines in a machine-interpretable form, and a specific, clear, standardised and precise communication with everyone. In this work, we generated a representation of histological and expert knowledge creating a histological ontology.

The ontology has information and make inferences which may improve the automatic classification of histological images. In this work, we proposed a process to refine the classification using an ontology taking into account the previous works —

automatic classification of histological images method and the histological ontology. This refinement process based on an ontology represent an alternative for obtaining more complete and consistent information which may not be accomplished by automatic classification process and histological ontology individually. This proposal helps to extract knowledge in a unified manner such that users do not perceive the heterogeneity of data sources but get the benefits out of them.

In the rest of the chapter, the main conclusions of this work and future work lines are presented.

6.2. General Conclusions

The main contribution of this dissertation is that, for the first time, we have demonstrate that it is possible to classify healthy tissues from the human cardiovascular system with a high accuracy.

Other, more specific, conclusions presented during this work are:

1. *We have proposed an approach for automatic recognition of fundamental tissues — epithelial, loose connective and muscle — on histology images of the human cardiovascular system using segmentation, clustering and cell nuclei classification.* We have obtained for cubic epithelial tissues a sensitivity of 0.79, 0.85 for cylindrical and 0.91 for flat. Furthermore, the experts gave our method an average score of 4.85 out of 5 in the recognition of loose connective tissue and 4.82 out of 5 for muscle tissue recognition. The results revealed that the proposed approach classified the fundamental tissues in a similar way to the conventional method employed by experts. Moreover, this proposal has the ability to recognise loose connective tissue even when immersed in muscle tissue.

2. *We have evaluated different texture features to describe histological images.* On the one hand, we have used LBP, LBPri and Haralick features — contrast, angular second moment, energy, correlation, entropy, and first and second correlation measures. On the other hand, we have used a concatenated combination of them such as LBP+LBPri and LBP+LBPri+Haralick. We have determined that, among those tested, the best texture descriptor is obtained by concatenating LBP with LBPri yielding a 90% hit rate.

3. *In the same line of work, we have proposed an approach for classifying cardiovascular tissues automatically and, in some cases, organs using texture information and a cascade SVM classifier.* SVM was selected for the classification after assessing different classifiers, RF and LDA classifier. The choice has yielded to a F-Score of 0.9385 using the cascade SVM, outperforming other classifiers, including RF and LDA which obtained F-Score of 0.870 and 0.823, respectively. We have

concluded that using the recognised tissues from single blocks, we were able to identify some organs. This approach can be easily extended to recognise every part of a complete histological image.

4. *We have evaluated the classification of tissues and organs in a histological image using the block-based recognition method.* We have obtained from 211 to 228 out of 300 blocks per image correctly classified. The method has yielded to an accuracy between 70.333% and 76.000% according to the histological image. The highest accuracy is obtained on the second step classification in the cascade SVM process with 37 and 53 blocks correctly classified and accuracy of 77.083% and 76.812% per image. We have concluded that using the block-based recognition method is possible to classify a histological image. Additionally, the accuracy measure obtained is over 70% compared with 90% obtained using the blocks that contains only a single kind of tissue.

5. *We have proposed a representation of histological and expert knowledge creating a histological ontology of the human cardiovascular system.* The ontology enables the integration of expert knowledge and histological knowledge with the purpose of processing, infering and obtaining new and more complete knowledge. The histological ontology was built from histological analysis perspective, which may be potentially used for teaching and medical practices. We have verified the completeness, duplication and consistency of the histological ontology with four different evaluations: (i) detecting pitfalls using the web tool *OOPS!* finding that this drawback is not occurring in the current approach; (ii) expert evaluation using surveys by two different sets of experts and according to the positive results obtained the ontology is usable; (iii) CQ verification using SPARQL queries and conforming to the answers obtained the task was completed successfully; (iv) verifying if the ontology is a heavyweight ontology considering that the histological ontology was enriched with axioms used to fix the semantic interpretation of concepts and relations.

6. *We have improved the previous classification using using the knowledge contained in the histological ontology.* With the prosed method, between 1 to 24 wrongly classified blocks per image were corrected. Additionally, the improvement process increases hits classification between 0.333% and 23.188% according to the histological image and its automatic classification. The results revealed that the proposed method increases hit rates in all cases.

7. *We have presented a method to recognise the epithelial tissue using the histological ontology and automatic classification.* With the proposed method, between 0 and 7 blocks per image were corrected. Additionally, the recognition of epithelial tissue process increases classification hits between 0% and 2.333% according to the areas which contain epithelial tissue. We have concluded that classification

improvement using a histological ontology enables us to infer which type of epithelium is present in a sample — using images with $10\times$ magnification. This fact give an additional step to pattern recognition.

8. *We have created our dataset and made it publicly available.* The dataset consists of three sub-datasets. The first dataset is composed of 400 images belonging to different organs and persons, 300 acquired at $40\times$ and the other 100 at $10\times$ objective. The second dataset consists of 3000 blocks — 600 per class. The third dataset is composed of 1500 blocks belonging to five histological images of different organs and persons acquired at $10\times$ objective and manually labelled. These datasets can be used to validate the results obtained in our work or to improve upon the proposed methods.

6.3. Outlook

In this section, we summarise the main research lines that remain open.

Firstly, the recognition and classification of fundamental tissues and organs suggests several lines of work are open to improve on and expand upon the results. Additionally, a histological atlas with additional functionalities could be developed.

1. The method for identifying the type of epithelial tissue according to the cell's morphology can be improved by differentiating cubic epithelial tissue. Cubic epithelial tissue is frequently misclassified as cylindrical tissue. This happens when cell nuclei are very close to each other and they are evaluated as a single cell. This could be achieved by using, for example, strategies for separating particles in order to improve the results.

2. The method of classifying the muscle tissues can be improved by identifying dense connective tissue. That muscle tissue has a high similitude with dense connective tissue that is sometimes difficult to differentiate even manually, it was one of the issues considered during the evaluation. One possible solution would be to use additional image features in order to overcome the current potential misclassification between muscle and dense connective tissues.

3. The method of classifying the fundamental tissues can be improved by exploring new classification techniques such as supervised deep learning algorithms.

4. In order to create a histological atlas with additional information which could be used in medical practice and in the learning process we could integrate the proposed methods as an additional functionality in BISCAR (Banco de Imágenes Histológicas del Sistema Cardiovascular) available at `http://biscar.univalle.edu.co`.

Secondly, histological ontology leaves some open lines of work such as:

- The histological ontology is limited to the human cardiovascular system. Further steps include extending the ontology to other systems using the methodology used in this work. This is possible taking into account that the ontology was implemented in a modular way — tissues, organs and systems.
- The development of a histological ontology is not the final goal. In addition, future work may be focused on other applications of the ontology, such as supporting research in different ways.

Thirdly, the classification improvement method using the histological ontology leaves some lines of works such as:

- The histological ontology contains additional information such as system, structures, composition, relations, regions, layers, sectors, tissues and cells which can be used in order to improve the results obtained by automatic classification. This could be achieved by, for instance, including rules which allow the recognition of dense connective tissue in the same way epithelial tissue is identified.
- The classification improvement process proposed is a general method which could be applied to different problems. In this way we could used the proposed method in other process of images classification in medicine or industry.

Capítulo 7

Conclusiones y Perspectiva

7.1. Resumen del Trabajo

En este trabajo se han presentado cuatro aportaciones principales relacionadas con el reconocimiento y clasificación de tejidos fundamentales: el reconocimiento de los tejidos fundamentales utilizando la información morfológica; la clasificación esos mismos tejidos y algunos órganos del sistema cardiovascular humano; la representación del conocimiento histológico y de los expertos; y la mejora en la clasificación de imagenes histológicas usando la ontología histológica previamente creada. Para conseguir estas soluciones se utilizaron técnicas de clasificación de imágenes y representación del conocimiento.

El reconocimiento y la clasificación de tejidos fundamentales y órganos, realizado en los procesos de aprendizaje y prácticas médicas diarias, se hace manualmente. Los expertos preparan muestras histológicas y las analizan a través de un microscopio o en un computador, utilizando las imágenes obtenidas con la tecnología digital — una cámara conectada a un microscopio para capturar imágenes. Por lo tanto, el proceso actual se enfrenta a muchos inconvenientes como: subjetividad, costos de tiempo, dificultad en su realización, falta de sentido práctico y deficiencia en el auto-aprendizaje. En este trabajo, inicialmente se realiza el reconocimiento y la clasificaión de los tejidos fundamentales del sistema cardiovascular humano, utilizando la información morfológica y tecnicas de procesamiento de imágenes. También se propone una clasificación automática de los tejidos cardiovasculares y, en algunos casos, se identifican los órganos usando técnicas de procesamiento de imágenes y algoritmos de aprendizaje automático. Además, se logra una mejora en los resultados obtenidos con la propuesta inicial mediante el uso de un método más general y robusto, de visión por computador y aprendizaje automático, el cual nos permite reconocer los órganos a través del mismo proceso.

El conocimiento histológico y el de los expertos son herramientas importantes en estudios y prácticas histológicas, los cuales son transmitidos de manera verbal o escrita. Además, en el proceso actual no es posible tener siempre la presencia de expertos, ni una interacción con los computadores, ni una comunicación específica, clara, precisa y estandarizada. En este trabajo, se propone la representación del conocimiento histológico y de expertos a traves de la creación de una ontología his-

tológica.

Una ontología contiene información y realiza inferencias que pueden mejorar la clasificación automática de imágenes histológicas. En este trabajo, hemos propuesto un proceso mediante el cual mejoramos la clasificación utilizando una ontología, teniendo en cuenta los trabajos previos — método de clasificación automática de imágenes histológicas y ontología histológica. Este proceso, permite obtener información más completa y consistente que no puede obtenerse por estas fuentes de datos de manera individual. Esta propuesta puede ayudar a extraer el conocimiento de manera unificada sin que los usuarios perciban la heterogeneidad de las fuentes de datos, pero obteniendo los beneficios de ellas.

En el resto del capítulo, presentamos las principales conclusiones de este trabajo y las futuras líneas de investigación posibles.

7.2. Conclusiones Generales

La principal aportación de este trabajo es que, por primera vez, hemos demostrado que es posible clasificar los tejidos sanos del sistema cardiovascular humano con una alta precisión.

Otras conclusiones más específicas presentadas en este trabajo son:

1. *Hemos propuesto un método para el reconocimiento automático de los tejidos fundamentales — epitelial, conectivo laxo y múscular — en imágenes histológicas del sistema cardiovascular humano usando segmentación, agrupación y clasificación de núcleos de células.* Hemos obtenido una sensibilidad de $0{,}79$ para el tejido epitelial cúbico, $0{,}85$ para el cilíndrico y $0{,}91$ para el plano. Por otra parte, los expertos calificaron nuestro método con una puntuación media de $4{,}85$ de 5 para el reconocimiento de tejido conectivo laxo y $4{,}82$ de 5 para el reconocimiento de tejido muscular. Los resultados revelaron que el enfoque propuesto clasifica los tejidos fundamentales de una manera similar al método convencional empleado por los expertos.

2. *Hemos evaluado diferentes características de textura para describir imágenes histológicas.* Por un lado, hemos utilizado LBP, LBPri y Haralick — contraste, segundo momento angular, energía, correlación, entropía, y primera y segunda medidas de correlación. Por otro lado, hemos utilizado diversas formas de concatenación entre ellas, tales como LBP+LBPri y LBP+LBPri+Haralick. Con base en las pruebas realizadas hemos determinado que el mejor descriptor de textura se obtiene mediante la concatenación de LBP con LBPri con tasas de acierto por encima del $90\,\%$.

3. *En la misma línea de trabajo, hemos propuesto un enfoque para la clasificación automática de los tejidos cardiovasculares y, en algunos casos, los órganos que utilizan*

información de textura y un clasificador SVM en cascada. Se seleccionó SVM para la clasificación después de evaluar diferentes clasificadores como RF y LDA, SVM obtuvo un F-Score de $0{,}939$ superando a RF con un F-Score de $0{,}870$ y LDA con un F-score de $0{,}823$. Hemos llegado a la conclusión que usando el reconocimiento basado en bloques se logra identificar en algunos casos el órgano, además este enfoque se puede extender fácilmente al reconocimiento de una imagen histológica completa.

4. *Hemos evaluado la clasificación de los tejidos y órganos en una imagen histológica utilizando un método de reconocimiento basado en bloques.* Hemos obtenido desde 211 a 228 bloques correctamente clasificados de 300 en una imagen completa, con una precisión entre $70{,}333\,\%$ y $76{,}000\,\%$ que varía de acuerdo a la imagen histológica para la clasificación realizada por la primera SVM en cascada. Sin embargo, la precisión más alta se obtiene en la segunda clasificación de nuestro SVM en cascada, donde se obtuvieron entre 37 y 53 bloques correctamente clasificados y una exactitud entre $77{,}083\,\%$ y $76{,}812\,\%$ por imagen. Hemos concluido que usando el método de reconocimiento basado en bloques es posible clasificar una imagen histológica completa.

5. *Hemos propuesto una representación del conocimiento histológico y el conocimiento de expertos con la creación de una ontología histológica del sistema cardiovascular humano.* La ontología histológica fue construida desde la perspectiva del análisis histológico, el cual puede ser potencialmente utilizado para la enseñanza y las prácticas médicas. Hemos verificado la integridad, la duplicación y la consistencia de la ontología utilizando cuatro evaluaciones: (i) la herramienta web *OOPS!*; (ii) evaluación de dos grupos de expertos mediante encuestas; (iii) verificación de las CQs mediante consultas en SPARQL; (iv) verificación del tipo de ontología, ontología liviana o de peso pesado.

6. *Hemos presentado un proceso que permite mejorar la clasificación previa utilizando utilizando conjuntamente el resultado de la clasificación basada en textura y el razonamiento procedente de la ontología histológica.* Hemos obtenido entre 1 y 24 bloques correctamente reclasificados por imagen, aumentando la tasa de aciertos entre $0{,}333\,\%$ y $23{,}188\,\%$ que varía según la imagen histológica y su clasificación automática.

7. *Hemos presentado un método para el reconocimiento del tejido epitelial utilizando la ontología histológica y la clasificación automática.* Hemos obtenido entre 0 y 7 bloques correctamente reclasificados por imagen, aumentando la tasa de aciertos entre $0\,\%$ y $2,333\,\%$ que varía según las áreas que contienen tejido epitelial. Hemos llegado a la conclusión de que el método de mejora de la clasificación automática usando la ontología histológica nos permite inferir qué tipo de epitelio está presente en una muestra — usando imágenes a $10\times$.

8. *Hemos creado y hecho público nuestros conjuntos de datos.* El conjunto de datos se compone de tres sub-conjuntos, que se describen a continuación. El primero se compone de 200 imágenes pertenecientes a diferentes órganos y personas, 100 adquiridas con un objetivo de $40\times$ y las 100 restantes a $10\times$. El segundo se compone de 3000 bloques — 600 por clase. El tercero se compone de 1500 bloques pertenecientes a cinco imágenes histológicas de diferentes órganos y personas adquiridas con un objetivo de $10\times$ y etiquetados de forma manual. Estos conjuntos de datos se pueden utilizar para validar los resultados obtenidos en nuestro trabajo o para proponer mejoras sobre los métodos propuestos.

7.3. Perspectiva

En esta sección, resumimos las principales líneas de trabajo que permanecen abiertas para cada una de las aplicaciones estudiadas.

En primer lugar, el reconocimiento y clasificación de los tejidos fundamentales y órganos sugiere varias líneas de trabajo abiertas para mejorar y ampliar los resultados obtenidos.

1. Mejora en la identificación del tipo de tejido epitelial teniendo en cuenta la morfología de la célula. El tejido epitelial cúbico es con frecuencia clasificado erroneamente como tejido cilíndrico; esto ocurre cuando los núcleos celulares están muy cercanos entre sí y se evalúan como una sola célula. Podría realizarse una mejora mediante el uso de, por ejemplo, estrategias para separar las partículas.

2. Mejora en la clasificación del tejidos muscular, el cual tiene una alta similitud con el tejido conectivo denso, siendo difícil diferenciarlos, incluso, manualmente. Podría realizarse una mejora utilizando características adicionales de la imagen.

3. El método de clasificación de tejidos fundamentales se puede mejorar mediante la exploración de nueva ténicas de clasificación tales como algoritmos de aprendizaje profundo supervisados.

4. Integración de los métodos propuestos como funcionalidades adicionales a BISCAR (Banco de Imágenes Histológicas del Sistema Cardiovascular)[1] con el fin de brindar información adicional que pueda ser utilizada en la práctica médica o en procesos de aprendizaje.

En segundo lugar, la ontología histológica deja algunas líneas abiertas de trabajo, tales como:

[1] Disponible en `http://biscar.univalle.edu.co`

1. Ampliación de la ontología histológica con la inclusión de los demas sistemas humanos.

2. Desarrollo de diversas aplicaciones de la ontología como apoyo a la investigación, teniendo en cuenta que el desarrollo de la ontología no es el objetivo final.

En tercer lugar, el método de mejora de la clasificación usando la ontología histológica deja algunas líneas de trabajos, tales como:

1. La ontología histológica contiene información adicional como: sistemas, estructuras, composición, relaciones, regiones, capas, sectores, tejidos y celulas, la cual puede ser utilizada para mejorar los resultados obtenidos por la clasificación automática. Podría realizarse, por ejemplo, la identificación del tejido conectivo denso, incluyendo reglas adicionales de la misma manera que se realizó la identificación del tejido epitelial.

2. El proceso de mejora de la clasificación es un método general que puede ser aplicado a diferentes problemas, de esta manera se podria utilizar el método en otras aplicaciones de la clasificación de imágenes en medicina e industria.

Bibliography

Abacha, A. B. and Zweigenbaum, P.: 2015, MEANS: A medical question-answering system combining NLP techniques and semantic web technologies, *Inf. Process. Manage.* **51**(5), 570–594.

Abdollahpour, Z., Samani, Z. R. and Moghaddam, M. E.: 2015, Image classification using ontology based improved visual words, *2015 23rd Iranian Conference on Electrical Engineering*, pp. 694–698.

Alén, S., Cernadas, E., Formella, A., Domínguez, R. and Saborido-Rey, F.: 2006, *Comparison of Region and Edge Segmentation Approaches to Recognize Fish Oocytes in Histological Images*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 853–864.

Bader-El-Den, M.: 2014, Self-adaptive heterogeneous random forest, *2014 IEEE/ACS 11th International Conference on Computer Systems and Applications (AICCSA)*, pp. 640–646.

Bertaud-Gounot, V., Duvauferrier, R. and Burgun, A.: 2012, Ontology and medical diagnosis., *Inform Health Soc Care* **37**(2), 51–61.

Bevilacqua, V., Pietroleonardo, N., Triggiani, V., Gesualdo, L., Di Palma, A. M., Rossini, M., Dalfino, G. and Mastrofilippo, N.: 2015, *Neural Network Classification of Blood Vessels and Tubules Based on Haralick Features Evaluated in Histological Images of Kidney Biopsy*, Springer International Publishing, Cham, pp. 759–765.

BioPortal: 2005, Medical subject headings. [Web; accessed on 26-06-2016 to `http://purl.bioontology.org/ontology/MESH/D046650`].

BioPortal: 2008a, Medical subject headings. [Web; accessed on 26-06-2016 to `http://purl.bioontology.org/ontology/MESH`].

BioPortal: 2008b, National cancer institute thesaurus. [Web; accessed on 26-06-2016 to `http://purl.bioontology.org/ontology/MESH/D054547`].

BioPortal: 2014a, Mouse adult gross anatomy ontology. [Web; accessed on 26-06-2016 to `http://bioportal.bioontology.org/ontologies/MA?p`].

BioPortal: 2014b, National cancer institute thesaurus. [Web; accessed on 26-06-2016 to `http://purl.bioontology.org/ontology/NCIT`].

BioPortal: 2014c, Robert hoehndorf version of mesh. [Web; accessed on 26-06-2016 to `http://purl.bioontology.org/ontology/RH-MESH`].

Breen, C., Khan, L., Kumar, A. and Wang, L.: 2002, Ontology-based image classification using neural networks.

Caicedo, J.: 2011, Features for histology images. [Web; accessed on 20-08-2016 to `http://www.informed.unal.edu.co/jccaicedo/docs/review.pdf`].

Caicedo, J. C., Roa, A. C. and González, F. A.: 2009, Histopathology image classification using bag of features and kernel functions, *Artificial Intelligence in Medicine, 12th Conference on Artificial Intelligence in Medicine, AIME 2009, Verona, Italy, July 18-22, 2009. Proceedings*, pp. 126–135.

Caicedo, J., González, F. and Romero, E.: 2011, Content-based histopathology image retrieval using a kernel-based semantic annotation framework, *Journal of Biomedical Informatics* **44**(4), 519–528.

Canada, B. A., Thomas, G. K., Cheng, K. C., Wang, J. Z. and Liu, Y.: 2008, Towards efficient automated characterization of irregular histology images via transformation to frieze-like patterns, *CIVR*, ACM, pp. 581–590.

Canada, B., Thomas, G., Cheng, K. and Wang, J.: 2011, Shiraz: an Automated Histology Image Annotation System for Zebrafish Phenomics, **51**, 401–440.

Canas, A., Leake, D. and Wilson, D.: 1999, Managing, mapping and manipulating conceptual knowledge, *In AAAI Workshop Technical Report WS-99-10: Exploring the Synergies of Knowledge Management & Case-Based Reasoning Menlo California: AAAI Press;.*

Canny, J.: 1986, A computational approach to edge detection, *IEEE Trans. Pattern Anal. Mach. Intell.* **8**(6), 679–698.

Carneiro, G., Peng, T., Bayer, C. and Navab, N.: 2015, Automatic detection of necrosis, normoxia and hypoxia in tumors from multimodal cytological images, *Image Processing (ICIP), 2015 IEEE International Conference on*, pp. 2429–2433.

Castro, A. G., Rocca-Serra, P., Stevens, R., Taylor, C., Nashar, K., Ragan, M. A. and Sansone, S.-A.: 2006, The use of concept maps during knowledge elicitation in ontology development processes – the nutrigenomics use case, *BMC Bioinformatics* **7**(1), 1–14.

Cateni, S. and Colla, V.: 2015, Improving the stability of sequential forward variables selection, *2015 15th International Conference on Intelligent Systems Design and Applications (ISDA)*, pp. 374–379.

Chankong, T., Theera-Umpon, N. and Auephanwiriyakul, S.: 2014, *Automatic Cervical Cell Classification Using Patch-Based Fuzzy Clustering and Minimum Average Correlation Energy Filter*, Springer International Publishing, Cham, pp. 164–167.

Chen, C., Ozolek, J. A., 0037, W. W. and Rohde, G. K.: 2011, A general system for automatic biomedical image segmentation using intensity neighborhoods., *Int. J. Biomedical Imaging* pp. 606857:1–606857:12.

Chen, S., Zhao, M., Wu, G., Yao, C. and Zhang, J.: 2012, Review article: Recent advances in morphological cell image analysis, *Computational and Mathematical Methods in Medicine* p. 10.

Cheng, L., Ye, N., Yu, W. and Cheah, A.: 2012, *A Bag-of-Words Model for Cellular Image Segmentation*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 209–222.

Claridge, E., Cotton, S., Hall, P. and Moncrieff, M.: 2002, *From Colour to Tissue Histology: Physics Based Interpretation of Images of Pigmented Skin Lesions*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 730–738.

Colantonio, S., Martinelli, M., Salvetti, O., Gurevich, I. B. and Trusova, Y. O.: 2008, Cell image analysis ontology, *Pattern Recognition and Image Analysis* **18**(2), 332–341.

Cruz, A., Spinel, C., Seligmann, D., Romero, E., Gonzalez, F., Diaz, G. and cols.: 2011, Sistema para la recuperación por contenido en un banco de imágenes médicas, *Revista de ciencia, educación, innovación y cultura apoyadas por Redes deTecnología Avanzada* **1**(1), 60–64.

d. A. Zampirolli, F., Stransky, B., Lorena, A. C. and d. M. Paulon, F. L.: 2010, Segmentation and classification of histological images - application of graph analysis and machine learning methods, *2010 23rd SIBGRAPI Conference on Graphics, Patterns and Images*, pp. 331–338.

Diamond, J., Anderson, N., Bartels, P., Montironi, R. and Hamilton, P.: 2004a, The use of morphological characteristics and texture analysis in the identification of tissue composition in prostatic neoplasia, *Human Pathology* **35**(9), 1121–1131.

Diamond, J., Anderson, N. H., Bartels, P. H., Montironi, R. and Hamilton, P. W.: 2004b, The use of morphological characteristics and texture analysis in the identification of tissue composition in prostatic neoplasia, *Hum. Pathol.* **35**(9), 1121–1131.

Doyle, S., Hwang, M., Shah, K., Madabhushi, A., Tomaszewski, J. and Feldman, M.: 2007, Automated grading of prostate cancer using architectural and textural image features, pp. 1284–87.

Eves, H.: 1971, *Mathematical circles revisited: a second collection of mathematical stories and anecdotes*, Prindle, Weber & Schmidt series in mathematics, Prindle, Weber & Schmidt.

Fakhrzadeh, A., Spörndly-Nees, E., Holm, L. and Luengo Hendriks, C. L.: 2013, *Epithelial Cell Segmentation in Histological Images of Testicular Tissue Using Graph-Cut*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 201–208.

Ficsor, L. and Molnar, B.: 2009, Automated disease classification of colon and gastric histological samples based on digital microscopy and advanced image analysis, *in* M. Hayat (ed.), *Gastrointestinal Carcinoma*, Vol. 3 of *Methods of Cancer Diagnosis, Therapy, and Prognosis*, Springer Netherlands, pp. 99–111.

Fischer, A., Jacobson, K., Rose, J. and Zeller, R.: 2008, Hematoxylin and eosin staining of tissue and cell sections, *Cold Spring Harb Protoc* .

Frangi, A. F., Niessen, W. J., Hoogeveen, R. M., van Walsum, T. and Viergever, M. A.: 1999, Model-based quantitation of 3d magnetic resonance angiographic images., *IEEE Trans. Med. Imaging* **18**(10), 946–956.

Fujii, T., Takahashi, M., Yamada, K. and Nakano, M.: 2013, *Precise Segmentation of Nuclei in Hepatic Histological Images*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 880–883.

Galaro, J., Judkins, A. R., Ellison, D., Baccon, J. and Madabhushi, A.: 2011, An integrated texton and bag of words classifier for identifying anaplastic medulloblastomas, *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 3443–3446.

Gandon, F.: 2002, *Distributed artificial intelligence and knowledge management: Ontologies and multiagent systems for a corporate semantic web*, PhD thesis, University of Nice - Sophia Antipolis, Doctoral School of Sciences and Technologies of Information and Communication, The address of the publisher.

Garnier, M., Alsheh Ali, M., Seguin, J., Mignet, N., Hurtut, T. and Wendling, L.: 2014, *Grading Cancer from Liver Histology Images Using Inter and Intra Region Spatial Relations*, Springer International Publishing, Cham, pp. 247–254.

Gharipour, A. and Liew, A. W. C.: 2015, Level set based segmentation of cell nucleus in fluorescence microscopy images using correntropy-based k-means clustering, *Digital Image Computing: Techniques and Applications (DICTA), 2015 International Conference on*, pp. 1–5.

Ghassabeh, Y. A., Rudzicz, F. and Moghaddam, H. A.: 2015, Fast incremental LDA feature extraction, *Pattern Recognition* **48**(6), 1999–2012.

Gómez-Pérez, A.: 2004, Ontology evaluation, *Handbook on Ontologies*, International Handbooks on Information Systems, Springer, pp. 251–274.

Grüninger, M. and Fox, M.: 1995, Methodology for the Design and Evaluation of Ontologies, *IJCAI'95, Workshop on Basic Ontological Issues in Knowledge Sharing, April 13, 1995*.

Guefack, V. D., Gounot, V. B., Duvauferrier, R., Bourde, A., Morelli, J. and Lasbleiz, J.: 2012, Ontology driven decision support systems for medical diagnosis - an interactive form for consultation in patients with plasma cell disease., *Quality of Life through Quality of Information* pp. 108–112.

Gurcan, M., Boucheron, L., Can, A., Madabhushi, A., Rajpoot, N. and Yener, B.: 2009, Histopathological image analysis: A review, *Biomedical Engineering, IEEE Reviews in* **2**, 147 –171.

Hafiane, A., Bunyak, F. and Palaniappan, K.: 2008, *Fuzzy Clustering and Active Contours for Histopathology Image Segmentation and Nuclei Detection*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 903–914.

Han, J., Breckon, T., Randell, D. and Landini, G.: 2012a, The application of support vector machine classification to detect cell nuclei for automated microscopy, *Machine Vision and Applications* **23**(1), 15–24.

Han, J. W., Breckon, T. P., Randell, D. A. and Landini, G.: 2012b, The application of support vector machine classification to detect cell nuclei for automated microscopy, *Machine Vision and Applications* **23**(1), 15–24.

Hansen, P. C.: 2010, *Discrete Inverse Problems*, Fundamentals of Algorithms, Society for Industrial and Applied Mathematics SIAM.

He, L., Long, L. R., Antani, S. and Thoma, G. R.: 2011, Multiphase level set model with local k-means energy for histology image segmentation, *Healthcare Informatics, Imaging and Systems Biology (HISB), 2011 First IEEE International Conference on*, pp. 32–39.

Heath, T. L. and Euclid: 1956, *The Thirteen Books of Euclid's Elements, Books 1 and 2*, Dover Publications, Incorporated.

Hernandez, A. I., Porta, S. M., Miralles, M., Garcia, B. F. and Bolumar, F.: 1990, La cuantificacion de la variabilidad en las observaciones clinicas, *Med Clin* pp. 424–429.

Herve, N., Servais, A., Thervet, E., Olivo-Marin, J. C. and Meas-Yedid, V.: 2011, Statistical color texture descriptors for histological images analysis, *Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on*, pp. 724–727.

Hitzler, P., Krötzsch, M., Parsia, B., Patel-Schneider, P. F. and Rudolph, S. (eds): 2009, *OWL 2 Web Ontology Language: Primer*, W3C Recommendation. Available at `http://www.w3.org/TR/owl2-primer/`.

Horridge, M., Knublauch, H., Rector, A., Stevens, R. and Wroe, C.: 2004, *A Practical Guide To Building OWL Ontologies With The Protege-OWL Plugin*, 1 edn, University of Manchester.

Izet, M.: 2008, E-learning as new method of medical education., *Acta Informatica Medica* **16**(2), 102–117.

Jitaree, S., Phinyomark, A., Thongnoo, K., Boonyapiphat, P. and Phukpattaranont, P.: 2013, Classifying breast cancer regions in microscopic image using texture analysis and neural network, *Biomedical Engineering International Conference (BMEiCON), 2013 6th*, pp. 1–4.

Kancherla, K. and Mukkamala, S.: 2013, Early lung cancer detection using nucleus segementation based features, *Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), 2013 IEEE Symposium on*, pp. 91–95.

Kang, S., Lee, C. Y., Gonçalves, M., Chisholm, A. D. and Cosman, P. C.: 2015, Tracking epithelial cell junctions in c. elegans embryogenesis with active contours guided by sift flow, *IEEE Transactions on Biomedical Engineering* **62**(4), 1020–1033.

Kanungo, T., Mount, D., Netanyahu, N., Piatko, C., Silverman, R. and Wu, A.: 2002, An efficient k-means clustering algorithm: analysis and implementation, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **24**(7), 881–892.

Kapoor, B. and Sharma, S.: 2010, A comparative study ontology building tools for semantic web applications, *International Journal of Web & Semantic Technology (IJWesT)* **1**(3), 1–13.

Kashif, M. N., Raza, S. E. A., Sirinukunwattana, K., Arif, M. and Rajpoot, N.: 2016, Handcrafted features with convolutional neural networks for detection of tumor cells in histology images, *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, pp. 1029–1032.

Kong, J., Sertel, O., Shimada, H., Boyer, K. L., Saltz, J. H. and Gurcan, M. N.: 2009a, Computer-aided evaluation of neuroblastoma on whole-slide histology images: Classifying grade of neuroblastic differentiation, *Pattern Recogn.* **42**(6), 1080–1092.

Kong, J., Sertel, O., Shimada, H., Boyer, K. L., Saltz, J. H. and Gurcan, M. N.: 2009b, Computer-aided evaluation of neuroblastoma on whole-slide histology images: Classifying grade of neuroblastic differentiation, *Pattern Recogn.* **42**(6), 1080–1092.

Kothari, S., Chaudry, Q. and Wang, M. D.: 2009, Automated cell counting and cluster segmentation using concavity detection and ellipse fitting techniques, *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pp. 795–798.

Kothari, S., Phan, J. H., Moffitt, R. A., Stokes, T. H., Hassberger, S. E., Chaudry, Q., Young, A. N. and Wang, M. D.: 2011, Automatic batch-invariant color segmentation of histological cancer images, *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pp. 657–660.

Kothari, S., Phan, J. H., Young, A. N. and Wang, M. D.: 2013, Histological image classification using biologically interpretable shape-based features, *BMC Medical Imaging* **13**(1), 1–17.

Krishnan, M. M. R., Shah, P., Ghosh, M., Pal, M., Chakraborty, C., Paul, R. R., Chatterjee, J. and Ray, A. K.: 2010, Automated characterization of sub-epithelial connective tissue cells of normal oral mucosa: Bayesian approach, *Students' Technology Symposium (TechSym), 2010 IEEE*, pp. 44–48.

Kylberg, G. and Sintorn, I.-M.: 2013, Evaluation of noise robustness for local binary pattern descriptors in texture classification., *EURASIP J. Image and Video Processing* **2013**, 17.

Lai, Y., Viswanath, S., Baccon, J., Ellison, D., Judkins, A. R. and Madabhushi, A.: 2011, A texture-based classifier to discriminate anaplastic from non-anaplastic medulloblastoma, *2011 IEEE 37th Annual Northeast Bioengineering Conference (NEBEC)*, pp. 1–2.

Lou, X., Koethe, U., Wittbrodt, J. and Hamprecht, F. A.: 2012, Learning to segment dense cell nuclei with shape prior, *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pp. 1012–1018.

Lu, B., Miao, C. and Wang, H.: 2010, Pixel level image fusion based on linear structure tensor, *Information Computing and Telecommunications (YC-ICT), 2010 IEEE Youth Conference on*, pp. 303–306.

Markkongkeaw, A., Phinyomark, A., Boonyapiphat, P. and Phukpattaranont, P.: 2013, Preliminary results of breast cancer cell classifying based on gray-level co-occurrence matrix, *Biomedical Engineering International Conference (BMEiCON), 2013 6th*, pp. 1–4.

Mazo, C., Trujillo, M. and Salazar, L.: 2012, An automatic segmentation approach of epithelial cells nuclei, *in* L. Alvarez, M. Mejail, L. Gomez and J. Jacobo (eds), *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, Vol. 7441 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 567–574.

Melnyk, G.: 2015, Algorithm of matching of microobjects with different shapes, *Information Technologies in Innovation Business Conference (ITIB), 2015*, pp. 31–34.

Meng, T., Lin, L., Shyu, M.-L. and Chen, S.-C.: 2010, Histology image classification using supervised classification and multimodal fusion, *Multimedia (ISM), 2010 IEEE International Symposium on*, pp. 145 –152.

Merkus, H.: 2009, *Particle Size Measurements: Fundamentals, Practice, Quality*, Particle Technology Series, Springer Netherlands.

Mohammed, E. A., Far, B. H., Mohamed, M. M. A. and Naugler, C.: 2013, Application of support vector machine and k-means clustering algorithms for robust chronic lymphocytic leukemia color cell segmentation, *e-Health Networking, Applications Services (Healthcom), 2013 IEEE 15th International Conference on*, pp. 622–626.

Nedzved, A. and Starovoitov, V.: 2010, Extraction of thin color pattern from images for histology investigation, *The 2010 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–7.

Nguyen, K., Bredno, J. and Knowles, D. A.: 2015, Using contextual information to classify nuclei in histology images, *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, pp. 995–998.

Nosal, E.: 2008, Flood-fill algorithms used for passive acoustic detection and tracking, *New Trends for Environmental Monitoring Using Passive Systems*, pp. 1–5.

Noy, N. F. and Mcguinness, D. L.: 2001, Ontology development 101: A guide to creating your first ontology, *Technical report*.

Ojala, T., Pietikainen, M. and Maenpaa, T.: 2002, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(7), 971–987.

Orlov, N., Delaney, J., Eckley, D., Shamir, L. and Goldberg, I.: 2009, Pattern recognition for biomedical imaging and image-guided diagnosis, *Life Science Systems and Applications Workshop, 2009. LiSSA 2009. IEEE/NIH*, pp. 120 –123.

Orlov, N., Shamir, L., Macura, T., Johnston, J., Eckley, D. M. and Goldberg, I. G.: 2008, Wndcharm: Multi-purpose image classification using compound image transforms, *Pattern Recogn. Lett.* **29**(11), 1684–1693.

Othmani, A., Meziat, C. and Loménie, N.: 2010, Ontology-driven image analysis for histopathological images, *Advances in Visual Computing - 6th International Symposium, ISVC 2010, Las Vegas, NV, USA, November 29-December 1, 2010. Proceedings, Part I*, pp. 1–12.

OWL Working Group, W.: 2009, *OWL 2 Web Ontology Language: Document Overview*, W3C Recommendation. Available at `http://www.w3.org/TR/owl2-overview/`.

Palacios, G. and Beltran, J. R.: 2007, Cell nuclei segmentation combining multiresolution analysis, clustering methods and colour spaces, *Machine Vision and Image Processing Conference, 2007. IMVIP 2007. International*, pp. 91–97.

Paslaru Bontas, E., Tietz, S., Tolksdorf, R. and Schrader, T.: 2004, *Generation and Management of a Medical Ontology in a Semantic Web Retrieval System*, Springer Berlin Heidelberg, Berlin, Heidelberg.

Paulson, P., Hohimer, R., Doucette, P., Harvey, W., Seedahmed, G., Petrie, G. and Martucci, L.: 2006, A methodology for integrating images and text for object identification, *Prospecting for Geospatial Information Integration*, pp. 1057–1066.

Peyret, R., Bouridane, A., Al-Maadeed, S. A., Kunhoth, S. and Khelifi, F.: 2015, Texture analysis for colorectal tumour biopsies using multispectral imagery, *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 7218–7221.

Pietikainen, M., Ojala, T. and Xu, Z.: 2000, Rotation-invariant texture classification using feature distributions, *Pattern Recognition* **33**, 43–52.

Poveda-Villalón, M., Gómez-Pérez, A. and Suárez-Figueroa, M. C.: 2014, Oops! (ontology pitfall scanner!): An on-line tool for ontology evaluation, *Int. J. Semant. Web Inf. Syst.* **10**(2), 7–34.

Prud'hommeaux, E. and Seaborne, A.: 2008, SPARQL Query Language for RDF, W3C Recommendation. `http://www.w3.org/TR/rdf-sparql-query/`.

Pudil, P., Novovičová, J. and Kittler, J.: 1994, Floating search methods in feature selection, *Pattern Recogn. Lett.* **15**(11), 1119–1125.

Rao, A. R. and Schunck, B. G.: 1991, Computing oriented texture fields., *CVGIP: Graphical Model and Image Processing* **53**(2), 157–185.

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C. and Muller, M.: 2011, proc: an open-source package for r and s+ to analyze and compare roc curves, *BMC Bioinformatics* **12**(1), 1–8.

Rogojanu, R., Bises, G., Smochina, C. and Manta, V.: 2010, Segmentation of cell nuclei within complex configurations in images with colon sections, *Intelligent Computer Communication and Processing (ICCP), 2010 IEEE International Conference on*, pp. 243–246.

Rohr, K.: 2001, *Landmark-based image analysis : using geometric and intensity models*, Computational imaging and vision, Kluwer Academic Publ. cop., Dordrecht, Boston, London.

Rubin, D. L., Moreira, D. A., Kanjamala, P. and Musen, M. A.: 2008, Bioportal: A web portal to biomedical ontologies., *AAAI Spring Symposium: Symbiotic Relationships between Semantic Web and Knowledge Engineering*, AAAI, pp. 74–77.

Ruiz, J., Mintzer, M. and Leipzig, R.: 2006, The impact of e-learning in medical education., *Academic Medicine* **81**(3), 207–212.

Sabou, M.: 2006, *Building Web Service Ontologies*, Phd thesis, Dutch Graduate School for Information and Knowledge Systems, Netherlands.

Sato, Y., Nakajima, S., Atsumi, H., Koller, T., Gerig, G., Yoshida, S. and Kikinis, R.: 1997, 3d multi-scale line filter for segmentation and visualization of curvilinear structures in medical images, *Proceedings of the First Joint Conference on Computer Vision, Virtual Reality and Robotics in Medicine and Medial Robotics and Computer-Assisted Surgery*, CVRMed-MRCAS '97, Springer-Verlag, London, UK, UK, pp. 213–222.

Schofield, P. N., Sundberg, J. P., Sundberg, B. A., McKerlie, C. and Gkoutos, G. V.: 2013, The mouse pathology ontology, mpath; structure and applications, *Journal of Biomedical Semantics* **4**(1), 1–8.

Sertel, O., Kong, J., Catalyurek, U. V., Lozanski, G., Saltz, J. H. and Gurcan, M. N.: 2008, Histopathological image analysis using model-based intermediate representations and color texture: Follicular lymphoma grading, *Journal of Signal Processing Systems* **55**(1), 169–183.

Sharma, H., Alekseychuk, A., Leskovsky, P., Hellwich, O., Anand, R., Zerbe, N. and Hufnagl, P.: 2012, Determining similarity in histological images using graph-theoretic description and matching methods for content-based image retrieval in medical diagnostics, *Diagnostic Pathology* **7**(1), 1–20.

Shir, O. M., Raz, V., Dirks, R. W. and Back, T.: 2007, *Classification of Cell Fates with Support Vector Machine Learning*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 258–269.

Simsek, A., Tosun, A., Aykanat, C., Sokmensuer, C. and Gunduz-Demir, C.: 2012, Multilevel segmentation of histopathological images using cooccurrence of tissue objects, *Biomedical Engineering, IEEE Transactions on* **59**(6), 1681–1690.

Smith, B., Arabandi, S., Brochhausen, M., Calhoun, M., Ciccarese, P., Doyle, S. and Gurcan, M.: 2015, Biomedical imaging ontologies: A survey and proposal for future work., *Journal of Pathology Informatics* **6**(37).

Song, Y., Cai, W., Feng, D. D. and Chen, M.: 2013, Cell nuclei segmentation in fluorescence microscopy images using inter- and intra-region discriminative information, *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 6087–6090.

Sonka, M., Hlavac, V. and Boyle, R.: 1998, *Image Processing, Analysis, and Machine Vision*, 2 edn, Chapman & Hall.

Studer, R., Benjamins, R. and Fensel, D.: 1998, Knowledge engineering: Principles and methods, *Data & Knowledge Engineering* **25**, 161–198.

Suarez-Figueroa, M. C. and Gomez-Perez, A.: 2009, NeOn Methodology for Building Ontology Networks: a Scenario-based Methodology, *in* S. . S. T. S. . I. C. on Software (ed.), *Proceedings of the International Conference on Software, Services & Semantic Technologies (S3T 2009)*.

Suganya, R. and Rajaram, S.: 2013, Feature extraction and classification of ultrasound liver images using haralick texture-primitive features: Application of svm classifier, *Recent Trends in Information Technology (ICRTIT), 2013 International Conference on*, pp. 596–602.

Tambasco, M., Costello, B., Kouznetsov, A., Yau, A. and Magliocco, A.: 2009, Quantifying the architectural complexity of microscopic images of histology specimens, *Micron* **40**(4), 486–494.

Tonkin, J. A., Rees, P., Brown, M. R., Errington, R. J., Smith, P. J., Chappell, S. C. and Summers, H. D.: 2011, Segmentation of epithelium in H&E stained odontogenic cysts, *Journal of Microscopy* **244**(3), 273–292.

Tosun, A. B., Kandemir, M., Sokmensuer, C. and Gunduz-Demir, C.: 2009, Object-oriented texture analysis for the unsupervised segmentation of biopsy images for cancer detection., *Pattern Recognition* **42**(6), 1104–1112.

Vasquez, H., Aguilera, A. and Tineo, L.: 2010, Ontologias medicas: una revision, *Tecnologia, Gerencia y Educacion* **11**, 9–29.

Veillard, A., Bressan, S. and Racoceanu, D.: 2012, Svm-based framework for the robust extraction of objects from histopathological images using color, texture, scale and geometry, *Machine Learning and Applications (ICMLA), 2012 11th International Conference on*, Vol. 1, pp. 70–75.

Wang, H.-T. and Tansel, A. U.: 2013, Composite ontology-based medical diagnosis decision support system framework, *Communications of the IIMA* **13**(2), 43–52.

Wang, J., MacKenzie, J. D., Ramachandran, R. and Chen, D. Z.: 2014, *Identifying Neutrophils in H&E Staining Histology Tissue Images*, Springer International Publishing, Cham, pp. 73–80.

Weickert, J. and Scharr, H.: 2002, A scheme for coherence-enhancing diffusion filtering with optimized rotation invariance, *JVCIR* **13**(1/2), 103–118.

Wu, G., Zhao, X., Luo, S. and Shi, H.: 2015, Histological image segmentation using fast mean shift clustering method, *BioMedical Engineering OnLine* **14**(1), 1–12.

Wu, Z., Jiang, S., Li, L., Cui, P., Huang, Q. and Gao, W.: 2010, Vicept: link visual features to concepts for large-scale image understanding, *Proceedings of the 18th International Conference on Multimedia 2010, Firenze, Italy, October 25-29, 2010*, pp. 711–714.

Yang, J., Jiang, Y.-G., Hauptmann, A. G. and Ngo, C.-W.: 2007, Evaluating bag-of-visual-words representations in scene classification, *Proceedings of the International Workshop on Workshop on Multimedia Information Retrieval*, MIR '07, ACM, New York, NY, USA, pp. 197–206.

Yang, Y., Wang, J. and Yang, Y.: 2012, Exploiting rotation invariance with svm classifier for microcalcification detection, *2012 9th IEEE International Symposium on Biomedical Imaging (ISBI)*, pp. 590–593.

Yu, F. and Ip, H. H. S.: 2008, Semantic content analysis and annotation of histological images, *Comput. Biol. Med.* **38**(6), 635–649.

Yu, Feiyang, Ip, H. and Horace, H. S.: 2008, Semantic content analysis and annotation of histological images, *Comput. Biol. Med.* **38**(6), 635–649.

Zeng, Z., Chen, S., Tang, S. and Yin, L.: 2015, Unsupervised segmentation of cell nuclei in cervical smear images using active contour with adaptive local region fitting energy modelling, *2015 8th International Conference on Biomedical Engineering and Informatics (BMEI)*, pp. 250–254.

Zeng, Z., Strange, H., Han, C. and Zwiggelaar, R.: 2013, *Unsupervised Cell Nuclei Segmentation Based on Morphology and Adaptive Active Contour Modelling*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 605–612.

Zhao, D., Chen, Y. and Correa, N.: 2005, Statistical categorization of human histological images, *Image Processing, 2005. ICIP 2005. IEEE International Conference* **3**, 628–631.

Annex A: Competency Questions

Table 1: Complete competency questions

Pregunta	Respuesta	Comentario
	Clasificación	
Debería decirme el sistema humano de interés	Sistema cardiovascular	Importante
Debería decirme los componentes del sistema cardiovascular	Sanguíneo y linfático	Importante
Debería decirme los tipos de órganos	Tubular y visceral	Adicional
Debería decirme los órganos del sistema cardiovascular que existen	Corazón, arterias, venas, capilares, vénula, arteriola y metarteriolas	Muy importante
Debería decirme los tipos de sistema conducente	Nodo sinuatrial, nodo atroventricular, haz de his y fibra de purkinje	Importante
Debería decirme las túnicas presentes en el corazón	Endocardio, miocardio y pericardio seroso	Importante
Debería decirme la composición del endocardio	Subendotelio, endocardio y subendocardio	Importante
Debería decirme la composición del subendocardio	Conectivo denso irregular	Importante
Debería decirme la composición del endocardio	Endotelio	Importante
Debería decirme la composición del subendotelio	Conectivo laxo	Importante
Debería decirme la composición del miocardio	Músculo estriado cardiaco	Importante

Continue in next page

Table 1 – Continuation from previous page

Pregunta	Respuesta	Comentario
Debería decirme la composición del pericardio seroso o epicardio seroso	Conectivo laxo y mesotelio	Importante
Debería decirme la composición del mesotelio	Epitelio plano simple	Importante
Debería decirme las regiones anatómicas presentes en el corazón	Ventrículo izquierdo, ventrículo derecho, atrio izquierdo y atrio derecho	Muy importante
Debería decirme los sectores presentes en el corazón	Válvulas, cuerdas tendinosas, sistema conducente del corazón y músculos	Muy importante
Debería decirme las válvulas de los sectores del corazón	Pulmonar, aórtica, tricúspide o atrioventricular derecha y mitral o atrioventricular izquierda	Muy importante
Debería decirme los sistemas conducentes de los sectores del corazón	Nodo atrioventricular, nodo sinusal, haz de his y fibra de purkinje	Muy importante
Debería decirme los músculos de los sectores del corazón	Papilares, pectinados y trabéculas cárneas	Muy importante
Debería decirme los tipos de arterias	Elásticas, musculares y arteriolas	Muy importante
Debería decirme las arterias elásticas	Aorta, subclavia, carótida común, ilíaca y femoral	Muy importante
Debería decirme las arterias musculares	Axilar, braquial, radial, ulnar, palmar, tibia, popitlea, plantar, etc. (no exhaustivo)	Muy importante
Debería decirme los tipos de venas	Gran calibre, mediano calibre, pequeño calibre y vénulas	Muy importante
Debería decirme los tipos de vénulas	Colectoras, postcapilares y musculares	Muy importante
Debería decirme las venas de gran calibre	Yugular, subclavia, cava superior, cava inferior, ilíaca y femoral	Muy importante

Continue in next page

Table 1 – Continuation from previous page

Pregunta	Respuesta	Comentario
Debería decirme las venas de mediano calibre	Axilar, braquial, radial, ulnar, palmar, tibia, popitlea, plantar, etc. (no exhaustivo, las mismas que las arterias musculares)	Muy importante
Debería decirme los tipos de capilares	Continuo, discontinuo, sinusoide y fenestrado	Muy importante
Debería decirme los tejidos fundamentales que existen	Tejido epitelial, conectivo, muscular y nervioso	Muy importante
Debería decirme la clasificación general de los epitelios	Glandulares y revestimiento	Muy importante
Debería decirme los tipos de tejidos epiteliales por número de capas	Simples, estratificados, pseudoestratificado y transicional	Muy importante
Debería decirme los tipos de tejido epitelial simple	Tejido epitelial plano simple, cúbico simple, cilíndrico simple y pseudoestratificado	Muy importante
Debería decirme los tipos de tejido epitelial estratificado	Tejido epitelial plano estratificado, cúbico estratificado, cilíndrico estratificado y transicional	Adicional
Debería decirme los tipos de tejido epitelial por especialización apical	Queratina, microvellosidades, estereocilios y cilios	Adicional
Debería decirme los tipos de tejido epitelial glandular	Glándulas exocrinas y endocrinas	Adicional
Debería decirme los tipos de tejido epitelial glandular endocrina de acuerdo a su morfología	Cordonal y folicular	Adicional
Debería decirme los tipos de tejido epitelial glandular exocrina según la morfología del adenómero	Tubular, acinar y tubulo acinar	Adicional
Debería decirme los tipos de tejido epitelial glandular exocrina según el número de conductos	Simple y compuesto	Adicional

Continue in next page

Table 1 – Continuation from previous page

Pregunta	Respuesta	Comentario
Debería decirme los tipos de tejido epitelial glandular exocrina según la naturaleza de la secreción	Serosa, mucosa y mixta	Adicional
Debería decirme los tipos de tejido epitelial glandular exocrina según el mecanismo de secreción	Merocrina, apocrina y holocrina	Adicional
Debería decirme los tipos de tejido conectivo	Embrionario, adulto propiamente dicho y especializado	Muy importante
Debería decirme los tipos de tejido conectivo adulto propiamente dicho	Tejido conectivo laxo, denso regular, denso irregular, especializado reticular y especializado adiposo	Muy importante
Debería decirme los tipos de tejido conectivo embrionario	Mesenquimático y conectivo mucoide	Adicional
Debería decirme los tipos de tejido conectivo especializado	Cartílago, óseo y sangre	Adicional
Debería decirme los tipos de tejido muscular	Tejido muscular liso y estriado	Muy importante
Debería decirme los tipos de tejido muscular estriado	Tejido muscular cardiaco y músculo esquelético	Muy importante
Debería decirme los tipos de tejido nervioso por su distribución anatómica	Sistema nervioso central y periférico	Muy importante
Debería decirme los tipos de tejido nervioso periférico	Somático y autónomo	Muy importante
	Propiedades	
Debería decirme las capas del pericardio	Visceral y parietal	Importante
Debería decirme las regiones de interés en corazón	Ventrículo izquierdo, ventrículo derecho, atrio izquierdo, atrio derecho y septos	Importante

Continue in next page

Table 1 – Continuation from previous page

Pregunta	Respuesta	Comentario
Debería decirme las estructuras del corazón	Válvula atrioventricular derecha o tricúspide, válvula atrioventricular izquierda o mitral, cuerdas tendinosas, músculos papilares, válvula aórtica, válvula pulmonar, músculos pectinados, nodo sinuatrial y nodo atroventricular	Importante
Debería decirme los tejidos fundamentales presentes en el corazón	Conectivo laxo, músculo estriado cardiaco y epitelio plano simple	Importante
Debería decirme las túnicas presentes en las venas	Íntima, media, adventicia y subendotelio (en grandes venas)	Importante
Debería decirme las partes de la túnica íntima	Endotelio, subendotelio y lámina elástica interna	Importante
Debería decirme la composición del endotelio	Epitelio plano simple	Importante
Debería decirme la composición del subendotelio	Conectivo laxo	Importante
Debería decirme la constitución de la túnica media	Músculo liso, fibra colágena y fibra elástica	Importante
Debería decirla la constitución de la túnica adventicia	Conectivo laxo	Importante
Debería decirme las estructuras presentes en las grandes venas	Vaso vasorum	Importante
Debería decirme los tejidos fundamentales presentes en venas de pequeño calibre	Epitelio plano simple, músculo liso y conectivo laxo	Importante
Debería decirme los tejidos fundamentales presentes en venas de mediano calibre	Epitelio plano simple, músculo liso y conectivo laxo	Importante
Debería decirme los tejidos fundamentales presentes en venas de gran calibre	Epitelio plano simple, músculo liso y conectivo laxo	Importante

Continue in next page

Table 1 – Continuation from previous page

Pregunta	Respuesta	Comentario
Debería decirme los tejidos fundamentales presentes en vénulas	Epitelio plano simple y conectivo laxo	Importante
Debería decirme los tejidos fundamentales presentes en capilares	Epitelio plano simple	Importante
Debería decirme las túnicas presentes en las arterias	Íntima, media y adventicia	Importante
Debería decirme las estructuras presentes en las arterias musculares	Fibra elástica, vaso vasorum y lámina elástica (interna y externa)	Importante
Debería decirme los tejidos fundamentales presentes en las arterias elásticas	Epitelio plano simple, músculo liso y conectivo laxo	Importante
Debería decirme los tejidos fundamentales presentes en las arterias musculares	Epitelio plano simple, músculo liso y conectivo laxo	Importante
Debería decirme las túnicas presentes en las arteriolas	Íntima, media y adventicia	Importante
Debería decirme los tejidos fundamentales presentes en las arteriolas	Epitelio plano simple, músculo liso y conectivo laxo	Importante
Debería decirme algunas células presentes en el tejido epitelial	Epitelial plana, epitelial cúbica y epitelial cilíndrica	Importante
Debería decirme algunas células presentes en el tejido muscular	Cardiomiocito, miocito liso y fibra de purkinje	Importante
Debería decirme algunas células presentes en el tejido conectivo	Adipocito y fibroblasto	Importante
Debería decirme los componentes del tejido nervioso	Glias y neuronas	Adicional
Debería decirme cuántas válvulas tiene la válvula tricuspide	3	Adicional
Debería decirme cuántas válvulas tiene la válvula mitral	2	Adicional

Continue in next page

Table 1 – Continuation from previous page

Pregunta	Respuesta	Comentario
Debería decirme las túnicas de los capilares	Únicamente túnica íntima	Importante
Debería decirme las túnicas de las vénulas	Únicamente túnica íntima	Importante
Debería decirme las túnicas de las arterias de pequeño calibre con sus subdivisiones	Íntima, media y adventicia	Importante
Debería decirme las túnicas de las arterias de mediano calibre con sus subdivisiones	Íntima, subendotelio, media y adventicia	Importante
Debería decirme las túnicas de las arterias de pequeño calibre con sus subdivisiones	Íntima, subendotelio y lámina elástica), media y adventicia	Importante
Debería decirme la composición del epitelio plano simple	Células epiteliales	Importante
Debería decirme la composición de las láminas elásticas	Fibras elásticas	Importante
Debería decirme la túnicas de las venas de gran calibre	Íntima, media y adventicia	Importante
Debería decirme la túnicas de las venas de mediano y pequeño calibre	Íntima, media y adventicia	Importante

Restricciones	
En el sistema cardiovascular el epitelio plano simple es un endotelio	Muy importante
Un tejido epitelial simple no puede ser estratificado	Muy importante
Un tejido epitelial simple sólo puede tener una morfología	Muy importante
Una fibra muscular es una célula	Importante

Continue in next page

Table 1 – Continuation from previous page

Pregunta	Respuesta	Comentario
Un capilar sólo tendrá presencia de endotelio		Importante
Un órgano podría tener máximo hasta 3 túnicas		Importante
	Inferencias	
Si un órgano tiene la túnica media delgada, la túnica adventicia gruesa y la región de luz amplia probablemente es una vena		Muy importante
Si un órgano tiene la túnica media gruesa y la región de luz pequeña probablemente es una arteria		Muy importante
Si un vaso sanguíneo pequeño tiene un diámetro pequeño será un capilar, si tiene un diámetro mayor será una vénula		Importante

Annex B: Histology Vocabulary Evaluation

En esta encuesta se evaluarán los diferentes conceptos histológicos del sistema cardiovascular, en particular evaluaremos macrocirculación, aunque en algunos casos se encontraran términos que abarcan más de lo necesario, con el fin de evaluar aspectos como:

Completitud: Los conceptos presentados cubren todos los relacionados con el sistema cardiovascular.

Duplicidad o Redundancia: Existen conceptos repetidos o sinónimos que se pueden agrupar en un único concepto.

Coherencia: Los términos conservan una relación lógica entre ellos de modo que no se produce contradicción ni oposición

Lee detenidamente cada pregunta y responde según el caso. Escriba la justificación sólo en caso de ser necesario.

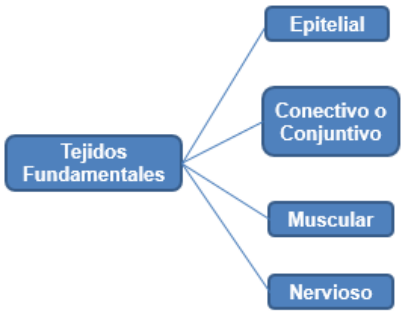


Figure 1: Mapa conceptual de los tejidos fundamentales a nivel general.

- Los conceptos presentados en la Figure 1: ¿Incluyen todos los tejidos fundamentales?
  ___Sí ___No
- ¿Existen conceptos repetidos o sinónimos que se pueden agrupar en un único concepto?
  ___Sí ___No
- ¿Existe una relación lógica entre los conceptos sin contradicción, ni oposición?
  ___Sí ___No

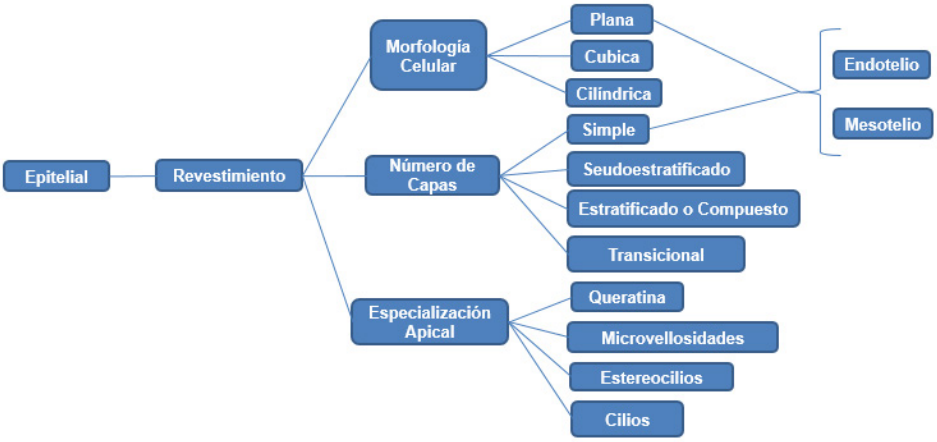


Figure 2: Mapa conceptual clasificación del tejido epitelial de revestimiento.

- Los conceptos presentados en la Figure 2: ¿Incluyen la clasificación completa del tejido epitelial de revestimiento?
  ___Sí ___No

- ¿Existen conceptos repetidos o sinónimos que se pueden agrupar en un único concepto?
  ___Sí ___No

- ¿Existe una relación lógica entre los conceptos sin contradicción, ni oposición?
  ___Sí ___No

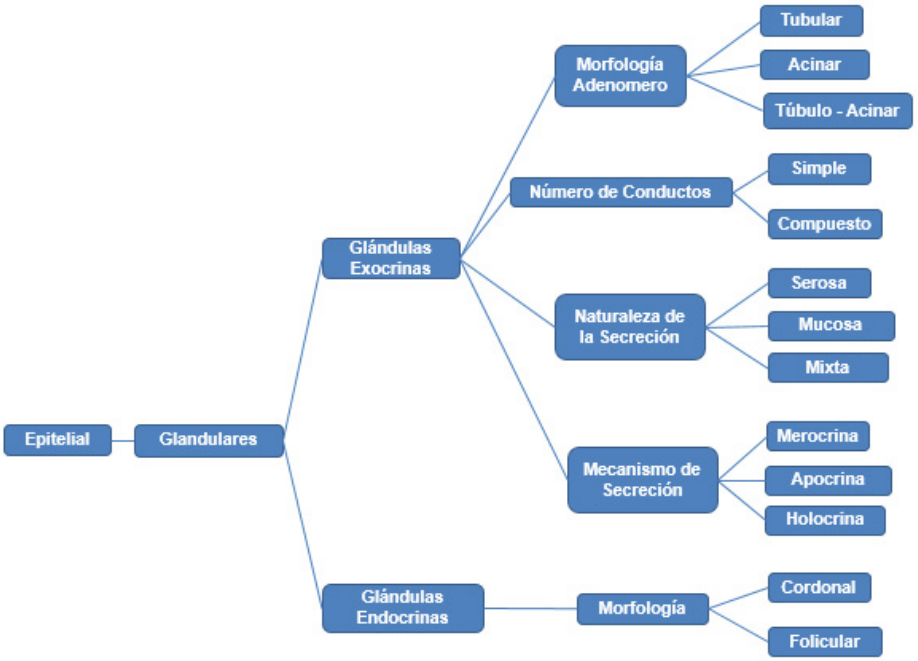


Figure 3: Mapa conceptual clasificación del tejido epitelial glandular.

- Los conceptos presentados en la Figure 3: ¿Incluyen la clasificación completa del tejido epitelial glandular?
  ___Sí ___No

- ¿Existen conceptos repetidos o sinónimos que se pueden agrupar en un único concepto?
  ___Sí ___No

- ¿Existe una relación lógica entre los conceptos sin contradicción, ni oposición?
  ___Sí ___No

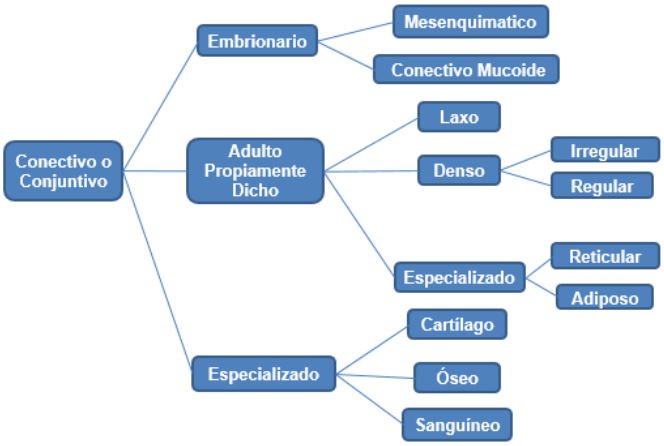


Figure 4: Mapa conceptual clasificación del tejido conectivo.

- Los conceptos presentados en la Figure 4: ¿Incluyen la clasificación completa del tejido conectivo?
  ___Sí ___No
- ¿Existen conceptos repetidos o sinónimos que se pueden agrupar en un único concepto?
  ___Sí ___No
- ¿Existe una relación lógica entre los conceptos sin contradicción, ni oposición?
  ___Sí ___No

Figure 5: Mapa conceptual clasificación morfológica del tejido muscular.

- Los conceptos presentados en la Figure 5: ¿Incluyen todos los tipos de tejidos musculares?
  ___Sí ___No
- ¿Existen conceptos repetidos o sinónimos que se pueden agrupar en un único concepto?
  ___Sí ___No
- ¿Existe una relación lógica entre los conceptos sin contradicción, ni oposición?
  ___Sí ___No

Figure 6: Mapa conceptual clasificación histológica del sistema circulatorio.

- Los conceptos presentados en la Figure 6: ¿Incluyen la clasificación completa del sistema circulatorio y sus órganos?
  ___Sí ___No
- ¿Existen conceptos repetidos o sinónimos que se pueden agrupar en un único concepto?
  ___Sí ___No
- ¿Existe una relación lógica entre los conceptos sin contradicción, ni oposición?
  ___Sí ___No

Figure 7: Mapa conceptual clasificación histológica de las arterias.

- Los conceptos presentados en la Figure 7: ¿Incluyen la clasificación histológica completa de las arterias?
  ___Sí ___No
- ¿Existen conceptos repetidos o sinónimos que se pueden agrupar en un único concepto?
  ___Sí ___No

- ¿Existe una relación lógica entre los conceptos sin contradicción, ni oposición?
  ___Sí ___No

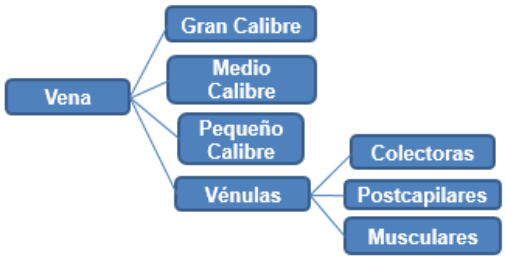


Figure 8: Mapa conceptual clasificación histológica de las venas.

- Los conceptos presentados en la Figure 8: ¿Incluyen la clasificación histológica completa de las venas?
  ___Sí ___No
- ¿Existen conceptos repetidos o sinónimos que se pueden agrupar en un único concepto?
  ___Sí ___No
- ¿Existe una relación lógica entre los conceptos sin contradicción, ni oposición?
  ___Sí ___No

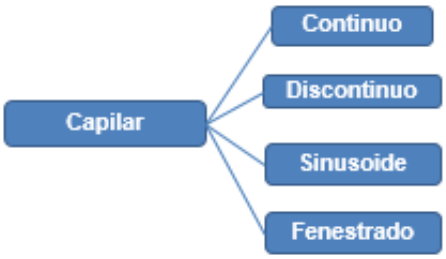


Figure 9: Mapa conceptual clasificación histológica de los capilares.

- Los conceptos presentados en la Figure 9: ¿Incluyen la clasificación histológica completa de los capilares?
  ___Sí ___No
- ¿Existen conceptos repetidos o sinónimos que se pueden agrupar en un único concepto?
  ___Sí ___No
- ¿Existe una relación lógica entre los conceptos sin contradicción, ni oposición?
  ___Sí ___No

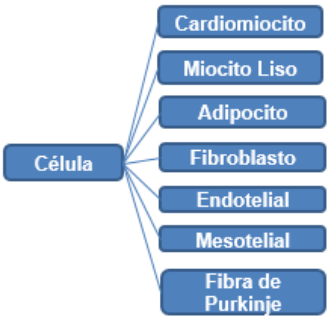


Figure 10: Principales células que se pueden observar en una placa correspondiente al sistema circulatorio.

- Los conceptos presentados en la Figure 10: ¿Incluyen las principales células que se pueden observar en una placa correspondiente al sistema circulatorio?
  ___Sí ___No
- ¿Existen conceptos repetidos o sinónimos que se pueden agrupar en un único concepto?
  ___Sí ___No
- ¿Existe una relación lógica entre los conceptos sin contradicción, ni oposición?
  ___Sí ___No

Figure 11: Clasificación histológica de las capas del corazón.

- Los conceptos presentados en la Figure 11: ¿Incluyen la clasificación completa de las capas en el corazón?
  ___Sí ___No
- ¿Existen conceptos repetidos o sinónimos que se pueden agrupar en un único concepto?
  ___Sí ___No
- ¿Existe una relación lógica entre los conceptos sin contradicción, ni oposición?
  ___Sí ___No

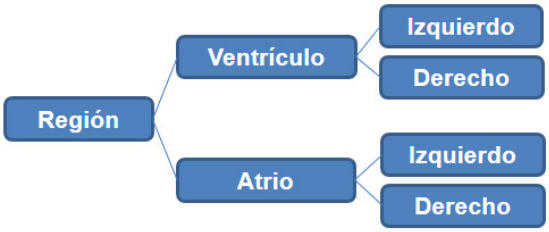


Figure 12: Clasificación de regiones anatómicas presentes en el corazón.

- Los conceptos presentados en la Figure 12: ¿Incluyen la clasificación completa de regiones presentes en el corazón?
  ___Sí ___No
- ¿Existen conceptos repetidos o sinónimos que se pueden agrupar en un único concepto?
  ___Sí ___No
- ¿Existe una relación lógica entre los conceptos sin contradicción, ni oposición?
  ___Sí ___No

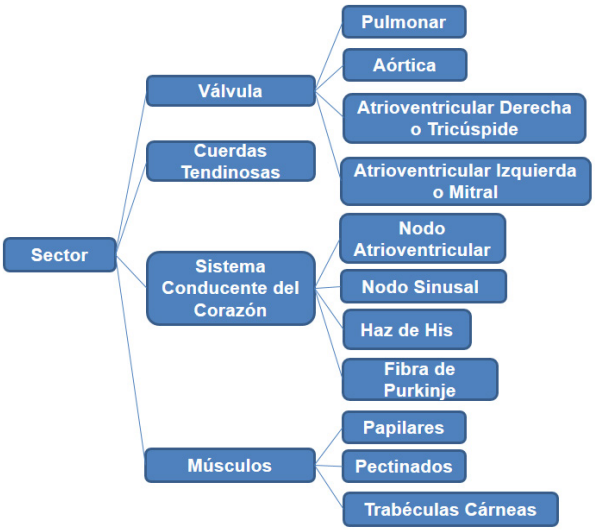


Figure 13: Clasificación de sectores anatómicos presentes en el corazón.

- Los conceptos presentados en la Figure 13: ¿Incluyen la clasificación completa de sectores anatómicos presentes en el corazón?
  ___Sí ___No
- ¿Existen conceptos repetidos o sinónimos que se pueden agrupar en un único concepto?

___Sí ___No

- ¿Existe una relación lógica entre los conceptos sin contradicción, ni oposición?
___Sí ___No

Figure 14: Clasificación histológica de las capas de los vasos sanguíneos.

- Los conceptos presentados en la Figure 14: ¿Incluyen la clasificación completa de las capas en los vasos sanguíneos?
___Sí ___No
- ¿Existen conceptos repetidos o sinónimos que se pueden agrupar en un único concepto?
___Sí ___No
- ¿Existe una relación lógica entre los conceptos sin contradicción, ni oposición?
___Sí ___No

Annex C: Research activities

Publications Related with this Manuscript

Journal Papers

- C. Mazo, M. Trujillo, E. Alegre and L. Salazar, *"Automatic Recognition of Fundamental Tissues on Histology Images of the Human Cardiovascular System,"* in Micron, vol 89, pp 1-8, ISSN: 0968-4328, ed 2016.
- C. Mazo, E. Alegre and M. Trujillo, *"Classification of Cardiovascular Tissues Using LBP Based Descriptors and a Cascade SVM,"* in Computer Methods and Programs in Biomedicine, 2016, [submitted].

Other Papers

- C. Mazo, M. Trujillo, and L. Salazar, *"Automatic Classification of Coating Epithelial Tissue,"* in Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications, vol. 8827 of Lecture Notes in Computer Science, pp. 311-318, Springer Verlag, ISSN: 1611-3349 ed, 2014.
- C. Mazo, M. Trujillo, and L. Salazar, *"Identificación Automática de Tejido Conectivo Laxo y Tejido Muscular,"* presented in X Congreso Colombiano de Morfología, Revista de investigaciones de la Universidad del Quindío Vol 26, pp. 108-109, ISSN 1794-631X, Armenia, Colombia, 2014.
- C. Mazo, M. Trujillo, and L. Salazar, *"Identifying Loose Connective and Muscle Tissues on Histology Images,"* in Progress in Pattern Recognition, Image Analysis, Computer Vision, and ApplicationsRuiz-Shulcloper, José and Sanniti di Baja, Gabriella, eds.), vol. 8259 of Lecture Notes in Computer Science, pp. 174-180, Springer Berlin Heidelberg, 2013.
- C. Mazo, M. Trujillo, and L. Salazar, *"An automatic segmentation approach of epithelial cells nuclei,"* in Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications (L. Alvarez, M. Mejail, L. Gomez, and J. Jacobo, eds.), vol. 7441 of Lecture Notes in Computer Science, pp. 567–574, Springer Berlin Heidelberg, 2012.

Conference Papers

- C. Mazo, M. Trujillo, and L. Salazar, *"Automatic identification of Fundamental Tissues of Cardiovascular System,"* presented in VII Simposio de Investigaciones. Cali – Colombia, 2014.
- C. Mazo, M. Trujillo, and L. Salazar, *"Histology Images Analisys,"* presented in IEEE EMBS International Summer School on Biomedical Imaging, Saint-Jacut-de-la-Mer, Francia, 2014.
- C. Mazo, M. Trujillo, and L. Salazar, *"Identificación Automática de los Tejidos Conectivo Laxo y Muscular Utilizando Información Morfológica,"* presented in XVIII Congreso Panamericano de Anatomía – XX Reunión Nacional de Morfología "Dr. Fernando Quiroz Pavía" – IX Simposio Ibero-Latinoamericano de Terminología, Huatulco – Mexico, 2013.
- C. Mazo, M. Trujillo, and L. Salazar, *"Identificación de Núcleos de las Células Epiteliales Asistido por Ordenador,"* presented in VIII Congreso Colombiano de Morfología., pp. 90, ISBN: 958-978-57694-1-0, Tunja – Colombia, 2012.
- C. Mazo and M. Trujillo, *"Etiquetado Automático de Imágenes de Histología,"* presented in VI Seminario Internacional de Procesamiento y Análisis de Imágenes Médicas SIPAIM, ISBN: 978-958-719-614-6, Bogota – Colombia, 2010.

BISCAR Project

- Santamaria M., Muñoz Y., Cuellar E., Mazo C., Scotti S., Salazar L., Trujillo M., *"BISCAR: Banco de Imágenes Histológicas del Sistema Cardiovascular,"* presented in X Congreso Colombiano de Morfología, Revista de investigaciones de la Universidad del Quindío Vol 26, pp. 104-105, ISSN 1794 -631X, Armenia, Colombia, 2014.

Summer schools

- IEEE EMBS International Summer School on Biomedical Imaging, Saint-Jacut-de-la-Mer, French, 36 hours of lectures, 12 hours of computer lab's sessions and 6 hours of poster sessions, 19-27th June 2014.

Awards and Honours

- One of the winers in the "Convocatoria para el apoyo a la financiación de Tesis Doctorales en temáticas relacionadas con necesidades empresariales. Convocatoria DOCTOR ULE-TCUE". *"Offered by fgulem, University of León, FUESCYL, Junta de Castilla y León, FEDER, European Union, RIS Castilla y León and t-cue,"* Jun. 2017.
- Scholarship "Finalización de Tesis Doctorales, en Régimen de Cotutela, en las Universidades de Castilla y León", *"Offered by AUIP,"* Jul. 2015.

- The best work in the category of Oral Presentation at thematic axes of anatomy, neuroanatomy, histology, embryology and anatomy applied by "BISCAR: desarrollo de un banco de imágenes histológicas del Sistema Cardiovascular" Santamaria M., Muñoz Y., Cuellar E., Mazo C., Scotti S., Salazar L., Trujillo M. presented by Salazar L., *"Offered by Asociación Colombiana de Morfología (ASCOM) in the X Congreso Colombiano de Morfología,"* Armenia - Colombia, Nov. 2014.
- Scholarship "Estudios de Doctorado en Colombia – COLCIENCIAS 2011", *"Offered by COLCIENCIAS,"* Jan 2012.

Other Publications

- M. Trujillo, C. Ballesteros and C. Mazo, *"Automatic Classification of Non-informative Frames in Colonoscopy Videos,"* presented in 6th Latin American Conference on Networked and Electronic Media. LACNEM - 2015, Facultad de Minas, Universidad Nacional sede Medellín, Medellin - Colombia, 2015.
- C. Mazo, V. Padilla, and M. Trujillo, *"Identificación Automática de Grupos Macérales y Minerales en Imágenes Digitales,"* presented in IX Congreso Nacional y IV Internacional de Ciencia y Tecnología del Carbón y Combustibles Alternativos, CONICCA - 2011, ISBN: 978-958-670-943-9, Cali – Colombia, 2011.

Research Projects

- Desarrollo del Banco de Imágenes Histologícas Sobre el Sistema Cardiovascular (BISCAR). Participation in research project. Universidad del Valle.
- Integrando Información Enriquecida Semánticamente como apoyo al diagnóstico de enfermedades de Hígado. Participation in research project. Universidad del Valle.
- Sistema Distribuido de Anotación Automática y Recuperación Semántica de Imágenes de Histología. Participation in research project. Universidad Nacional de Colombia, Centro Extremeño de Tecnologías Avanzadas CETA-CIEMAT and Universidad del Valle.

Software Products

- C. Mazo, and O. Bedoya, *"PESPAD: Predicción de la Estructura Secundaria de la Proteína basada en Árboles de Decisión,"* Register Libro 13, Tomo 30, Partida 355, Colombia, Sept. 2011.

Other Teaching Activities

Supervising final career projects, Universidad del Valle, Cali - Colombia

- Student: Cristian Ballesteros, *"Project title: Eliminación de Fotogramas no Informativos en un Sistema de Anotación de Videos de Colonoscopia."* Jan. 2015 - Dec. 2015
- Student: Jorge Rodriguez, *"Project title:Creación de Imágenes de Alta Resolución con Imágenes Histológicas del Sistema Cardiovascular."* Jan. 2014 - Dec. 2014

Other Awards and Honours

- Scholarship "Programa Jóvenes Investigadores e Innovadores "Virginia Gutiérrez de Pineda"", *"Offered by COLCIENCIAS,"* Dec. 2009.
- Selected as Best Student of the System Engineering and Computer School at Universidad del Valle according to GPA, *"Scholarship for the best students of the College of Engineering,"* Cali - Colombia, 2007.
- Selected as Third Best Student of the System Engineering and Computer School at Universidad del Valle according to GPA, *"Scholarship for the best students of the College of Engineering,"* Cali - Colombia, 2006.

Annex D

SUMMARY OF THE THESIS IN SPANISH

RESUMEN DE LA TESIS EN CASTELLANO

En cumplimiento del punto 7° de la normativa complementaria del Real Decreto 778/1998, de 30 de Abril y de las normas para la aplicación del mismo, aprobadas por acuerdo de la Junta de Gobierno de fecha 10 de mayo de 1999, se adjunta un resumen en castellano de cada uno de los capítulos de esta tesis doctoral para que pueda admitirse a trámite.

1 Introducción

1.1 Motivación

Un problema de visión por computador, con imágenes histológicas, consiste en identificar los tejidos fundamentales y reconocer los patrones formados por las estructuras espaciales entre ellos para inferir un órgano, siendo éste un reto para los investigadores y un problema abierto (Zhao et al., 2005). Dado que un órgano es una estructura 3D y una imagen es una representación 2D de una muestra histológica, inferir información a partir de una representación 2D de un órgano 3D es un problema inverso (Hansen, 2010). Además, la información sobre las funciones fisiológicas de un órgano no están contenidas en una imagen; esta falta de información genera soluciones inestables haciendo de esté un problema mal planteado.

Adicionalmente, la representación del conocimiento histológico y el conocimiento de los expertos es otro reto que se debe afrontar en este contexto. La formalización del conocimiento humano y su procesamiento, dentro de las máquinas en una forma que permita su interpretación, es una de las problemáticas abordadas en esta tesis para resolver tareas complejas como apoyo a la enseñanza, las prácticas médicas o lograr la interacción en lenguaje natural.

Esta tesis doctoral presenta algunas particularidades de clasificación automática de imágenes histológicas y modelado del conocimientos histológico a través de cuatro propuestas diferentes:

- Reconocimiento de los tejidos fundamentales — epitelial, conectivo y muscular — utilizando técnicas de procesamiento de imágenes.
- Clasificación de los tejidos — epitelial, conectivo y muscular — y órganos cardiovasculares — corazón, arteria muscular, arteria elástica y vena de gran calibre — usando descriptores basados en *Local Binary Pattern* (LBP) y Máquinas de Soporte Vectorial (SVM) en cascada.
- Construcción de una ontología histológica.
- Mejora de la clasificación automática usando la ontología histológica del sistema cardiovascular humano.

Estas propuestas pueden ser muy útiles para reforzar el proceso de aprendizaje de los histólogos, biólogos, patólogos y otros profesionales de la salud (Izet, 2008). El reconocimiento automatizado de tejidos beneficiaria a los estudiantes en: aumento del número de casos a analizar, promover el auto-aprendizaje en el campus y facilitar el aprendizaje en línea o remoto a través de sistemas de *E-Learning* (Ruiz et al., 2006). Todo esto se traduce en profesionales mejor formados con una menor inversión social y económica. Otra ventaja de este tipo de sistemas es que podrían permitir la anotación en los bancos o conjuntos de imágenes obtenidas por la tecnología digital automática — una cámara conectada a un microscopio para capturar

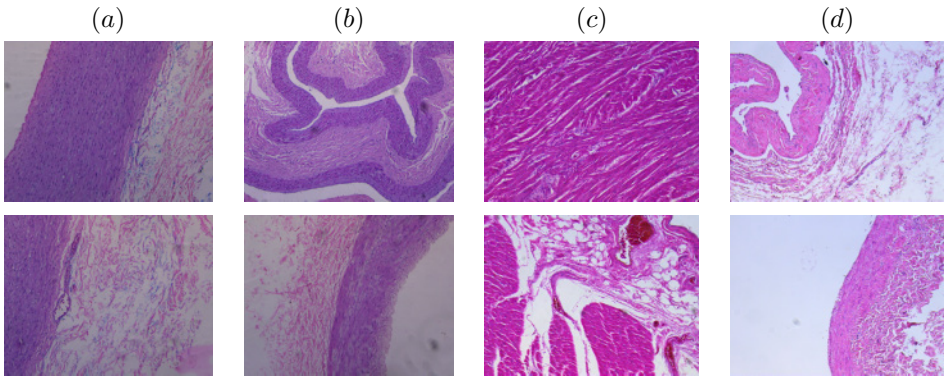


Figura 15: Patrones en diferentes órganos del sistema cardiovascular. (a) Arteria carótida. (b) Arteria coronaria. (c) Corazón. (d) Vena de gran calibre.

imágenes — disponibles en hospitales o distribuidos a través de dispositivos de almacenamiento de histólogos. El etiquetado automático soluciona algunos problemas presentes en anotación manual como la subjetividad, los costos de tiempo y dificultad (Hernandez et al., 1990). Por último, las tres aplicaciones antes mencionadas son tareas de gran relevancia que plantean retos actuales de problemas de visión por computador. Estas dos razones fundamentales llevaron a la selección de estas propuestas. Un resumen de la motivación de cada aplicación se presenta a continuación.

1.1.1 Reconocimiento y Clasificación de los Tejidos Fundamentales y los Órganos

En histología, un órgano tiene estructuras espaciales complejas. La Figura 15 muestra las estructuras espaciales de cuatro órganos diferentes: la arteria carótida, la arteria coronaria, el corazón y la vena de gran calibre. Se puede observar cómo los tejidos fundamentales interactúan para formar patrones específicos y distintivos en cada órgano, además es importante tener en cuenta las diferencias observadas en un mismo órgano por el cambio de la zona de captura o de la muestra. Las Figuras 1.1(a) y 1.1(b) muestran algunos patrones específicos en las arterias separadas en tres capas bien definidas: una capa delgada cercana a la región de luz, una capa compacta y sucinta que es más gruesa en medio de las otras dos capas y, finalmente, una capa porosa con estructuras más separados. La Figura 1.1(c) muestra una estructura homogénea y compacta presente en el corazón. La Figura 1.1(d) muestra algunos patrones específicos de la vena de gran calibre, la cual se separa en tres capas similares a las que presentan las arterias con dos diferencias claramente definidas: la capa de en medio es más delgada y la capa porosa es más gruesa.

Algunos métodos han sido utilizados para proponer diversas soluciones en el reconocimiento de células, tejidos y órganos tales como: estrategias de aprendizaje supervisado, características de textura Zhao et al. (2005), procesos de segmentación Herve et al. (2011) y modelos de Markov Yu et al. (2008). Sin embargo, la principal diferencia, según nuestro conocimiento, es que los tejidos fundamentales sanos no han sido clasificados y no hay trabajos

centrados en el sistema cardiovascular. El reconocimiento de tejidos fundamentales y órganos, asistido por ordenador, presenta diversos problemas, ello debido a la íntima relación entre si y a la baja definición en los bordes de los tejidos y órganos. En este trabajo, nosotros presentamos dos enfoques diferentes para el reconocimiento y clasificación de los tejidos fundamentales, utilizando técnicas basadas en el procesamiento de imágenes, características de textura y algoritmos de aprendizaje automático para el reconocimiento de patrones.

1.1.2 Mejora de la Clasificación Automática de Imágenes Histológicas usando una Ontología Histológica del Sistema Cardiovascular Humano

La representación del conocimiento permite describir la información del mundo real, de manera que un sistema informático puede utilizarla para resolver tareas complejas. En este caso, es necesario considerar cómo los seres humanos solucionan problemas y representan el conocimiento con el objetivo final de formalizarlo; una forma de lograr este objetivo es la creación de una ontología histológica. Está ontología permitiría una comprensión común entre las personas con sistemas informáticos y la reutilización y análisis del conocimiento histológico y de expertos — supuestos, deducción y razonamiento. La construcción de una ontología histológica es una tarea compleja que requiere la colaboración de ingenieros y expertos en el dominio, debido a la complejidad del ámbito médico y los lenguajes de descripción formal.

Por otro lado, las fuentes de datos heterogéneas producen diferentes tipos o representaciones de datos que no se pueden operar o tratar de la misma manera. En este trabajo, tenemos una ontología histológica y una clasificación automática de los tejidos fundamentales y órganos del sistema cardiovascular. El objetivo principal es utilizar la ontología histológica para mejorar el proceso de clasificación automática y lograr la reducción de los márgenes de error o incertidumbre obtenidos cuando se utiliza cada fuente de información de manera independiente. Este trabajo constituye una nueva forma de realizar integración de múltiples fuentes heterogéneas de información en el contexto histológico con el fin de identificar patrones e inferir la organización de los tejidos fundamentales y la identificación de los órganos y demás estructuras presentes en una muestra. Un diagrama para el enfoque del proceso de identificación se presenta en la Figura 16.

1.2 Objetivos

El principal objetivo de este trabajo es proponer y evaluar un método para la clasificación automática de los tejidos fundamentales y modelar el conocimiento histológico y de expertos para el sistema cardiovascular humano.

Dado dicho objetivo general, se definen los siguientes objetivos especificos:

1. Proponer un método de reconocimiento automático de los tejidos fundamentales utilizando técnicas de procesamiento de imágenes e información morfológica del tejido.
2. Proponer un método de clasificación automático de los tejidos y los órganos cardiovasculares basado en bloques, utilizando descriptores de textura y algoritmos de aprendizaje automático.
3. Proponer una representación del conocimiento histológico y de los experto usando modelos que capturen las características esenciales de las relaciones y su utilidad.

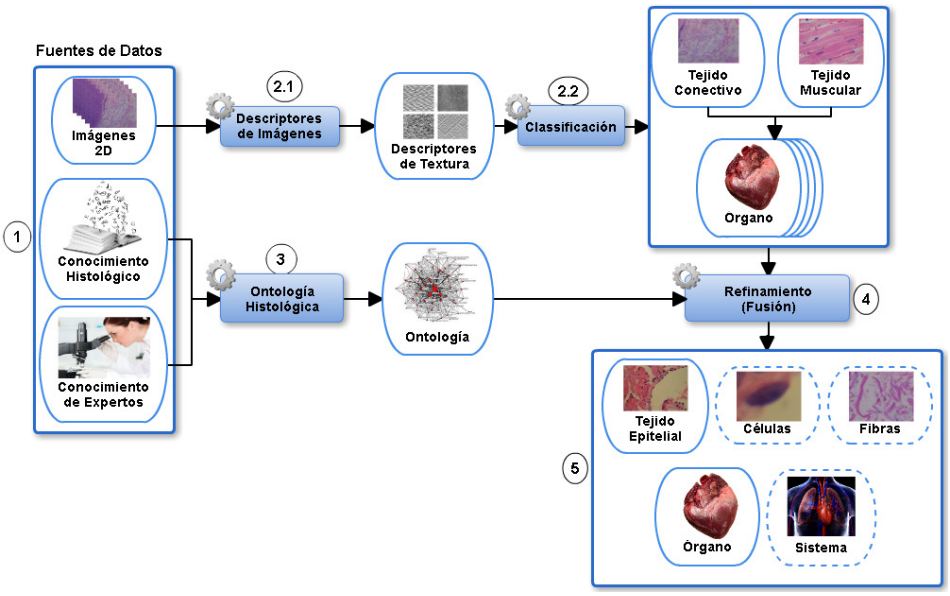


Figura 16: Proceso de mejora de la clasificación automática de imágenes histológicas usando una ontología del sistema cardiovascular humano. (1) Fuente de datos. (2.1) Proceso de descripción de la imagen usando características de textura. (2.2) Proceso de clasificación automática de los tejidos fundamentales y de órganos. (3) Proceso para la construcción de una ontología histológica. (4) Procesos de mejora y clasificación del tejido epitelial. (5) Clasificación final.

4. Combinar la clasificación automática basada en textura con la ontología histológica para mejorar los resultados obtenidos.

1.3 Contribuciones Principales

Las principales contribuciones de esta tesis doctoral pueden resumirse como se indica a continuación:

1. *Un método para el reconocimiento y clasificación automática de los tejidos fundamentales, utilizando la información morfológica*. La clasificación del tejido epitelial — plano, cúbico y cilíndrico —, el reconocimiento de los tejidos conectivo laxo y muscular.
2. *Un método para clasificar los tejidos fundamentales basado en su información de textura*. Este método reconoce el músculo cardíaco del corazón, tejido conectivo laxo — venas, arterias y corazón — y el músculo liso de la arteria muscular, de la vena de gran calibre y de la arteria elástica, con tasas de aciertos superiores al 90 %.
3. *Una ontología histológica del sistema cardiovascular humano*.
4. *Un proceso de mejora de la clasificación automática, utilizando el conocimiento contenido en la*

ontología histológica creada. Esta propuesta aumenta las tasas de aciertos de la clasificación automática.

1.4 Organización del Resto del Documento

En el Capítulo 2 se realiza un breve resumen de la revisión del estado del arte. En el Capítulo 3 se presentan los métodos para el reconocimiento y clasificación automática de los tejidos fundamentales, utilizando información morfológica, incluyendo la experimentación y los resultados obtenidos. En el Capítulo 4 se describe una propuesta para la clasificación automática de los tejidos cardiovasculares basada en información de textura y un clasificador SVM en cascada. En el Capítulo 5 se presentan: (i) una ontología histológica humana para representar el conocimiento histológico y de los expertos; (ii) un proceso de mejora de la clasificación automática utilizando la ontología histológica. Finalmente, en el Capítulo 6 se exponen las conclusiones de esta tesis doctoral y las líneas futuras de trabajo.

2 Revisión del Estado del Arte

En las últimas décadas ha habido un trabajo sustancial en el campo de la visión computacional, que aborda el problema de reconocimiento de los tejidos fundamentales y órganos en imágenes histológicas. Ademas, la representación computacional del conocimiento ha tenido un gran auge con el fin de obtener soluciones más completas y consistentes.

2.1 Selección de Características y Algortimos de Aprendizaje aplicado en Imágenes Histologicas

En esta subsección se presenta una breve descripción de las diferentes características y algoritmos de aprendizaje automático de imágenes histológicas.

La extracción de características y la clasificación son dos pasos importantes en las técnicas de procesamiento de imágenes, debido a que son una forma de entender la percepción humana. Gurcan et al. (2009) presenta algunas de las características más utilizadas en el análisis de imágenes de histopatología tales como: funciones elípticas, forma, textura, dimensión fractal y las características Wavelet, entre otras. Además, presenta algoritmos heurísticos que se han desarrollado para mejorar la precisión de la clasificación tales como: la selección secuencial hacia adelante (SFS) (Cateni and Colla, 2015) y la selección secuencial hacia atrás (SBS) (Pudil et al., 1994). SFS funciona añadiendo secuencialmente las características que mejoran la clasificación; mientras, SBS comienza con todo el conjunto de características y elimina de forma secuencial características buscando mejorar la clasificación. Después de la selección de características, en algunos trabajos, se realiza un proceso de reducción de dimensionalidad, para estos casos Gurcan et al. (2009) evalúa tres métodos bien conocidos: análisis de componentes principales (PCA), análisis de componentes independientes (ICA) y el análisis discriminante lineal (LDA).

Caicedo (2011) presenta un breve resumen de los principales algoritmos y estrategias para la extracción de características en imágenes de histología. Esta revisión de la literatura cubre imágenes histológicas de tejidos como cervical, piel, tracto gastrointestinal, neural

y próstatico. Este documento se divide en seis partes que desarrollan el método de análisis e identificación de imágenes histológicas: segmentación de la imagen, extracción de características (color y textura), arquitectura, morfología, transformación y representación de la imagen (algoritmos de aprendizaje). Sin embargo, este trabajo no cubre imágenes histológicas del sistema cardiovascular, además no se detalla la manera en que se utiliza la información morfológica como parte del procesamiento de imágenes.

Por otro lado, trabajos como Cruz et al. (2011) utilizan características de color, textura y bordes con el objetivo de realizar recuperación por contenido de tejidos en imágenes histológicas. Los resultados obtenidos muestran precisiones en la recuperación entre $67\,\%$ y $80\,\%$. Orlov et al. (2009) usa descriptores basados en coeficientes polinómicos (Chebyshev, Chebyshev-Fourier y Zernike), textura (Haralick, Gabor, y Tamura) y un histograma multiescala para la identificación de dos conjuntos de imágenes: tumores malignos en biopsias de ganglios linfáticos y tumores benigno de melanoma. Los resultados experimentales arrojaron la tasa de clasificación más alta para el primer conjunto de imágenes con un $97\,\%$ y el segundo conjunto con un $93\,\%$. Sin embargo, aunque se conoce el porcentaje de exactitud, no se dan detalles sobre del conjunto en el que se realizaron las pruebas. Meng et al. (2010) propone una aplicación con un modelo para la clasificación de imágenes de histología en general utilizando clasificación supervisada. Este modelo está construido con un vector de características de tamaño 505, el cual incluye color (color dominante, histograma de color y momento en el color) y textura (histograma de bordes, co-ocurrencia, textura Wavelet, textura Tamura, textura Gabor y LBP). Para los experimentos y resultados utilizó validación cruzada con tres conjuntos de datos de referencia de muestras histológicas superando a otros clasificadores conocidos con $92\,\%$ en comparación con resultados reportados de $85\,\%$. Sin embargo, los conjuntos de datos utilizados tienen similitudes entre ellos, lo cual no da una buena idea del comportamiento en conjuntos de datos con mayor variabilidad.

Caicedo et al. (2011) propone un sistema de anotación y recuperación de imágenes histológicas de consulta por ejemplo o por conceptos semánticos para el diagnóstico de un tipo particular de cáncer conocido como carcinoma de células basales. El sistema propuesto contiene tres etapas: la extracción de características (histograma de escala de grises, histograma de características invariantes, LBP, histograma de color RGB, SIFT e histograma de Tamura), la combinación de características con métodos basados en *kernel* (SVM) y la anotación automática de la imagen. Finalmente, se determina que las características de mayor poder discriminativo son LBP y Tamura, con porcentajes de exactitud que oscilan entre $44\,\%$ y $77\,\%$ dependiendo del término evaluado. Canada et al. (2011) propone un sistema de recuperación por contenido para la anotación automática de imágenes con alteraciones histológicas en el ojo de larvas del pez cebra. La imagen se divide en bloques de 64×64 píxeles, con el fin de identificar anomalías presentes, se extraen un total de 54 características para cada bloque de imagen compuesto de: matriz de co-ocurrencia, Lacunarity, características morfológicas en niveles de grises, modelo de Markov y Daubechies Wavelet. Las pruebas se realizaron con un conjunto de 176 imágenes, de las cuales se utilizaron 100 para el entrenamiento y 76 para las pruebas. Los resultados de la aplicación se evalúan usando las etiquetas *correcta, incorrecta o aceptable*. En diferentes tipos de pruebas los resultados sobre los conteos obtenidos muestran una precisión entre $62\,\%$ y $98\,\%$ para *aceptables*. Aunque es importante señalar que el proceso de formación puede ser sensible al número de imágenes utilizadas. En este estudio el número

de imágenes puede no ser suficiente para mantener o aumentar la tasa de propagación de los modelos.

2.2 Reconocimiento de Tejidos Fundamentales

Los trabajos existentes en la literatura para el reconocimiento de tejidos fundamentales sanos o con patologías se han abordado con diferentes técnicas. En Chen et al. (2011) se utiliza la intensidad de los píxeles dentro de una vecindad para asignar una clase a cada píxel de una imagen usando una estrategia de aprendizaje supervisado. El reconocimiento en imágenes de resonancia magnética e imágenes microscópicas de histopatología, de este trabajo, permite identificar huesos, cartílagos, grasa, y separar el fondo. El reconocimiento de teratoma en imágenes tumorales obtuvo una precisión entre $59\,\%$ y $91\,\%$, donde el color representa información útil para la clasificación de ciertos elementos en las imágenes histológicas. A pesar de ello, un método que tiene en cuenta únicamente la información de color podría tener fallos debido a que una muestra histológica puede variar según la tinción utilizada y el proceso de laboratorio para su obtención.

Algunos trabajos utilizan características de textura para identificar los tejidos fundamentales. Por ejemplo, Diamond et al. (2004b) propone un método utilizando características de textura de Haralick para identificar la composición del tejido de neoplasia prostática y clasificarlos como normal, estroma o adenocarcinoma de próstata. La evaluación se realizó con un conjunto de subregiones de tamaño 100×100 píxeles, pertenecientes a 12 casos de los cuales cuatro son para entrenamiento y ocho para pruebas, y se obtuvo un promedio de $79{,}3\,\%$ subregiones correctamente clasificadas. Sin embargo, el número de casos es reducido y el principal enfoque del método es la identificación de una patología en particular. Simsek et al. (2012) describe una propuesta basada en la frecuencia de co-ocurrencias para realizar una segmentación no supervisada de regiones de tejido canceroso del colon. En primer lugar, introducen un nuevo conjunto de características de textura de alto nivel para representar el conocimiento previo de una relación espacial particular de los componentes del tejido. En segúndo lugar, obtienen múltiples segmentaciones con un particionamiento de varios niveles. Finalmente, utilizan el algoritmo K-means para definir los objetos de los tejidos y la matriz de co-ocurrencias sobre estos objetos. Los resultados experimentales en un conjunto de 200 imágenes de tejido de colon obtuvieron un F-score mayor de $90\,\%$. Sin embargo, este trabajo se aplica específicamente a cáncer de colon en imágenes de tejidos con características particulares.

Dentro de la revisión de literatura realizada no hemos encontrado ningún trabajo que se centre en la segmentación de los tejidos fundamentales sanos o que tengan en cuenta el sistema cardiovascular en particular.

Por otra parte, algunos trabajos de procesamiento de imágenes en histopatología reconocen los tejidos fundamentales indirectamente. Dentro de las técnicas utilizadas se encuentran descriptores de texturas (Chen et al., 2012), Máquinas de Soporte Vectorial (SVM) (Chen et al., 2012), procesos de segmentación para la detección de glándulas utilizando umbrales de color, la posición de las células y el contorno (Ficsor and Molnar, 2009), entre otros. Sin embargo, este tipo de resultados indirectos no son exactos, ya que los autores están interesados en obtener otro tipo de información de las imágenes procesadas por ejemplo glándulas o reconocimiento

del tejido canceroso.

El reconocimiento del tejido epitelial requiere del análisis de la morfología de las células para identificar el tipo de epitelio de una muestra histológica y para detectar diferentes patologías. Por lo anterior, en el estado del arte se han dedicado muchos esfuerzos para el reconocimiento de los núcleos de las células, ya que es una pieza clave para la determinación de las estructuras biológicas.

En cuanto a la segmentación de los núcleos de las células, muchos trabajos emplean enfoques de técnicas de agrupamiento y descriptores de imágenes (Song et al., 2013; Lou et al., 2012). Por ejemplo, en Tonkin et al. (2011) segmentan regiones epiteliales de imágenes utilizando un algoritmo basado en cortes de grafos binarios, teniendo en cuenta las probabilidades obtenidas a partir del histograma de color. Se utilizó un conjunto de imágenes de entrenamiento de 38 y uno de prueba de 35, que contenían muestras de cuatro tipos de quistes odontogénicos. Los resultados arrojan una sensibilidad de $91{,}5 \pm 17\,\%$ y una especificidad de $85{,}1 \pm 18{,}6\,\%$. Los dentígeros y queratoquistes odontogénicos arrojaron una sensibilidad de $91{,}9 \pm 6{,}15\,\%$ y $96{,}1 \pm 1{,}98\,\%$ y una especificidad de $97{,}4 \pm 2{,}15\,\%$ y $98{,}7 \pm 3{,}16\,\%$, respectivamente. Este método se puede emplear en condiciones patológicas con tejidos similares, tales como la piel y la mucosa, en las cuales existe una distinción clara entre el tejido epitelial y conectivo. Kong et al. (2009b) también propuso un sistema de pronóstico asistido por ordenador para el neuroblastoma, un tipo de cáncer del sistema nervioso, el cual clasifica las muestras en *favorables o desfavorables* basado en la morfología de los tejidos. En este trabajo se utiliza la información de textura por medio de estadísticas de la matriz de co-ocurrencias y LBP, además realizan una modificación al clasificador K-nn. Los resultados fueron obtenidos sobre un conjunto de 43 muestras de tejidos, 32 muestras identificadas con estroma pobre y las muestras restantes fueron identificadas con estroma rico, presentando una precisión general de clasificación de $88{,}4\,\%$. Sin embargo, este método se aplica en un estado patológico específico.

Otra forma de reconocer los núcleos de las células es utilizando *contornos activos adaptativos*; este es el caso de Zeng et al. (2015) donde se propone un método basado en un modelado de *contornos activos adaptativos* para segmentar los núcleos de las células a partir de imágenes de frotis cervicales. Los resultados fueron evaluados teniendo en cuenta las tasas de verdaderos positivos y *Differential Scanning Calorimetry* (DSC) con los cuales se obtuvo una media de $0{,}87$ y $0{,}85$, respectivamente. No obstante, este método fue evaluado únicamente con imágenes de frotis cervicales y no proporciona información acerca de la forma de los núcleos celulares. Por otro lado, las técnicas de aprendizaje automático también se han utilizado en el reconocimiento de núcleos celulares, por ejemplo Han et al. (2012a) presentan un enfoque utilizando SVM y características de borde de Laplace. En esta propuesta se identifican los fibroblastos NIH/3T3 en un conjunto de 75 imágenes de muestras con Hematoxilina Eosina y 25 imágenes de muestras con Hematoxilina únicamente. Finalmente se obtuvo una tasa promedio de detección por encima del $90\,\%$. Sin embargo, en este trabajo se utilizó un solo tipo de núcleos de células y esto puede afectar la selección de características y los resultados obtenidos.

Todos los trabajos mencionados usan diferentes tinciones y tipos de imágenes, además sus objetivos se centran en la segmentación de los núcleos teniendo en cuenta el área, contornos o conteo de células.

2.3 Identificación de Órganos

Un órgano puede identificarse utilizando el reconocimiento de los tejidos fundamentales y algunos criterios de análisis como la ubicación, el tipo y las características especiales de acuerdo con su función o morfología. El estudio y el reconocimiento automático de patrones, que hacen único a cada órgano, es un problema abierto debido a que un órgano posee estructuras complejas.

Algunos trabajos abordan la identificación de órganos utilizando características de color y textura (Herve et al., 2011), por ejemplo Zhao et al. (2005) proponen un modelo estadístico utilizando características de textura de Gabor para identificar diez órganos humanos — glándula suprarrenal, corazón, riñón, hígado, pulmón, páncreas, bazo, testículos, tiroides y útero. La validación del método se realizó con 778 imágenes histológicas obteniendo una tasa de precisión entre $44\,\%$ y $93\,\%$ que varía en función del órgano a identificar. Sin embargo, este trabajo tiene una gran variabilidad de los porcentajes de exactitud según el órgano identificado; el corazón es el único órgano del sistema cardiovascular con una precisión de $80\,\%$. Una propuesta similar es presentada en Yu et al. (2008) donde se busca identificar cinco órganos del tracto gastrointestinal — esófago, estómago, intestino delgado, intestino grueso y ano. En este trabajo se presenta un nuevo método estocástico en 2D para el análisis semántico del contenido de imágenes histológicas, llamado *Hidden Markov Model*, usando un enfoque de clasificación por bloques de tamaño 64×64 píxeles. Finalmente, se obtiene un vector característico que contiene la medida de energía de Gabor y el valor promedio de grises. Los resultados obtenidos muestran una precisión entre $59\,\%$ y $82\,\%$ que varía según el órgano identificado. No obstante, este trabajo tiene una alta complejidad computacional y el conjunto de datos es de 200 imágenes histológicas, 40 imágenes por cada órgano, siendo un conjunto de datos limitado que puede afectar los resultados obtenidos.

2.4 Ontología de Histología Humana

Muchas ontologías y taxonomías están disponibles en formato electrónico con licencias de código abierto. Algunas de las taxonomías médicas más conocidas son: Galeno (conceptos clínicos básicos, relaciones y conceptos complejos), UMLS (*Unified Medical Language System*), MeSH (*Medical Subject Heading*), *Kingsbury Center for Cancer Care Glossary*, MedicineNet diccionario médico, Glosario de Términos Médicos Técnicos y Populares, CD (*International Classification of Diseases*), entre otros (Vasquez et al., 2010).

Se realizó una investigación y análisis de diferentes ontologías médicas teniendo en cuenta los enfoques histológicos y anatómicos desde los cuales se identifican partes que se pueden reutilizar para nuestro propósito (Rubin et al., 2008). BioPortal (2005) contiene algunos términos histológicos, sin embargo esta ontología tiene una orientación diferente a nuestra investigación que se ve reflejada en el orden específico de los términos, puesto que algunos están situados al azar. BioPortal (2008b) y BioPortal (2008a) tienen términos histológicos similares a los requeridos en nuestra investigación, un ejemplo son las instancias relacionadas al tejido epitelial. A pesar de ello, la unión y relación de estos conceptos están diseñadas por rutas diferentes teniendo en cuenta los vasos sanguíneos. Estas ontologías contienen una gran cantidad de conceptos que no se describen en detalle, dejando algunos caminos sin concluir en algunas instancias, como es el caso del tejido muscular. Adicionalmente, los conceptos están

vinculados unidireccionalmente permitiendo conectar sólo estructuras grandes a pequeñas, pero no al contrario. Por otra parte, algunas relaciones que permiten llegar a un concepto no son intuitivas o lógicas, lo cual se ve reflejado en que el usuario debe hacer un mayor esfuerzo para identificar la ruta o tener conocimientos más amplios para encontrar posibles rutas y hallar un término. BioPortal (2014c) contiene el sistema cardiovascular y sus órganos, siendo una ontología completa y muy cercana al objetivo de nuestra investigación. No obstante, carece de términos relevantes para nuestro propósito y tiene grandes deficiencias con respecto a los tejidos fundamentales. De forma similar, BioPortal (2014b) contiene conceptos similares a los requeridos en nuestra investigación como algunas celulas y tejidos. Sin embargo, esta es una ontología histopatológica humana, la cual contiene conceptos de enfermedades o anormalidades. Además, esta ontología no contiene los órganos del sistema cardiovascular y en la clasificación de los tejidos considera otro enfoque particular — retinal, mamaria, uretra, entre otros — y algunos términos pueden ser referenciados como conceptos individuales. BioPortal (2014a) es una ontología de ratón con un enfoque de anatomía macroscópica adulta, por esta razón no contiene términos microscópicos ni un orden histológico. Esta ontología incluye algunos términos similares de los órganos y sistemas que pueden ser referenciadas como conceptos individuales en nuestra ontología.

2.5 Clasificación basada en Ontología

Las ontologías y taxonomías contienen conocimientos representados con información estructural y semántica. Los enfoques que utilizan estas herramientas en el proceso de clasificación automática se dividen en dos: (i) modelar la relación entre la información visual y semántica (Wu et al., 2010; Yang et al., 2007) y (ii) utilizar estas ontologías y taxonomías en el algoritmo de clasificación (Othmani et al., 2010; Smith et al., 2015; Abdollahpour et al., 2015; Paulson et al., 2006; Breen et al., 2002). Nosotros nos centraremos en el segundo grupo para llevar a cabo el proceso de clasificación, utilizando imágenes y ontologías.

En Othmani et al. (2010) se propone utilizar el razonamiento de alto nivel y la formalización del conocimiento del software basado en ontologías para hacer la anotación de imágenes más eficiente e interactiva. El objetivo principal del procesamiento de imágenes de bajo nivel es describir los objetos biológicos generales — los núcleos, el lumen y las áreas invasivas — en las imágenes histopatológicas. En esta etapa se utilizan umbralización, operaciones morfológicas y un método basado en contornos. Posteriormente, utilizan una ontología anatómica para mejorar la tasa de especificidad y sensibilidad usando un lenguaje de consulta SPARQL. Los resultados muestran que el método propuesto detecta todas las mitosis, pero la detección contiene muchos falsos positivos. Además, el algoritmo con restricciones geométricas es más específico, pero disminuye la sensibilidad. Sin embargo, este trabajo se centra en imágenes histopatológicas y utiliza una ontología anatómica. Smith et al. (2015) proporciona un estudio de ontologías de imágenes biomédicas, actualmente en desarrollo. En esta revisión se describen los retos particularmente enfrentados de las ontologías en los campos de imágenes histopatológicas y de el análisis de imágenes. Además, sugiere una estrategia para abordar estos desafíos. Este trabajo presenta el uso de ontologías en los procesos de anotación, investigación y bancos biológicos. Por otra parte, realiza una revisión crítica de las principales contribuciones que se han aportado a las ontologías y estándares relacionados con ellas en el

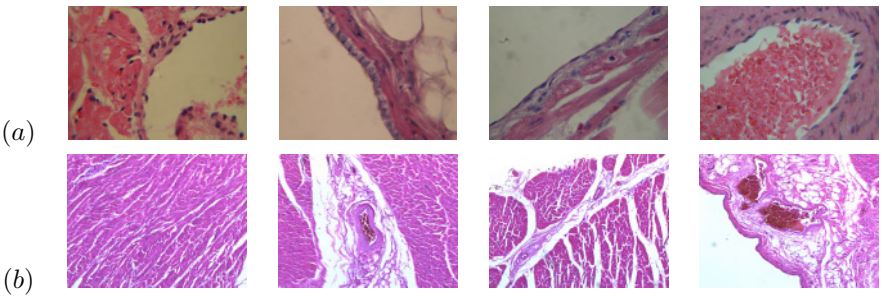


Figura 17: Imágenes histológicas. (a) Objetivo $40\times$. (b) Objetivo $10\times$.

dominio de imágenes. Ademas, propone una iniciativa, teniendo en cuenta lo siguiente elementos: la adquisición de imágenes, el protocolo de muestras, parámetros de procesamiento de imágenes y los niveles de organización. Por último, se presenta un ejemplo de una ontología basada en histopatología cuantitativa de imágenes. A pesar de ello, esta revisión se enfoca en ontologías basadas en imágenes histopatológicas.

3 Reconocimiento de Tejidos Fundamentales

3.1 Conjunto de Imágenes

Se han utilizado muestras histológicas de diferentes órganos siguiendo un protocolo de laboratorio para controlar el proceso de tinción con Hematoxilina Eosina y Tricrómica de Masson. Se definió un protocolo de captura de imágenes con el fin de reducir errores en el reconocimiento automático, teniendo en cuenta algunas características como la configuración del microscopio, configuración del software, manipulación de la muestra y captura de las imágenes. Hemos creado y publicado un conjunto de 400 imágenes histológicas de diferentes órganos y personas, 300 adquiridas usando un objetivo de $40\times$ — 100 por cada tipo de tejido epitelial — y 100 usando uno de $10\times$[2]. Nosotros usamos las imágenes pertenecientes al proyecto *Desarrollo del Banco de Imágenes Histológicas sobre el Sistema Cardiovascular* (BISCAR), CI-2714 Vicerectoria de Investigación de la Universidad del Valle. Las imágenes fueron obtenidas con un microscopio Leica *DM750-M* con una resolución de 2048×1536 píxeles y se almacenaron en formato *PNG*. El microscopio tiene un objetivo ocular con un factor de aumento de $10\times$ y un campo de visión de 20, obteniendo magnificaciones de 400 y 100 para los objetivos de $40\times$ y $10\times$, respectivamente.

El equipo de expertos de histología está compuesto por seis miembros del grupo de investigación Teblami de la Universidad del Valle. Los algoritmos fueron implementados en *C++*, usando la librería *CImg* en un computador con 4 núcleos de procesamiento y *4Gb RAM*[3]. La Figura 17 muestra ejemplos de imágenes histológicas usando los objetivos de $40\times$ and $10\times$.

[2]El conjunto de imágenes está disponible en `http://biscar.univalle.edu.co/?page_id=1003`

[3]La aplicación de escritorio está disponible en `http://biscar.univalle.edu.co/?page_id=1049`

3.1.1 Motivación

Cada tejido fundamental tiene patrones únicos que nos permiten identificarlo y diferenciarlo de los demás, por ejemplo el tejido epitelial de revestimiento se encuentra ubicado en dos lugares específicos, la epidermis y el lumen — la región interior de los órganos tubulares —, por esta razón siempre se encuentra cerca de las regiones de luz. Tres tipos del tejido epitelial — plano, cúbico y cilíndrico — se identifican teniendo en cuenta la forma y posición de los núcleos celulares. Las células epiteliales cúbicas tienen forma de esfera mientras que las células planas y cilíndricas tienen formas de elipse, siendo la posición de las células la manera de diferenciarlas; las células planas son paralelas a las regiones de luz mientras que las células cilíndricas son perpendiculares a las regiones de luz. El tejido muscular tiene una estructura homogénea y compacta, con dirección y organización variable de acuerdo con la región o el corte de la muestra. El tejido conectivo laxo tiene estructuras separadas o dispersas. No obstante, la apariencia de un tejido puede variar incluso en un mismo órgano por el cambio con respecto a la zona de captura, corte o muestra. En cuanto a las características visuales, el color no es discrimiante en las imágenes histológicas, debido a que los cambios en la tinción puede invalidar los resultados. La Figura 18 contiene ejemplos de los tejidos epitalial, muscular y conectivo laxo.

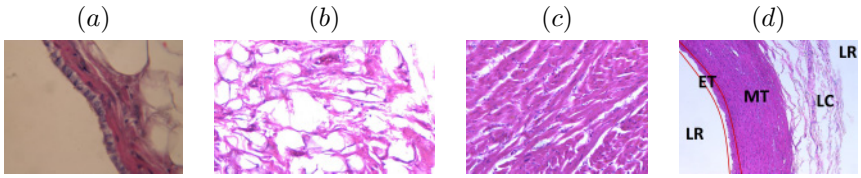


Figura 18: Ejemplos de los tejidos fundamentales. (a) Epitelial. (b) Conectivo laxo. (c) Muscular. (d) Imagen histológica. LR representa regiones de luz; LC representa región de tejido conectivo laxo; MT representa región de tejido muscular; ET representa región de tejido epitelial marcado entre las líneas rojas.

3.1.2 Método

En esta sección se propone un método que utiliza técnicas de procesamiento de imágenes para extraer información de la morfología del tejido — composición, ubicación y las relaciones espaciales — para posteriormente clasificar las imágenes de tejidos con base en las características obtenidas. La Figura 19 ilustra el método propuesto. (1) En la parte izquierda y derecha se presentan imágenes tomadas con objetivos de $40\times$ y $10\times$, respectivamente. (2) En ambos casos, hemos utilizado tres imágenes obtenidas de una otra histológica mediante la extracción del tensor de estructura (Lu et al., 2010) y los canales rojo y verde. (3) Cada posición de píxel se ha prepresetado con un vector de características de tres elementos, con los valores indicados anteriormente. Además, utilizamos el algoritmo K-means (Kanungo et al., 2002) tomando como entrada el conjunto de los vectores de características. (4) Hemos obtenido como resultado tres conjuntos diferentes en cada caso, las regiones negras corresponden

a las zonas segmentadas y los caracteres a y b harán referencia a las imágenes de $40\times$ y $10\times$, respectivamente: (4.1.a) y (4.1.b) corresponden al tejido conectivo laxo que es reconocido inmediatamente después de realizar la agrupación con K-means y, por lo tanto, no requiere procesamiento adicional; (4.2.a) y (4.2.b) corresponden a las regiones de luz; (4.3.a) son los núcleos de las células y (4.3.b) es el tejido muscular. Para reconocer el tejido epitelial, únicamente con imágenes a $40\times$, son necesarios tres pasos adicionales: (i) El reconocimiento de células epiteliales utilizando el algoritmo *Flood-fill* y el tamaño de las regiones en (4.2.a) obteniendo el resultado (6.a). Hemos determinado los píxeles pertenecientes al tejido epitelial en función de la distancia entre las regiones de luz en (4.3.a) y las células obtenidas en los resultados anteriores. (ii) Clasificado de las células epiteliales de acuerdo con la forma y la posición de los núcleos celulares utilizando la proporción de circularidad y la relación de proyección (8.a); (iii) Clasificación del tejido epitelial basado en la frecuencia de los núcleos celulares. Por otra parte, hemos realizado el reconocimiento del tejido muscular en dos pasos adicionales: (i) eliminación de detalles irrelevantes o áreas pequeñas usando erosión y umbraliación en (4.3.b) para obtener (6.b), el cual representa el tejido muscular con células sanguineas; (ii) extracción de las células sanguíneas para obtener el resultado final (8.b).

3.1.3 Experimentos y Resultados

Hemos reconocido los tejidos epitelial — plano, cúbico y cilíndrico —, conectivo laxo y muscular. Seis expertos realizaron la evaluación del método propuesto en dos etapas independientes : (i) evaluación de la clasificación del tejido epitelial teniendo en cuenta las medidas de sensibilidad y especificidad; (ii) evaluación de la clasificación de los tejidos conectivo laxo y muscular utilizando una escala del 1 al 5, donde 5 indica que la identificación es la mejor.

La Tabla 2 muestra los resultados de la evaluación cuantitativa para la clasificación de los tipos de tejido epitelial. Los resultados obtenidos revelan que la tasa de aciertos más alta se logró con la identificación del tejido epitelial plano. Por otro lado, el error más frecuente se presenta cuando el tejido cúbico se clasifica como tejido cilíndrico. Esto ocurre cuando los núcleos celulares están muy cercanos y se evalúan como una sola célula.

La Figura 20 contiene una representación gráfica de la mediana de la evaluación por los expertos considerando cuatro características diferentes sobre la segmentación de los tejidos conectivo laxo y muscular en el conjunto de imágenes de prueba. En la segmentación automática de tejido conjuntivo laxo se evaluaron dos aspectos del enfoque propuesto: (i) la capacidad de reconocer el tejido conectivo laxo y (ii) la capacidad de diferenciar el tejido conectivo laxo del tejido muscular. En la segmentación automática del tejido muscular se evaluaron dos aspectos: (i) la capacidad de reconocer el tejido muscular y (ii) la capacidad de diferenciar el tejido muscular del tejido conectivo laxo. Estos resultados muestran que el tejido conectivo laxo es reconocido a pesar de su íntima relación con el tejido muscular — el tejido conectivo laxo forma capas delgadas que rodean el tejido muscular, siendo en ocasiones difícil de delimitar, incluso manualmente. Además, el reconocimiento del tejido muscular se ve afectado por la alta similitud que tiene con el tejido conectivo denso siendo dificil su diferenciación, incluso manualmente.

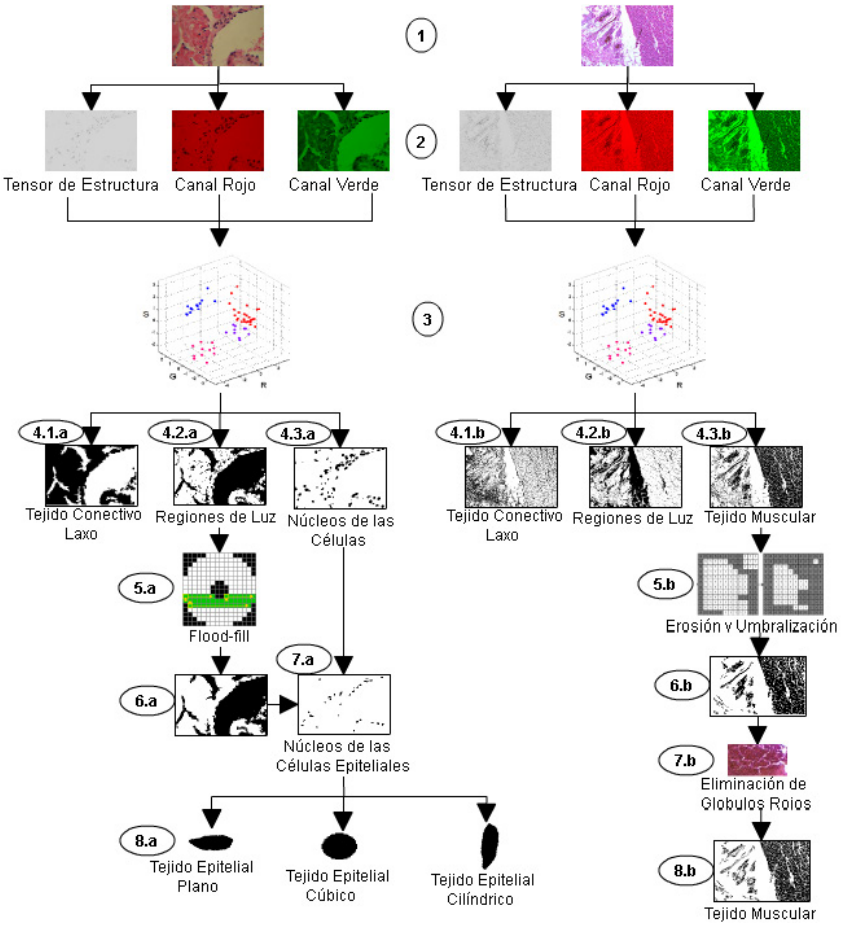


Figura 19: Ilustración del método propuesto para el reconocimiento automático de los tejidos fundamentales. El proceso para imágenes tomadas con los objectivos de $40\times$ y $10\times$ se presentan en la parte izquierda y derecha, respectivamente.

Tabla 2: Evaluación de desempeño de la clasificación del tejido epitelial.

Matriz de Confusión	Plano	Cúbico	Cilíndrico	Total
Verdadero Positivo	31	26	28	85
Falso Positivo	12	8	10	30
Falso Negativo	3	7	5	15
Verdadero Negativo	54	59	57	170
Sensibilidad	0.91	0.79	0.85	0.85
Especificidad	0.81	0.88	0.85	0.85

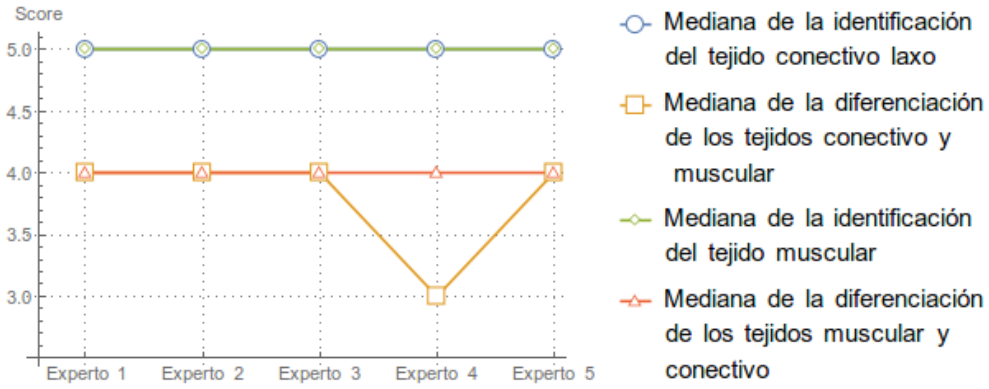


Figura 20: Resultados obtenidos para el reconocimiento de los tejidos conectivo laxo y muscular.

3.2 Clasificación de los Tejidos Cardiovasculares utilizando Descriptores basados en LBP y una SVM en Cascada

3.3 Conjunto de Imágenes

Las muestras de tejido y el protocolo de captura de imágenes se describen en la Subsección 0.3.1. Hemos creado y publicado un conjunto de 1500 bloques, adquiridos usando un objetivo de $10\times$ de imágenes histológicas de diferentes órganos y personas[4]. Los algoritmos fueron implementados en *C++*, usando la librería *CImg* en un computador con 4 núcleos de procesamiento y *4Gb RAM*.

3.3.1 Motivación

Un órgano puede ser identificado conociendo los tejidos que están presentes en una muestra histológica. Por ejemplo, el tejido muscular es similar en diferentes órganos, tiene una estructura homogénea y compacta, con dirección y organización variable de acuerdo con la región y el corte de la muestra; el tejido conectivo laxo tiene estructuras separadas o dispersas. Posibles cambios en la zona de captura, corte o muestra puede generar grandes cambios en la apariencia de un tejido, incluso de un mismo órgano. Los patrones espaciales observados indican que los descriptores de textura pueden proporcionar información relevante para el reconocimiento de los tejidos. La Figura 21 contiene ejemplos de bloques con un único tejido de tamaño 100×100 píxeles. Cada fila muestra patrones específicos de los diferentes tejidos por órganos: (a) tejido muscular cardíaco del corazón, (b) tejido muscular liso de la arteria muscular, (c) tejido muscular liso de la arteria elástica, (d) tejido muscular liso de la vena de gran calibre y (e) tejido conectivo laxo. Igualmente se puede observar que existen similitud intra-clase y diferencias entre clases de los bloques del mismo tejido.

[4]El conjunto de imágenes está disponible en `http://biscar.univalle.edu.co/?page_id=1003`

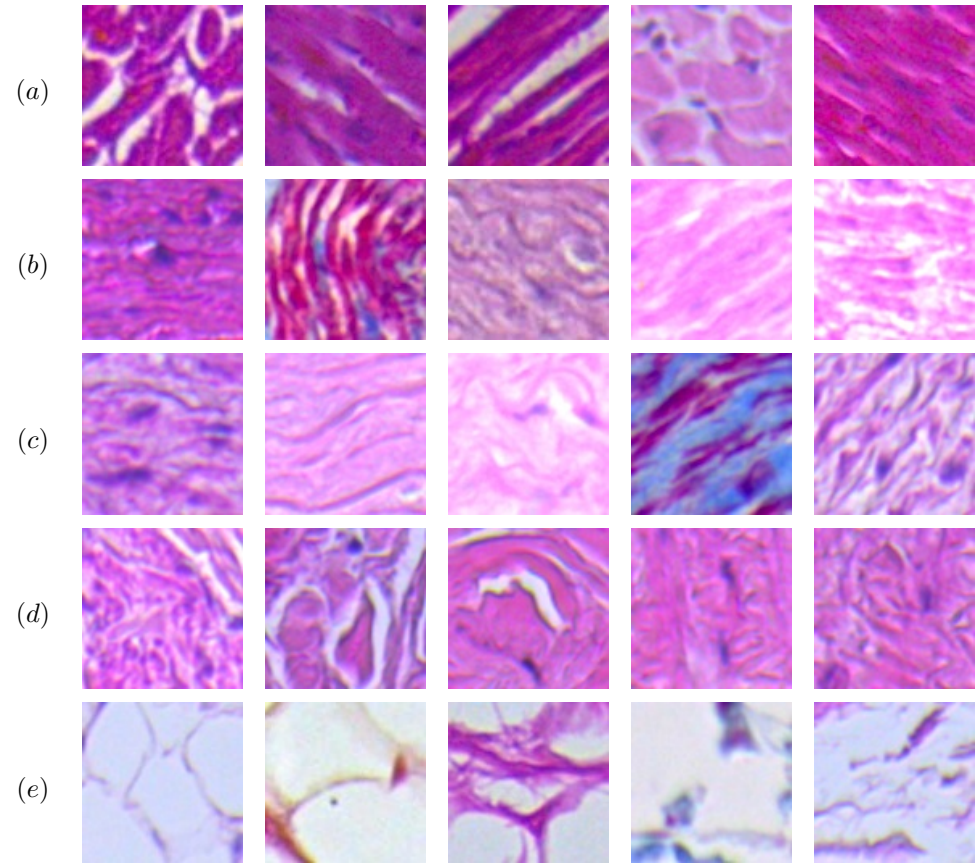


Figura 21: Ejemplos de regiones de interés. (a) Tejido muscular cardíaco del corazón. (b) Tejido muscular liso de la arteria muscular. (c) Tejido de músculo liso de la arteria elástica. (d) Tejido muscular liso de la vena de gran calibre. (e) Tejido conectivo laxo.

3.3.2 Método

La clasificación propuesta se ejecuta en tres pasos fundamentales: (i) dividimos una imagen en bloques; (ii) extraemos la información de cada bloque utilizando el descriptor de textura obtenido de la concatenación de LBP y LBPri; (iii) clasificamos los bloques, utilizando nuestro descriptor de textura con una SVM en cascada.

La Figura 22 muestra un esquema general de nuestra propuesta. 1.) En primer lugar, obtenemos bloques de tamaño 100×100 píxeles a partir de las imágenes histológicas. Un bloque contiene un solo tipo de tejido con información discriminante, haciendo posible el reconocimiento de los tejidos. 2.) Realizamos un proceso de extracción de características para obtener información relevante de cada bloque, después de evaluar diferentes descriptores de

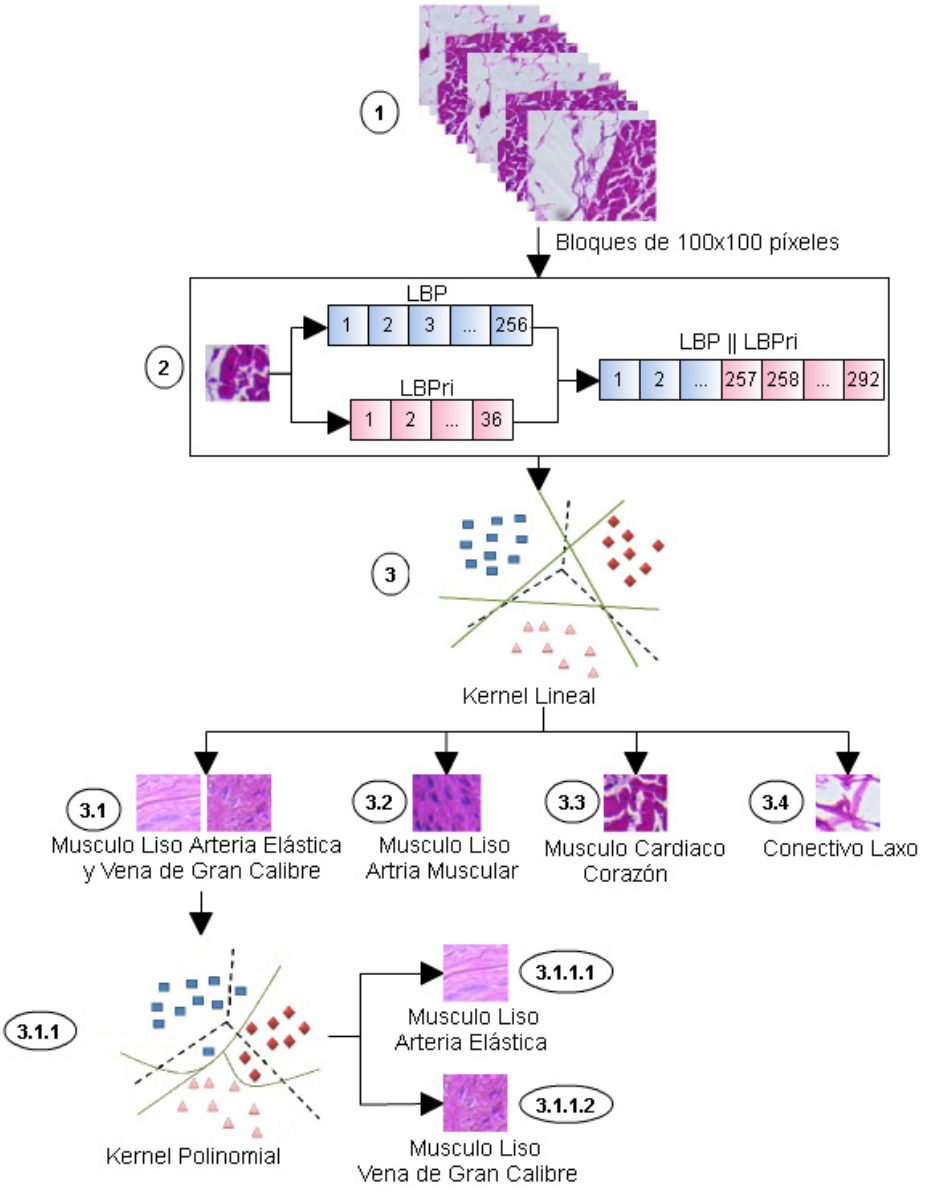


Figura 22: Enfoque propuesto para la clasificación automática de los tejidos fundamentales asociados a un órgano: (1) bloque de tamaño 100×100 píxeles. (2) LBP y LBPri para cada bloque, el vector de características es generado por la concatenación de ambos histogramas. (3) Classificación usando SVM con *kernel* lineal. (3.1), (3.2), (3.3) y (3.4) Bloques clasificados. (3.1.1) Clasificación usando SVM con *kernel* polinomial para separar el primer grupo en (3.1.1.1) y (3.1.1.2).

textura y combinaciones entre ellos. Nosotros proponemos el uso de LBP y LBPri para representar micro-patrones locales de manera eficiente, lo cual es posible gracias a la solidez de LBP y sus variaciones. El proceso de extracción de características genera un vector de tamaño 292, obtenidos mediante la concatenación de LBP (256) y LBPri (36). 3.) Realizamos un proceso de clasificación utilizando una SVM en cascada, debido a que sus resultados superan los resultados obtenidos con SVM (Yang et al., 2012), RF (Bader-El-Den, 2014) y LDA (Ghassabeh et al., 2015). Primero clasificamos los bloques usando una SVM con *kernel* lineal en una de las siguientes cuatro clases: 3.1) la primera clase corresponde al músculo liso de la vena de gran calibre y la arteria elástica; 3.2) la segunda clase es el músculo liso de la arteria muscular; 3.3) la tercera clase es el músculo cardíaco del corazón; y 3.4) la cuarta clase corresponde al tejido conectivo laxo. 3.1.1) Separamos los órganos identificados en la primera clase utilizando una SVM con *kernel* polinomial: 3.1.1.1) corresponde al músculo liso de la arteria elástica y 3.1.1.2) corresponde al músculo liso de la vena de gran calibre.

3.3.3 Experimentos y Resultados

Los resultados obtenidos del proceso de clasificación se evaluaron teniendo en cuenta el número de bloques correctamente clasificadas. La Figura 23 muestra ejemplos de imágenes histológicas con los resultados obtenidos en la clasificación automática, sus aciertos y errores. En los resultados automáticos cada clase es representada con un color distintivo: (i) el músculo cardíaco del corazón corresponde al verde, (ii) el tejido conectivo laxo corresponde al azul, (iii) el músculo liso de la arteria muscular corresponde al violeta, (iv) el músculo liso de la vena de gran calibre corresponde al amarillo, (v) el músculo liso de la arteria elástica corresponde al naranja y (vi) las regiones de luz corresponde al fucsia. Por otro lado, los aciertos y errores se representan con los colores verde y rojo, respectivamente.

La Figura 24 muestra una representación gráfica de las tasas de aciertos y fallos obtenidos con el proceso de clasificación automática, por bloques, en el conjunto de imágenes de prueba. Hemos obtenido entre 211 y 228 bloques correctamente clasificadas por imagen, cada imagen tiene 300 bloques. La precisión obtenida está entre $70{,}333\,\%$ y $76{,}000\,\%$ de acuerdo a la imagen histológica. Cabe resaltar que la mayor precisión se obtiene en el segundo paso de la clasificación en cascada, con, entre 37 a 53 bloques correctamente clasificados y una exactitud de $76{,}812\,\%$ a $77{,}083\,\%$ por imagen. Esto se debe a que esta fase considera sólo dos posibles clases y un menor número de bloques. La precisión obtenida está por encima del $70\,\%$ en comparación con el $90\,\%$ obtenido en las pruebas basadas en bloques, esto ocurre debido a que en una imagen histológica los bloques puede contener más de un tejido.

3.4 Mejora de la Clasificación basada en una Ontología de Histológia Humana

3.5 Conjunto de Imágenes

Se han utilizado muestras histológicas de diferentes órganos con Hematoxilina Eosina; el protocolo de captura de imágenes se describe en la Subsección 0.3.1. Hemos creado y publicado un conjunto de 1500 bloques de cinco imágenes histológicas, adquiridas usando un objeti-

vo de $10\times$[5]. La ontología histológica está implementada en el lenguaje *OWL*, usando Protégé. Los algoritmos fueron implementados en *C++*, usando la librería *CImg* en un computador con 8 núcleos de procesamiento y 8*Gb RAM*.

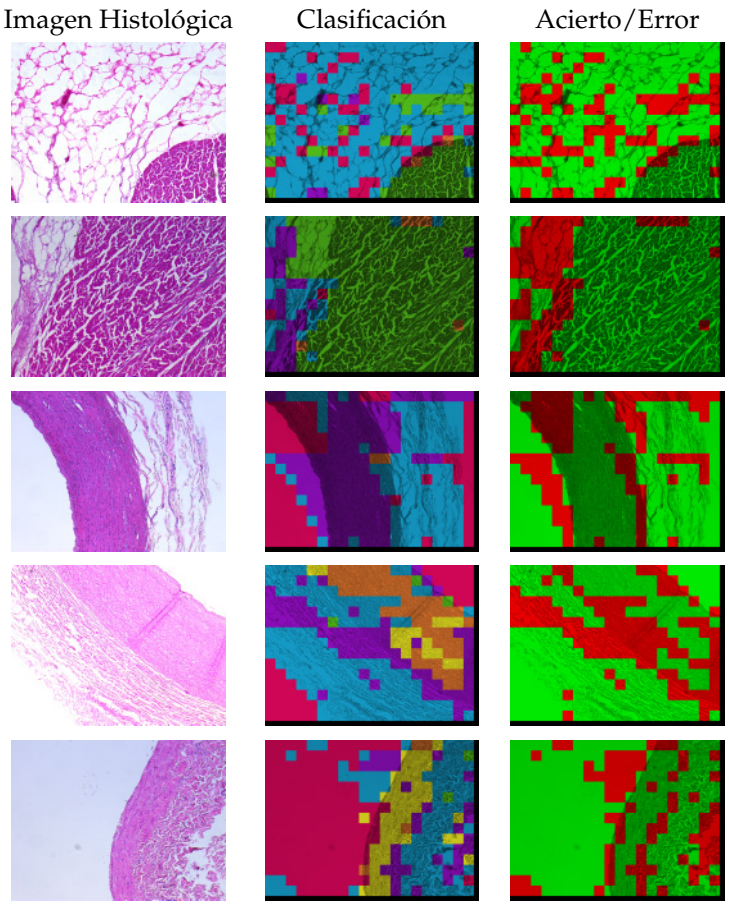


Figura 23: Resultados de la clasificación automática de una imagen histológica. En la primera columna las imágenes histológicas. En cada fila de arriba a abajo: *Img-He* e *Img-He1* representan las imágenes del corazón; *Img-MA* representa la imagen de la arteria muscular; *Img-EA* representa la imagen de la arteria elástica; y *Img-LV* representa la imagen de la vena de gran calibre. En la segunda columna se observa la clasificación automática. En la tercera columna se muestran los aciertos y errores de la clasificación automática.

[5]El conjunto de imágenes está disponible en `http://biscar.univalle.edu.co/?page_id=1003`

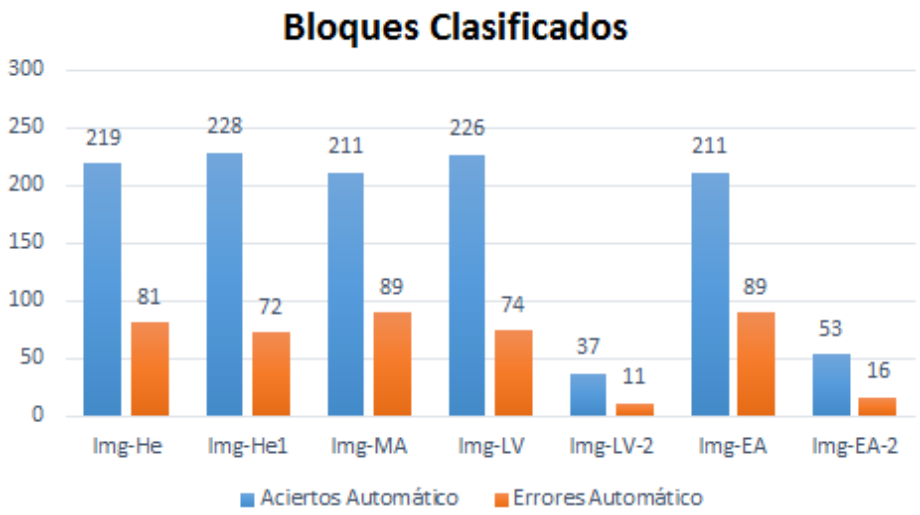


Figura 24: Aciertos y fallos usando el proceso de clasificación automática. *Img-He* e *Img-He1* representan las imágenes del corazón; *Img-MA* representa la imagen de la arteria muscular; *Img-EA* representa la imagen de la arteria elástica; y *Img-LV* representa la imagen de la vena de gran calibre. El nombre de cada imagen con *-2* al final corresponde a la segunda clasificación en el proceso en cascada SVM.

3.5.1 Método

Hemos desarrollado una ontología histológica del sistema cardiovascular humano para la representación del conocimiento histológico y del conocimiento de los expertos, usando una métodología basada en Castro et al. (2006), que tiene en cuenta los siguientes pasos: (i) capturar el conocimiento histológico y de los expertos, (ii) identificar las ontologías reutilizables, (iii) construir los modelos informales de la ontología de manera iterativa, (iv) formalización de la ontología y (v) evaluación. Adicionalmente, hemos propuesto un método de mejora de la clasificación automática de tejidos y órganos usando la ontología histológica para obtener un conocimiento nuevo y más completo que el arrojado con cada una de las fuentes de datos de manera independiente.

La Figura 25 muestra un esquema general de nuestra propuesta que consta de cinco pasos: 1) obtenemos una imagen a partir de una muestra histológica. (2) Realizamos la clasificación mediante técnicas de procesamiento de imágenes. En este caso, utilizamos el método de clasificación propuesto en el Capítulo 4. (3) obtenemos la clasificación de los bloques. (4) Realizamos el proceso de mejora de la clasificación, usando la ontología histológica y la clasificación automática. (4.1) Cálculo de estadísticas de aciertos para cada clase. (4.2) Ontología histológica. (5) Bloques reclasificadas.

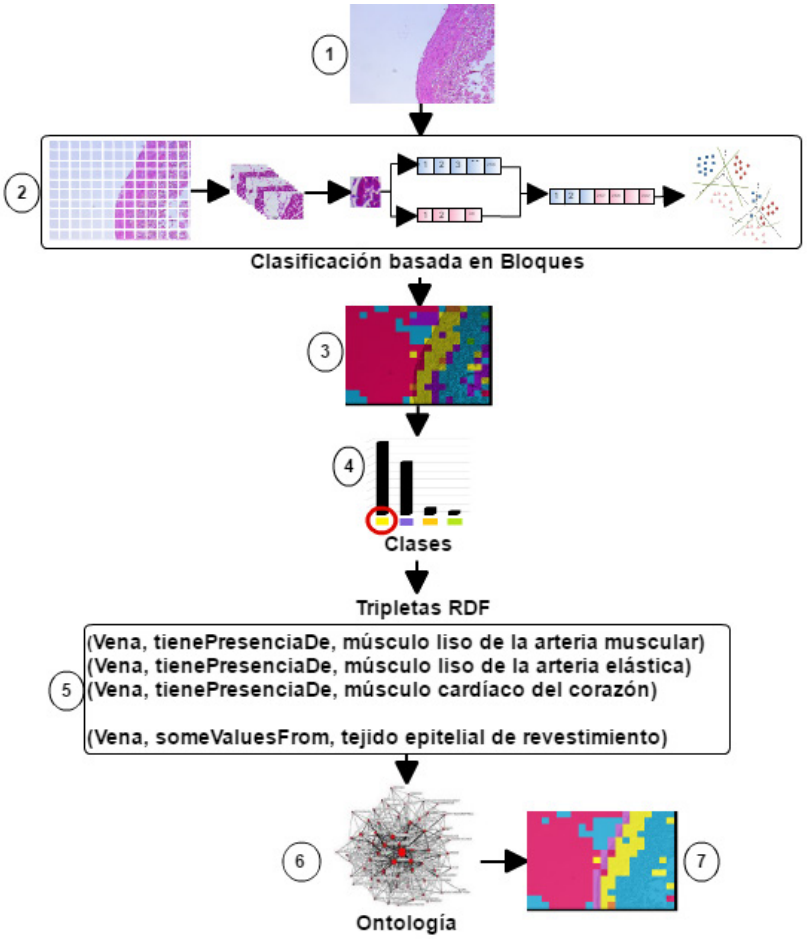
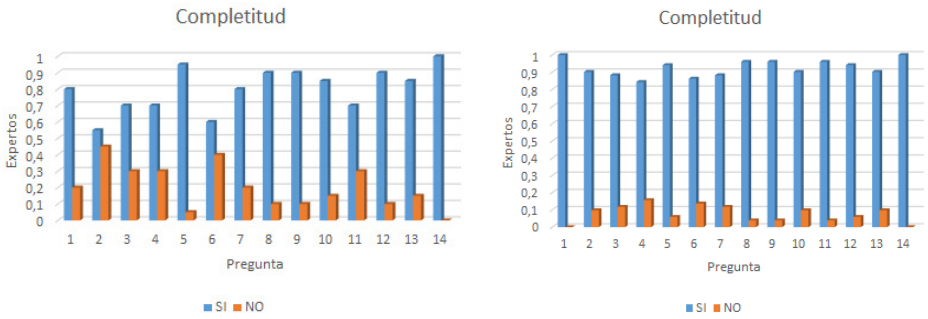


Figura 25: Enfoque propuesto para el proceso de mejora de la clasificación usando una imagen de ejemplo. (1) Imagen histológica. (2) Clasificación basada en características de textura y una SVM en cascada. (3) Bloques clasificados. (4) Ocurrencias para cada clase discriminante. Amarillo representan el músculo liso de la vena de gran calibre, violeta representa el músculo liso de la arteria muscular, naranja representa el músculo liso de la arteria elástica y verde representa el músculo cardíaco del corazón. (5) Tripletas RDF. (6) ontología histológica. (7) Bloques reclasificados.

3.5.2 Experimentos y Resultados

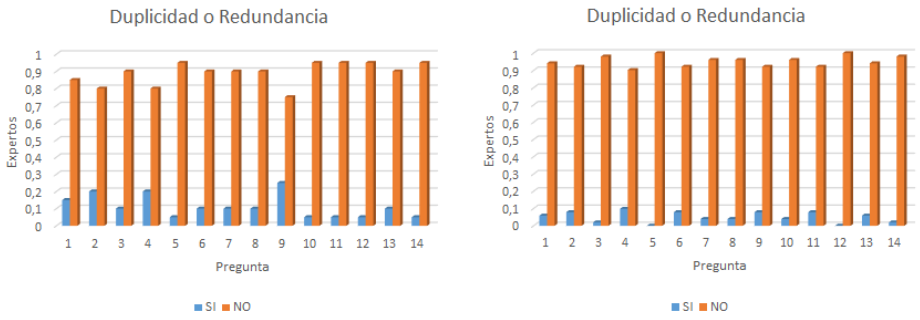
Inicialmente evaluamos la ontología por medio de dos encuestas a dos grupos de expertos diferentes. La encuesta inicial se realizó con el fin de obtener una evaluación a la primera versión de nuestra ontología, la cual fue mejorada siguiendo las recomendaciones de los expertos. Esta encuesta la realizaron 20 estudiantes de Medicina y Cirugía de $5to$ semestre de la Universidad del Valle. La segunda encuesta la realizaron 51 expertos de América Latina con

diferentes especialidades, de los cuales 32 tienen más de 10 años de experiencia en su área. Los resultados de las encuestas se resumen en las Figura 26, Figura 27 y Figura 28.



(a) (b)

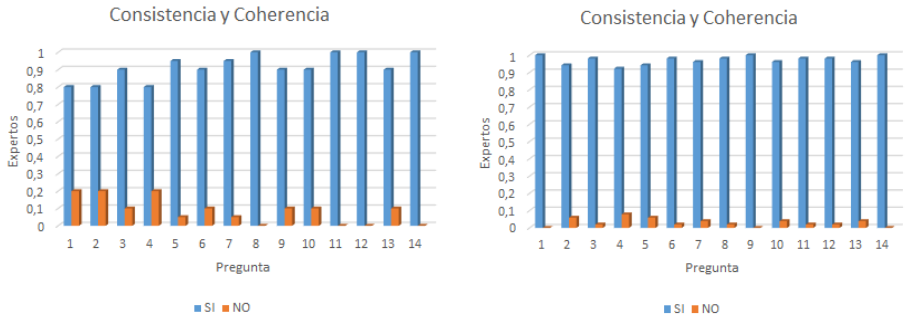
Figura 26: Resultados sobre completitud. (a) Primera encuesta. (b) Segunda encuesta.



(a) (b)

Figura 27: Resultados sobre duplicidad o redundancia. (a) Primera encuesta. (b) Segunda encuesta.

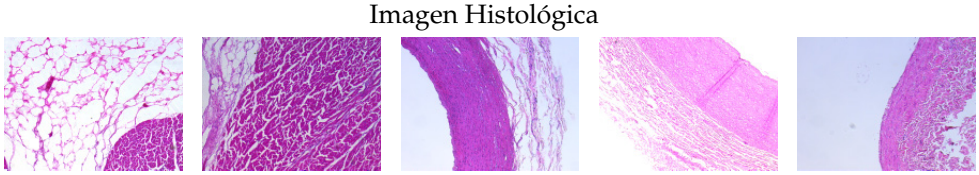
Las evaluaciones realizadas muestran una mejora sustancial en la segunda versión de la ontología en comparación con la primera. Estos resultados son cruciales para nosotros, ya que fueron proporcionados por profesionales con mucha experiencia en la histología. En conclusión, el hecho de que se obtuvieron resultados satisfactorios, recomienda la publicación de la ontología para su uso.



(a) (b)

Figura 28: Resultados sobre coherencia. (a) Primera encuesta. (b) Segunda encuesta.

Finalmente, hemos realizado una evaluación de los dos métodos para el proceso de mejora de la clasificación automática de tejidos y órganos. El conjunto de imágenes histológicas seleccionadas se presenta en la Figura 29, mientras que la Figura 30 incluye los resultados de la clasificación automática y la mejora con los dos métodos propuestos teniendo en cuanta sus aciertos y errores. En la Figura 30 la clasificación automática y los resultados del proceso de mejora representan cada clase con un color distintivo: (i) el músculo cardíaco del corazón con verde, (ii) el tejido conectivo laxo con el azul, (iii) del músculo liso de la arteria muscular con violeta, (iv) el músculo liso de la vena grande con amarillo, (v) el músculo liso de la arteria elástica con naranja, (vi) el tejido epitelial con rosado y (vii) las regiones de luz con fucsia. Por otro lado, aciertos y errores resultados se representan en colores verde y rojo, respectivamente. Se puede observar que los aciertos aumentan y hay una dismunición de los fallos utilizando el método de mejora de la clasificación.



Img-He *Img-He1* *Img-MA* *Img-LV* *Img-EA*

Figura 29: Imágenes histológicas. *Img-He* e *Img-He1* representan las imágenes del corazón; *Img-MA* representa la imagen de la arteria muscular; *Img-EA* representa la imagen de la arteria elástica; y *Img-LV* representa la imagen de la vena de gran calibre.

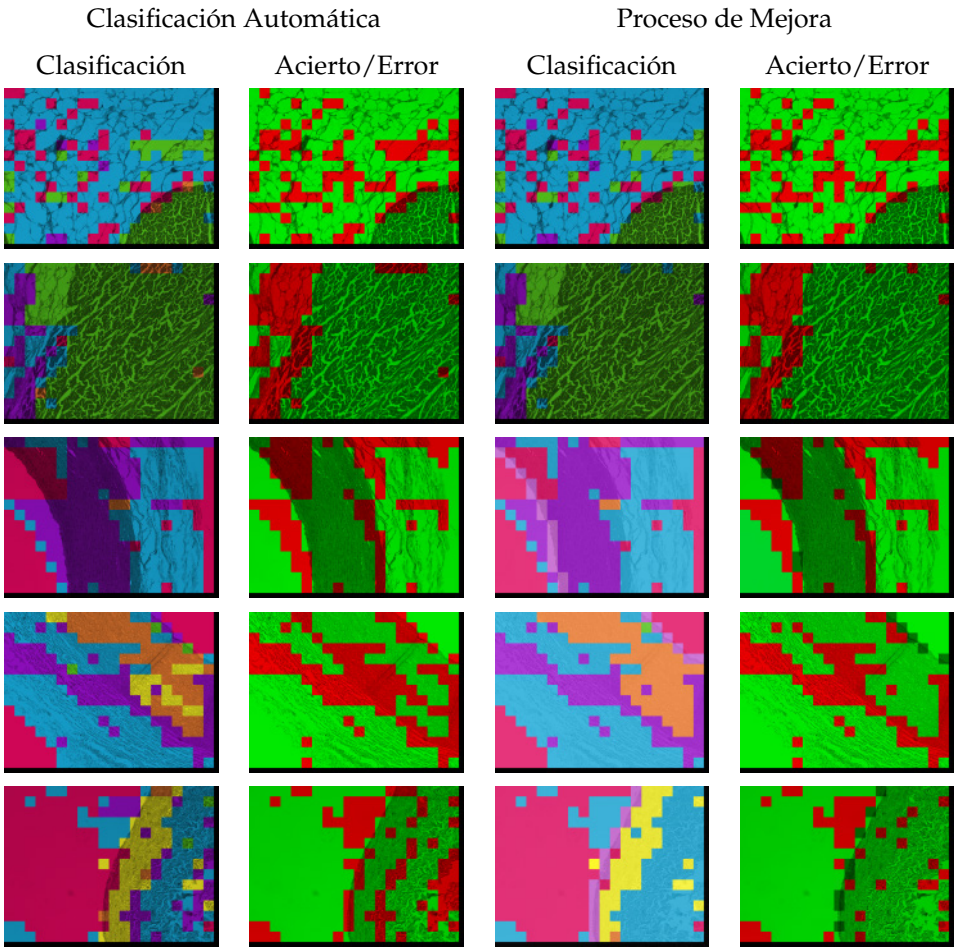


Figura 30: Resultados del proceso de mejora de la clasificación automática. En la primera columna la clasificación automática. En la segunda columna aciertos y fallos de clasificación automática. En la tercera columna la clasificación mediante los procesos de mejora. En la cuarta columna, aciertos y fallos de los procesos de mejora. En cada fila una imagen histológica se representa, de arriba a abajo: *Img-He*, *Img-He1*, *Img-MA*, *Img-EA*, and *Img-LV*.

La Figura 31 contiene una representación gráfica de los aciertos y fallos de la clasificación automática y el proceso de mejora por bloques en el conjunto de imágenes de prueba. La Figura 32 muestra el aumento en las tasas de acierto con el proceso de mejora de la clasificación. Los bloques correctamente reclasificadas están entre 1 y 24 por imagen, aumentando las tasas de clasificación entre 0,333 % y 23,188 % que varía según la imagen histológica y su clasificación automática. Por otro lado, la Figura 33 contiene una representación gráfica de los aciertos y fallos adicionando el reconocimiento del tejido epitelial por bloques en el conjunto

de imágenes de prueba. La Figura 34 muestra el aumento en las tasas de acierto para el reconocimiento del tejido epitelial. Los bloques correctamente reclasificadas están entre 0 y 7 por imagen aumentando las tasas de clasificación entre 0 % y 2,333 % que varía de acuerdo a la existencia y área del tejido epitelial en la imagen.

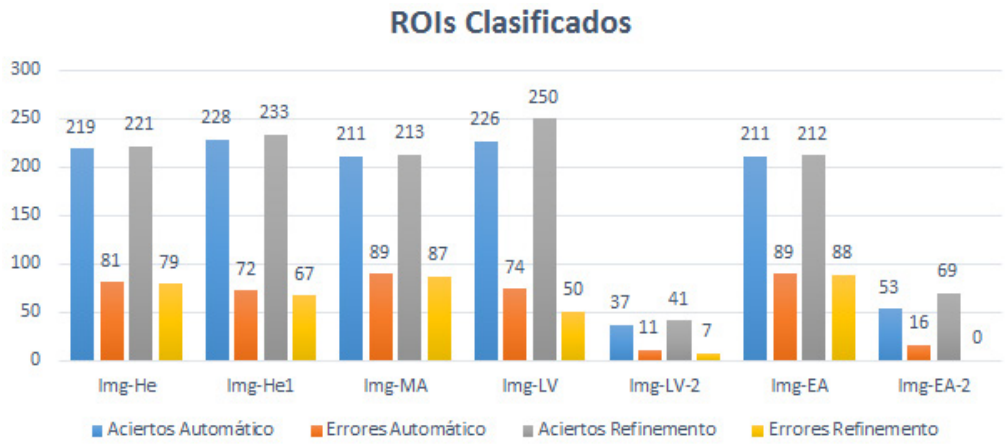


Figura 31: Aciertos y fallos por bloques con la clasificación automatica y el proceso de mejora. El nombre de cada imagen con -2 al final corresponde a la segunda clasificación en el proceso en cascada SVM.

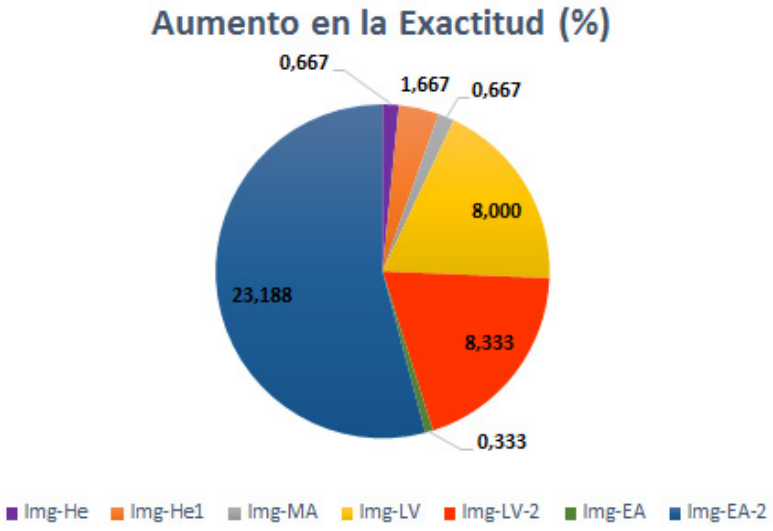


Figura 32: Incremento en las tasas de acierto con el proceso de mejora de la clasificación. El nombre de cada imagen con -2 al final corresponde a la segunda clasificación en el proceso en cascada SVM.

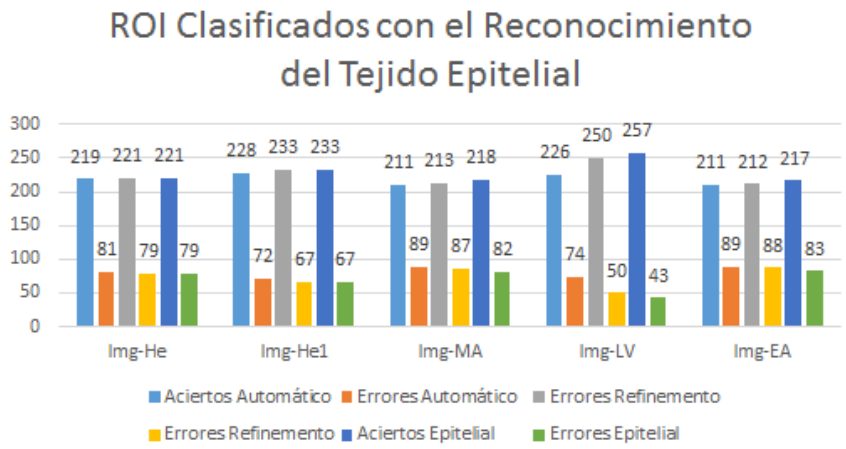


Figura 33: Aciertos y fallos por bloques con el reconocimiento del tejido epitelial.

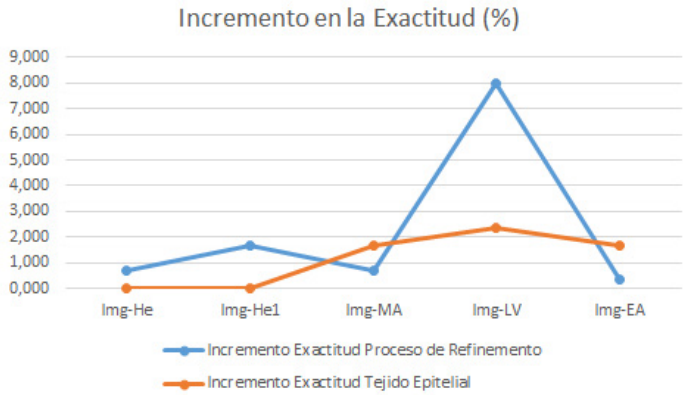


Figura 34: Incremento en las tasas de acierto con el reconocimiento del tejido epitelial. *Img-He* e *Img-He1* representan las imágenes del corazón; *Img-MA* representa la imagen de la arteria muscular; *Img-EA* representa la imagen de la arteria elástica; e *Img-LV* representa la imagen de la vena de gran calibre.

Hemos probado que utilizando el método de mejora para la clasificación incrementando las tasas de acierto, aumentando su confiabilidad, y logrando reconocer el tejido epitelial en imágenes tomadas a $10\times$.