

Departamento de Ciencias de la Computación e Inteligencia Artificial
Konputazio Zientziak eta Adimen Artifiziala Saila



Universidad del País Vasco
Euskal Herriko Unibertsitatea

Clasificadores Bayesianos en la Selección Embrionaria en Tratamientos de Reproducción Asistida

por

Dinora A. Morales Vega

Tesis presentada al Departamento de Ciencias de la Computación e Inteligencia Artificial de la Universidad del País Vasco para optar al título de doctora en Informática

Directores: Pedro Larrañaga Múgica y Endika Bengoetxea Castro

Donostia - San Sebastián, diciembre de 2008

Agradecimientos

La realización de esta tesis fue posible gracias al apoyo de varias personas y desde estas líneas me gustaría expresarles mi agradecimiento.

En primer lugar quiero agradecer especialmente a mis directores Pedro Larrañaga y Endika Bengoetxea, por su valioso apoyo y comprensión en todo este tiempo y sobre todo su sabia dirección. Ha sido un verdadero placer trabajar con ambos.

Como olvidarme de José A. Lozano, Iñaki Inza y Alex Mendiburu, por la dirección del grupo de investigación, fraternal y profesional. Gracias por su apoyo y comprensión.

Me gustaría agradecer a mis compañeros del laboratorio, Roberto Santana, Aritz Pérez, Borja Calvo, Rubén Armañanzas, Juan Diego Rodríguez, Carlos Echegoyen, Guzmán Santafé, Ramón Sagarna, José Luis Flores y Teresa Miquélez gracias por las charlas y los momentos amenos. Trabajar con gente excepcional como ustedes ha sido una experiencia muy enriquecedora.

Quisiera agradecer a Marisa Merino, Miguel García, Yosu Franco-Iriarte y Maite Estevan de la Clínica del Pilar ubicada en la ciudad de San Sebastián, por su tiempo y su contribución con los datos e imágenes para la realización de esta tesis. Especialmente a Yosu por su paciencia, disponibilidad y sobre todo su orientación e introducción en la problemática de la selección embrionaria. Como olvidarme de Mónica Fresnada, a quien agradezco su colaboración en el proyecto.

También quisiera agradecer la contribución de los doctores G. Patrizi y C. Manna, por la provisión de datos de las imágenes de embriones, previamente utilizados en Patrizi y col., (2004).

Quisiera hacer extensivo mi agradecimiento a Elena Bidondo por su apoyo desde que llegue a la facultad y también por su agradable compañía.

En el plano personal, quiero agradecer especialmente al M.D. Joseba Barcena por su atención, confianza y profesionalismo.

A mis hermanos, especialmente a Benjamín, Carlos y Mary. Sin olvidarme de todas mis amigas que siempre están cerca. Gracias a todos porque me han ayudado a mantenerme firme en el camino de la vida.

Por último pero no menos importante, quiero agradecer a mis hijos Ulises y Diana Urretabizkaya su comprensión, cooperación y sobre todo porque a pesar de que nuestra vida se cimbró, han seguido adelante manteniendo el tipo como unos campeones. Gracias pequeños, hemos logrado una meta más.

Quiero dedicar esta tesis a la memoria de Rafael, que siempre estará cerca de mi corazón, porque siempre estuvo ahí para acompañarme y brindarme lo necesario para alcanzar todas mis metas. La culminación de esta etapa de mi vida, no sería posible sin haber contado con su amor y constante ayuda.

Dinora

Este trabajo ha sido parcialmente soportado por Etortek, Saiotek y la subvención para grupos de investigación consolidados 2007-2012 (IT-242-07) del Gobierno Vasco, TIN2005-03824 y del proyecto Consolider Ingenio 2010 - CSD2007-00018 (Ministerio de Educación y Ciencia, Español) y de la Red en biomedicina computacional COMBIOMED (Instituto de Salud Carlos III).

Resumen

Durante los últimos años se ha incrementado la aplicación de tecnologías computacionales, basadas en la minería de datos y el aprendizaje automático en diferentes áreas como la biología o la medicina. Un ejemplo es la aplicación de diferentes técnicas de la inteligencia artificial en un problema médico complejo como es la infertilidad en la pareja a través de sistemas de apoyo en los tratamientos de fertilización in-vitro. Actualmente, la infertilidad es un problema social que afecta a personas en todo el mundo, sin embargo su tratamiento varía de un país a otro, ya que cada clínica de reproducción asistida está sujeta a la legislación de cada país. Situación que genera la falta de homogeneidad en los protocolos y tratamientos médicos existentes para tratarla. En particular, el proceso de la selección embrionaria, es una de las fuentes de mayor subjetividad e incertidumbre en el tratamiento de reproducción asistida, ya que actualmente la selección embrionaria depende completamente de la formación y experiencia del embriólogo.

Debido a la incertidumbre inherente en la selección embrionaria en tratamientos de reproducción asistida en esta tesis se propone la adaptación de los clasificadores Bayesianos como soporte a la decisión del experto en la tarea de la selección embrionaria en tratamientos de reproducción asistida, combinando la experiencia del embriólogo y la información relacionada con la caracterización morfológica del embrión así como la información clínica de la paciente y del tratamiento. Por medio del proceso de la selección de un subconjunto de variables para la inducción de algunos de los clasificadores Bayesianos se exploró la confiabilidad de los clasificadores relacionando las variables seleccionadas con el criterio aplicado por el experto en el proceso de la selección embrionaria. Se amplió la información disponible hasta ahora de los tratamientos de reproducción humana asistida con medidas extraídas de la imagen del embrión entre ellas la medición del grosor de la membrana de la zona pelúcida. En esta tesis se aporta la medición automática del grosor de la membrana de la zona pelúcida para mejorar la precisión de los clasificadores Bayesianos en el desarrollo de un sistema de apoyo a la decisión del experto en el problema de la selección embrionaria.

Las aportaciones de esta disertación abarcan la inducción de clasificadores Bayesianos, la propuesta de un multi-clasificador basado en el esquema conocido como *stacked-generalization* adaptado al dominio continuo, la adaptación de los clasificadores Bayesianos con un aprendizaje sensible al coste y la mejora de la precisión de los clasificadores Bayesianos, por medio de procesos de selección de variables.

Abstract

In the recent years the use of computational technologies based on data mining and automatic learning techniques have been extensively applied to different disciplines such as biology and medicine. In those domains the application of artificial intelligence as well as data mining techniques can be helpful to build decision-support systems. As an example of this, we have the case of applying artificial intelligent techniques into the complex medical problem of infertility for in vitro fertilization treatment, to improve the decisions to be made by embryologists and practitioners.

Infertility is nowadays considered an important social problem that has been subject of special interest by medical doctors and biologists. Due to ethical reasons, different legislative regulations apply in each country regarding human assisted reproduction techniques such as in-vitro fertilization, which leads to different protocols and treatments. An essential problem in human assisted reproduction is the selection of suitable embryos to transfer in a patient. Given the uncertainty of the treatment, this thesis proposes an intelligent decision support system based on supervised classification by Bayesian classifiers that proves to provide a more accurate selection than nowadays procedures. Current treatments rely fully on the expertise and experience of embryologists to select the most promising embryos that will form the batch to be transferred to the woman's uterus.

We propose to take into consideration a reduced subset of feature variables representing embryo morphology and clinical data of patients, and from this information to induce Bayesian classification models. Results obtained applying the wrapper and filter techniques to choose the subset of variables, as well as the performance of Bayesian classifiers applied to this problem, are presented. In order to improve the performance of Bayesian classifiers, we propose a system to automatically measure zona pellucida thickness variation, and to add this information to the Bayesian classifier. This is a potentially reliable parameter to evaluate the ability of embryo hatching and, it is directly related with the implantation rate.

In this dissertation we introduce a new multi-classification system using Gaussian networks to combine the outputs (probability distributions) of standard machine learning classification algorithms. Our method considers these outputs as inputs for a superior-level and to apply a stacking scheme to provide a meta-level classification result.

The contributions presented in this thesis cover Bayesian classifiers induction, the new Gaussian-stacking multi-classifier, the adaptation of Bayesian classifiers to sensible cost learning and the improvement of Bayesian classifiers by feature subset selection algorithms.

Laburpena

Azken urtetan teknologia konputazionalen aplikazioak asko zabaldu dira, hauen artean datu-meatzaritza eta ikasketa automatikoa erabiltzea biologia edo medikuntza bezalako arlotan. Honen adibidea da adimen artifizialeko teknika desberdinaren aplikazioa in-vitro ernaltze tratamenduen arazo mediku konplexuan.

Gaur egun ernalezintasuna mundu guztian arazo soziala bihurtu da, baina honen tratamendua asko aldatzen da estatu batetik bestera, klinika bakoitza bere menpeko legislazioaren berezitasunei egokitutako behar delako. Egoera honek tratamendu eta protokolo medikoen homogeneizatzeko ezintasuna sortarazten du. Bereziki, enbrioia aukeratzeko prozedura da tratamendu hauetan subjektibotasun eta ziurgabetasun iturri nagusienetarikoa, gaur egun enbrioien aukeraketa enbriologoen prestakuntza eta esperientziaren guztiz menpekoa baita.

Lagunduriko ugalketa tratamenduen ziurgabetasuna murrizteko asmoz, tesi honetan sailkatzaila Bayesiarren erabilpena proposatzen da adituaren erabakiaren euskarri gisa, horrela elkarrekin konbinatua ahal izateko enbriologoaren esperientzia, enbrioaren ezaugarri morfologikoen karakterizazio-informazioa, eta bikote nahiz tratamenduaren informazio klinikoa. Sailkatzaila Bayesiar batzuen indukzioaren ondorioz aldagaien azpimultzo bat aukeratu eta gero, sailkatzaila hauen fidagarritasuna aztertu da aukeratutako aldagaiak enbriologo adituek erabiltzen dituzten parametroekin alderatuz. Honez gain, tesi honetan lagunduriko ugalketan orain arte erabiltzen den informazioa zabalduta da enbrioien irudietatik ateratako neurriak erabiliz, hots *zona pellucida* deritzon mintzaren zabaleraren neurri zehatzen bidez. Horretarako, mintz honen neurketa automatikoa egin ahal izateko mekanismoa proposatzen da tesi honetan, sailkatzaila Bayesiarren erabakigarritasuna hobetzeko asmoz.

Lan honen ekarpen nagusienak oso zabalak badira ere, ondokoak azpimarratu daitezke: sailkatzaila Bayesiarren indukzioa, *stacked-generalization* eskeeman oinarritutako multi-sailkatzale baten proposamena balio-barruti jarraira aplikatu ahal izateko, sailkatzaila Bayesiarren adaptazioa ikasketan kostuaren menpekotasuna ezarriz, eta sailkatzaila Bayesiarren eraginkortasunaren hobekuntza aldagaien aukeraketa prozedurak erabiliz.

Índice general

Parte I Introducción

1.	Introducción	3
1.1.	Objetivos	5
1.1.1.	Objetivos relacionados con diferentes técnicas aplicadas para la mejora de resultados	5
1.2.	Organización de la tesis	6
1.3.	Publicaciones	8

Parte II Clasificación supervisada

2.	Introducción a la clasificación supervisada	11
2.1.	Introducción	11
2.2.	Problema de clasificación supervisada	11
2.3.	Validación de clasificadores	14
2.3.1.	Matriz de confusión	14
2.3.2.	Métodos de validación de clasificadores	16
2.3.3.	Comparación de paradigmas de clasificación	18
2.3.4.	Nivel de concordancia entre clasificadores	22
2.4.	Selección de variables	23
3.	Clasificadores Bayesianos	27
3.1.	Introducción	27
3.2.	Naive Bayes	28
3.3.	Selective naive Bayes	30
3.4.	Semi naive Bayes	31
3.5.	Naive Bayes aumentado a árbol	33
3.6.	Clasificador Bayesian k-dependiente	34
3.7.	Aproximaciones wrapper y filter para la inducción de clasificadores Bayesianos	35

XII Índice general

4.	Combinación de clasificadores	37
4.1.	Introducción	37
4.2.	Arquitecturas de combinación de clasificadores	37
4.3.	Métodos de combinación de múltiples clasificadores	39
4.4.	Estado del arte de combinación de clasificadores basados en el esquema de pila	42
4.5.	Multi-clasificador <i>Gaussian-stacking</i>	43
4.6.	Paradigmas de clasificación aplicados como clasificadores-base .	44

Parte III Selección embrionaria en tratamientos de reproducción humana asistida

5.	El problema de la selección embrionaria	49
5.1.	Introducción	49
5.2.	Estado del arte de la inteligencia artificial aplicada a tratamientos de reproducción asistida	50
5.3.	Tasa de éxito y procedimiento de transferencia	51
5.4.	Caracterización morfológica de cigotos y embriones en tratamientos de reproducción asistida	52
5.5.	Planteamiento de la selección embrionaria como problema de clasificación supervisada	58
5.6.	Aportación de la medición del grosor de la zona pelúcida en imágenes de embriones humanos por medio de contornos activos	59
5.6.1.	Contornos activos	61
5.6.2.	Inicialización automática del contorno activo	62
5.6.3.	Medición del grosor de la zona pelúcida	63
5.6.4.	Resultados	63

Parte IV Recomendaciones metodológicas para la selección embrionaria en tratamientos de reproducción asistida

6.	Multi-clasificador <i>Gaussian-stacking</i> aplicado a la selección embrionaria	69
6.1.	Validación del multi-clasificador <i>Gaussian-stacking</i>	70
6.1.1.	Diseño experimental	70
6.1.2.	Resultados experimentales a partir de clasificadores-base	71
6.1.3.	Resultados experimentales a partir del multi-clasificador <i>Gaussian-stacking</i>	71
6.2.	Multi-clasificador <i>Gaussian-stacking</i> aplicado al problema supervisado de la selección embrionaria	74
6.2.1.	Base de datos	74
6.2.2.	Diseño experimental	75

Índice general XIII

6.2.3. Resultados experimentales	75
7. Selección del mejor embrión para su transferencia en tratamientos de reproducción humana asistida	77
7.1. Base de datos	77
7.2. Diseño experimental	78
7.3. Resultados experimentales	79
8. Selección de un conjunto de embriones en tratamientos de fertilización in-vitro	83
8.1. Diseño experimental	83
8.2. Resultados experimentales	84
8.2.1. Selección de un subconjunto de características (FSS) ..	85
8.2.2. Simulación de la experiencia por medio del clasificador naive Bayes	86
9. Mejora de la clasificación en la selección embrionaria en base a información extraída de la imagen del embrión	89
9.1. Selección embrionaria en base a características morfológicas evaluadas por el experto y los datos clínicos del tratamiento de fertilización in-vitro	90
9.1.1. Diseño experimental	90
9.1.2. Resultados experimentales	91
9.2. Mejora de la clasificación en base a la información extraída de la imagen del embrión, características morfológicas evaluadas por el experto y datos clínicos	92
9.2.1. Base de datos	92
9.2.2. Diseño experimental	94
9.2.3. Resultados experimentales	94

Parte V Conclusiones y trabajo futuro

10. Conclusiones y trabajo futuro	101
10.1. Conclusiones	101
10.2. Trabajo futuro	103

Parte VI Apéndices

A. Técnicas del pre-proceso de la imagen	109
A.1. Introducción	109
A.2. Pre-proceso de la imagen del embrión	110
A.3. Generación del mapa de bordes de la imagen	111

XIV Índice general

B. Técnicas de extracción automática de características de textura de la imagen del embrión	113
B.1. Introducción	113
B.2. Extracción de medidas de la imagen	113
C. Caracterización morfológica del embrión por medio de momentos centrales	117
C.1. Introducción	117
C.2. Momentos centrales	117
Referencias	123

Lista de figuras

2.1.	Esquemas de las aproximaciones <i>wrapper</i> y <i>filter</i>	24
3.1.	Ejemplos de estructuras de clasificadores Bayesianos.	29
3.2.	Pseudocódigo del algoritmo FSSJ [Pazzani, 1997].	31
3.3.	Pseudocódigo del algoritmo BSEJ [Pazzani, 1997].	32
3.4.	Pseudocódigo de la inducción del clasificador TAN por medio del algoritmo de Chow-Liu [Friedman y col., 1997].	34
3.5.	Pseudocódigo del algoritmo para inducir el clasificador <i>kDB</i> [Sahami, 1996].	35
4.1.	Ejemplo de configuraciones de combinación de clasificadores. .	38
4.2.	Esquema del multi-clasificador <i>Gaussian-stacking</i>	44
5.1.	Esquema del cigoto y el embrión.	53
5.2.	Esquema de la rutina de selección embrionaria realizada por el experto.	55
5.3.	Ejemplo de imágenes de cigotos reales.	56
5.4.	Esquema con la representación gráfica de los blastómeros y su fragmentación.	57
5.5.	Ejemplos de la segmentación automática y manual de la zona pelúcida.	63
5.6.	Segmentación automática de imágenes de embriones humanos con presencia de artefactos y diferentes formas por medio de los contornos activos.	64
7.1.	Imágenes reales de embriones adquiridas en el Centro Genesis (Roma, Italia).	78
7.2.	Gráficas de las curvas ROC de diferentes clasificadores Bayesianos aplicados en la selección del mejor embrión.	80

XVI Lista de figuras

8.1.	Modelo semi naive Bayes para la clasificación de un conjunto de embriones.	84
8.2.	Estructura del clasificador naive Bayes, simulando la experiencia del experto.	87
9.1.	Ejemplo de imágenes de embriones adquiridas en la Clínica del Pilar, Donostia-San Sebastián.	93
9.2.	Prueba Kruskal-Wallis de los clasificadores Bayesianos.	96
A.1.	Esquema del proceso de extracción automática de las características morfológicas del embrión.	110
A.2.	Ejemplo de mejora de la imagen del embrión.	111

Lista de tablas

2.1.	Matriz de confusión.	15
2.2.	Matriz de McNemar.	21
2.3.	Escala para medir el grado de concordancia del índice Kappa.	22
5.1.	Ejemplo de los vectores de características de tratamientos de fertilización in-vitro.	59
5.2.	Resultados de la medición del grosor de la zona pelúcida por medio de <i>snakes</i>	64
6.1.	Lista de bases de datos seleccionadas del repositorio UCI.	70
6.2.	Resultados de siete clasificadores clásicos aplicados a trece problemas de clasificación supervisada.	72
6.3.	Resultados de cinco multi-clasificadores Gaussian-stacking.	73
6.4.	Variables contenidas en la base de datos recabada de la Clínica del Pilar, San Sebastián.	74
6.5.	Resultados de cinco multi-clasificadores Gaussian-stacking aplicados al problema de selección embrionaria en tratamientos de fecundación in-vitro.	76
7.1.	Resultados de diferentes clasificadores Bayesianos aplicados a la selección del embrión más viable a partir de las características extraídas de la imagen.	81
7.2.	Áreas bajo la curva ROC de los clasificadores Bayesianos aplicados a la selección del embrión más viable.	81
8.1.	Resultados de los clasificadores Bayesianos aplicados a la selección de un conjunto de embriones.	85
8.2.	Valor predictivo de implantación y otras medidas de la selección de un conjunto de tres embriones por medio de diferentes clasificadores Bayesianos.	85
8.3.	Lista de variables, resultado de FSS.	86

XVIII Lista de tablas

9.1.	Resultados control de cinco clasificadores Bayesianos aplicados a 81 casos en la selección de un conjunto de embriones.	91
9.2.	Matrices de confusión de cinco clasificadores Bayesianos. Resultados control.	91
9.3.	Resultados control de 81 casos en la clasificación de un conjunto de tres embriones por medio de clasificadores Bayesianos.	92
9.4.	Lista de variables extraídas de la imagen del mejor embrión. ..	93
9.5.	Resultados de cinco clasificadores Bayesianos en base a información extraída de la imagen del mejor embrión de un conjunto de tres embriones transferidos en FIV.	94
9.6.	Matrices de confusión de los clasificadores Bayesianos aplicados a un conjunto de tres embriones con información adicional extraída de la imagen entre otra información sugerida por el experto.	94
9.7.	Medidas de la clasificación de un conjunto de embriones, en base a información adicional extraída de la imagen por medio de clasificadores Bayesianos.	95
9.8.	Lista de variables predictoras obtenidas por medio del proceso de la selección de un subconjunto de variables (FSS) para la inducción de diferentes clasificadores Bayesianos en base a información extraída de la imagen del embrión y otros datos. ..	97
C.1.	Matriz de rotación.	121

Parte I

Introducción

1

Introducción

Durante los últimos años en diferentes áreas como la biología o la medicina, se ha incrementado la aplicación de tecnologías computacionales, basadas en la minería de datos y el aprendizaje automático entre otras áreas de conocimiento de la inteligencia artificial. En el caso de la medicina se han aplicado diferentes técnicas de la inteligencia artificial en el diagnóstico, monitorización de pacientes, diseño de tratamientos, elaboración de aplicaciones de enseñanza asistida por ordenador así como sistemas de apoyo a la toma de decisiones.

En el amplio campo de las ciencias médicas y entre los diferentes problemas médicos con índole social se encuentra el problema de infertilidad en la pareja. La infertilidad se ha considerado un problema complejo y durante décadas ha generado un interés especial tanto en médicos como biólogos desarrollándose así un amplio campo de investigación. Principalmente, las investigaciones tienden a plantear nuevos tratamientos y procedimientos de reproducción humana asistida, así como la aplicación de nuevas tecnologías y la integración de la inteligencia artificial. En este sentido, desde la década de los 70's donde [Steptoe y Edwards, 1978, Edwards y col., 1980] culminaron diez años de trabajo con el nacimiento del primer niño producto del tratamiento de fertilización in-vitro, las investigaciones en tratamientos y técnicas de reproducción humana asistida se han combinado para incrementar las tasas de implantación y nacimientos considerando el bienestar y el cuidado de las pacientes y a su vez disminuyendo la tasa de embarazos múltiples.

Debido a la complejidad del problema de infertilidad en diferentes clínicas del mundo se aplican diversos procedimientos y tratamientos clínicos para tratarla. Sin embargo, los tratamientos de reproducción asistida en general constan de varias fases como son: la valoración del médico, la determinación del tipo de infertilidad, el tratamiento para estimular la ovulación, así como la elección de la técnica de reproducción asistida para la fertilización.

Existen varios aspectos a considerar para alcanzar el éxito del tratamiento de reproducción asistida en la etapa posterior a la fertilización. Entre ellos se encuentra la formación y criterio del embriólogo para la observación, evaluación y catalogación de los cigotos y embriones, así como la selección de los

embriones y la elección entre los más prometedores (conjunto de embriones con el mayor potencial para su desarrollo e implantación en el útero materno).

Los sistemas inteligentes aplicados en los procesos del tratamiento de reproducción asistida, permiten combinar el conocimiento de los expertos así como las técnicas de aprendizaje automático y minería de datos. En esta tesis se propone la adaptación de los clasificadores Bayesianos ampliamente utilizados en la literatura al problema de la selección embrionaria en tratamientos de reproducción asistida tratando de optimizar los resultados de predecir si un tratamiento será o no exitoso como apoyo a la decisión del embriólogo.

Una de las características de los clasificadores Bayesianos es su intuitiva representación gráfica, la cual permite al experto la visualización de la relación entre las variables del dominio del problema y las probabilidades marginales y condicionales del modelo para una mejor comprensión de la incertidumbre inherente al mismo. Como resultado del proceso de clasificación se provee la distribución de probabilidad de la variable clase. Estas características de los clasificadores Bayesianos resultan interesantes para el grupo de médicos y biólogos.

En esta tesis se propone un nuevo multi-clasificador basado en redes Gaussianas condicionales basado en el conocido paradigma *stacked generalization* introducido en [Wolpert, 1992].

En la literatura se encuentran diferentes adaptaciones del paradigma *stacked*, en las cuales se han aplicando como meta-clasificador algoritmos como vecinos más cercanos, regresión lineal multi respuesta, árboles de decisión, multi respuesta basada en árboles de decisión, todos ellos descritos en la Sección 4.4. Sin embargo, en todos estos multi-clasificadores es necesario la discretización previa de los datos para ser aplicados como entrada del meta-clasificador. En contraste, la aplicación de un meta-clasificador basado en redes condicionales Gaussianas permite el manejo de datos continuos como discretos, con lo cual las distribuciones de probabilidad de las clases, producto de la predicción de los clasificadores a combinar en un nivel anterior, no necesitan la discretización de los datos, evitando así la pérdida de información que este proceso conlleva.

Una de las aplicaciones potenciales de la investigación realizada en esta tesis es aplicar la metodología en base a las redes Bayesianas, en el desarrollo de un sistema de apoyo para la formación de nuevos expertos, con el objetivo de estandarizar su criterio en el proceso de evaluación y selección embrionaria así como la disminución de la incertidumbre acerca de la viabilidad de implantación de los embriones transferidos en tratamientos de reproducción asistida. La incertidumbre inherente al problema de la selección embrionaria en gran medida es generada por la falta de la homogeneidad del criterio de los embriólogos en la evaluación y selección embrionaria así como por la falta de un sistema de apoyo a la decisión en el proceso crucial de la selección embrionaria.

A continuación se describen los objetivos que se han planteado para el desarrollo de esta tesis.

1.1. Objetivos

Los trabajos que se llevan a cabo en esta tesis se enmarcan dentro de dos grandes áreas de conocimiento: el aprendizaje automático por medio de la clasificación supervisada y la biomedicina, concretamente en la selección embrionaria en tratamientos de reproducción asistida. El proyecto pretende como objetivo general estudiar empíricamente y proponer nuevos métodos de las aproximaciones de clasificadores basados en redes Bayesianas aplicados al problema de selección embrionaria en tratamientos de fertilización in-vitro.

Los principales objetivos que se han considerado en el desarrollo de este trabajo se describen a continuación:

- Desarrollar un multi-clasificador basado en redes Gaussianas condicionales basado en el conocido paradigma *stacked generalization* (Capítulo 4).
- Adaptar el problema de la selección embrionaria como problema de clasificación supervisada (Capítulo 5).
- Medición del grosor de la membrana de la zona pelúcida para mejorar la precisión de los clasificadores Bayesianos en el problema de la selección embrionaria (Capítulo 5).
- Adaptar el multi-clasificador propuesto al problema de la selección embrionaria (Capítulo 6).
- Adaptar los paradigmas clásicos de clasificadores Bayesianos con un aprendizaje sensible al coste del error (Capítulo 7).
- Emular el conocimiento del experto por medio del clasificador naive Bayes en base a la lista de variables predictoras seleccionadas por el proceso de selección de variables (Capítulo 8).
- Adaptar los clasificadores Bayesianos al problema de la selección embrionaria (Capítulos 7, 8 y 9).
- Aplicar el proceso de selección de un subconjunto de variables predictoras para la inducción de los clasificadores Bayesianos, en la selección embrionaria de un conjunto de embriones en base a sus características morfológicas evaluadas por el embriólogo así como los datos clínicos de la paciente y del tratamiento de fertilización in-vitro (Capítulos 8 y 9).
- Relacionar la lista de variables predictoras seleccionadas para la inducción de los diferentes clasificadores Bayesianos con el criterio de los expertos en el procedimiento de selección embrionaria y el tratamiento de fertilización in-vitro (Capítulos 8 y 9).

1.1.1. Objetivos relacionados con diferentes técnicas aplicadas para la mejora de resultados

Con el propósito de mejorar la precisión de los clasificadores Bayesianos se propuso enriquecer la información acerca de la morfología del embrión. En particular la medida de variación del grosor de la membrana que rodea al

embrión, llamada zona pelúcida, es considerada una de las principales características para predecir la viabilidad de implantación del mismo. Para ello se definieron los siguientes objetivos, todos ellos realizados en el Capítulo 9:

- Aplicación de contornos activos para la segmentación automática y medición de la variación del grosor de la zona pelúcida.
- Caracterización de la morfología del embrión a partir de la imagen.
- Verificación de que la información extraída de la imagen del embrión mejora la tarea de clasificación.

1.2. Organización de la tesis

La organización de la tesis se estructura en seis partes con un total de diez capítulos y tres apéndices de la siguiente forma:

En la primera parte, Capítulo 1, se hace una introducción general a la tesis y a los principales objetivos considerados en esta disertación así como las publicaciones realizadas.

La segunda parte, consta de tres capítulos en los cuales se introduce el problema de la clasificación supervisada, los clasificadores Bayesianos clásicos y la propuesta del multi-clasificador basado en el esquema *stacked generalization* aplicando redes condicionales Gaussianas como meta-clasificador, denominando *Gaussian-stacking*. El Capítulo 2 introduce de forma general el aprendizaje con clasificación supervisada y los conceptos necesarios para aplicar diferentes paradigmas de clasificación y se introduce el proceso de la selección de un subconjunto de variables. El Capítulo 3 introduce los clasificadores Bayesianos clásicos así como las adaptaciones de los clasificadores Bayesianos con enfoques *wrapper* y *filter*. También se hace una revisión del estado del arte de la clasificación supervisada con clasificadores Bayesianos adaptados al dominio continuo. El Capítulo 4 presenta una introducción general a los sistemas de combinación de clasificadores e introduce una descripción del estado del arte de los multi-clasificadores basados en el esquema conocido como *stacked generalization*. Se describe el multi-clasificador propuesto en esta tesis, basado en el esquema del *stacked* utilizando redes condicionales Gaussianas en el nivel de meta-aprendizaje así como algunos de los paradigmas de clasificación supervisada ampliamente utilizados en la literatura, los cuales fueron seleccionados como clasificadores-base para la construcción del multi-clasificador.

La tercera parte, Capítulo 5, introduce el problema de estudio de esta tesis, la selección embrionaria en tratamientos de fertilización in-vitro. Se hace una revisión del estado del arte de las técnicas de inteligencia artificial aplicadas al problema de selección embrionaria en tratamientos de fertilización in-vitro, así como relacionados con los tratamientos de reproducción asistida. Se plantea el problema de la selección embrionaria en tratamientos de reproducción humana asistida como un problema de clasificación supervisada. En este capítulo se describe la técnica de contornos activos aplicada a la segmentación de la membrana de la zona pelúcida así como el método propuesto para

la medición automática del grosor de la zona pelúcida a partir de la imagen del embrión.

La cuarta parte consta de cuatro capítulos en los cuales se presentan las aportaciones de esta tesis al problema de la selección embrionaria desde el campo de la clasificación supervisada por medio de redes Bayesianas, así como la validación del multi-clasificador *Gaussian-stacking*.

El Capítulo 6 describe los experimentos realizados para la validación del multi-clasificador *Gaussian-stacking* propuesto en esta tesis. Se introduce la adaptación del multi-clasificador basado en el esquema de pila (*stacked*), aplicando como meta-clasificador un clasificador Bayesiano adaptado al dominio continuo, para predecir el éxito del tratamiento de reproducción asistida a partir de las distribuciones de probabilidad de diferentes paradigmas de clasificación supervisada. Este proceso se efectúa en base a la catalogación y caracterización de los embriones hecha por el experto, así como la información clínica.

El Capítulo 7 muestra los resultados obtenidos a partir de la adaptación del aprendizaje sensible al coste del error de los clasificadores Bayesianos para la selección del embrión más apto para su transferencia en tratamientos de fecundación in-vitro, en base a la caracterización morfológica del embrión a partir de la imagen.

El Capítulo 8 presenta los resultados obtenidos de la adaptación de los clasificadores basados en redes Bayesianas al problema de la selección embrionaria. Se muestra una serie de variables, extraídas mediante diferentes procesos de selección de un subconjunto de variables, aplicados para la inducción de los clasificadores Bayesianos.

El Capítulo 9 describe dos estudios enfocados a la selección de un conjunto de tres embriones para su transferencia en tratamientos de fertilización in-vitro. El primer estudio se plantea como control para medir la mejora de la predicción del tratamiento de reproducción asistida en base a mejorar el proceso de la selección de variables y por ende la tarea de clasificación a partir de la información relacionada con las características morfológicas del embrión extraídas de la imagen.

La quinta parte, Capítulo 10 describe las aportaciones metodológicas desarrolladas en esta tesis así como las relacionadas con el problema de la selección embrionaria en tratamientos de reproducción humana asistida y las conclusiones. Finalmente se describe brevemente las líneas futuras de trabajo.

La última parte contiene tres apéndices con información adicional de las técnicas aplicadas para la mejora de los resultados de la tarea de clasificación supervisada de los clasificadores Bayesianos en el problema de la selección embrionaria en tratamientos de fertilización in-vitro.

Apéndice A describe las etapas anteriores a la segmentación automática por medio de los contornos activos.

Apéndice B introduce una serie de medidas para la caracterización de la textura de las estructuras morfológicas del embrión a partir de la imagen calculadas en base a la matriz de co-ocurrencia.

Apéndice C introduce los momentos centrales para la caracterización de la forma de las estructuras morfológicas del embrión a partir de la imagen.

1.3. Publicaciones

Como resultado de esta tesis se obtuvieron las siguientes publicaciones en el ámbito internacional y una de ellas en el nacional:

- **D.A. Morales**, E. Bengoetxea, P. Larrañaga, M. García, Y. Franco, M. Fresnada, M. Merino. (2008). Bayesian classification for the selection of in-vitro human embryos using morphological and clinical data. *Computer Methods and Programs in Biomedicine*, 90, 104-116.
- **D.A. Morales**, E. Bengoetxea, P. Larrañaga. Selection of human embryos for transfer by Bayesian classifiers.(2008). *Computers in Biology and Medicine*, 38 (11-12), 1177-1186.
- **D.A. Morales**, E. Bengoetxea, P. Larrañaga. Combining multi-classifiers with Gaussian network for selection of in-vitro human embryos using morphological and clinical data. *Data Mining and Medical Knowledge Management: Cases and Applications*, 307-331. P. Berka, J. Rauch, D. Zighed (eds.). IGI Global Inc. In press.
- **D.A. Morales**, E. Bengoetxea, P. Larrañaga. Automatic segmentation of zona pellucida in human embryo images applying an active contour model. *Medical Image Understanding and Analysis* 2008. J. Hoey and S. McKenna (eds.), pp. 209-213. Dundee, Scotland.
- **D.A. Morales**, E. Bengoetxea, P. Larrañaga, M. García, Y. Franco-Iriarte, M. Fresnada, M. Merino. (2007). Selection of human embryos in IVF by Bayesian classifiers. CEDI 2007, II *Simposio de Inteligencia Computacional* (SICO 2007), pp. 251-260. Spain.

Parte II

Clasificación supervisada

Introducción a la clasificación supervisada

2.1. Introducción

En el amplio campo de conocimiento de la inteligencia artificial se encuentran las áreas del aprendizaje automático y de la minería de datos, las cuales incluyen el área de *clasificación*. Las técnicas comprendidas en este campo han sido aplicadas ampliamente en medicina, biomédica, bioinformática y visión por computador, entre otros.

El área de la clasificación se divide en clasificación supervisada y clasificación no supervisada. La clasificación no supervisada consiste en el descubrimiento subyacente a un modelo de clasificación dado un conjunto de datos. La clasificación supervisada consiste en construir un modelo de clasificación a partir de un conjunto de datos de entrenamiento etiquetado previamente con un número de clases r ya conocidas, que sea capaz de generalizarlos y clasificar correctamente nuevas instancias.

En este capítulo se presenta el problema de la clasificación supervisada así como algunos conceptos necesarios para su aplicación.

2.2. Problema de clasificación supervisada

En la clasificación supervisada se aplican modelos o algoritmos capaces de aprender a partir de un conjunto de instancias o casos denotados por $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathcal{R}^n$. Cada instancia es un vector de características etiquetado con una variable clase C . La clase real es denotada por $c_i \in C$ donde $\Omega_c = \{c_1, c_2, \dots, c_r\}$ denota todos los posibles valores de la clase. Formalmente el problema de clasificación supervisada consiste en asignar un valor del conjunto de la variable clase C a una nueva instancia $\mathbf{x} = (x_1, x_2, \dots, x_n)$. Un clasificador puede verse como una función, γ , la cual asigna la etiqueta correspondiente a cada una de las instancias:

$$\gamma : (x_1, x_2, \dots, x_n) \rightarrow \{c_1, c_2, \dots, c_r\} \quad (2.1)$$

Cada problema de clasificación tiene asociado una matriz de coste del error generado en la clasificación $cost(k, c)$, donde c es la clase real y k es la clase estimada por el clasificador. En el caso de que el coste del error k sea siempre el mismo se dice que tenemos una función de coste simétrica 0/1. Esta función de coste ha sido ampliamente utilizada en clasificación y se define de la siguiente manera:

$$cost(k, c) = \begin{cases} 0 & \text{si } k = c \\ 1 & \text{en otro caso} \end{cases} \quad (2.2)$$

Uno de los objetivos en la clasificación supervisada es diseñar un clasificador que discrimine las clases excluyentes y minimice el error total de clasificación por medio de una función de coste. Un estimador $p(c|\mathbf{x})$ que minimice el error total en base a una función de coste definida como:

$$\gamma(\mathbf{x}) = \arg \min_k \sum_{c=1}^r cost(k, c)p(c|\mathbf{x}) \quad (2.3)$$

asigna a una nueva instancia $\mathbf{x} = (x_1, x_2, \dots, x_n)$ la clase c con el coste total mínimo.

Existen diferentes paradigmas de clasificación como son los árboles de decisión, redes neuronales, las máquinas de soporte vectorial, tabla de decisiones, vecinos más cercanos y redes Bayesianas entre otros. En el caso particular de los clasificadores Bayesianos, de acuerdo con [Duda y Hart, 1973] estos minimizan el error total teniendo en cuenta la distribución de probabilidad conjunta, la cual es calculada *a priori* a partir del conjunto inicial de datos:

$$p(c, x_1, \dots, x_n) = p(c|x_1, \dots, x_n)p(x_1, \dots, x_n) \quad (2.4)$$

Dada la distribución de probabilidad y la ecuación de coste del error (2.3), las redes Bayesianas que minimizan el error total de clasificación están definidas de la siguiente forma:

$$\gamma(\mathbf{x}) = \arg \min_k \sum_{c=1}^r cost(k, c)p(c|x_1, \dots, x_n) \quad (2.5)$$

En el caso particular de la función de coste simétrica 0/1, el clasificador Bayasiano asigna un valor de la clase C con el valor máximo *a posteriori* dada una instancia, a partir de la siguiente función:

$$\gamma(\mathbf{x}) = \arg \max_c p(c|x_1, \dots, x_n) = \arg \max_c p(c, x_1, \dots, x_n) \quad (2.6)$$

En problemas donde los costes del error en la clasificación son diferentes, la probabilidad *a posteriori* es comparada con un valor del umbral τ para predecir el valor de la clase \mathbf{x} . Usualmente se escoge un valor del umbral τ que minimice el coste del error [Maloof, 2003], por medio de la siguiente función de clasificación:

$$\gamma(\mathbf{x}) = c \Leftrightarrow p(C = c|\mathbf{x}) > \tau \quad (2.7)$$

Uno de los métodos de *aprendizaje sensible al coste del error* es variando los valores del umbral de decisión τ . [Breiman y col., 1984] establecen que variar el umbral de decisión τ es similar a variar la probabilidad a priori de la clase en cada conjunto de entrenamiento o el coste de error. [Duda y col., 2001] demuestran que el aprendizaje sensible al coste del error, variando el umbral de decisión, minimiza la tasa de error de Bayes, definida por la intersección entre las dos distribuciones de probabilidad de la clase. En el Capítulo 7 se adaptaron los clasificadores Bayesianos con un aprendizaje sensible al coste del error, los cuales fueron aplicados para la selección del embrión más viable para su transferencia en tratamientos de fertilización in-vitro.

En particular en el problema de la selección de un conjunto de embriones el efecto del coste del error podría ser considerado diferente en la predicción del tratamiento con la implantación de dos o tres embriones, si se tiene en cuenta que las consecuencias de los embarazos múltiples representa un riesgo para la paciente. En el caso de la selección embrionaria el costo de un falso positivo (un embrión no prometedor es clasificado erróneamente provocando la ausencia de embarazo) es más alto que el de un falso negativo (un embrión prometedor es descartado). No obstante, como admiten los embriólogos, no existe una manera comúnmente aceptada para estimar el coste de cada uno de estos errores y es por esta razón que en esta tesis no fue posible contar con una matriz de coste del error y se decidió no utilizar una cualquiera ya que esto afecta directamente el aprendizaje de los clasificadores. El Capítulo 5 describe el problema de la selección embrionaria en tratamientos de reproducción asistida así como el planteamiento de la selección de un conjunto de embriones como un problema de clasificación supervisada.

Los paradigmas de clasificación pueden ser inducidos de dos formas dependiendo de la distribución de los datos a aprender: aprendizaje generativo o discriminativo [Dawid, 1976, Rubinstein y Hastie, 1997].

Los modelos generativos aprenden el modelo de la probabilidad conjunta de las variables predictoras $\mathbf{X} = (X_1, \dots, X_n)$ y la variable clase C . A continuación el clasificador predice la clase de una nueva instancia $\mathbf{x} = (x_1, x_2, \dots, x_n)$ usando la regla de Bayes para calcular la probabilidad a posteriori de la variable clase dados los valores de las variables predictoras $P(c|\mathbf{x})$.

En los modelos generativos se obtienen los parámetros del modelo mediante el aprendizaje de la distribución de probabilidad conjunta, $P(c, \mathbf{x})$, los cuales pueden obtenerse maximizando el logaritmo de la función de verosimilitud

conjunta. Los clasificadores Bayesianos normalmente, aprenden los parámetros usando métodos generativos. Modelos como el análisis discriminante o el clasificador naïve Bayes son ejemplos clásicos de clasificadores generativos.

Por otra parte, los clasificadores discriminativos son los que obtienen sus parámetros modelando directamente la distribución de la clase dadas las variables predictoras, $P(c|\mathbf{x})$. En este caso los parámetros del modelo de clasificación pueden ser obtenidos maximizando el logaritmo de la función log-verosimilitud condicionada. La regresión logística y las redes neuronales son ejemplos de clasificadores discriminativos.

En esta tesis se aplicaron los modelos generativos concretamente los clasificadores Bayesianos (introducidos en el Capítulo 3) los cuales presentan un mejor comportamiento al ser inducidos a partir de la distribución de probabilidad de los datos.

A continuación se revisan algunos aspectos a considerar con el propósito de abordar posteriormente los problemas de clasificación supervisada.

2.3. Validación de clasificadores

Existen diversos criterios para la evaluación de los algoritmos de clasificación. Su elección dependerá del dominio del problema. Por ejemplo, *el porcentaje de casos bien clasificados* mide la bondad (precisión) del clasificador. La bondad de un clasificador es una estimación de la probabilidad de la clasificación correcta de una instancia elegida al azar [Kohavi, 1995b].

En clasificación, es aconsejable inducir un modelo a partir de un conjunto de datos llamado conjunto de entrenamiento y otro conjunto de datos llamado conjunto de prueba, el cual se aplica en la fase de clasificación. Con ello se evitan los resultados denominados optimistas. Es importante que para estimar la precisión de un clasificador se utilice un método con poca varianza.

2.3.1. Matriz de confusión

La matriz de confusión detalla el resultado de la clasificación. En la diagonal principal se reportan los casos correctamente clasificados y en la opuesta se detallan los errores de la predicción. Las columnas representan las clases presentes en los datos y las filas las clases en las que son predichas las instancias. En la tarea de clasificación con dos valores, dado un clasificador y una instancia se producen cuatro valores de salida como son: *verdadero positivo*, si la instancia es clasificada correctamente y su clase pertenece a la positiva; *verdadero negativo* se genera cuando la instancia es correctamente clasificada con la clase de valor negativo; *falso positivo* cuando la instancia es de la clase negativa y es clasificada como clase positiva y por último *falso negativo* se presenta cuando el clasificador clasifica erróneamente una instancia de la clase

		clase real	
		predicha	verdadero falso
predicha	verdadero	verdaderos positivos	falsos positivos
	falso	falsos negativos	verdaderos negativos

Tabla 2.1. Matriz de confusión.

positiva como un caso de clase negativa. En base a estas cuatro salidas se puede construir una tabla de contingencia, o matriz de confusión, representando al conjunto de datos de prueba.

A partir de la matriz de confusión se pueden extraer algunas medidas para comprender la distribución y naturaleza de los errores cometidos por el clasificador.

- La *sensibilidad* de un clasificador representa la fracción de verdaderos positivos y se calcula de la siguiente forma:

$$\text{sensibilidad} = \frac{\text{verdaderos positivos}}{\text{verdaderos positivos} + \text{falsos negativos}}$$

- La *especificidad* representa la proporción de verdaderos negativos y se calcula de la siguiente forma:

$$\text{especificidad} = \frac{\text{verdaderos negativos}}{\text{verdaderos negativos} + \text{falsos positivos}}$$

- El *valor predictivo de la clase positiva* representa la precisión del clasificador en términos del porcentaje de casos positivos correctamente clasificados y es calculado de la siguiente forma:

$$\text{valorpredictivopositivo} = \frac{\text{verdaderos positivos}}{\text{verdaderos positivos} + \text{falsos positivos}}$$

- El *valor predictivo de la clase negativa* se calcula de la siguiente forma:

$$\text{valorpredictivonegativo} = \frac{\text{verdaderos negativos}}{\text{verdaderos negativos} + \text{falsos negativos}}$$

Otras dos medidas a tener en cuenta en la validación son el ratio de error verdadero y el ratio de error aparente.

El *ratio de error verdadero* de un clasificador es la tasa del error al clasificar mal los casos del conjunto de prueba que no han sido incluidos en el conjunto de entrenamiento utilizado para su inducción. Resulta ser una medida honesta de la estimación de la tasa de error.

Ratio de error aparente, es la tasa de error obtenido por un modelo al clasificar las instancias incluidas en su inducción. En particular esta medida tiende a ser muy optimista ya que los datos utilizados para su construcción se ajustan mejor al modelo que las instancias no incluidas en la inducción, generándose un fenómeno de sobre entrenamiento o también llamado *overfitting* y su error tiende a ser sesgado (*biased*). Para obtener un clasificador no sesgado, es recomendable probar el clasificador con un conjunto de datos no incluidos en su inducción. En el caso de que se cuente con pocos casos para la prueba, el error estimado tiende a tener una varianza grande.

2.3.2. Métodos de validación de clasificadores

Existen diferentes métodos de validación como son el método H (*Holdout*), remuestreo (*random subsampling*) y validación cruzada (*cross-validation*), entre otros. En este trabajo se aplicó ampliamente la validación cruzada para la inducción y validación de los diferentes clasificadores utilizados en esta tesis. A continuación se explican brevemente los métodos de validación mencionados.

Holdout

El método *holdout* (método H) divide aleatoriamente el conjunto de datos en dos subconjuntos. Uno de ellos con $2/3$ partes para el grupo de entrenamiento, el cual es utilizado para inducir el clasificador, y el segundo grupo con el $1/3$ de datos restante lo forma el grupo de prueba.

Remuestreo

El método de remuestreo (*random subsampling*) es una variante del método H, y consiste en aplicar el método H múltiples veces sobre diferentes subconjuntos del grupo de entrenamiento y grupo de prueba.

Validación cruzada

En el método de *k*-rodajas validación-cruzada (*k-fold cross-validation*) [Stone, 1974], la base de datos es particionada aleatoriamente en k subconjuntos, aproximadamente del mismo tamaño, donde $k - 1$ subconjuntos constituyen el *conjunto de entrenamiento* y el restante el *conjunto de prueba*. El clasificador es inducido a partir del conjunto de entrenamiento. Este proceso se repite k veces y cada ocasión, el clasificador prueba con un conjunto de prueba diferente. La estimación de la precisión del clasificador se calcula promediando

la exactitud de cada uno de los k procesos de clasificación y la desviación estándar es calculada de la misma manera. La desviación estándar debería ser aproximadamente la misma, independiente del número de subconjuntos en base al trabajo presentado por [Kohavi, 1995b].

[Stone, 1974] propuso un método para mejorar el resultado de éxito del clasificador obtenido con k -rodajas validación-cruzada llamado validación cruzada estratificada. Los subconjuntos son estratificados de tal forma que contienen la misma proporción de casos respecto a las clases que el conjunto de datos.

Existe un caso especial del método de k -rodajas validación-cruzada llamado dejar-uno-fuera (*leave-one-out*). En este método de validación la base de datos es particionada k veces, siendo k el número de casos originales n . Se forma un conjunto de entrenamiento con los $n - 1$ casos dejando uno caso fuera, el cual será utilizado en la clasificación. [Efron, 1983] aclara que este método es el menos sesgado pero presenta una varianza alta, la cual en algunos casos presenta una estimación de la precisión del clasificador no deseada. Este tipo de validación, generalmente es aplicado a bases de datos con relativamente pocos casos.

Bootstrapping

El método *bootstrapping* introducido por [Efron y Tibshirani, 1993] se utiliza para la validación de clasificadores y está basado en la selección de una muestra de datos con reemplazamiento. El muestreo con reemplazamiento consiste en extraer elementos de una población de forma que, tras cada extracción, el elemento extraído se vuelve a introducir y puede volver a ser seleccionado.

A partir de un conjunto de casos, se seleccionan B muestras aleatorias independientes, con reemplazamiento, de tamaño N cada una de las muestras conforma un grupo de entrenamiento. Cada conjunto de entrenamiento es utilizado para inducir el clasificador, el cual es testado con los casos no seleccionados como grupo de prueba. En el estimador *0.632bootstrapping*, la probabilidad de que un caso no sea seleccionado como parte del conjunto de entrenamiento es: $1 - \frac{1}{N}$, si N es suficientemente grande, dicha probabilidad tiende a $e^{-1} \simeq 0,368$. Por tanto, el número esperado de casos distintos incluidos en el conjunto de entrenamiento es: $(1 - 0,368)N = 0,632N$, es decir contiene el 62.3 % de casos. Este método presenta poco sesgo (con un número de casos grande) y menos varianza que el método k -rodajas de validación-cruzada.

Área bajo la curva ROC

Las curvas ROC son frecuentemente utilizadas en el campo médico, biomédico así como en bioinformática para evaluar modelos computacionales diseñados para el soporte de toma de decisiones, pronóstico y diagnóstico clínico [Lasko y col., 2005]. Las curvas ROC fueron introducidas en el campo de la minería

de datos por [Spackman, 1989] y surgieron de la teoría de la información en los años 50. El análisis de la curva ROC se ha convertido en una poderosa herramienta para la evaluación y comparación de paradigmas de clasificación. Ejemplo de ello son los trabajos presentados por [Provost y col., 1998, Provost y Fawcett, 2001].

La curva ROC permite la visualización del comportamiento del clasificador a través de un sistema de coordenadas en el que se representa la sensibilidad en el eje y versus 1-especificidad en el eje x . El gráfico se construye a partir de distintos valores de corte en la probabilidad de la clase del clasificador, los cuales son representados por los diferentes puntos de la curva (1-especificidad, sensibilidad). La curva ROC se construye a partir de variar el valor del umbral $\tau \in [0, 1]$ cuyo valor sirve como punto de corte para la clasificación en base a la Ecuación (2.2). Normalmente se escoge un valor del umbral τ que minimice la función de coste del error (2.3).

El cálculo del área bajo la curva AUC (area under the ROC curve) es un índice ampliamente utilizado [Hanley y McNeil, 1982]. Dicho índice resume el comportamiento y precisión del clasificador. Existen diferentes métodos para calcular el área bajo la curva ROC como son: el método no-paramétrico, paramétrico y por reemplazamiento. El método de la regla trapezoidal propuesto por [Bamber, 1975] es un método no-paramétrico, que suele utilizarse para el cálculo del área bajo la curva ROC y es equivalente a la prueba estadística de Mann-Whitney.

El valor del área bajo la curva ROC está entre 0.5 (asociado a la diagonal del cuadrado de la gráfica) y 1 (el cual corresponde al punto superior izquierdo) este último asociado a un buen comportamiento del clasificador. La precisión del clasificador aumenta a medida que la curva se desplaza desde la diagonal hacia el vértice superior izquierdo.

En los últimos años se ha incrementado el uso del análisis del área bajo la curva ROC en la comunidad de minería de datos, especialmente en problemas con bases de datos no balanceadas y funciones de error no simétricas, generando el área de investigación en el aprendizaje supervisado denominado sensible al coste [Maloof, 2003, Lasko y col., 2005].

2.3.3. Comparación de paradigmas de clasificación

En los métodos de k -rodajas validación-cruzada o los métodos de reemplazo la estimación del porcentaje de casos bien clasificados es la medida de precisión del clasificador, la cual se calcula a partir del promedio de todos los valores de éxito de la clasificación. Cuando se comparan dos algoritmos diferentes, puede darse el caso de que un algoritmo es mejor que el otro con respecto al promedio del porcentaje de casos bien clasificados.

Se han propuesto diferentes pruebas estadísticas para medir la diferencia entre dos paradigmas de clasificación y determinar, si dicha diferencia es estadísticamente significativa. Una de las pruebas estadísticas entre dos algoritmos es *k-fold cross-validation paired t-test* propuesto por [Dietterich, 1998].

Esta prueba asume normalidad en los datos. En el caso de que los datos no cumplan esta premisa de normalidad, conviene utilizar pruebas no paramétricas, como la prueba estadística de [Wilcoxon, 1945] o la prueba estadística conocida como prueba U de [Mann y Whitney, 1947], especialmente cuando se tienen muestras pequeñas ($n < 10$).

Prueba de Wilcoxon de los rangos con signo

La prueba estadística de Wilcoxon, también conocida como la prueba de Wilcoxon de los rangos con signo [Wilcoxon, 1945], es un método no paramétrico, alternativo a la prueba t de Student, para comparar dos muestras relacionadas y determinar si existen diferencias entre ellas.

La prueba de Wilcoxon de los rangos con signo se utiliza para contrastar datos pareados. En este caso la hipótesis nula del test se formula de la manera siguiente: la media de las distribuciones de probabilidad del porcentaje de casos bien clasificados es igual en el algoritmo A y en el B .

La prueba de Wilcoxon de los rangos con signo consiste en calcular la diferencia de las observaciones $p_i = a_i - b_i$ de los algoritmos a comparar para $i = 1, \dots, k$ casos. Posteriormente se calcula y ordena el valor absoluto de las diferencias de las observaciones $|p_i|$ y se les asigna el correspondiente rango R_i . Finalmente se calcula la suma de rangos positivos (R_i^+) y la suma de rangos negativos (R_i^-). La prueba de Wilcoxon es calculada de la siguiente forma:

$$\begin{aligned} W^+ &= \sum_{i=1}^k \left(\phi(p_i > 0)R_i + \frac{1}{2}\phi(p_i = 0)R_i \right) \\ W^- &= \sum_{i=1}^k \left(\phi(p_i < 0)R_i + \frac{1}{2}\phi(p_i = 0)R_i \right) \\ W &= \min(W^+, W^-) \end{aligned} \quad (2.8)$$

donde $\phi(X)$ es una función que presenta el valor de 1 si X es verdadera y el valor de 0 si no lo es. Cuando k es pequeño, los valores críticos para W , son tabulados como incrementos de k . El test estadístico tiende bajo la hipótesis nula, a una distribución normal con $\mu_W = \frac{k(k+1)}{4}$ y $\sigma_W = \sqrt{\frac{k(k+1)(2k+1)}{24}}$.

Prueba de Mann-Whitney para muestras independientes

La prueba de Mann-Whitney es también una prueba no-paramétrica para comparar si dos muestras provienen de distribuciones diferentes. Es por ello que el proceso de validación cruzada de ambos algoritmos, A y B , puede hacerse con particiones de subconjuntos independientes.

Se parte de la hipótesis nula basada en que ambas muestras A_1, \dots, A_n y B_1, \dots, B_m , provienen de distribuciones de probabilidad del porcentaje de casos bien clasificados con una media igual.

La prueba de Mann-Whitney ordena conjuntamente todos los valores de precisión de los algoritmos A y B en sentido creciente, asignándoles un rango, corrigiendo los empates con el rango medio.

Sea R_A y R_B la suma de rangos para las estimaciones del porcentaje de casos bien clasificados obtenidos por el algoritmo A y B respectivamente. La prueba de Mann-Whitney calcula los valores U_A y U_B de la siguiente forma:

$$\begin{aligned} U_A &= \eta_1 \eta_2 + \frac{\eta_1(\eta_1 + 1)}{2} - R_A \\ U_B &= \eta_1 \eta_2 + \frac{\eta_2(\eta_2 + 1)}{2} - R_B \end{aligned} \tag{2.9}$$

donde η_1 y η_2 denotan el tamaño de la muestra aplicado al algoritmo A y al algoritmo B respectivamente. En el caso de muestras pequeñas, se elige el valor del estadístico U más pequeño y se compara con los valores críticos de U Mann-Whitney. Mientras que en el caso de muestras mayores a 25, el test estadístico tiende bajo la hipótesis nula a una distribución normal con $\mu_U = \frac{\eta_1 \eta_2}{2}$ y $\sigma_U = \sqrt{\frac{\eta_1 \eta_2 (\eta_1 + \eta_2 + 1)}{12}}$.

Existen otras pruebas no paramétricas como la prueba de Kruskal-Wallis [Kruskal y Wallis, 1952] para comparar K muestras no relacionadas, la prueba de Friedman [Friedman, 1937] para comparar K muestras relacionadas (la cual puede considerarse como una extensión de la prueba de Wilcoxon para el caso de más de dos muestras), el coeficiente de correlación de Spearman para rangos [Spearman, 1904], etc. En este trabajo se aplicó la prueba de Kruskal-Wallis para comparar los algoritmos (más de 2) aplicados en un mismo estudio.

Prueba de Kruskal-Wallis para comparar K muestras

La prueba no paramétrica propuesta por [Kruskal y Wallis, 1952] es una extensión de la prueba de los rangos con signo de Wilcoxon para más de dos grupos y una versión no-paramétrica del análisis de la varianza de un factor. La prueba de Kruskal-Wallis no asume normalidad en los datos, en oposición al tradicional ANOVA.

La prueba de Kruskal-Wallis realiza la comparación de las medias de los g grupos (clasificadores) de forma simultánea. Primeramente se ordenan todos los datos de la muestra de menor a mayor. En el caso de datos que se repiten, se asigna el rango promedio a cada uno de ellos. La prueba de Kruskal-Wallis es calculada de la siguiente forma:

$$KW = (N - 1) \frac{\sum_{i=1}^g n_i (\bar{r}_{i\cdot} - \bar{r})^2}{\sum_{i=1}^g \sum_{j=1}^{n_i} (r_{ij} - \bar{r})^2} \quad (2.10)$$

con

$$\bar{r}_{i\cdot} = \frac{\sum_{j=1}^{n_i} r_{ij}}{n_i}$$

$$\bar{r} = (N + 1)/2$$

donde g , denota el número de grupos, siendo n_i el número de observaciones en el grupo i , r_{ij} representa el rango de la observación j en el grupo i y N denota el número total de medidas entre todos los grupos. Finalmente, el *p-value* es aproximado por $\Pr(\chi^2_{g-1} \geq KW)$. La distribución de KW puede ser distinta de la chi-cuadrado cuando se tiene en alguno de los grupos un número de observaciones inferior a 5.

[Demšar, 2006] concluye que no existe un método estándar para comparar el rendimiento entre algoritmos de aprendizaje y que las pruebas no paramétricas son las pruebas más adecuadas en el caso de que no podamos garantizar que los resultados siguen una distribución normal ni que haya homogeneidad de la varianza. Recomienda usar el test de Wilcoxon [Wilcoxon, 1945] en caso de que tengamos dos algoritmos a comparar.

En este trabajo se aplicaron pruebas estadísticas no-paramétricas para comparar dos algoritmos de clasificación diferentes, inducidos con una misma base de datos, siguiendo las recomendaciones de [Demšar, 2006], en el caso de que los datos no estén normalmente distribuidos.

Prueba de McNemar

La prueba no paramétrica de McNemar [Dietterich, 1998] es una medida de la discrepancia entre los recuentos esperados y los observados y puede considerarse como una prueba de signos.

B A	acuerdo	error
acuerdo	n_{11}	n_{10}
error	n_{01}	n_{00}

Tabla 2.2. Matriz de McNemar en la que se registra el conteo de discrepancia entre el valor de la clase predicho y el valor de la clase real en la clasificación realizada por los clasificadores A y B .

Para evaluar un par de clasificadores A y B existen claramente cuatro tipos de pares de observaciones como se muestra en la Tabla 2.2, en la que se indica el conteo del número de casos en que el clasificador A es correcto pero B no, se denota mediante n_{10} , y el número de casos en que B es correcto pero A no lo es, es denotado por n_{01} . Cuando A y B tienen un comportamiento

semejante no se espera ninguna discrepancia entre ambos clasificadores. Es ese caso se esperaría que las dos entradas fuera de diagonal de la Tabla 2.2 fuesen $\frac{1}{2}(n_{01} + n_{10})$. La prueba de McNemar es calculada por medio del estadístico:

$$T = \frac{(n_{01} - n_{10})^2}{n_{01} + n_{10}} \quad (2.11)$$

Si se aplica la corrección de continuidad (recomendable sobre todo si el tamaño de muestra es pequeño o hay celdas con frecuencias pequeñas), la fórmula anterior se modifica ligeramente:

$$T = \frac{(|n_{01} - n_{10}| - 1)^2}{n_{01} + n_{10}} \quad (2.12)$$

donde T está distribuida aproximadamente como la distribución χ^2 con 1 grado de libertad. Se necesita calcular $\chi^2_{1,0,95}$ y compararlo con T . Haciéndolo así, con un nivel de significación del 5 % se puede rechazar la hipótesis nula si $T > 3,8414$.

2.3.4. Nivel de concordancia entre clasificadores

El índice Kappa introducido por [Cohen, 1960] es una medida de fiabilidad que proporciona una escala de concordancia entre un par de clasificadores al clasificar respectivamente N casos en K categorías mutuamente exclusivas de la variable clase C . El índice Kappa únicamente tiene en consideración si hay o no acuerdo, es decir, si se clasifica o no la instancia en la misma clase.

Kappa	grado de acuerdo
< 0	sin acuerdo
(0 - 0,2]	insignificante
(0,2 - 0,4]	bajo
(0,4 - 0,6]	moderado
(0,6 - 0,8]	bueno
(0,8 - 1]	muy bueno

Tabla 2.3. Escala de [Landis y Koch, 1977] para medir el grado de concordancia del índice Kappa.

En el caso de que exista independencia entre ambos métodos de clasificación, la probabilidad de que coincidan clasificando al mismo sujeto como positivo será entonces el producto de las dos probabilidades (sucesos independientes). Aplicando el mismo razonamiento se calcula la probabilidad de que se produzca un acuerdo entre los métodos al clasificar a un caso como negativo, y entonces la probabilidad de acuerdo en cualquiera de las dos clasificaciones será la suma de ambos valores. Con el fin de determinar hasta qué punto la concordancia observada es superior a la que es esperable obtener

por puro azar, se define el índice de concordancia Kappa K de la siguiente manera:

$$K = \frac{p(a) - p(b)}{1 - p(b)}$$

donde $p(a)$ es un estimador de la probabilidad de que dos clasificadores coincidan, y $p(b)$ es la probabilidad de que el acuerdo sea aleatorio. El índice K tiene un valor de 1 cuando un par de clasificadores están en total acuerdo, y los valores menores que cero (negativos) de K cuando no hay acuerdo en la clasificación de la misma instancia por un par de clasificadores. En base a la escala propuesta por [Landis y Koch, 1977] para valorar el grado de acuerdo entre los clasificadores (véase la escala de la Tabla 2.3), también se puede considerar un grado de concordancia insignificante en un rango de $(0 - 0, 2]$, un grado bajo en un rango de $(0, 2 - 0, 4]$, un grado de concordancia moderado, bueno y muy bueno en los rangos $(0,4 - 0,6]$, $(0,6 - 0,8]$ y $(0,8 - 1]$ respectivamente.

2.4. Selección de variables

En el dominio médico como en otros campos de investigación, es común tener un volumen importante de información el cual es manejado desde el campo de la minería de datos y el aprendizaje automático por medio de la *selección de un subconjunto de variables* denominado FSS (*feature subset selection*) [Kohavi y John, 1997]. El problema de la selección de un subconjunto de variables, puede plantearse en términos de analizar las variables del dominio del problema y determinar si aportan información que disminuya la incertidumbre sobre las demás variables, ya que no necesariamente todas las variables implicadas son *relevantes*. Por medio del proceso de la selección de variables se reduce el número de variables del dominio del problema al descartar las variables con información *irrelevante o redundante*.

Se considera que una variable predictora es *irrelevante* cuando el conocimiento del valor de la misma no aporta información nueva a la variable clase. Una variable predictiva es *redundante* cuando su valor puede ser calculado o determinado a partir de otras variables predictoras. Ambos tipos de variables pueden decrementar la precisión del clasificador.

Las ventajas que supone la aplicación de la selección de un subconjunto de variables para la inducción de los clasificadores son la mejora de la precisión del clasificador así como el tiempo de cómputo, lo que permite crear modelos parsimoniosos, es decir que si se dispone de dos modelos que explican suficientemente bien los datos, se debe escoger el modelo más simple de los dos. Otra ventaja, es la reducción de costo que representa la adquisición de información innecesaria y finalmente, al plantearse un problema menos complejo se reduce la complejidad de los modelos de clasificación y a su vez se facilita la comprensión de los modelos por los expertos.

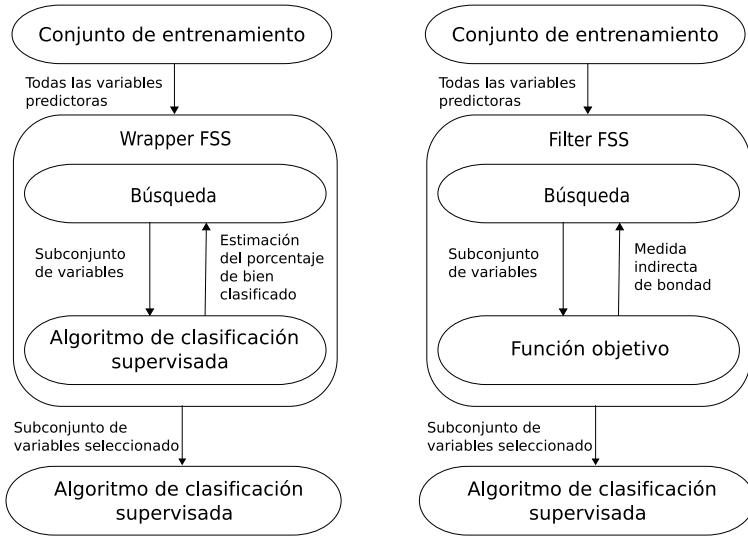


Figura 2.1. Esquemas generales de las aproximaciones *wrapper* y *filter* en el proceso de selección de un subconjunto de variables (FSS).

Una de las características deseables en un clasificador es que sea exacto, y es este, precisamente uno de los objetivos de la selección del subconjunto de variables. Para la construcción del clasificador previamente un algoritmo deberá seleccionar el subconjunto de variables de entre los casos del conjunto de entrenamiento y descartar algunas variables predictoras que resultan ser irrelevantes y/o redundantes.

Los modelos de clasificación supervisada pueden dividirse en dos categorías dependiendo de la técnica de selección de variables que se aplique: enfoque directo o *wrapper* y enfoque indirecto o *filter*.

Enfoque directo o wrapper

En los años 90s [John y col., 1994] reportan que cuando el objetivo es maximizar la precisión del clasificador, el proceso de la selección de variables no debe restringirse únicamente al aprendizaje de los datos si no que también se deben considerar las características del clasificador. En el enfoque *wrapper*, el algoritmo de clasificación es usado para evaluar la mejora en la precisión del modelo inducido con el subconjunto de variables seleccionadas.

En la aproximación directa o *wrapper* [Kohavi y John, 1997] cada subconjunto de variables es determinado por un algoritmo de búsqueda, el cual está implícito en la construcción del clasificador. El algoritmo de búsqueda es guiado por una función de mérito o porcentaje de casos bien clasificados

obtenida en base a la validación cruzada de k particiones del conjunto de entrenamiento.

En esta aproximación se han utilizado diferentes algoritmos de búsqueda heurísticos como *sequential forward selection* y *sequential backward elimination* [Kittler, 1978]. Estos algoritmos determinísticos, dadas unas condiciones iniciales, alcanzan una solución final, que no siempre es la óptima. Como alternativa se han propuesto los algoritmos de búsqueda no-determinísticos como son el algoritmo de búsqueda voraz [Debuse y Rayward-Smith, 1999], los algoritmos genéticos [Yang y Honavar, 1998] y algoritmos de estimación de distribuciones [Inza y col., 2001].

Enfoque indirecto o filter

Como se mencionó anteriormente el objetivo general del proceso de la selección de un conjunto de variables para inducir el modelo es mejorar la precisión del clasificador. Funciones como logaritmo de la verosimilitud (*log-likelihood*) de los datos y las medidas basadas en la información (como la entropía condiciona a la clase o información mutua) han sido ampliamente utilizadas para selección de variables. Estas medidas son conocidas como *medidas filter* y el proceso de la selección de un conjunto de variable basadas en estas medidas es conocido como aproximación indirecta o *filter*.

La aproximación indirecta o *filter* [Blum y Langley, 1997, Lewis, 1998] tiene en cuenta las características de los datos y aplica medidas *filter* para determinar las relaciones existentes entre pares de variables. Inicialmente se lleva a cabo la selección de un subconjunto de variables para después inducir el clasificador. Existen otras medidas *filter* como la entropía de Shanon, distancia Euclídea, métrica Matusita, divergencia de Kullback-Leiber o métrica Bhattacharyya, todas ellas descritas en [Ben-Bassat, 1982] para variables dicotómicas.

Las diferencias entre las aproximaciones *wrapper* y *filter* presentadas en esta sección se resumen en la Figura 2.1.

La información mutua o *mutual information* [Shannon, 1948] es una de las medidas *filter* más ampliamente utilizadas. En base a la teoría de la información, calcula la relación que existe entre una variable y otra por medio de la verosimilitud de los datos. Esta medida indica la cantidad de incertidumbre que el conocimiento de una variable predictora es capaz de despejar con respecto al estado en el que se encuentre una segunda variable.

El ranking de variables es otra medida *filter* y es una de las métricas para establecer una escala de importancia (rankings) de las variables. Estos rankings son realizados de forma univariada, es decir, tan solo se tiene en cuenta la relación existente entre la variable predictora que está siendo analizada y la variable clase. Siempre las primeras variables de la escala han de ser las más importantes dentro del conjunto de variables a seleccionar.

Clasificadores Bayesianos

3.1. Introducción

Los modelos gráficos probabilísticos son modelos especialmente diseñados para manejar la incertidumbre, la cual caracteriza a la mayoría de problemas de la vida real. Esta característica, conjuntamente con su representación gráfica, la cual resulta intuitiva para el experto, facilita la comprensión subyacente a la relación entre las variables del modelo. Esto explica porqué expertos los han aplicado en diferentes problemas del dominio médico [Sierra y col., 2001, Blanco y col., 2005].

Las redes Bayesianas [Pearl, 1988] son grafos acíclicos dirigidos donde los nodos representan variables y los arcos representan las relaciones condicionales de (in)dependencia entre las variables. Las redes Bayesianas están basadas en modelos gráficos probabilísticos donde cada variable aleatoria sigue una función de probabilidad condicional dado un valor específico de sus padres.

En el caso de que las variables predictoras sean continuas se emplean las redes Gaussianas [Heckerman y col., 1995] y en el caso de que se tengan variables discretas y continuas se aplican las *redes condicionales Gaussianas* [Lauritzen, 1996, Bottcher, 2004].

Los clasificadores Bayesianos son modelos particulares de las redes Bayesianas que se aplican en la tarea de clasificación supervisada con datos discretos. Han sido ampliamente utilizados en el campo de la minería de datos y el aprendizaje automático obteniendo buenos resultados. Se han propuesto diferentes clasificadores Bayesianos dependiendo de las relaciones de dependencia contempladas entre las variables predictoras.

En esta sección se describirán los clasificadores Bayesianos ampliamente utilizados en la literatura. Paradigmas como naive Bayes [Minsky, 1961], selective naive Bayes [Langley y Sage, 1994], semi naive Bayes [Pazzani, 1997] así como los enfoques directo e indirecto para la inducción de los mismos, naive Bayes aumentado a árbol (TAN) [Friedman y col., 1997] y el clasificador Bayesiano k -dependiente (k DB) [Sahami, 1996] diseñados especialmente para problemas con variables predictoras discretas de clasificación supervisada. En

el caso de variables predictoras continuas se presentan las redes condicionales Gaussianas [Lauritzen, 1996, Bottcher, 2004] para la inducción de los clasificadores Bayesianos mencionados anteriormente adaptados por [Pérez y col., 2006]. En este trabajo no se consideran multiredes Bayesianas, se recomienda consultar [Larrañaga, 2003] para mayor información.

Se puede establecer una jerarquía en base a la complejidad de la estructura de los clasificadores Bayesianos, dependiendo de las relaciones permitidas entre todas las variables predictoras. En dicha jerarquía el clasificador naive Bayes es el modelo más simple entre los clasificadores Bayesianos, dado a que se asume la independencia condicional entre variables predictoras dada la variable clase.

3.2. Naive Bayes

El clasificador *naive Bayes* [Minsky, 1961] es el modelo de clasificación construido bajo la premisa de que todas las variables predictoras son condicionalmente independientes dada la variable clase C . Sin embargo en el dominio de problemas reales esta asunción no se cumple en numerosas ocasiones. A pesar de ello el paradigma naive Bayes en algunas ocasiones ha mostrado un buen funcionamiento en problemas de este tipo [Domingos y Pazzani, 1997, Hand y Yu, 2001].

El clasificador *naive Bayes* [Minsky, 1961] aplica el teorema de Bayes para predecir para cada instancia \mathbf{x} , la clase $c \in C$ con la máxima probabilidad *a posteriori*:

$$p(c|\mathbf{x}) \propto p(c, \mathbf{x}) = p(c) \prod_{i=1}^n p(x_i|c) \quad (3.1)$$

donde $p(x_i|c)$ representa la probabilidad condicional de $X_i = x_i$ dado que $C = c$. Como resultado el clasificador naive Bayes predice en base a:

$$c^* = \arg \max_c p(c) \prod_{i=1}^n p(x_i|c) \quad (3.2)$$

En este paradigma de clasificación todas las variables predictoras están incluidas en el modelo. La Figura 3.1(a) muestra un ejemplo de la estructura de un clasificador naive Bayes con cinco variables predictoras.

Los clasificadores Bayesianos se caracterizan por representar la distribución de probabilidad conjunta $p(x_1, \dots, x_n, c)$ entre todas las variables. La estructura completa de los clasificadores Bayesianos pueden ser definida por la factorización $p(c, \mathbf{x}) = p(c)p(\mathbf{x}|c)$.

Los clasificadores Bayesianos basados en redes condicionales Gaussianas están definidos por la factorización $f(c, \mathbf{x}) = p(c) \prod_{i=1}^n f(x_i|\mathbf{pa}_i, c)$ donde \mathbf{pa}_i representa un valor de \mathbf{Pa}_i un conjunto de variables que junto con la variable clase C son padres de X_i . Suele ser habitual utilizar una variable

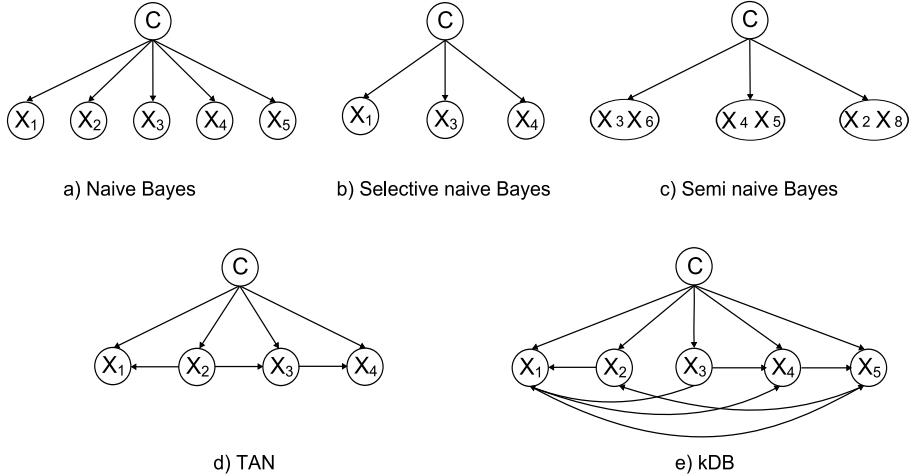


Figura 3.1. Ejemplos de estructuras de clasificadores Bayesianos.

aleatoria normal (para cada valor de C) para modelar el comportamiento de la variable X_i . Se asume que la factorización de cada variable dado sus padres sigue una distribución normal y por lo tanto la función de densidad $f(x_i|\mathbf{pa}_i, c)$ representa la probabilidad de la variable dado sus padres. Es decir, $\forall c \text{ y } \forall i \in 1, \dots, n$ asumimos que $f(X_i|C = c) \sim N(\mu_i^c, \sigma_i^c)$. La probabilidad a posteriori se calcula de la siguiente manera:

$$p(c|\mathbf{x}) \propto p(c, x_1, \dots, x_n) \propto p(c, \mathbf{x}) = p(c) \prod_{i=1}^n f_{N(\mu_i^c, \sigma_i^c)}(x_i|c) \quad (3.3)$$

Filter Raking Naive Bayes

El clasificador filter ranking naive Bayes (FRankingNB) fue propuesto por [Pérez y col., 2006] como una mejora del modelo naive Bayes clásico por medio de la selección de variables continuas, en base al cálculo de la información mutua de cada variable predictora y la variable clase $I(X_i, C)$. A continuación se calculan las probabilidades en base a la frecuencia de las variables y se ordenan dichas cantidades de información mutua, siendo las primeras las de mayor frecuencia. Es necesario fijar un valor o umbral de corte τ con respecto a la distribución $I(X_i, C)$ para llevar a cabo el proceso de selección de variables predictoras.

Únicamente las variables que superan el valor, $I(X_i, C) \geq \tau$, son incluidas en el modelo. Entonces la estructura del clasificador naive Bayes es construida con las primeras m variables en la lista ordenada en base a la frecuencia ($m = 1, 2, \dots, n$). De esta forma se eliminan las variables irrelevantes del modelo

final del clasificador naive Bayes. La probabilidad a posteriori es definida de la siguiente forma:

$$f(c, \mathbf{x}) = p(c) \prod_{i=1}^m f(x_i|c) \quad (3.4)$$

3.3. Selective naive Bayes

A pesar del éxito del clasificador naive Bayes en la mayoría de problemas la precisión de los algoritmos de clasificación decrece debido a las variables irrelevantes del dominio del problema o bien debido a variables redundantes.

El clasificador *selective naive Bayes* [Langley y Sage, 1994, Kohavi y John, 1997] aplica la selección del subconjunto de variables (FSS) al clasificador naive Bayes y es capaz de detectar variables redundantes o irrelevantes, aunque es similar al clasificador naive Bayes en el sentido de que no detecta dependencias entre las variables predictoras.

En la literatura se encuentran dos formas de inducir el clasificador selective naive Bayes: una de ellas es iniciar con un conjunto vacío de variables predictoras para posteriormente integrar al conjunto de variables alguna de ellas, una a una. Otra forma es ir incorporando de una a una variables predictoras del conjunto formado por todas las variables del dominio del problema en cada iteración hasta que no haya mejora en la clasificación.

El algoritmo *wrapper* para la selección de variables hacia adelante es un ejemplo de este último caso para inducir el clasificador selective naive Bayes, el cual se inicia con un conjunto de variables vacío al cual se va añadiendo en cada iteración la variable que mejore el comportamiento del clasificador calculado en base al porcentaje de casos bien clasificados. El algoritmo deja de incluir variables cuando no se alcanzan mejores resultados de clasificación.

La factorización de la distribución de probabilidad conjunta para el clasificador selective naive Bayes se define de la siguiente manera:

$$p(c, \mathbf{x}) = p(c) \prod_{X_i \notin \mathbf{X}_F} p(x_i) \prod_{X_j \in \mathbf{X}_F} p(x_j|c) \quad (3.5)$$

donde \mathbf{X}_F es el subconjunto de variables predictoras seleccionadas. En el caso de los modelos del clasificador selective naive Bayes, la probabilidad condicional de la clase dadas las variables predictoras está definida como:

$$p(c|\mathbf{x}) = p(c|\mathbf{x}_F) \propto p(c, \mathbf{x}_F) = p(c) \prod_{X_j \in \mathbf{X}_F} p(x_j|c) \quad (3.6)$$

La Figura 3.1(b), muestra un ejemplo de la estructura de un clasificador selective naive Bayes, donde a una instancia $\mathbf{x} = (x_1, x_2, x_3, x_4, x_5)$, se le puede asignar la variable clase en base a la siguiente definición:

```

FSSJ
Paso 1: Sea la variable,  $S$ , una lista vacía
Paso 2: Repetir hasta que no se presente ningún incremento en el comportamiento del clasificador
  2.1: Seleccionar la mejor opción
    (a): Considerar cada variable predictora que no esté en  $S$  como una nueva variable
          condicionalmente independiente de la variable clase
    (b): Considerar la fusión de cada variable predictora que no esté en  $S$ 
          con la variable predictora incluida en  $S$ 
  2.2: Incluir la variable seleccionada a  $S$ 

```

Figura 3.2. Pseudocódigo del algoritmo FSSJ [Pazzani, 1997].

$$c^* = \arg \max_c p(c)p(x_1|c)p(x_3|c)p(x_4|c) \quad (3.7)$$

3.4. Semi naive Bayes

El algoritmo selective naive Bayes es capaz de detectar variables redundantes e irrelevantes, sin embargo no considera la dependencia entre las variables predictoras y asume la independencia incondicional con respecto a la variable clase entre las variables que construyen el modelo. Comúnmente, en la mayoría de los problemas reales existe relación entre las variables del dominio del problema. Por esta razón, es necesario tener en cuenta otro tipo de clasificadores Bayesianos capaces de superar la asunción de independencia condicional entre las variables predictoras como es el caso del clasificador *semi naive Bayes* [Kononenko, 1991] que es capaz de detectar la relación entre las variables predictoras. [Kononenko, 1991] propone la definición de un nuevo tipo de nodo en la estructura del clasificador *semi naive Bayes* –una variable unión– construida a partir del producto cartesiano de un subconjunto de variables. Este nuevo nodo permite superar la asunción de independencia condicional entre las variables [Pazzani, 1997]. Cada nodo unión representa una nueva variable que considera todas las dependencias entre todas las variables predictoras originales que lo forman.

El enfoque propuesto por [Pazzani, 1997] para la inducción del clasificador semi naive Bayes consiste en construir un clasificador naive Bayes a través de un algoritmo voraz capaz de detectar variables irrelevantes y variables dependientes entre si. En este trabajo, Pazzani propone dos algoritmos voraces de búsqueda: FSSJ (*forward sequential selection and joining*) que realiza una búsqueda hacia delante y BSEJ (*backward sequential elimination and joining*), que realiza una búsqueda hacia atrás. La búsqueda voraz de cada algoritmo está guiada por la estimación del porcentaje de casos bien clasificados de la solución evaluada por validaciones *10-fold cross validation* o *leave-one-out validation* [Stone, 1974].

La inducción del clasificador semi naive Bayes con el algoritmo voraz FSSJ, se inicia con un conjunto de variables vacío al cual se le añaden variables en cada iteración o nuevas variables creadas por el producto cartesiano con

BSEJ

```

Paso 1: Sea  $S$  una lista de variables que contiene todas las variables predictoras
Paso 2: Repetir hasta que no se produzca mejora en la clasificación
  2.1: Seleccionar la mejor opción
    (a): Considerar el reemplazo de cada par de variables predictoras
        contenidas en  $S$  con una nueva variable producida por el
        producto cartesiano de las dos variables
    (b): Considerar la eliminación de cada variable predictora contenida en  $S$ 
  2.2: Eliminar o reemplazar la(s) variable(s) seleccionadas de  $S$ 

```

Figura 3.3. Pseudocódigo del algoritmo BSEJ [Pazzani, 1997].

las variables existentes en el conjunto, hasta mejorar el comportamiento del clasificador en términos de la precisión del mismo. Todas las instancias son clasificadas con la clase más probable a posteriori. En la Figura 3.2 se muestra el pseudocódigo del algoritmo FSSJ.

El segundo algoritmo, BSEJ, como se puede observar en la Figura 3.3 comienza con un clasificador con todas las variables predictoras considerándolas condicionalmente independientes y sigue con una aproximación semejante a la anterior pero en dirección contraria, hacia atrás. El algoritmo BSEJ considera todas las posibles configuraciones del clasificador, las evalúa utilizando la validación dejando un caso fuera o bien aplicando k-subconjuntos de validación cruzada y presenta el mejor resultado. Termina cuando no se produce mejora en el comportamiento del clasificador.

La Figura 3.1 (c) se muestra un ejemplo de una posible configuración de la estructura del clasificador semi naive Bayes. Considerando esta estructura la instancia $\mathbf{x} = (x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8)$ un modelo semi naive Bayes asigna la clase c de la siguiente manera:

$$c^* = \arg \max_c p(c)p(x_3, x_6|c)p(x_4, x_5|c)p(x_2, x_8|c) \quad (3.8)$$

[Pérez y col., 2006] propusieron una adaptación al algoritmo [Kononenko, 1991] para inducir el clasificador semi naive Bayes en el dominio continuo con redes condicionales Gaussianas. Las redes Gaussianas son inducidas con un conjunto de variables continuas, las cuales componen las nuevas variables de unión. Estas nuevas variables fusionadas siguen una distribución normal [Anderson, 1958]. La función de densidad de la variable unión condicionada a la variable clase viene dada por:

$$f(\mathbf{y}_k|c) = (2\pi)^{-\frac{1}{2}m_k} \left| \sum_k^c \right|^{-\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{y}_k - \boldsymbol{\mu}_k^c)^t (\Sigma_k^c)^{-1} (\mathbf{y}_k - \boldsymbol{\mu}_k^c)} \quad (3.9)$$

donde m_k es el número de variables que componen la nueva variable predictora fusionada, la matriz de covarianza de \mathbf{Y}_k condicionada a la clase C es representada por Σ_k^c , y $\boldsymbol{\mu}_k^c$ es el vector condicionado al valor de la clase de la variable fusionada \mathbf{Y}_k .

3.5. Naive Bayes aumentado a árbol

Naive Bayes aumentado a árbol (TAN) [Friedman y col., 1997] es otro tipo de clasificador Bayesiano que permite dependencias entre las variables predictoras. La principal restricción entre las dependencias de las variables es que cada variable predictora únicamente puede tener un máximo de dos padres: la variable clase C y alguna otra variable predictora X_1, \dots, X_n .

Para construir la estructura del clasificador naive Bayes aumentado a árbol se necesita previamente aprender las dependencias entre las diferentes variables predictoras X_1, \dots, X_n . En [Friedman y col., 1997] se presenta un algoritmo en el cual se tiene en cuenta la cantidad de información mutua condicionada a la variable clase, en lugar de la cantidad de información mutua en la que se basa el algoritmo [Chow y Liu, 1968]. La cantidad de información mutua entre las variables discretas (X_i, X_j) condicionada a la variable clase C se define como:

$$I(X_i, X_j | C) = \sum_c \sum_{x_i} \sum_{x_j} p(x_i, x_j, c) \log \frac{p(x_i, x_j | c)}{p(x_i | c)p(x_j | c)} \quad (3.10)$$

Tal y como puede verse en la Figura 3.4 el algoritmo para la construcción del modelo TAN requiere de varios pasos. Primeramente se calculan las cantidades de información mutua para cada par de variables condicionadas a la variable clase $I(X_i, X_j | C)$. Una vez calculados estos valores, la estructura de la red Bayesiana del clasificador TAN se construye en dos fases. La primera se inicia con la construcción de un grafo no dirigido completo con n nodos, uno por cada variable predictora, X_1, \dots, X_n , en el cual el peso de cada arista viene dado por la cantidad de información mutua entre las dos variables unidas por la arista condicionada a la variable clase C . El algoritmo asigna por tanto un peso $I(X_i, X_j | C)$ a las aristas que las conectan X_i y X_j . El árbol es construido con las aristas con mayor información mutua condicionada. Este proceso es repetido anexando a la estructura de la red las aristas con mayor información mutua a no ser que se formen ciclos. En este caso las aristas son descartadas, en cuyo caso la siguiente arista con mayor peso es analizada. El procedimiento termina cuando $n - 1$ aristas han sido seleccionadas. Finalmente el grafo no dirigido se convierte en uno dirigido escogiendo una variable predictora de forma aleatoria como padre.

En la segunda fase la estructura del clasificador es aumentada a una estructura del clasificador naive Bayes agregando el nodo de la variable clase C , y conectando con un arco el nodo clase C a cada variable predictora X_i ($i = 1, \dots, n$).

La Figura 3.1(d) muestra un ejemplo de una estructura del clasificador TAN donde a una instancia $\mathbf{x} = (x_1, x_2, x_3, x_4)$ le asigna la clase c de la siguiente manera:

$$c^* = \arg \max_c p(c)p(x_1|c, x_2)p(x_2|c)p(x_3|c, x_2)p(x_4|c, x_3) \quad (3.11)$$

```

TAN
Paso 1: Calcula  $I(X_i, X_j | C)$  con  $i < j$ ,  $i, j = 1, \dots, n$ 
Paso 2: Construir un grafo no dirigido completo cuyos nodos correspondan a las
variables predictoras  $X_1, X_2, \dots, X_n$ . Asignar a cada arista conectando las variables
 $X_i$  y  $X_j$  un peso dado por  $I(X_i, X_j | C)$ 
Paso 3: A partir del grafo completo anterior y siguiendo el algoritmo de Kruskal
construir un árbol expandido de máximo peso
Paso 4: Transformar el árbol no dirigido resultante en uno dirigido escogiendo
una variable como nodo raíz para a continuación dirigir el resto de las aristas
Paso 5: Construir un modelo TAN añadiendo un nodo etiquetado como  $C$  y posteriormente
un arco desde  $C$  a cada variable predictora  $X_i$ 

```

Figura 3.4. Pseudocódigo de la inducción del clasificador TAN por medio de la adaptación del algoritmo de Chow-Liu [Friedman y col., 1997].

Como una extensión, [Pérez y col., 2006] proponen una aproximación diferente para la construcción del clasificador naive Bayes aumentado a árbol denominado *wrapper tree augmented naive Bayes*, el cual está basado en la modificación del algoritmo propuesto por [Keogh y Pazzani, 1999] al dominio continuo. Proponen el algoritmo de búsqueda voraz denominado ascensión de la colina (*hill climbing*) para realizar la búsqueda en todo el espacio de posibles estructuras del clasificador, la cual es guiada por el porcentaje de bien clasificados. Para adaptar el algoritmo TAN al dominio continuo los autores proponen calcular la información mutua entre cada par de variables X_i y X_j condicionadas a la variable clase C de la siguiente manera:

$$I(X_i, X_j | C) = \frac{1}{2} \sum_{c=1}^r p(c) \log(1 - \rho_c^2(X_i, X_j)) \quad (3.12)$$

Siendo ρ_c el coeficiente de correlación entre la variable X_i y la variable X_j condicionado a la variable clase C .

3.6. Clasificador Bayesiano k -dependiente

De todos los clasificadores revisados anteriormente solo el modelo TAN –y en menor medida el semi naive Bayes– son capaces de considerar las dependencias entre las variables, aunque TAN tiene la restricción de considerar dos padres para cada variable (una variable como padre además de la variable clase).

[Sahami, 1996] presenta un algoritmo que supera la restricción del máximo de dos padres de cada variable predictora del clasificador TAN, denominado *clasificador Bayesiano k -dependiente* (k DB).

El clasificador k DB permite que cada variable predictora pueda tener k padres sin contar la variable clase. Para inducir el clasificador se debe determinar previamente un valor de k por el usuario, donde k representa el máximo número de padres por variable.

Inicialmente se calcula la información mutua de cada variable X_i y la clase $I(X_i, C)$. Las dependencias entre las variables dada la clase son calculadas por

kDB

Paso 1: Para cada variable predictora X_i , calcular la información mutua con respecto a la clase C , $I(X_i, C)$

Paso 2: Calcular la información mutua condicionada a la clase $I(X_i, X_j | C)$ para cada par de variables con $i < j$, $i, j = 1, \dots, n$

Paso 3: Inicializar una lista vacía de variables, S

Paso 4: Inicializar la red del clasificador k dependiente a construir, kDB , con un único nodo, el de la variable clase, C

Paso 5: Repetir hasta que S incluya todas las variables predictoras

- 5.1: Seleccionar de entre las variables que no están en S , aquella variable X_{max} con mayor cantidad de información mutua respecto a la clase C , $I(X_{max}, C)$
- 5.2: Añadir un nodo a kDB , $X \notin S$ representado por X_{max}
- 5.3: Añadir un arco del nodo C a X_{max} en kDB
- 5.4: Añadir $m = \min(|S|, k)$ arcos desde las m diferentes variables $X_j \in S$ con los mayores valores $I(X_{max}, X_j | C)$
- 5.5: Añadir X_{max} a S

Paso 6: Calcular las probabilidades condicionadas necesarias para especificar la red kDB

Figura 3.5. Pseudocódigo del algoritmo para inducir el clasificador kDB [Sahami, 1996].

la información mutua condicionada, $I(X_i, X_j | C)$ con $i < j$, $j = 2, \dots, n$, como sigue:

$$I(X_i, X_j | C) = \sum_{i=1}^n \sum_{j=1}^m \sum_{c=1}^r p(x_i, y_j, c_r) \log \frac{p(x_i, y_j | c_r)}{p(x_i | c_r)p(y_j | c_r)} \quad (3.13)$$

Tal y como puede verse en la Figura 3.5 la estructura del clasificador inicia únicamente con el nodo de la variable clase C . En cada iteración el algoritmo selecciona el nodo a incluir en la estructura con la mayor información mutua $I(X_i, C)$. A continuación se anexa, la arista del nodo C al nodo de la variable X_i , para más tarde calcular la información mutua condicionada $I(X_i, X_j | C)$ para todas las posibles nuevas aristas desde el nodo X_j hacia los nodos de la estructura. Todas las aristas son arcos ordenados de mayor a menor siendo anexados a la estructura los k arcos con mayor información mutua condicionada. En la Figura 3.1(e) se muestra un ejemplo de una posible estructura del clasificador kDB inducido con esta aproximación considerando un valor de $k=2$ y la instancia $\mathbf{x} = (x_1, x_2, x_3, x_4, x_5)$, la cual es asignada a la clase c de la siguiente manera:

$$c^* = \arg \max_c p(c)p(x_1 | c, x_3)p(x_2 | c, x_1, x_5)p(x_3 | c)p(x_4 | c, x_1, x_3)p(x_5 | c, x_1, x_4) \quad (3.14)$$

3.7. Aproximaciones wrapper y filter para la inducción de clasificadores Bayesianos

En esta sección se describe brevemente la adaptación de los clasificadores selective naive Bayes, TAN y kDB propuestos en [Blanco y col., 2005], los cuá-

les han sido aplicados al problema de la selección embrionaria en tratamientos de reproducción asistida (Capítulo 8).

Bajo la hipótesis de independencia entre la variable predictora X_i y la variable clase C , es conocido que $2NI(X_i, C)$ –donde N denota el tamaño de la base de datos– sigue asintóticamente una distribución χ^2 con $(r_i - 1)(r_0 - 1)$ grados de libertad [McKay, 2003, Cover y Thomas, 1991]. Dada, la información mutua de la variable predictora y el valor de la clase el estadístico χ^2 adaptado por [Blanco y col., 2005] puede verificar si la cantidad de información mutua es significativa. Sin embargo, no se puede aplicar igual que la prueba χ^2 para $2NI(X_i, X_j|C)$ dado que esta distribución no sigue una distribución χ^2 .

La aproximación *filter* del clasificador *selective naive Bayes* (FSNB), se induce únicamente con el subconjunto de variables para las cuales $2NI(X_i, C)$ han superado la prueba χ^2 , desapareciendo las variables irrelevantes del modelo final. Sin embargo la aproximación selective naive Bayes no considera las dependencias entre las variables del dominio del problema.

La aproximación *filter* del paradigma naive Bayes aumentado a árbol (FTAN) es otra adaptación propuesta en [Blanco y col., 2005]. El clasificador FTAN es construido en base a un subconjunto de variables $2NI(X_i, C)$, las cuales han superado el valor $\chi^2_{(r_i-1)(r_0-1);1-\alpha}$ y al mismo tiempo, la estructura del árbol se construye siguiendo el algoritmo propuesto por [Chow y Liu, 1968].

La aproximación *filter* del clasificador Bayesiano k dependiente (FkDB) es similar al FTAN. Se calcula la información mutua y la información mutua condicionada a la variable clase. La estructura del clasificador FkDB es construida con las variables cuyos valores correspondientes de $2NI(X_i, C)$ superaron la prueba estadística χ^2 . Los arcos correspondientes a $2N_c I(X_i, X_j|C = c)$ que superan la prueba para todos los valores de la variable clase $c \in C$ son agregados a la estructura, donde N_c representa el número de casos. Finalmente, en los modelos FTAN y FkDB, los arcos entre la variable clase C y todas las variables presentes en la estructura final del modelo son incluidos.

Combinación de clasificadores

4.1. Introducción

En la literatura se encuentran diversos paradigmas de clasificación supervisada que han sido aplicados a diferentes tipos de problemas. Sin embargo, no existe un clasificador capaz de abarcar el dominio completo de problemas reales y que presente un comportamiento óptimo y confiable en todos ellos. En base a la experiencia de combinar el conocimiento y criterio de los expertos para la toma de decisiones y elección de la respuesta final [Dietterich, 1998] se propone la idea de combinar la predicción de diversos paradigmas de clasificación para aumentar la precisión de la tarea de clasificación con respecto a la precisión obtenida de aplicar un único clasificador a determinado problema.

En este capítulo se introducen algunas alternativas propuestas en la literatura sobre tipos de arquitecturas y métodos para la combinación de clasificadores. Se introduce el estado del arte de las adaptaciones al modelo *stacked* para continuar con la introducción del multi-clasificador *Gaussian-stacking* como una aportación metodológica de esta tesis. Finalmente se describen brevemente los paradigmas de clasificación supervisada aplicados como clasificadores-base para la construcción del multi-clasificador.

4.2. Arquitecturas de combinación de clasificadores

Existen diferentes arquitecturas para configurar la combinación de clasificadores como son: serie, paralela e híbrida o jerárquica, las cuales difieren en como se combinan las salidas de la predicción. La Figura 4.1 muestra un ejemplo de los esquemas de combinación de clasificadores. A continuación se describe cada una de ellas.

- Serie o cascada: La clasificación realizada por los diferentes clasificadores se lleva a cabo de forma secuencial. Primero se inicia con los clasificadores más simples y/o menos costosos computacionalmente y su salida es la entrada del siguiente clasificador hasta obtener la clasificación final.

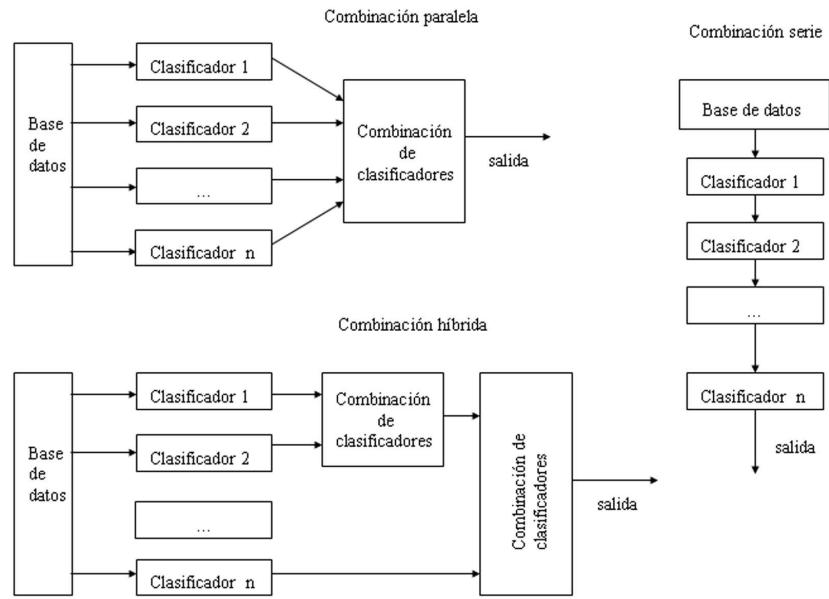


Figura 4.1. Ejemplo de configuraciones de combinación de clasificadores. Configuración en serie, paralela e híbrida.

- **Paralela:** En este tipo de arquitectura se seleccionan las salidas del proceso de clasificación de cada uno de los clasificadores o se les añade un peso antes de ser combinados.
- **Híbrida o jerárquica:** Combina los clasificadores de una forma estructurada, similar a los árboles de clasificación. En las configuraciones anteriores, cada nodo se asocia con un clasificador, teniendo diferentes niveles para la combinación de las salidas de clasificación.

Para combinar las decisiones de varios clasificadores y obtener una decisión final en base a ellas, es necesario aplicar una función de decisión. En la literatura encontramos funciones de decisión muy variadas, las cuales han sido catalogadas en base al tipo de los parámetros de entrada, que en este caso son las salidas de predicción de los clasificadores a combinar.

Los paradigmas de clasificación se pueden combinar, según la naturaleza de la respuesta o resultado de clasificación que ofrecen los clasificadores individualmente en: métodos de votación (clase), lista de preferencia de clase (ranking) y métodos basados en distribuciones de probabilidad de la clase. Como ejemplos de funciones de discriminación entre clases dentro de los métodos de votación, tenemos los siguientes: regla del más votado o mayoría simple, mayoría absoluta y voto por eliminación. El método contador de Borda y la

simple transferencia de votos o votación alternativa, son ejemplos de funciones que hacen uso de un valor de confianza, el cual está sujeto a un ranking de preferencia. Existen métodos basados en valores de confianza como son las distribuciones de probabilidad de la clase. Los métodos probabilísticos basados en el teorema de Bayes y reglas asociadas a probabilidades como: la regla del producto, regla de la suma, regla del mínimo, regla del máximo, regla de la media y aproximación Dempster-Shaper son ejemplos de funciones discriminantes aplicadas para la combinación de clasificadores [Xu y col., 1992].

En la combinación de clasificadores se puede considerar la fiabilidad o confianza que ofrece cada clasificador. La fiabilidad de un clasificador puede expresarse con un peso al que corresponde una ponderación en función del proceso de clasificación. Las funciones que hasta ahora hemos mencionado pueden utilizarse de forma pesada, es decir, incluyendo el cálculo de la fiabilidad global de la clase. Los métodos de votación pesada, mayoría por votación restringida y mayoría pesada mejorada son ejemplos del análisis de confianza de un clasificador.

4.3. Métodos de combinación de múltiples clasificadores

En el amplio campo de investigación de la minería de datos existe un área especializada en la innovación de métodos para combinar las predicciones de múltiples clasificadores, llamados clasificadores-base. Estos métodos de combinación de clasificadores están caracterizados por diferentes arquitecturas o algoritmos de aprendizaje, los cuales proveen de información complementaria en base a sus regiones de decisión para la clasificación. Esta información es combinada por un clasificador en un nivel superior denominado meta-clasificador o multi-clasificador.

Existen distintos métodos para la combinación de clasificadores dependiendo del tipo de clasificadores a combinar como: los métodos basados en hibridaciones, métodos de combinación del mismo clasificador base y métodos de combinación de clasificadores basados en diferentes paradigmas de clasificación. A continuación se mencionan algunos ejemplos de estos métodos.

Combinación de clasificadores a partir de hibridaciones

En los métodos basados en hibridaciones de algoritmos para la combinación de clasificadores, el meta-clasificador se induce teniendo en cuenta dos o más paradigmas de clasificación. En la literatura se encuentran algunos ejemplos de combinación de clasificadores basados en el método de hibridación como son:

- *Lazy Bayesian rules.* Algoritmo propuesto por [Zheng y Webb, 1998], el cual obtiene un clasificador naive Bayes para cada caso a clasificar.

- *Naive Bayes tree.* Método propuesto por [Kohavi, 1996], en el que se induce un árbol de clasificación cuyas hojas son clasificadores naïve Bayes, los cuales son aplicados posteriormente para clasificar los casos que alcancen dichos nodos.
- *Logistic model trees.* Es un algoritmo basado en árboles de clasificación propuesto por [Landwehr y col., 2003], en cuyas hojas se lleva a cabo la clasificación por medio de una regresión logística.

Combinación de clasificadores provenientes del mismo paradigma

Se han desarrollado dos métodos para la combinación de clasificadores provenientes del mismo paradigma de clasificación:

- *Bagging (bootstrap aggregating)* propuesto por [Breiman, 1996a] combina clasificadores obtenidos del mismo paradigma de clasificación a partir de k muestras con reemplazo de un conjunto de datos (método de remuestreo *bootstrap*). Cada clasificador se induce de manera individual, luego se induce el meta-clasificador y se aplica el voto por mayoría para combinar las salidas de los clasificadores inducidos anteriormente. En caso de empate la clase es elegida al azar.
- *Boosting* propuesto por [Freund y Schapire, 1996] combina múltiples clasificadores del mismo paradigma complementándose unos a otros. A diferencia del algoritmo *bagging*, que genera los clasificadores de manera independiente, en *boosting* la distribución de muestreo comienza siendo uniforme y en cada iteración el paradigma agrega un nuevo clasificador. A todos los clasificadores les asigna un peso inicial igual. Cada vez que se genera un clasificador, se cambian los pesos de los nuevos ejemplos usados para inducir el siguiente clasificador, es decir, el nuevo clasificador es inducido considerando el comportamiento de los anteriores en donde la contribución final de cada clasificador está basada en su bondad. Finalmente utiliza el voto por la mayoría, para combinar las salidas de los modelos individuales. El ejemplo más aplicado en la literatura es Adaboost (ADAptive BOOSTing).

Combinación de clasificadores provenientes de diferentes paradigmas

Se han desarrollado diferentes métodos para la *combinación de clasificadores provenientes de diferentes paradigmas* de clasificación. Formalmente denotaremos a un conjunto de clasificadores de distinta base como $D = \{D_1, D_2, \dots, D_h\}$ siendo h el número de clasificadores diferentes.

Ejemplos clásicos de este tipo de métodos son: fusión de etiquetas, fusión de rankings, fusión de salidas continuas, cascada y método de pila.

Método fusión de etiquetas

La combinación de clasificadores en base al método de fusión de etiquetas aplica algoritmos basados en mayoría de votos propuestos por [Littlestone y Warmuth, 1989] como son: pluralidad de voto, mayoría simple, pluralidad de voto con umbral y voto por mayoría con peso.

Método cascada

El método cascada propuesto por [Gama y col., 2000] ordena una secuencia de clasificadores diferentes en serie ($D_1, D_2, \dots, D_h, D_{h+1}, \dots$) en función ascendente del espacio, tiempo, coste de representación, complejidad, etc. Se usa el clasificador D_h para clasificar la nueva instancia $\mathbf{x} = (x_1, \dots, x_n)$ en el caso de que clasificadores anteriores D_1, \dots, D_{h-1} no hayan proporcionado la suficiente confianza al clasificar dicha instancia.

Método de pila general

El modelo de pila (*stacked generalization*) -o simplemente *stacking*- fue introducido por [Wolpert, 1992] como un esquema para combinar múltiples paradigmas de clasificación que han sido aprendidos bajo clasificación supervisada.

En el marco del modelo *stacked*, la configuración del multi-clasificador consta de dos niveles: en el primer nivel llamado nivel-base se tiene un conjunto de diferentes clasificadores $D = \{D_1, D_2, \dots, D_h\}$ denominados clasificadores-base. Cada instancia es un vector de n características etiquetado con una variable clase C . La clase real es denotada por C con $\Omega_c = \{c_1, c_2, \dots, c_r\}$ donde c_1, c_2, \dots, c_r son todos los posibles valores de la clase. En el nivel superior se aplica un paradigma de clasificación denominado meta-clasificador o multi-clasificador.

La inducción del multi-clasificador se inicia con el aprendizaje de los clasificadores base D y sus predicciones en el caso de una nueva instancia $\mathbf{x} = (x_1, \dots, x_n)$ que pertenezca a la clase $c \in C$. Estas distribuciones de la clase conforman una nueva *meta-base de datos* obtenida mediante k -subconjuntos de validación cruzada. Un meta-vector definido como $P_{hi}(\mathbf{x}_n)$ denota la i -ésima probabilidad de la clase C para la instancia n -ésima del h -ésimo clasificador D_h . Concatenando todas las distribuciones de probabilidad de los D_h clasificadores-base con el valor de la clase real $C = c$ se construye una nueva meta-base de datos definida como:

$$MBD = f(P_{h=1,n=1,r=1}, \dots, P_{h=1,n=1,r=2}, \dots, P_{h,n,r}), C_r$$

En una segunda fase se induce el meta-clasificador aplicando a la meta-base de datos MBD una validación cruzada de k -subconjuntos, con el propósito de evitar el sobre ajuste y sobre entrenamiento del multi-clasificador.

4.4. Estado del arte de combinación de clasificadores basados en el esquema de pila

En la literatura se encuentran diferentes adaptaciones del esquema de pila básico (*stacking*). Por ejemplo, en [Merz, 1999] se introduce una variante del modelo *stacking* anexando a la meta-base de datos un análisis de correspondencias con el propósito de detectar las correlaciones entre los clasificadores-base, eliminando dichas correlaciones de la meta-base de datos, en el cual se aplica el paradigma de clasificación de vecinos más cercanos como meta-clasificador.

Otra variante del método de pila fue propuesta por [Ting y Witten, 1999] para combinar las distribuciones de probabilidad de clasificadores-base en lugar de considerar únicamente el valor numérico de la clase predicho por cada clasificador. El algoritmo aplica el análisis de correspondencias para modelar la relación entre los ejemplos de aprendizaje y el modo en el que son clasificados por los clasificadores-base. En el meta-nivel propone el algoritmo MLR (*multi-response linear regression*), el cual es una adaptación del algoritmo de regresión lineal introducido por [Breiman, 1996b], transformando un problema de clasificación en diferentes problemas de predicción binaria. Esta consiste en definir una ecuación lineal para cada valor de la variable clase, teniendo un valor de 1 si la clasificación es correcta, o 0 si no lo es.

[Todorovski y Džeroski, 2002] introdujeron un nuevo método de aprendizaje en el meta-nivel, los árboles de clasificación (MDTs) en el esquema de pila en el cual reemplazan las predicciones del valor de la clase de los nodos terminales (hojas) por los valores devueltos por los clasificadores-base. Los datos aplicados en el meta-nivel tienen las distribuciones de probabilidad de cada clasificador-base multiplicadas por la entropía y la probabilidad máxima, reflejando la confiabilidad de los diferentes clasificadores-base.

[Seewald, 2003] sugiere una modificación en el modelo propuesto por [Ting y Witten, 1999], creando una meta-base de datos a partir de la distribución de probabilidad de la clase de cada clasificador-base, considerando únicamente los valores de probabilidad para la clase a tratar. Esta variación del esquema de pila propuesta por [Ting y Witten, 1999] reduce el número de atributos en el meta-nivel independientemente del número de clases.

[Džeroski y Ženko, 2004] proponen una meta-base de datos extendida la cual está compuesta con los datos de la distribución de probabilidad y características extras como la entropía y la probabilidad máxima obtenida por cada clasificador-base. Proponen un nuevo paradigma de clasificación basado en árboles de clasificación llamado *multi-response model trees* como meta-clasificador. Los autores comparan diferentes variaciones del esquema *stacking* y concluyen que el meta-clasificador *multi-response model trees* obtiene mejores resultados que el propuesto por [Seewald, 2003].

Las diferentes adaptaciones del método de pila que se encuentran en la literatura [Ting y Witten, 1999, Todorovski y Džeroski, 2002, Seewald, 2003] aplican como meta-clasificador diferentes paradigmas de clasificación a una

meta-base de datos la cual también varía en términos de que algunos incluyen la distribución de probabilidad de las clases, anexando la entropía, la máxima probabilidad obtenida por cada clasificador-base o bien la probabilidad de distribución de la clase predicha. Actualmente, en el caso de las distribuciones de probabilidad de las clases de los clasificadores-base necesitan ser discretizadas para posteriormente ser aplicadas a un método que las combinen, con la pérdida de información que este proceso de discretización conlleva.

En esta tesis se propone como aportación una nueva variante del esquema de pila *-stacking-* capaz de manejar datos continuos como las distribuciones de probabilidad de las clases predichas por los clasificadores-base utilizando una red condicional Gaussiana como meta-clasificador. Dicho esquema ha sido denominado *Gaussian-stacking*.

4.5. Multi-clasificador *Gaussian-stacking*

En esta sección se propone la aportación metodológica para la combinación de diversos paradigmas de clasificación denominados clasificadores-base, aplicando una red condicional Gaussiana como meta-clasificador, basado en la variación del método de pila generalizado *-stacking-* propuesto en [Ting y Witten, 1999].

El proceso de inducción del meta-clasificador basado en redes condicionales Gaussianas *-Gaussian-stacking-* consta de tres fases: selección de variables, aprendizaje estructural y aprendizaje paramétrico. Aplicando los métodos FSS de selección de variables indirecto *-filter-* o de envoltura *-wrapper-* se realiza la selección de un subconjunto de variables, con el propósito de mejorar el comportamiento y precisión del multi-clasificador. En el caso particular del método envoltura, se realiza en paralelo con el aprendizaje estructural del clasificador.

El aprendizaje estructural de las redes Gaussianas usualmente conlleva un proceso de búsqueda. El espacio de búsqueda está compuesto por todas las estructuras posibles para obtener el modelo más apropiado teniendo en cuenta las dependencias condicionales entre todas las variables. El aprendizaje estructural del modelo es guiado por una función que evalúa las dependencias entre las variables predictoras. Como ejemplo de funciones de puntuación o *score* son aplicables algunas como BIC (Bayesian information criterion) propuesta por [Schwarz, 1978] y BDe (Bayesian Dirichlet equivalent) propuesta por [Geiger y Heckerman, 1994]. Como alternativa, la función EDGE (edge exclusion test) [Dempster, 1972, Smith y Whittaker, 1998] detecta las independencias condicionales por medio de pruebas de hipótesis.

Finalmente en la última etapa, el aprendizaje paramétrico consiste en estimar los parámetros a partir de los datos, con el propósito de modelar la dependencia entre las variables predictoras, las cuales están representadas por una distribución Gaussiana. Es importante señalar que la función de densidad conjunta no está restringida a los valores entre 0 y 1. Para evitar el *overfitting*

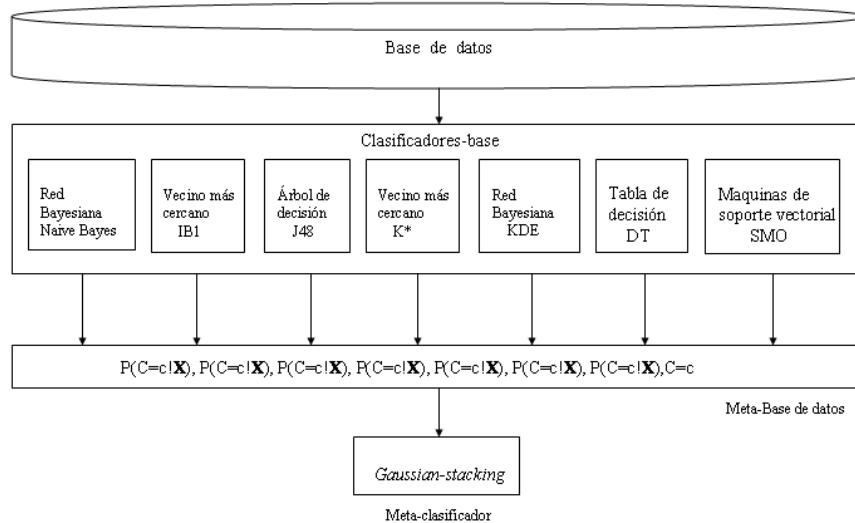


Figura 4.2. Esquema del multi-clasificador *Gaussian-stacking*. El esquema muestra la nueva meta-base de datos creada a partir de siete distribuciones de probabilidad de la clase para cada instancia del grupo de entrenamiento además del valor de la clase real.

del multi-clasificador *Gaussian-stacking* se genera una meta-base de datos a partir de k-subconjuntos por validación cruzada [Stone, 1974] concatenando las distribuciones de probabilidad de la clase de cada clasificador-base anexando también la clase real. La Figura 4.2 muestra un esquema del multi-clasificador basado en el esquema *stacked* con una red condicional Gaussiana como meta-clasificador.

A continuación se describen algunos de los paradigmas de clasificación supervisada, los cuales han sido ampliamente utilizados en la literatura y en particular en esta tesis fueron los que se aplicaron para la construcción del multi-clasificador *Gaussian-stacking*.

4.6. Paradigmas de clasificación aplicados como clasificadores-base

En los campos de la estadística y del aprendizaje automático se han desarrollado diferentes paradigmas de clasificación los cuales se pueden catalogar en familias. Algunas de las familias de clasificadores ampliamente utilizadas

en el aprendizaje automático son: los árboles de clasificación, los clasificadores basados en reglas de decisión, los clasificadores basados en casos, las redes neuronales y las redes Bayesianas. Cabe aclarar que esta catalogación no incluye clasificadores de aprendizaje no supervisado.

Para la construcción del multi-clasificador *Gaussian-stacking* se eligieron siete diferentes paradigmas de clasificación supervisada. Los paradigmas de clasificación así como el número de clasificadores-base fue determinado en base al estudio propuesto por [Seewald, 2003], conservando el principio de diversidad entre ellos. Se eligieron entre los clasificadores k -vecinos más cercanos, árboles de clasificación, tablas de decisión y máquinas de soporte vectorial así como los clasificadores Bayesinos (introducidos en el Capítulo 3). A continuación se describen brevemente los paradigmas de clasificación supervisada, utilizados en este trabajo para la construcción del multi-clasificador.

K-vecinos más cercanos

Entre los diferentes métodos de aprendizaje supervisado, el clasificador del vecino más próximo es un método intuitivo que pertenece a los métodos basados en casos que emplean criterios de vecindad para la clasificación. Este clasificador se caracteriza porque no hay inducción del modelo sino que todos los casos del conjunto de entrenamiento se almacenan en memoria. A partir de la regla del vecino más cercano, el clasificador asignará a un caso nuevo la clase del caso más próximo almacenado en memoria. En el caso de los k vecinos más cercanos en base a una métrica determinada se calcula la distancia entre un caso nuevo y cada caso del conjunto de entrenamiento. Teniendo en cuenta las k distancias del grupo de entrenamiento se le asigna al nuevo caso la clase mayoritaria presente en los k vecinos. La elección de la métrica (e.j. Euclídea, de Mahalanobis, etc.) es determinante ya que los resultados pueden variar significativamente en un mismo conjunto de datos.

Un ejemplo ilustrativo es el clasificador IB1 [Aha y Kibler, 1991] el cual es una variante del clasificador k vecinos más cercanos que emplea la distancia Euclídea para encontrar el caso más cercano, perteneciente al grupo de entrenamiento. Otra variante es el clasificador K* basado en casos [Cleary y Trigg, 1995], el cual asocia pesos a los casos vecinos del conjunto de entrenamiento en base a la similaridad con las características del nuevo caso, asignando la clase en función de su proximidad a la instancia a clasificar.

Árboles de clasificación

Los árboles de clasificación [Breiman y col., 1984] también son conocidos como árboles de decisión [Quinlan, 1986]. El modelo consiste en una serie de decisiones organizadas jerárquicamente en forma de árbol. Dicha estructura está formada por un conjunto de nodos, donde cada nodo que no es hoja contiene una condición sobre un atributo. Para clasificar un nuevo caso se parte

del nodo raíz del árbol y, comprobando en cada nodo el valor del atributo asociado a dicho nodo, se va descendiendo por la rama correspondiente al valor del atributo para el caso a clasificar.

El clasificador C4.5 fue propuesto por Quinlan [Quinlan, 1986], también es conocido como J48, constituye una mejora del algoritmo ID3 [Quinlan, 1986]. Las mejoras principales que introduce C4.5 respecto ID3 son la alta eficacia con la que trata los valores perdidos, el manejo de atributos continuos considerando técnicas de discretización y la corrección de la tendencia del ID3 de seleccionar los atributos con muchos valores distintos. El aprendizaje de esta estructura se lleva a cabo mediante algoritmos de partición (divide y vencerás).

Tablas de decisión

El clasificador de tablas de decisión [Kohavi, 1995a] está basado en la inducción de un conjunto de reglas de decisión con el propósito de obtener un conjunto de hipótesis que expliquen un determinado sistema. Las tablas de decisión son matrices tabuladas que representan un conjunto de condiciones y sus correspondientes acciones. Se aplican algoritmos basados en reglas de decisión del tipo *si-entonces* y conjuntos de reglas a las cuales se asocia una acción.

El clasificador de tablas de decisión busca en un espacio determinado por las características incluidas en la tabla y asigna a una nueva instancia la clase mayoritaria de los casos incluidos en la tabla de decisión con los que coincidió. En el caso de que la nueva instancia no coincide con ninguna contenida en la tabla de decisión, el clasificador asigna la clase mayoritaria.

Máquinas de soporte vectorial

Los clasificadores de máquinas de soporte vectorial fueron introducidos inicialmente para clasificación binaria por [Vapnik, 1995]. El modelo busca un discriminante lineal basado en una función kernel que maximice la distancia a los ejemplos fronterizos de las distintas clases, construyendo un hiperplano que separe en mayor medida las dos clases. Cuando los datos no son linealmente separables, como lo es en la mayoría de problemas reales, el clasificador máquina de soporte vectorial genera en base a ciertos kernels, espacios de características de muy alta dimensión donde se asume que los datos del conjunto de entrenamiento son linealmente separables. El clasificador introduce una serie de restricciones para permitir ciertos errores de clasificación. En cada restricción se tiene una variable holgura, que indica cuánto se supera la condición de clasificar bien ese ejemplo. El entrenamiento de las máquinas de soporte vectorial requiere de la optimización de la función para la inducción del modelo.

En el Capítulo 6 se presenta la experimentación realizada con el propósito de validar el método *Gaussian-stacking* y su adaptación al problema biomédico de la selección embrionaria en tratamientos de fertilización in-vitro.

Parte III

**Selección embrionaria en tratamientos de
reproducción humana asistida**

El problema de la selección embrionaria

5.1. Introducción

Actualmente la infertilidad es considerada un problema social y es diagnosticada cuando una pareja después de un año no logra un primer embarazo de forma natural, la cual se conoce como la infertilidad primaria. Existe otro caso de infertilidad en la pareja, denominada infertilidad secundaria, caracterizándose por el hecho de que la pareja, después de un embarazo previo, no consigue en condiciones normales obtener un nuevo embarazo. Las causas de infertilidad pueden ser de origen femenino como masculino o incluso presentarse en ambos. La fertilización in-vitro (FIV) o la inyección intracitoplasmática del esperma (ICSI) son algunas de las técnicas que se emplean en la reproducción humana asistida a partir de las cuales es posible la fecundación, selección y transferencia de embriones al útero de la paciente, haciendo posible la implantación y el desarrollo del embrión.¹

Los procedimientos y las técnicas empleadas para la evaluación y control del desarrollo de los embriones varían de un país a otro a causa de razones éticas y legislativas, lo que conlleva a diferentes protocolos para la manipulación de los embriones y restricciones en cuanto a aspectos como el período de tiempo para el seguimiento de los embriones más prometedores después de la fertilización. Por ejemplo, en España la ley sufrió una modificación en 2005, cuando el parlamento restringió a tres el número máximo de embriones a transferir a la paciente con el objetivo de reducir los embarazos con nacimiento múltiple y también se modificó la restricción del número de óvulos a ser fecundados en cada ciclo, quedando dicho número a criterio del responsable del equipo biomédico de la unidad de reproducción asistida de

¹ En los tratamientos de reproducción asistida el término *transferencia* se refiere al procedimiento clínico para insertar los embriones seleccionados en el útero materno, mientras que el término *implantación* está relacionado con el éxito de lograr el embarazo (alguno de los embriones transferidos ha logrado adherirse a la pared del útero).

cada clínica. En Italia, por ejemplo se permite fecundar tres óvulos y transferir todos los fecundados, no permitiéndose la criopreservación de embriones, situación permitida en España. En algunos países como Suiza y Francia entre otros de Europa, la práctica general es transferir un solo embrión y la legislación entre esos países también considera diferentes aspectos. Por lo que actualmente la práctica de la reproducción asistida varía de un país a otro considerablemente dificultando la comparación de resultados y conclusiones que ayuden a entender la complejidad del problema médico.

En las técnicas de reproducción asistida se busca maximizar la probabilidad de que nazca un niño sano y por otro lado se busca minimizar el riesgo de los embarazos múltiples. Siendo estos los motivos por los cuales el problema de la selección embrionaria en los tratamientos de fecundación in-vitro es tan importante.

La selección embrionaria forma parte de las rutinas de las unidades de reproducción asistida en todas las clínicas del mundo, la cual está basada en la observación, evaluación y catalogación de los embriones realizada por el embriólogo y tiene como objetivo el determinar cuales de entre los embriones en desarrollo son los más viables para ser transferidos el segundo o tercer día después de la fecundación. La formación y capacitación del embriólogo resulta decisiva en la fecundación in-vitro así como en la selección de los embriones ya que el factor tiempo influye en la evaluación y el biólogo no dispone de un periodo largo para realizarla, por lo que su criterio resulta muy importante para el éxito del tratamiento.

5.2. Estado del arte de la inteligencia artificial aplicada a tratamientos de reproducción asistida

Debido al alto índice de subjetividad en el criterio del embriólogo e incertidumbre del éxito del propio tratamiento de reproducción asistida, en los últimos años se proponen en la literatura varios ejemplos de técnicas de inteligencia artificial para mejorar la tasa de éxito (embarazo) en los procedimientos de fertilización in-vitro. A continuación se describen brevemente los ejemplos más representativos de publicaciones científicas más recientes en este campo:

- Saith y col.,(1998) analizan el potencial de la contribución de la inteligencia artificial en el proceso de la selección embrionaria y proponen un modelo de minería de datos basado en árboles de decisión para investigar la relación entre las características de los embriones, los ovocitos y los folículos en la predicción del éxito en tratamientos de reproducción asistida.
- Jurisica y col., (1998) presentan un sistema inteligente basado en un sistema de razonamiento de casos para el apoyo a la decisión en los tratamientos en los tratamientos de reproducción asistida.
- Trimarchi y col., (2003) proponen un sistema basado en técnicas de minería de datos, concretamente en el algoritmo C5.0 para la inducción de árboles de clasificación.

- Patrizi y col., (2004) presentan un algoritmo de reconocimiento de patrones para la selección de embriones a partir de imágenes de embriones, caracterizados con los primeros seis momentos estadísticos. Este algoritmo reconoce objetos dado un número de clases, formulando una regla general para el reconocimiento.
- Manna y col., (2004) comparan la precisión del algoritmo de reconocimiento introducido por [Patrizi y col., 2004] con la tarea de reconocimiento realizada por un grupo de expertos.
- En este trabajo se propuso la adaptación de las redes Bayesianas para realizar la predicción de los embriones más viables para su transferencia en tratamientos de reproducción humana asistida y la predicción del éxito del tratamiento de fertilización in-vitro.

5.3. Tasa de éxito y procedimiento de transferencia

En las clínicas de reproducción humana asistida, se considera que un tratamiento de fertilización in-vitro es exitoso cuando ocurre la implantación de alguno de los embriones transferidos dando inicio al embarazo. Las clínicas pueden reportar su *tasa de éxito* en función de los embarazos clínicos (comprobados por medio de ultrasonido) o en base a los nacimientos registrados. Es importante resaltar que el desarrollo del embarazo y el nacimiento son considerados como problemas ginecológicos diferentes al tratamiento de reproducción asistida. Siendo la implantación del embrión el resultado exitoso del tratamiento de reproducción asistida independientemente de si se obtiene el nacimiento o no.

Debido a los tratamientos de reproducción asistida, las parejas que tienen muy baja probabilidad de lograr un embarazo espontáneo (no mayor al 1% en muchos casos) aumentan sus expectativas de éxito hasta alcanzar un 25% de probabilidad de embarazo por cada intento. Esto las pone en coincidencia con los índices de embarazo natural de quienes no tienen problemas reproductivos, los cuales se ubican en un rango del 20% y el 30% por cada intento.

La tasa de implantación (éxito) en tratamientos de reproducción asistida, en algunos países Europeos y Estados Unidos, está en un rango del 29% y 38%. En el caso de tratamientos de reproducción asistida aplicados a mujeres menores de 36 años se obtiene una tasa de éxito del 38%. La tasa de éxito del tratamiento de FIV para las mujeres de 36 a 39 años es del 29%. Y, en las mujeres mayores de 40 años, el tratamiento de FIV tiene una tasa de éxito muy pequeña mejorando aproximadamente hasta un 13% con donación de óvulos [Windt y col., 2004, HFEA, 2005, Andersen y col., 2008]. Estos porcentajes de éxito son orientativos ya que como se ha mencionado anteriormente la falta de homogeneidad en los tratamientos y procedimiento médicos, así como en los protocolos implementados en las unidades de fertilización in-vitro, todos ellos regidos por la legislación de cada país, dificulta la estandarización de la tasa de éxito entre las clínicas de reproducción asistida.

Los factores que más influyen en la variación de la tasa de embarazo son la edad de la mujer, la calidad de los óvulos extraídos, la calidad espermática, el número de embriones transferidos, el origen de la infertilidad así como los tratamientos médicos y la falta de criterios concluyentes de criterios en el proceso de selección embrionaria. En algunos países como Estados Unidos, Suiza, Francia, Escandinavia y Países Bajos han llegado a obtener tasas de éxito del 40 % siendo este el resultado más optimista sin tener en cuenta el grupo de mujeres con más de 40 años de edad [Giorgetti y col., 2007, Keenan y col., 2007].

En los países donde la legislación permite transferir dos embriones y un máximo de tres, normalmente se selecciona la mejor combinación de los embriones con la mayor viabilidad de desarrollo y con la mayor posibilidad de implantarse en el útero materno. El número de embriones a transferir lo determina el equipo de expertos en base a la edad de la paciente y del diagnóstico médico. La transferencia de los embriones es un procedimiento que lleva a cabo el ginecólogo y el éxito del tratamiento de reproducción asistida se asume cuando alguno de los embriones transferidos se ha implantado en el útero materno.

Los protocolos aplicados para la selección embrionaria en todas las unidades de reproducción asistida del mundo están basados en el criterio de los embriólogos, los cuales tratan de homologarse a través de la formación y entrenamiento, para reducir la variabilidad del procedimiento de la selección embrionaria. Siendo la subjetividad del proceso de selección embrionaria una de las motivaciones del presente trabajo, se planteó la selección embrionaria como un problema de clasificación supervisada descrito en la Sección 5.5 y se aplicaron diversos algoritmos de clasificación basados en redes Bayesianas.

Baczkowski y col., (2004) proponen la consideración de los siguientes factores para mejorar la tasa de éxito (implantación del embrión) como son: la edad de la paciente, la presencia de dimorfismos en el cigoto, la calidad del semen, el tiempo que tarda en iniciar la división celular, el número de embriones trasferidos, el grosor del endometrio y el número de ciclos previos. En el tratamiento de reproducción asistida, también es necesario tener en cuenta otro tipo de decisiones como elegir la técnica de fertilización (FIV o ICSI), el número de embriones a transferir, el diagnóstico genético y finalmente la composición del medio de cultivo.

En la siguiente sección se describen las características morfológicas de los cigotos y embriones así como los protocolos habitualmente más utilizados para su selección.

5.4. Caracterización morfológica de cigotos y embriones en tratamientos de reproducción asistida

Entre las diferentes técnicas de reproducción asistida se encuentra la fertilización in-vitro (FIV) en la que el esperma previamente tratado es depositado

en el mismo plato de cultivo que contiene los ovocitos proporcionando las condiciones óptimas para que se dé la fecundación. Otra técnica alternativa, antes mencionada, es la inyección intracitoplasmática (ICSI), en la que el biólogo inyecta un espermatozoide al óvulo. Ambas técnicas son ampliamente utilizadas en la práctica de reproducción asistida no presentándose diferencia en el proceso posterior de evaluación y selección embrionaria.

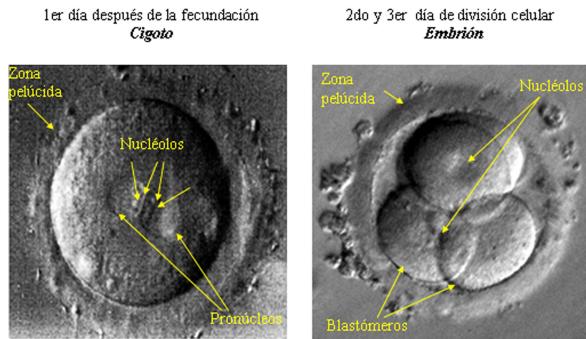


Figura 5.1. Esquema del cigoto y el embrión.

En el tratamiento de fertilización in-vitro se administra un tratamiento médico a la paciente con el objetivo de estimular la glándula pituitaria e incrementar la secreción de folículos y al mismo tiempo incrementar el nivel de hormonas. La muestra de semen, que se obtiene de forma separada, es analizada y tratada (lavada) para la posterior fecundación in-vitro. Los embriones se desarrollan fuera del cuerpo uterino en un medio controlado durante las primeras horas.

El proceso de fecundación puede definirse como una secuencia de acontecimientos moleculares coordinados en los que participan los gametos masculino y femenino. Consiste en la unión de las dos células reproductoras, de sexos contrarios -los gametos- hasta que se fusionan los núcleos y parte del citoplasma. Es un proceso complicado que conduce a la formación de una célula, el *cigoto* o huevo y que comienza con la penetración de un espermatozoide en un óvulo. En la fecundación no participa todo el espermatozoide, sino sólo el núcleo y el centrosoma; ambos corpúsculos se dirigen al núcleo femenino y el primero acaba por fusionarse con él, mientras el centrosoma se divide en dos, originándose las esferas atractivas que se colocan en los polos del cigoto para la primera división del desarrollo embrionario, la cual comienza con la segmentación. A partir del segundo día después de la fecundación se inicia la división del *embrión* [Carlson, 2005].

Con el propósito de poner en contexto la información que se introduce en este capítulo para la descripción de las características morfológicas del cigoto

y del embrión la Figura 5.1 muestra un esquema de algunas de sus estructuras celulares. El *cigoto* es la célula que se forma como resultado de la unión de los gametos femenino y masculino las primeras horas después de la fecundación. El *embrión*, inicia su división celular aproximadamente a las 24 horas después de la fecundación. En el esquema del cigoto se observa la zona pelúcida y dos pronúcleos en los que se encuentran los nucléolos y en cada uno de ellos los puntos representan los nucléolos. Siendo la distribución de los nucléolos una de las características consideradas en la evaluación y catalogación del cigoto. El esquema del embrión muestra la zona pelúcida, los blastómeros y los nucléolos. La zona pelúcida es la membrana que rodea tanto al cigoto como al embrión y una de sus características es la variación de su grosor, siendo el grado de la variación del grosor una variable importante para predecir la posibilidad de implantación del embrión en la pared del útero. El embrión puede presentar cierto grado de fragmentación así como que los blastómeros sean multinucleados o no. Estas y otras características como el aspecto del citoplasma son observadas por los biólogos a través de un microscopio invertido con contraste de Hoffman. Los cigotos y embriones son catalogados en base a sus características morfológicas, las cuales están relacionadas directamente con la *tasas de éxito* del tratamiento [Giorgetti y col., 1995, Hill y col., 1989, Veek, 1999, Scott, 2003].

Protocolo y criterios de observación, evaluación y selección embrionaria

En el campo de la biología, en particular en lo que a división celular se refiere, especialmente las primeras horas de división celular después de la fertilización in-vitro resultan especialmente importantes para la evaluación de cigotos y la selección de los embriones más prometedores a transferir a la paciente. La Figura 5.2 muestra un esquema de los períodos de tiempo posteriores a la fertilización in-vitro en los que se observan y evalúan los cigotos y embriones.

Evaluación y catalogación del cigoto

La primera evaluación del cigoto se inicia en las primeras 16-18 horas después de la fecundación, y en la misma el biólogo asigna a cada cigoto una categoría basándose en un sistema de catalogación concreto. En la literatura se encuentran diferentes sistemas para catalogar al cigoto. Cada sistema tiene en cuenta diferentes criterios basados en las características morfológicas del cigoto como son el tamaño y el número de nucléolos, así como su alineación con respecto a la unión de los pronúcleos [Tesarik y Greco, 1999, Scott y col., 2000, Fisch y col., 2001, Terriou y col., 2001, Zollner y col., 2002]. Un ejemplo de estos sistemas es el propuesto por [Scott y col., 2000]. La Figura 5.3 muestra un ejemplo de imágenes reales de cigotos donde se observa la distribución de los nucléolos, en cada pronúcleo.

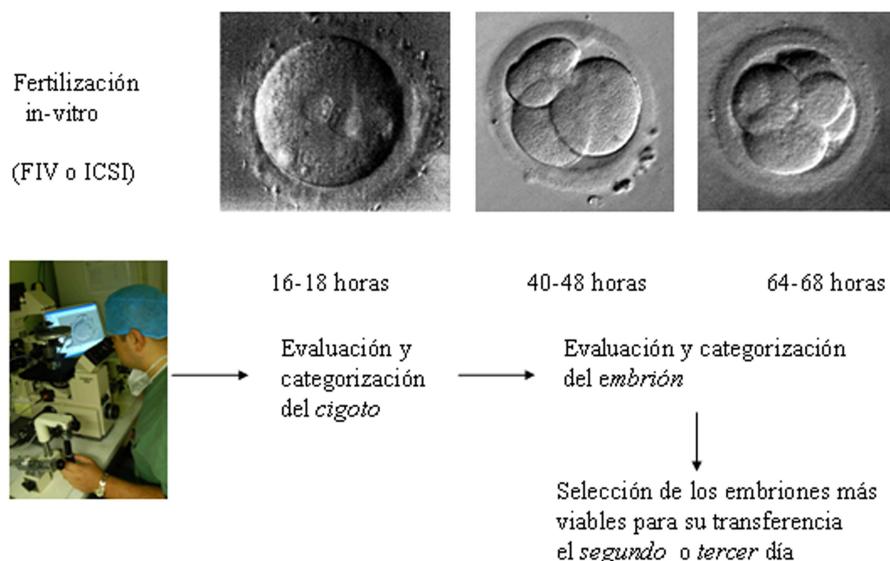


Figura 5.2. Esquema de la rutina de observación, evaluación y selección embrionaria desde las primeras 16-18 horas después de la fecundación y a las 40-48 y 64-68 horas.

El sistema propuesto por [Scott y col., 2000] para la catalogación de cigotos, es un sistema estándar ampliamente utilizado, el cual tiene cuatro categorías (Z1,Z2,Z3 y Z4). Cada categoría está determinada por la distribución de los nucléolos así como el tamaño de los pronúcleos, siendo las categorías Z1 y Z2 las más prometedoras para el desarrollo de un embrión sano. Especialmente la primera de ellas. A continuación se describe cada una de las categorías.

- Z1: Pronúcleos de igual tamaño. Igual tamaño y número de nucléolos alineados en ambos lados de la unión de los pronúcleos.
- Z2: Pronúcleos de igual tamaño. Igual tamaño y número de nucléolos espardidos en ambos pronúcleos.
- Z3: Pronúcleos de igual tamaño. Igual número de nucléolos de tamaño igual o no, los cuales se encuentran alineados en uno de los pronúcleos y localizados en la unión de los mismos. En el segundo pronúcleo los nucléolos se encuentran espardidos.
- Z4: Pronúcleos de diferente tamaño con los nucléolos espardidos.

Evaluación y catalogación del embrión

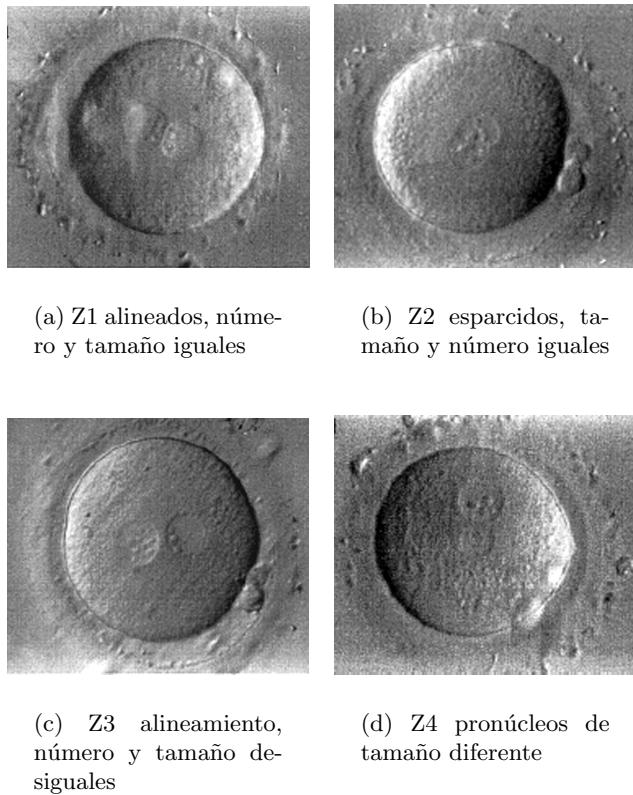


Figura 5.3. Ejemplo de imágenes de cigotos reales de la base de datos de la Clínica del Pilar (San Sebastián, Guipúzcoa), catalogados en base a la escala de [Scott y col., 2000]. El tipo Z1 es el de mayor viabilidad para desarrollarse en un embrión sano y el tipo Z4 el de menor viabilidad.

A partir del segundo día después de la fecundación, la escala de Scott no se puede aplicar por lo que se necesita un sistema diferente para la catalogación de embriones. En la literatura se encuentran algunos estudios [Puissant y col., 1987, Schulman y col., 1993, Giorgetti y col., 1995, Van Royen y col., 1999, Veek, 1999] en los que se menciona que la *calidad embrionaria* (viabilidad para su desarrollo) puede evaluarse en base a las estructuras morfológicas de los embriones pudiendo ser utilizadas como bio-marcadores para la catalogación de los mismos.

Los embriones son evaluados y seleccionados en los intervalos de tiempo de las 40-48 y 64-68 horas después de la fertilización. La selección de los embriones se realiza en base a su catalogación de acuerdo a un sistema determinado, por ejemplo, por el sistema de catalogación de embriones propuesto por [Veek,

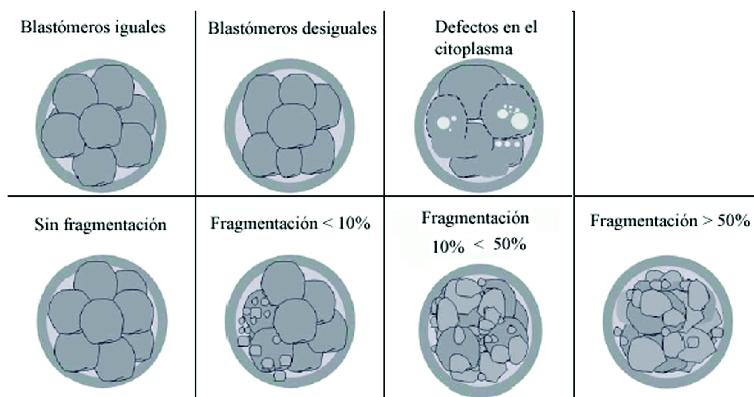


Figura 5.4. Esquema con la representación gráfica de los blastómeros y su fragmentación [Baczkowski y col., 2004].

1999] o como alternativa al sistema propuesto por [Mills, 1992]. Ambos sistemas consideran las características morfológicas óptimas para que el embrión continúe un desarrollo sano.

Un ejemplo ampliamente utilizado en la práctica médica es el sistema introducido por [Veek, 1999], el cual está basado en la caracterización morfológica del embrión empleando parámetros específicos como son el número de células, el porcentaje de fragmentación, el tamaño de los blastómeros así como la existencia de blastómeros multinucleados o no. De forma gráfica se puede observar en la Figura 5.4 el tamaño entre los blastómeros, el aspecto del citoplasma y el grado de fragmentación de los diferentes tipos de embriones.

A continuación se describen cuatro tipos diferentes de embriones en base a la catalogación propuesta en [Veek, 1999]. El Tipo I es el de mayor posibilidad para desarrollarse en un embrión.

Tipo I: 0 % de fragmentación con blastómeros homogéneos de igual tamaño.

Tipo II: < 10 % de fragmentación con blastómeros homogéneos de igual tamaño.

Tipo III: este tipo está dividido en dos subtipos: Tipo III-A: 10-25 % de fragmentación con blastómeros homogéneos de igual tamaño. Tipo III-B: 25-50 % de fragmentación con blastómeros de diferente tamaño.

Tipo IV: > 50 % de fragmentación con blastómeros que pueden tener el mismo tamaño o no.

Tipo V: 100 % de fragmentación.

Como se mencionó anteriormente cada clínica de reproducción asistida y en particular las unidades de fertilización in-vitro aplican sus propios protocolos dentro del programa que permite la legislación de cada país, considerando los diferentes sistemas de catalogación de cigotos y de embriones como los descritos anteriormente. Esta diversidad de protocolos e información relacionada

con el tratamiento de reproducción asistida propicia la falta de homogeneidad y compatibilidad en los datos, dificultando el estudio de la selección embrionaria, desde el campo de la minería de datos y el aprendizaje automático.

Los capítulos 6, 7, 8 y 9 describen diferentes adaptaciones de los clasificadores Bayesianos así como del multi-clasificador *Gaussian-stacking* aplicadas al problema de la selección embrionaria en tratamientos de reproducción asistida.

5.5. Planteamiento de la selección embrionaria como problema de clasificación supervisada

En primer lugar se describen algunas limitaciones de la información relacionada con el tratamiento de reproducción asistida y posteriormente su consideración para el planteamiento de la *selección embrionaria en tratamientos de reproducción humana asistida* como un problema de clasificación supervisada.

En base a la práctica médica general de las clínicas de reproducción asistida en países donde la legislación lo permite, como es el caso de la unidad de reproducción asistida que facilitó los datos para la realización de esta tesis, se transfiere un grupo de dos o tres embriones. En particular en esta tesis se consideró la selección de un conjunto de tres embriones transferidos el segundo o tercer día después de la fecundación in-vitro.

La condición a priori de tener un grupo de tres embriones transferidos al mismo tiempo en los tratamientos de reproducción humana asistida, no permite determinar cuál de los tres embriones es el que se ha implantado. La única forma de identificar al embrión implantado sería realizando una biopsia a cada embrión antes de trasferirlos y este procedimiento somete al embrión a un gran estrés por lo que no se realiza.

A partir de esta situación, remarcamos la asunción que hace el experto al considerar que el embrión con mayor valor en la escala de catalogación del conjunto de tres embriones, será el que se ha implantado. Sin embargo, no se puede asegurar de que la asunción del experto sea siempre la correcta, por lo que se diseñó un planteamiento para el problema de clasificación supervisada con un conjunto de tres embriones y no para cada embrión, ya que aportaría más incertidumbre a los clasificadores empleados.

[Matorras y col., 2005] confirman el hecho de que cuando se transfiere más de un embrión al mismo tiempo, no siempre se implantan todos y demuestran que pueden ayudarse entre sí para que alguno se implante porque se crea un medio favorable. En base a esta premisa, se planteó el análisis de los casos de la *selección embrionaria en un conjunto de embriones*, respetando el orden asignado por el embriólogo en su selección, siendo el primero el de mayor viabilidad para implantarse.

Cabe mencionar que debido a la limitación de saber cuál de los embriones del conjunto transferido se ha implantado, se asumió como *tratamiento*

exito, pese a la implantación de dos o tres embriones (siendo muy poco frecuente la implantación de los tres embriones), debido a no poder estimar el coste de la implantación de dos o tres de los embriones transferidos. Debido a esta limitación no se aplicó la clasificación supervisada con la correspondiente matriz de coste del error.

Caso	Características morfológicas del embrión			Datos clínicos	Clase
1	Emb1	Emb2	Emb3	...	1
2	Emb1	Emb2	Emb3	...	0
3	Emb1	Emb2	Emb3	...	1
...
n	Emb1	Emb2	Emb3	...	0

Tabla 5.1. Ejemplo de los vectores de características de tratamientos de fertilización in-vitro conformado por un conjunto de 3 embriones, datos clínicos de la paciente y del tratamiento. El valor de la clase 1 se asignó a casos donde ocurrió la implantación y el valor de la clase 0 a los casos en los cuales no se dio la implantación.

Las variables consideradas en el tratamiento de fertilización in-vitro, la información clínica así como las características morfológicas de los embriones son detalladas en la Sección 6.2.

Formalmente denotaremos por $\mathbf{x} = (x_1, \dots, x_n)$ el vector de características de un conjunto de embriones transferidos así como las variables clínicas de la paciente y del tratamiento de fertilización in-vitro.

El problema de la selección embrionaria se ha planteado como un problema de clasificación supervisada creando un modelo que asigne a cualquier nuevo caso de fertilización in-vitro $\mathbf{x} = (x_1, \dots, x_n)$ un valor de variable clase C . El valor de la variable clase está denotado por c y tiene valores $\{implantado, no-implantado\}$, los cuales son clasificados como $\{0,1\}$ asignándose el valor 1 si ha ocurrido la implantación y el valor 0 en caso contrario. La Tabla 5.1 ilustra cómo se caracteriza un caso de la base de datos en base a tratamientos con grupos de tres embriones transferidos siguiendo este planteamiento.

5.6. Aportación de la medición del grosor de la zona pelúcida en imágenes de embriones humanos por medio de contornos activos

Como se mencionó en el Capítulo 1, uno de los objetivos de este trabajo es medir la variación del grosor de la zona pelúcida (VGZP) por medio de los contornos activos, segmentando la zona pelúcida ya que se conoce que la variación del grosor de la membrana está directamente relacionada con la capacidad de implantación del embrión [Cohen y col., 1989, Palmstierna y col., 1998, Gabrielsen y col., 2000].

En esta sección se describe la aplicación de los contornos activos, como una técnica para ampliar la información proporcionada por los expertos acerca de la medida del grosor de la zona pelúcida, que debido a la subjetividad de la evaluación y la percepción del ojo humano es difícil de determinar un grado de variación del grosor de la zona pelúcida [Morales y col., 2008a].

El método de contornos activos es sensible a los cambios de contraste en los píxeles del contorno de la membrana de la zona pelúcida, que se caracteriza por tener un contorno difuso. Además los embriones pueden presentar diferentes formas (e.j. circular, elípticas), VGTP y artefactos como restos de células de granulosa o cuerpos de espermatozoides presentes en la pared exterior de la membrana. Estas características pueden generar una segmentación errónea de la zona pelúcida.

Un problema de importancia fundamental en el análisis de la imagen es la detección de bordes. Los contornos caracterizan las fronteras de los objetos en la imagen. El proceso de la localización de esos contornos se llama segmentación. Existen diferentes y muy variadas técnicas que se han aplicado a la segmentación de imágenes médicas. En [Pham y col., 2000, Withey y Koles, 2007] se hace una revisión amplia de ellas. Usualmente, los métodos clásicos de segmentación se dividen en métodos basados en segmentación de regiones y métodos basados en bordes. En [Gonzales y Woods, 2002] se puede encontrar una descripción amplia de estas técnicas. Los algoritmos basados en segmentación de regiones usualmente buscan zonas de píxeles conectados con características similares. La imagen se divide en regiones que son homogéneas respecto a una determinada propiedad como: brillo, patrón de textura, etc. En imágenes complejas como las imágenes médicas puede ser necesario obtener regiones homogéneas con solapamiento. Finalmente la segmentación se realiza en base a la información de alto nivel en función de la propiedad dominante que se emplea.

Los métodos basados en la detección de bordes buscan los píxeles con el valor de gradiente más alto. Generalmente este tipo de métodos requieren de una etapa posterior de preprocesso, por ejemplo la umbralización o la unión de los puntos encontrados pertenecientes a los bordes de los objetos en la imagen. En esta tesis se propone la aplicación de los contornos activos para la delimitación completa del contorno de la membrana de la zona pelúcida.

La técnica de los contornos activos se caracteriza por la capacidad de deformación, cualidad que permite delimitar el contorno de una estructura celular a pesar de la falta de homogeneidad y cambios de nivel de grises que caracterizan a las imágenes médicas. Existe una amplia variedad de trabajos que presentan el método de contornos activos o *snake*, introducido por [Kass y col., 1987], como el método de segmentación más robusto para segmentar imágenes médicas debido a su capacidad de localizar especialmente contornos difusos, formados por píxeles con baja intensidad. Como ejemplo se encuentran los siguientes trabajos [McInerney y Terzopoulos, 1996, Xu y col., 2000, Boscolo y col., 2002].

En la literatura se encuentran los siguientes trabajos relacionados con la segmentación de la zona pelúcida [Karlsson y col., 2004, Karlsson y col., 2005] en los cuales se propone un método de segmentación automático basado en el cálculo del área conjuntamente con un procedimiento probabilístico para segmentar la zona pelúcida. Esta aproximación consta de dos etapas: en la primera, se remueve el fondo de la imagen para encontrar la pared exterior de la membrana y después en una segunda etapa, separan el conjunto de blastómeros que forman al embrión de la zona pelúcida para segmentar la pared interior de la membrana. Aplicaron una máscara para remover los artefactos más grandes de las paredes de la membrana y utilizaron una función de penalización para remover artefactos pequeños como los restos de células de granulosa. Sin embargo, este método no resulta muy exacto en presencia de grandes artefactos, ya que el algoritmo asume que forman parte de la zona pelúcida y en el caso de la medición de la variación del grosor de la zona pelúcida, resulta una medida errónea.

Por medio de las deformaciones dinámicas de los contornos activos, la curva final es capaz de adaptarse exactamente a las paredes de la membrana y no resulta afectado por los artefactos presentes en la zona pelúcida ya que en este trabajo se aplicaron operadores morfológicos para remover los artefactos más grandes.

Los contornos activos, inicialmente necesitan que se defina una posición inicial, a partir de la cual comienza la deformación de la curva de forma dinámica hasta minimizar una función de energía. En esta tesis se propone *la inicialización automática del snake*. La inicialización automática permite al *snake*, adaptarse a cualquier forma que presente el embrión, además de adaptarse a cualquier irregularidad presente en la membrana. Debido a la forma circular o elíptica de la zona pelúcida, se implementó únicamente la forma básica del *snake* propuesta por [Kass y col., 1987] por lo que no fue necesario la implementación de otras adaptaciones al método de los contornos activos como el propuesto por [Xu y Prince, 1998], ya que los contornos de las paredes exterior e interior de la membrana de la zona pelúcida no presentan grandes concavidades.

5.6.1. Contornos activos

Los contornos activos o *snakes*, también conocidos como *modelos deformables* introducidos por [Kass y col., 1987], son curvas que se deforman bajo la influencia de fuerzas internas y externas. Tradicionalmente un contorno activo, es una curva definida por $\mathbf{z}(s) = (x(s), y(s))$ donde x e y son las coordenadas y s es un parámetro que presenta valores en el rango $[0,1]$. El contorno activo se deforma en el dominio del espacio de la imagen para minimizar la siguiente función de energía:

$$E_{\text{snake}} = \int_0^1 (E_{\text{int}}(z(s)) + E_{\text{ext}}(z(s))) ds \quad (5.1)$$

donde E_{int} es la energía interna definida como $E_{\text{int}} = \alpha|z'(s)|^2 + \beta|z''(s)|$, donde z' y z'' corresponden a la primera y segunda derivada de $z(s)$ respectivamente, α y β son parámetros de peso que sirven para controlar la tensión y rigidez de la curva (*snake*).

La fuerza externa E_{ext} es una función de potencial, la cual es diseñada para que el contorno activo sea atraído por el contorno del objeto u otras características de interés. El diseño de las fuerzas externas es muy importante, ya que determinan directamente la precisión y el comportamiento del *snake*. En este trabajo se necesita identificar los píxeles con valores altos de gradiente, los cuales representan los píxeles con un cambio brusco de nivel de gris.

Si se considera la imagen como una función continua $f(x, y)$, su derivada tiene un máximo local en la dirección del contorno. Es por esto que las técnicas empleadas para la detección de contornos se basan en la medida del gradiente.

La función de potencial, caracterizada por fuerzas de potencial Gaussianas en todo el dominio de la imagen, es definida como:

$$P(x, y) = -\gamma|\nabla[G_\sigma(x, y) * I(x, y)]|^2 \quad (5.2)$$

donde ∇ y $*$ son los operadores de gradiente y convolución respectivamente, γ es un parámetro de peso, $G_\sigma(x, y)$ es un filtro Gaussiano con una desviación estándar σ , e $I(x, y)$ corresponde a la imagen.

Un contorno activo $z(s)$ que minimice la función de energía E_{snake} debe satisfacer la ecuación de Euler, la cual puede considerarse como una fuerza de balance definida por $F_{\text{int}} + F_{\text{ext}} = 0$ donde $F_{\text{int}} = \alpha z''(s) - \beta z'''(s)$ y $F_{\text{ext}} = -\nabla E_{\text{ext}}$. Para calcular estas ecuaciones, la curva paramétrica del *snake* se calcula dinámicamente tratando a z en función del tiempo $z(s, t)$.

5.6.2. Inicialización automática del contorno activo

La inicialización automática del contorno activo, propuesta en este trabajo, consiste en definir un círculo lo más cercano posible al borde exterior de la zona pelúcida. La ubicación del círculo se calculó promediando el centro de 60 imágenes. El radio del círculo se calculó desde el centro propuesto a un punto en el mapa de bordes en el eje x . En el caso de que el centro calculado no coincida o no esté lo suficientemente cerca del centro de la imagen original, se calculó un nuevo centro a partir del mapa de bordes, especialmente en las imágenes de embriones con forma elíptica o irregular.

Para la aplicación de la técnica de los contornos activos es necesario aplicar una serie de técnicas para el preprocessamiento de la imagen así como para la generación de un mapa de bordes, al cual se aplican los contornos *snakes*. En el Apéndice A se describen las etapas anteriores a la segmentación automática.

5.6.3. Medición del grosor de la zona pelúcida

Se realizó un estudio con 76 imágenes de los embriones con mayor viabilidad de desarrollarse en un embrión sano (Tipo I) de un grupo de tres embriones transferidos en el tratamiento de reproducción humana asistida de los cuales en 34 casos se registró su implantación y los 42 casos restantes pertenecen al grupo de los tratamientos donde no se logró la implantación.

La medición de la VGZP se realizó automáticamente por medio de contornos activos, los cuales coincidieron con las paredes interior y exterior de la zona pelúcida. El muestreo de VGZP se realizó en 360 grados, un grado por cada punto.

Se realizó una experimentación empírica sobre el ajuste de la curva del contorno activo, con varios valores de α y β para controlar el grado de elasticidad y rigidez respectivamente obteniéndose con los valores de $\alpha = 0,5$ y $\beta = 0,2$ un mejor ajuste del *snake* al contorno de la zona pelúcida.

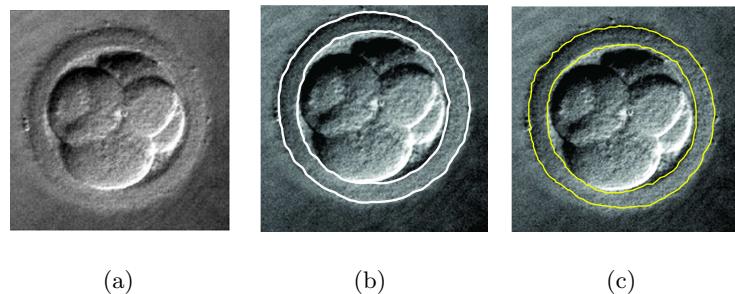


Figura 5.5. (a) Imagen original, (b) Segmentación automática de la zona pelúcida por medio de *snakes* y (c) Segmentación manual de la zona pelúcida.

La Figura 5.5(a) muestra la imagen original del embrión, en la Figura 5.5 (b) se puede observar la segmentación final de la zona pelúcida, y la Figura 5.5(c) muestra la zona pelúcida segmentada manualmente.

5.6.4. Resultados

Con el propósito de cuantificar la precisión de la localización de los contornos de la membrana con la segmentación automática, se segmentó manualmente la zona pelúcida de 20 imágenes de embriones, las cuales fueron supervisadas por el biólogo responsable de la selección de embriones de la Unidad de Reproducción Asistida de la Clínica del Pilar. Este grupo de imágenes incluyen variaciones en la forma del embrión así como restos de células de granulosa en la zona pelúcida como se muestra en la Figura 5.6.

Para medir la precisión de la segmentación automática de la zona pelúcida, aplicando el método de contornos activos, se midió la coincidencia de los contornos de la segmentación automática y manual midiendo la coincidencia MP entre dos regiones, la cual es definida como:

$$MP = \frac{np(SM \cap SS)}{n(SM \cup SS)} \quad (5.3)$$

donde SM es la imagen segmentada manualmente por el experto y SS denota la imagen segmentada automáticamente por los contornos activos; n es el número de píxeles en la región.

La Figura 5.5(a) muestra la imagen original del embrión, en la Figura 5.5 (b) se puede observar la segmentación final de la zona pelúcida, y la Figura 5.5(c) muestra la zona pelúcida segmentada manualmente.

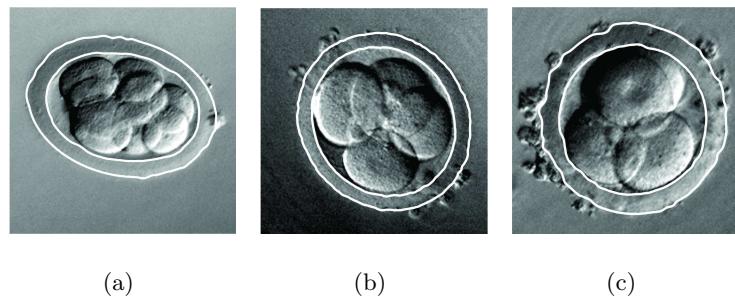


Figura 5.6. Segmentación automática de imágenes de embriones humanos con presencia de artefactos y diferentes formas. (a) Forma elíptica, (b) artefactos en la zona pelúcida y (c) zona pelúcida no uniforme.

Resultado FIV	Número de ciclos	Media VGZP
implantación	34	23.58 ± 6.28
no-implantación	42	16.31 ± 5.71

Tabla 5.2. Valor de la media de la variación del grosor de la membrana, zona pelúcida (VGZP) entre los grupos de imágenes de embriones implantados y el grupo de embriones que no se implantaron.

También se analizó la relevancia de la variación en el grosor de la zona pelúcida, entre el grupo de embriones implantados y el grupo de embriones no implantados, incluyendo únicamente las imágenes de los embriones Tipo I de acuerdo a la catalogación del embrión propuesta por [Veek, 1999]. De acuerdo

a la medida de variación del grosor de la zona pelúcida definida en [Cohen y col., 1989] la variación del grosor de la zona pelúcida se calcula de la siguiente forma:

$$ZP_{var} = [(ZP_{max} - ZP_{media})/ZP_{media}] \times 100 \quad (5.4)$$

donde ZP_{max} es la zona con mayor grosor y ZP_{media} es la media del grosor de la zona. La variación del grosor de la zona pelúcida se calculó en base a esta medida.

La Tabla 5.2 muestra los resultados de la media de VGZP obtenida a partir de la segmentación automática por modelos de contornos activos de 76 imágenes de embriones, con una inicialización automática. Se obtuvo una media (desviación standard) de VGZP de 23.58 (6.28) para el grupo de embriones implantados y para el grupo de no implantados obtuvimos una media (desviación standard) de 16.31 (5.71). En base a los resultados obtenidos se puede establecer su relación con la conclusión de [Gabrielsen y col., 2000] con respecto a la relevancia de la medida de la variación del grosor de la zona pelúcida en el tercer día de cultivo, para predecir el éxito del tratamiento de fertilización in-vitro.

Parte IV

**Recomendaciones metodológicas para la
selección embrionaria en tratamientos de
reproducción asistida**

Multi-clasificador *Gaussian-stacking* aplicado a la selección embrionaria

Este Capítulo está organizado de la siguiente manera. En la primera Sección 6.1 se realizan un par de experimentos para validar el método del multi-clasificador *Gaussian-stacking* (Capítulo 4) y en la sección final se aplica el multi-clasificador al problema de la selección embrionaria. La Sección 6.1.1 describe un par de experimentos realizados con el propósito de validar el método propuesto en esta tesis del multi-clasificador *Gaussian-stacking*. La Sección 6.1.2 contiene los resultados del *primer* estudio, en el que se aplicaron algunos de los paradigmas de clasificación supervisada más representativos en el aprendizaje automático los cuales fueron descritos en la Sección 4.6. La Sección 6.1.3 reporta los resultados del *segundo* estudio obtenidos por cada uno de multi-clasificadores *Gaussian-stacking* inducidos, los cuales fueron comparados con los obtenidos por los clasificadores base, reportados en la Sección 6.1.2.

Finalmente en la última parte de este capítulo se describe un tercer estudio. En la Sección 6.2 se aplicó el multi-clasificador *Gaussian-stacking* a un problema médico complejo de índole social como es la infertilidad. En particular se aplicaron cinco diferentes multi-clasificadores *Gaussian-stacking* en la selección de un conjunto de embriones transferidos a la paciente en tratamientos de fertilización in-vitro para predecir su éxito en términos de la implantación.

Todos los paradigmas de clasificación incluidos como clasificadores-base se encuentran en el entorno de WEKA (*Waikato Environment for Knowledge Analysis*) el cual contiene una colección de diversos paradigmas de clasificación contenidos en el estado del arte de la minería de datos. Una de las principales ventajas de WEKA es la variedad de herramientas para la transformación de bases de datos y su discretización así como poder evaluar diferentes clasificadores estadísticamente, incluyendo varias medidas de validación de los modelos de clasificación ampliamente aplicados en el campo de minería de datos y aprendizaje automático. El entorno de WEKA se encuentra documentado en [Witten y Frank, 2005].

6.1. Validación del multi-clasificador *Gaussian-stacking*

El multi-clasificador *Gaussian-stacking* combina diferentes paradigmas de clasificación, denominados *clasificadores-base*, por medio de una red condicional Gaussiana aplicada como meta-clasificador. Con el propósito de validar el método del multi-clasificador *Gaussian-stacking* se desarrollaron los siguientes experimentos.

6.1.1. Diseño experimental

Base de Datos	número de instancias	número de clases	número de variables discretas/continuas	número de atributos
balance	625	3	0/4	4
breast-cancer	286	2	9/0	9
breast-W	699	2	9/0	9
diabetes	768	2	0/8	8
heart-statlog	270	2	6/7	13
hepatitis	155	2	13/6	19
ionosphere	351	2	0/34	34
iris	150	3	0/4	4
sonar	208	2	0/60	60
tic-tac-toe	958	2	9/0	9
vote	435	2	16/0	16
waveform	5000	3	0/21	21
wine	178	3	0/13	13

Tabla 6.1. Lista de bases de datos seleccionadas del repositorio UCI aplicadas ampliamente en el campo de la minería de datos.

En este estudio se evaluó individualmente el comportamiento de diferentes paradigmas de clasificación supervisada clásicos como: naive Bayes, IB1, J48, K*, naive Bayes KDE, tablas de decisión (DT) y algoritmos de máquinas de soporte vectorial (SMO), los cuales fueron aplicados como *clasificadores-base*.

Los *clasificadores-base* se aplicaron a trece bases del repositorio de datos UCI [Asuncion y Newman, 2007] como son: balance, breast-cancer, breast-W, diabetes, heart-statlog, hepatitis, ionosphere, iris, sonar, tic-tac-toe, vote, waveform y wine. La Tabla 6.1 muestra una lista de las bases de datos incluidas en este estudio así como las propiedades de cada una, como son: número de casos, número de clases, número de variables discretas o continuas y número de atributos.

Se aplicó el método de k -subconjuntos validación-cruzada con el propósito de evitar el problema de sobre ajuste de los datos (*overfitting*) de los clasificadores con un valor de $k = 10$. El porcentaje de bien clasificados se calculó a

partir del promedio de los 10 procesos de clasificación. Todas las bases de datos con datos continuos fueron discretizadas aplicando el método de discretización de igual frecuencia [Dougherty y col., 1995], con 10 intervalos.

Con el propósito de medir la exactitud y el rendimiento del multi-clasificador *Gaussian-stacking*, en el *segundo* estudio se aplicaron individualmente, cinco redes Gaussianas, como meta-clasificador para combinar siete diferentes paradigmas de clasificación (clasificadores-base) aplicando a trece bases de datos del repositorio de UCI. En base a los resultados que se muestran en la Sección 6.1.3 se validó cada uno de multi-clasificadores *Gaussian-stacking* comparándose con los obtenidos por los clasificadores base, reportados en la Sección 6.1.2.

6.1.2. Resultados experimentales a partir de clasificadores-base

A continuación se describen los resultados obtenidos por los paradigmas de clasificación supervisada que posteriormente fueron aplicados como clasificadores-base. En la Tabla 6.2 se muestra la precisión de cada clasificador en términos del porcentaje de bien clasificados así como su desviación estándar. Para cada base de datos el clasificador con mayor porcentaje de casos bien clasificados ha sido resaltado con negrillas.

La Tabla 6.2 muestra que el clasificador naive Bayes KDE obtuvo la mayor precisión en el caso de bases de datos con variables continuas como balance, ionosphere, iris y wine, así como bases de datos con variables tanto discretas como continuas como es el caso de las siguientes bases de datos: statlog, hepatitis, y breast W. En el caso de otros problemas de clasificación con datos continuos como por ejemplo diabetes y waveform el clasificador SMO obtuvo el mayor porcentaje de casos bien clasificados. El clasificador IB1 presentó el mejor comportamiento en el caso de las bases de datos sonar y tic-tac-toe, las cuales contienen datos continuos y discretos respectivamente. En el caso de las bases de datos breast cancer y vote ambas con datos discretos el clasificador J48 obtuvo el mayor porcentaje de casos bien clasificados.

6.1.3. Resultados experimentales a partir del multi-clasificador *Gaussian-stacking*

En la siguiente etapa se realizó un *segundo* experimento incluyendo los mismos siete clasificadores del experimento anterior de uso habitual en la literatura y de diversa naturaleza (naive Bayes, IB1, J48, K*, KDE, DT y SMO) como clasificadores-base para construir el multi-clasificador *Gaussian-stacking*.

Una vez más consideramos las mismas trece bases de datos utilizadas en el experimento anterior para aplicar cada una a los clasificadores-base. Construimos un nuevo vector de características compuesto por la concatenación de las distribuciones de probabilidad de la clase de cada clasificador anexando el valor de la clase real. Finalmente se construyó una meta-base de datos la

Base de datos	naive Bayes	IB1	J48	K
balance	90.40±1.72	86.56±2.70	76.64±3.77	88.48±2.27
breast-cancer	71.67±7.90	72.37±9.48	75.52±5.57	73.42±9.30
breast-W	95.99±1.63	95.13±3.44	94.56±3.62	95.42±2.68
diabetes	76.30±5.52	70.18±4.69	73.82±5.66	69.14±2.62
heart-statlog	83.70±7.02	75.18±8.56	76.66±9.08	75.18±9.41
hepatitis	84.51±8.13	80.64±8.43	83.87±7.24	81.93±8.55
ionosphere	82.62±5.46	86.32±4.58	91.45±3.27	84.61±6.21
iris	96.00±4.66	95.33±5.48	96.00±5.62	94.66±5.25
sonar	67.78±9.29	86.53±7.00	71.15±7.10	84.61±7.85
tic-tac-toe	69.62±3.06	98.95±0.98	84.55±3.93	96.86±1.63
vote	90.11±4.55	92.41±3.53	96.32±3.42	93.33±4.34
waveform	76.00±1.49	75.12±1.13	75.36±2.25	77.02±1.12
wine	98.87±2.34	98.31±3.74	93.82±3.14	98.31±3.74

Base de datos	naive Bayes - KDE	DT	SMO
balance	91.36±1.37	74.56±3.89	87.68±2.51
breast-cancer	71.67±7.90	72.72±7.57	69.58±7.54
breast-W	97.56±1.79	95.42±2.67	96.99±2.07
diabetes	74.60±6.38	73.30±3.63	77.34±4.06
heart-statlog	84.44±4.55	82.96±7.44	84.07±5.53
hepatitis	85.80±7.98	81.29±11.08	85.16±7.82
ionosphere	91.73±2.05	89.45±4.47	88.60±4.26
iris	96.66±4.71	92.66±5.83	96.00±4.66
sonar	73.07±9.44	74.51±8.17	75.96±7.79
tic-tac-toe	69.62±3.06	78.28±1.86	98.32±1.49
vote	90.11±4.55	95.17±5.08	96.09±3.40
waveform	81.76±1.49	73.58±2.84	86.18±1.14
wine	98.87±2.34	96.62±2.89	98.31±2.68

Tabla 6.2. Resultados de siete clasificadores clásicos aplicados a trece problemas de clasificación supervisada.

cual fue utilizada para la clasificación del nivel superior del esquema *stacked* aplicando.

En el meta-nivel, se aplicó diferentes multi-clasificadores *Gaussian-stacking* los cuales fueron descritos en el capítulo anterior como son los clasificadores *filter-ranking-naive Bayes*, *wrapper-selective naive Bayes*, *wrapper-semi naive Bayes*, *wrapper-tree augmented naive Bayes* y *wrapper-k-dependence Bayesian classifier* adaptados al dominio continuo [Pérez y col., 2006].

Con el propósito de medir la mejora de la precisión del multi-clasificador propuesto en esta tesis (*Gaussian-stacking*), se comparó el mayor porcentaje de casos bien clasificados de los diferentes clasificadores-base (naive Bayes, IB1, J48, K, naive Bayes KDE, tablas de decisión y máquinas de soporte

Base de datos	filter-ranking NB	wrapper-selectiveNB	seminaive Bayes	wrapper-TAN	wrapper-kDB
balance	93.92 ± 1.71	95.68 ± 2.01	95.67 ± 2.98	95.68 ± 3.28	95.67 ± 1.61
breast-cancer	75.51 ± 9.47	75.46 ± 5.83	75.60 ± 10.99	75.49 ± 6.50	75.50 ± 2.85
breast-W	97.13 ± 2.39	97.27 ± 1.75	97.28 ± 1.85	97.28 ± 1.62	97.13 ± 1.43
diabetes	78.78 ± 4.49	78.25 ± 5.18	78.26 ± 3.92	78.12 ± 4.39	78.25 ± 5.23
heart-statlog	85.55 ± 7.30	85.55 ± 6.07	84.07 ± 4.69	85.18 ± 6.82	84.07 ± 9.52
hepatitis	84.50 ± 8.12	85.74 ± 8.33	85.66 ± 7.17	87.12 ± 9.42	85.12 ± 5.85
ionosphere	93.73 ± 3.98	94.00 ± 3.49	93.73 ± 5.23	93.73 ± 4.37	93.73 ± 4.57
iris	96.00 ± 4.42	96.00 ± 4.42	94.66 ± 5.81	96.00 ± 4.42	96.00 ± 5.33
sonar	87.07 ± 7.40	88.07 ± 8.82	87.47 ± 7.17	87.47 ± 7.60	87.54 ± 7.14
Tic-tac-toe	98.53 ± 0.95	98.85 ± 1.27	99.58 ± 0.51	98.85 ± 0.98	98.53 ± 1.06
vote	96.80 ± 3.54	96.55 ± 2.12	96.32 ± 2.75	96.78 ± 1.12	96.76 ± 2.98
waveform	84.50 ± 1.52	84.69 ± 1.86	84.92 ± 1.38	84.59 ± 1.70	84.68 ± 1.79
wine	98.88 ± 2.22	99.44 ± 1.66	99.44 ± 1.66	99.44 ± 1.66	99.44 ± 1.66

Tabla 6.3. Resultados de cinco multi-clasificadores Gaussian-stacking.

vectorial) reportados en la Tabla 6.2 con los resultados obtenidos de los cinco meta-clasificadores *Gaussian-stacking* propuestos en este trabajo los cuales se presentan en la Tabla 6.3.

En la Tabla 6.3 se resalta en negrita el porcentaje más alto de casos bien clasificados entre los cinco multi-clasificadores *Gaussian-stacking*. En todos los casos los cinco multi-clasificadores obtienen mejor precisión que cada uno de los clasificadores-base para cada dominio presentado en la Tabla 6.2. Estos resultados demuestran lo propuesto por [Dietterich, 1998] donde reporta que el método de pila obtiene mejores resultados que cualquiera de los clasificadores-base. Por ejemplo, si consideramos la base de datos de diabetes y el clasificador-base *SMO* siendo el que obtuvo 77.34 ± 4.06 el mayor porcentaje de casos bien clasificados reportado en la Tabla 6.2 el cual resulta inferior a cualquiera de los resultados obtenidos por los cinco multi-clasificadores *Gaussian stacking* propuestos. El clasificador *filter-ranking naive Bayes* obtuvo el porcentaje de casos bien clasificados más alto con un 78.78 ± 4.49 y en el caso opuesto es el del clasificador *wrapper-TAN* que obtuvo el porcentaje de casos bien clasificados más bajo con un 78.12 ± 4.39 .

En base a los resultados reportados en la Tabla 6.3, queda validado el modelo del multi-clasificador *Gaussian-stacking* y se demuestra que el meta-clasificador basado en redes Gaussianas es un método adecuado para combinar las distribuciones de probabilidad de la clase de diferentes paradigmas de clasificación.

Con el propósito de comparar el comportamiento de los cinco multi-clasificadores *Gaussian-stacking* (filter-ranking naive Bayes, wrapper-selective naive Bayes, seminaive Bayes, wrapper-TAN y wrapper-kDB) aplicados a trece bases de datos del repositorio UCI, se aplicó la prueba no paramétrica de Kruskal-Wallis y con un valor de $p = 0,9986$ no se encuentra una diferencia significativa entre ellos.

6.2. Multi-clasificador *Gaussian-stacking* aplicado al problema supervisado de la selección embrionaria

El propósito del *tercer* estudio es aplicar diferentes multi-clasificadores basados en el esquema *Gaussian-stacking*, los cuales son descritos en el Capítulo 4 al complejo problema biomédico de la selección de los mejores embriones en tratamientos de reproducción asistida, y predecir la viabilidad de implantación de algún embrión del conjunto transferido, el cual ha sido planteado como un problema de clasificación supervisada.

6.2.1. Base de datos

La base de datos fue obtenida de 63 expedientes clínicos del programa de fertilización in-vitro de la unidad de reproducción de la Clínica del Pilar en San Sebastián, Guipúzcoa durante el periodo de julio de 2003 a diciembre de 2005. Los 63 casos fueron escogidos de un total de 89 ciclos de tratamientos de fertilización in-vitro de pacientes entre 27-46 años. Los tratamientos en los que se transfirieron dos embriones se descartaron, ya que la condición necesaria que empleamos para seleccionar los casos incluidos en el presente estudio fueron los tratamientos en los que se transfirió un grupo de tres embriones.

Variable	Rango de valores
Ciclo actual	Numérico
Tratamientos previos	Numérico
Edad	Numérico
Donante	{Si,No}
Calidad del semen	{Buena,Media,Defectuosa}
Causa femenina	{Si,No}
Causa masculina	{Si,No}
Tipo de esterilidad	{Primaria,Secundaria}
Número de células	Numérico
Catalogación del cigoto	{Z1,Z2,Z3,Z4}
Catalogación del embrión	{Tipo I, Tipo II, Tipo III, Tipo IV, Tipo V}
Tamaño de los blastómeros	{Igual, Desigual, Defectos en el citoplasma}
Grado de fragmentación	{0 %,(0-10) %,(10-25) %,(25-50) %,(50-100) %}
Grosor de la zona pelúcida	{Delgada,Gruesa,Muy Gruesa}
Blastómeros multinucleados	{Si,No}
Número de embriones transferidos	Numérico
Número de embriones congelados	Numérico
Calidad en la transferencia de embriones	{Buena,Media,Defectuosa}
Día de transferencia	Numérico

Tabla 6.4. Lista de variables y su correspondiente rango de valores, incluidos en la primera base de datos obtenida de la Clínica del Pilar, San Sebastián. Utilizada en los Capítulos 6 y 8.

El protocolo aplicado en la Clínica del Pilar para la selección embrionaria está basado primeramente en la catalogación del sistema propuesto por [Scott

y col., 2000] para la catalogación del cigoto y posteriormente en la evaluación y catalogación morfológica del embrión basado en el sistema propuesto en [Veek, 1999], ambos descritos en el Capítulo 5.

Dichas evaluaciones fueron registradas en la base de datos junto con las variables relacionadas con datos clínicos y del tratamiento como son la edad de la paciente, el número de ciclos previos, el ciclo actual, la calidad del esperma, causa de infertilidad femenina o masculina, infertilidad primaria o secundaria, número de embriones transferidos, si dichos embriones eran descongelados o frescos, etc. La Tabla 6.4 muestra la lista de variables incluidas en la base de datos y sus respectivos valores.

6.2.2. Diseño experimental

Se aplicaron los algoritmos *wrapper-TAN*, *wrapper-kDB*, *wrapper-selective naive Bayes* y *wrapper-semi naive Bayes*, como meta-clasificadores para inducir diferentes multi-clasificadores *Gaussian-stacking*. La búsqueda del mejor modelo fue guiada en todos los casos siguiendo la medida de mejora en la clasificación en base al porcentaje de casos bien clasificados. La única excepción fue el modelo *filter-ranking naive Bayes*, en el que la búsqueda del mejor modelo está basada en el criterio de información mutua.

Los datos aplicados para la inducción de cada meta-clasificador fueron obtenidos de las distribuciones de probabilidad de la variable clase, producto de la clasificación de siete *clasificadores-base* (naive Bayes, IB1, J48, K*, naive Bayes KDE, decision table (DT) y máquina de soporte vectorial (SMO)). A esta información se añadió el resultado del tratamiento de fertilización in-vitro, como la variable clase. Se aplicó a cada clasificador-base la base misma base de datos, la cual contiene 18 casos de tratamientos que resultaron exitosos registrándose la implantación y 45 casos en los que no se logró (no-implantados). Cada caso contiene un conjunto de tres embriones transferidos. Se aplicó una validación cruzada dejando-uno-fuera en cada clasificador-base, evitando de esta manera el problema de sobre ajuste de los datos (*overfitting*).

6.2.3. Resultados experimentales

En esta Sección se muestran los resultados obtenidos al aplicar cinco diferentes multi-clasificadores *Gaussian-stacking* al problema de la selección de un conjunto de tres embriones para ser transferidos en tratamientos de fertilización in-vitro.

La Tabla 6.5 muestra el porcentaje de casos bien clasificados y su desviación estándar (SD). El algoritmo *wrapper-semi naive Bayes* obtuvo la máxima precisión del conjunto de meta-clasificadores con un porcentaje de casos bien clasificados del 71.66 ± 17.89 mientras que el clasificador *filter-ranking naive Bayes* obtuvo una precisión del 58.09 ± 20.45 .

<i>Clasificador</i>	<i>Precisión±SD</i>
Gaussian-stacking	
<i>filter-ranking naive Bayes</i>	<i>58.09±20.45</i>
<i>wrapper selective naive Bayes</i>	<i>71.19±15.52</i>
<i>wrapper semi naive Bayes</i>	<i>71.66±17.89</i>
<i>wrapper-TAN</i>	<i>71.19±14.65</i>
<i>wrapper-kDB</i>	<i>71.19±18.76</i>

Tabla 6.5. Resultados de cinco multi-clasificadores Gaussian-stacking aplicados al problema de selección embrionaria en tratamientos de fecundación in-vitro.

Como prueba adicional con el propósito de comparar el grado de acuerdo entre clasificadores-base, se realizó un análisis de concordancia entre pares de clasificadores-base, aplicando la prueba estadística Kappa. El par de clasificadores-base que obtuvieron el más alto índice Kappa fueron los clasificadores naive Bayes y SMO con un índice Kappa de 0.5414, y los algoritmos J48 y K* un valor Kappa de 0.5280. El par de clasificadores que obtuvieron el más bajo índice Kappa fueron los algoritmos DT y SMO con -0.0073. Los índices Kappa obtenidos se pueden interpretar en base a la escala propuesta por [Landis y Koch, 1977] en el que se corresponden con un grado de concordancia moderado entre pares de clasificadores, y en el caso de que el par de clasificadores con índice Kappa negativo, demuestra que hay un desacuerdo sistemático entre el par de clasificadores.

El trabajo presentado en este capítulo es avalado por [Morales y col., 2009].

Selección del mejor embrión para su transferencia en tratamientos de reproducción humana asistida

Introducción

El propósito de este estudio es aplicar los modelos clásicos de los clasificadores Bayesianos al problema de selección embrionaria, eligiendo al mejor embrión para su transferencia en tratamientos de fertilización in-vitro así como la predicción de la viabilidad de implantación del embrión, el cual ha sido planteado como un problema de clasificación supervisada.

Este estudio difiere de todos los planteados en los Capítulos 6, 8 y 9 en los cuales se planteó la selección de un conjunto de embriones en base a información clínica y morfológica de los embriones transferidos en el tratamiento de reproducción in-vitro.

7.1. Base de datos

Tanto las imágenes de embriones en nuestros experimentos, como su caracterización por momentos centrales se obtuvieron por el estudio publicado en la literatura por [Patrizi y col., 2004], en el cual los datos y su validación son aproximados de diferente manera. Los tratamientos de reproducción asistida fueron conducidos en el Centro Genesis, Roma, Italia durante el periodo entre enero 1998 y diciembre 2001. La base de datos incluye 275 ciclos de inyección de esperma intra citoplasmática (ICSI) realizados en 195 pacientes [Patrizi y col., 2004].

Para este estudio fueron seleccionados y fotografiados 249 embriones. Imágenes de embriones en la etapa de cuatro células fueron consideradas entre 40 y 50 horas después de la fertilización y antes de la transferencia. Los embriones fueron catalogados de acuerdo a la escala [Mills, 1992]. La Figura 7.1 muestra algunos ejemplos de imágenes de los embriones seleccionados.

Se aplicaron los momentos centrales para extraer la información morfológica del embrión a partir de la imagen. Los momentos centrales son una técnica

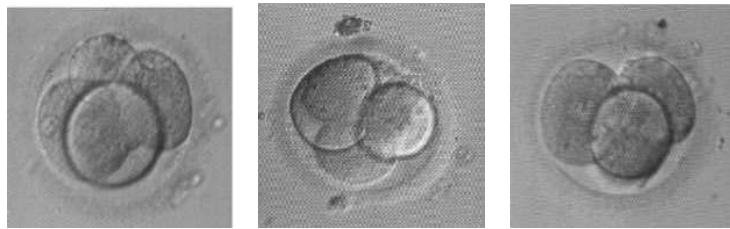


Figura 7.1. Imágenes reales de embriones adquiridas en el Centro Genesis (Roma, Italia), catalogados en base a la escala de Mills [Mills, 1992] como embriones con una calidad estándar.

matemática que permite la caracterización de la forma de los objetos contenidos en la imagen. En especial en imágenes médicas donde los contornos de las estructuras suelen ser difusos conteniendo variaciones sutiles en la intensidad suelen ser aplicados los momentos centrales. La intensidad de los píxeles de la imagen indica la homogeneidad de la imagen, si tiene zonas oscuras o claras y cómo están distribuidas. En el Apéndice C se explican con detalle los momentos centrales.

Se calcularon los primeros seis momentos centrales a partir de la distribución de los píxeles en las direcciones horizontal y vertical. El primer momento central se descartó, ya que su valor es cero. Los siguientes diez valores se utilizaron para formar un vector de características para cada embrión añadiéndole el valor de la clase real. La clase real tiene un valor de 1 si ha ocurrido la implantación y el valor 0 en caso contrario.

7.2. Diseño experimental

Se adaptaron los clasificadores Bayesianos con un aprendizaje sensible al coste del error variando el umbral de decisión τ [Breiman y col., 1984, Duda y col., 2001]. Los clasificadores aplicados en este capítulo son: el clasificador naive Bayes, selective naive Bayes, semi naive Bayes, TAN y kDB así como los clasificadores con un enfoque de envoltura para inducir los modelos *wrapper-TAN* y *wrapper-kDB*, los cuales fueron introducidos en el Capítulo 3.

Debido a que las variables predictoras son continuas, se aplicó un algoritmo de igual frecuencia [Dougherty y col., 1995] en tres intervalos, para su discretización. Se adaptaron los clasificadores Bayesianos sensibles al coste del error en el marco de trabajo del software Elvira [Elvira Consortium, 2002].

Con el propósito de validar los clasificadores Bayesianos, se aplicó una validación cruzada con diez subconjuntos estratificados y se comparó el comportamiento de los clasificadores en base al análisis del área bajo la curva ROC. En base al método de la regla trapezoidal propuesta por [Bamber, 1975] se

calculó del área bajo la curva ROC. Este método es equivalente a la prueba estadística de Mann-Whitney. Se consideraron diez valores del área bajo la curva con un intervalo de confianza del 95 % para cada una de las curvas ROC obtenidas.

Se aplicó al proceso de clasificación el *aprendizaje sensible al coste* variando los valores del umbral de decisión τ para la clasificación, introducido en la Sección 2.2. En este estudio se han tenido en cuenta 1000 valores para el umbral de decisión τ . Para cada valor del umbral de decisión, se aplicó la Ecuación (2.2) adaptada a la etapa de clasificación de los clasificadores Bayesianos.

Como se mencionó en el Capítulo 3, los algoritmos TAN y kDB aplican el criterio de información mutua para guiar la búsqueda del mejor modelo. En [Blanco y col., 2005] se presenta una alternativa para la inducción de estos modelos, introduciendo los clasificadores *wrapper-TAN* y *wrapper-kDB*, donde la búsqueda en el espacio de las posibles estructuras, de los modelos TAN y kDB, está guiada por la medida de mejora de la clasificación en base al porcentaje de casos bien clasificados.

7.3. Resultados experimentales

Con el propósito de evaluar la precisión de los clasificadores Bayesianos aplicados a la selección del mejor embrión en base a su caracterización morfológica por medio de los momentos centrales se presentan en la Tabla 7.1 los resultados obtenidos aplicando una validación-cruzada con 10 subconjuntos estratificados. El clasificador *wrapper-TAN* obtuvo la mayor precisión con un porcentaje de casos bien clasificados de 91.25 ± 11.86 , siendo el clasificador semi naive Bayes el que menor porcentaje de éxito presenta, con un porcentaje de casos bien clasificados de 78.48 ± 02.62 .

La Figura 7.2 muestran las curvas ROC obtenidas por los clasificadores Bayesianos aplicados en este estudio: naive Bayes, selective naive Bayes, semi naive Bayes, TAN, kDB, *wrapper-TAN* y *wrapper-kDB*. Cada curva ROC se obtuvo con cien valores muestreados de los 1000 valores del umbral de clasificación τ , mostrándose asimismo diez barras de error con un intervalo de confianza del 95 %.

En la Figura 7.2 se observa que los clasificadores *wrapper-TAN*, *wrapper-kDB*, TAN y kDB tienen el área bajo la curva más alta. El clasificador kDB alcanza la mejor sensibilidad con un 82 %, con una especificidad del 80 %. El clasificador *wrapper-TAN* alcanza una sensibilidad más alta con un 92 % y una especificidad del 60 %. El mejor clasificador con una relación sensibilidad y especificidad del 90 % y 74 % respectivamente es el clasificador *wrapper-kDB*.

La Tabla 7.2 muestra el área bajo las curvas ROC y sus intervalos de confianza del 95 %. Los clasificadores *wrapper-TAN*, *wrapper-kDB*, kDB y TAN tienen la más alta AUC con los valores 0.9994, 0.9991, 0.9819 y 0.9454 respectivamente. Los clasificadores *semi naive Bayes*, *selective naive Bayes*

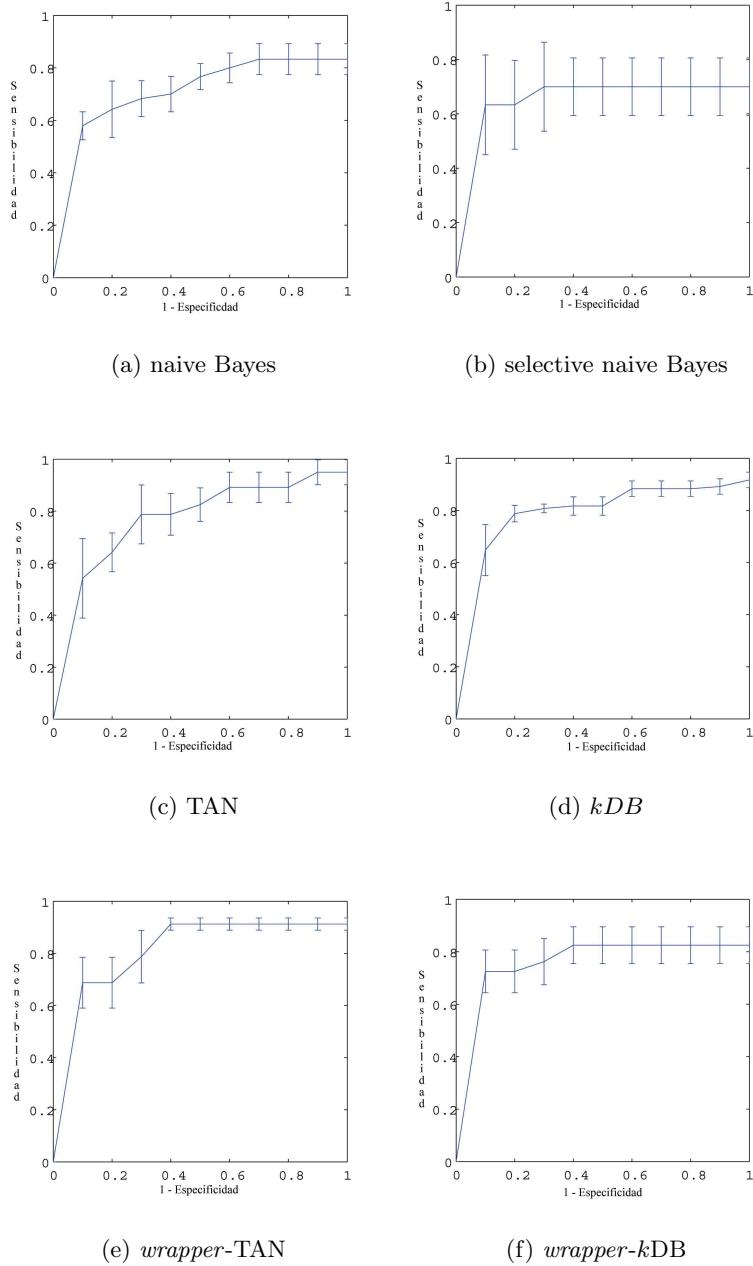


Figura 7.2. Gráficas de las curvas ROC de diferentes clasificadores Bayesianos aplicados a la selección del mejor embrión, con intervalos de confianza del 95 % basados en 10 ejecuciones.

<i>Clasificador</i>	<i>Precisión±SD</i>
<i>naive Bayes</i>	<i>85.49±17.57</i>
<i>selective naive Bayes</i>	<i>83.33±16.10</i>
<i>semi naive Bayes</i>	<i>78.48± 2.62</i>
<i>TAN</i>	<i>89.17±15.80</i>
<i>kDB</i>	<i>88.33±12.85</i>
<i>wrapper-TAN</i>	<i>91.25±11.86</i>
<i>wrapper-kDB</i>	<i>87.63±11.24</i>

Tabla 7.1. Resultados del porcentaje de casos bien clasificados de diferentes clasificadores Bayesianos aplicados a la selección embrionaria a partir de las características extraídas de la imagen.

y *naive Bayes*, obtuvieron 0.7008, 0.8483 y 0.8918 áreas bajo la curva ROC, respectivamente.

<i>Clasificador</i>	<i>AUC</i>
<i>naive Bayes</i>	<i>0.8918±0.0718</i>
<i>selective naive Bayes</i>	<i>0.8483±0.0175</i>
<i>semi naive Bayes</i>	<i>0.7008±0.0643</i>
<i>TAN</i>	<i>0.9454±0.0785</i>
<i>kDB</i>	<i>0.9991±0.0525</i>
<i>wrapper-TAN</i>	<i>0.9994±0.0589</i>
<i>wrapper-kDB</i>	<i>0.9819±0.0264</i>

Tabla 7.2. Áreas bajo la curva ROC y sus intervalos de confianza del 95 % para cada uno de los siete diferentes clasificadores Bayesianos aplicados a la selección del embrión más viable para su transferencia en tratamientos de fertilización in-vitro.

Finalmente, con el propósito de poder comparar los resultados obtenidos por clasificadores Bayesianos se realizó otro experimento en el que se aplicó un estimador *log-likelihood* con penalización (*ridge*) [Le Cessie y Van Houwelingen, 1992] de regresión logística al problema de selección embrionaria, el cual ha sido ampliamente aplicado en la literatura [Verberg y col., 2007, della Ragonne y col., 2007, Van Peperstraten y col., 2008].

Brevemente se comentará que la regresión logística es una herramienta estadística que permite modelar la relación entre una o más variables con un conjunto de variables predictoras independientes o covariables, ya sean cualitativas o cuantitativas. En el caso de una variable cualitativa, dicotómica se aplica la regresión logística binaria o binomial y en el caso de analizar más de dos valores se aplica la regresión logística multinomial. Al aplicar una penalización (*ridge*) al algoritmo de verosimilitud (*log-likelihood*) se evita el problema de la multicolinealidad (vectores de características casi idénticas) incluyendo solo los coeficientes de regresión con un valor mayor (relevantes).

El uso de penalidades en regresión logística fue propuesto por [Le Cessie y Van Houwelingen, 1992], siendo la variable clase de tipo ordinal.

Para la clasificación con el método de regresión logística con penalización se utilizaron los valores predeterminados en el entorno de trabajo WEKA [Witten y Frank, 2005]. La precisión del estimador fue del 82.23 %, con un área bajo la curva ROC de 0.535 en base a una validación cruzada de 10 subconjuntos.

Comparando los resultados registrados en la literatura para el problema de la selección embrionaria, obtenemos un buen desempeño de los clasificadores Bayesianos, en especial para *wrapper-TAN* en base a la selección de variables obtuvo una precisión del 91 %. La exactitud más baja fue de 78 % con un clasificador semi naive Bayes, lo cual es aún comparable al 82.33 % obtenido con los mismos datos usando regresión logística, y también al 81.79 % del estudio en [Patrizi y col., 2004] con la colección de datos original usando un método de validación diferente. En lo que respecta a otros estudios, [Saith y col., 1998] obtuvieron un 74 % de exactitud aplicando el algoritmo de árboles de clasificación.

En base a los resultados obtenidos en este estudio se puede concluir que los clasificadores TAN y *kDB* son los modelos que reflejan las dependencias condicionales entre los momentos invariantes de segundo y tercer orden, lo que evidencia la complejidad del problema y la relevancia de considerar la selección de variables para poder mejorar la exactitud de la clasificación. Siendo el modelo *wrapper-kDB* el mejor clasificador en términos de la relación sensibilidad y especificidad. El trabajo presentado en este capítulo se publicó en [Morales y col., 2008b].

Selección de un conjunto de embriones en tratamientos de fertilización in-vitro

Introducción

En este capítulo se plantea un estudio empírico en base a la precisión de los clasificadores Bayesianos descritos previamente en el Capítulo 3 y su potencial aplicación para seleccionar la mejor combinación de embriones humanos para ser transferidos en un grupo de tres en tratamientos de fertilización in-vitro.

Este estudio difiere con el planteado en el Capítulo 7 en el que se seleccionaba un único embrión en base a sus características morfológicas extraídas de la imagen provenientes de una clínica diferente.

8.1. Diseño experimental

Se realizó un estudio retrospectivo en el que fue necesario hacer algunas asunciones debido a la naturaleza de los datos y las limitaciones al obtenerlos. Se aplicó la base de datos descrita en la Sección 6.2.1, la cual contiene 18 casos de tratamientos exitosos (en los que se logró la implantación) y 45 casos no exitosos. Debido a que los clasificadores Bayesianos aplicados no manejan variables continuas y discretas (como es el caso de la base de datos), fue necesaria la discretización de los datos por medio del algoritmo de igual frecuencia [Dougherty y col., 1995]. Las variables fueron discretizadas en dos intervalos. Para validar cada uno de los clasificadores Bayesianos se aplicó el método de validación-cruzada dejando-uno-fuera.

La precisión de los clasificadores Bayesianos (*naive Bayes, selective naive Bayes, semi naive Bayes, TAN, kDB, FSNB, FTAN y FkDB*) fue evaluada en base al porcentaje de casos bien clasificados y se aplicó la prueba no paramétrica de McNemmar para evaluar la diferencia de la precisión entre cada par de clasificadores.

8.2. Resultados experimentales

Los resultados de este experimento se describen en términos del porcentaje de casos bien clasificados, los cuales se encuentran resumidos en la Tabla 8.1 mientras que en la Tabla 8.2 se muestran los resultados del estudio en términos de la sensibilidad, la especificidad, el valor de predicción de implantación y del valor predictivo de no-implantación.

Nótese que el clasificador semi naive Bayes obtuvo el porcentaje de casos bien clasificados más alto con un 71.43 % con una sensibilidad del 22.22 % y una especificidad del 91.11 % presentando un valor de predicción para los casos implantados del 50.00 % y un 74.54 % en la predicción de los casos no-implantados. Los clasificadores naive Bayes, selective naive Bayes, FSNB, y FTAN obtuvieron una precisión del 68.25 %, mientras que el clasificador naive Bayes obtuvo un 38.89 % de sensibilidad, 80.00 % de especificidad y un valor de predicción de implantación del 43.75 %. El clasificador Bayesiano *kDB* fue el que obtuvo la menor precisión con un 60.32 % de casos bien clasificados.

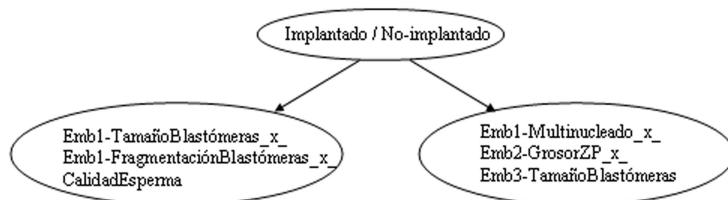


Figura 8.1. Modelo semi naive Bayes para la clasificación de un conjunto de embriones en tratamientos de fertilización in-vitro.

La Figura 8.1 muestra la estructura del clasificador semi naive Bayes, el cual obtuvo la mayor precisión en la selección de un conjunto de embriones en tratamientos de fertilización in-vitro. Su estructura está formada por dos nodos: El primero incluye el producto cartesiano de las variables predictoras *Embrión1-Tamaño Blastómeros*, *Embrión1-Fragmentación Blastómeros*, y *Calidad del esperma*. El segundo nodo está formado por el producto cartesiano de las variables predictoras *Embrión1-Multinucleado*, *Embrión2-Grosor de la zona pelúcida* y *Embrión3-Tamaño de Blastómeros*. Estas variables predictoras son acorde respecto a la asunción de los expertos de incluir las características morfológicas del embrión así como la información clínica del tratamiento [Ziebe y col., 1997, Alikanin y col., 1999, Gabrielsen y col., 2001].

Al aplicar la prueba de McNemar para evaluar la diferencia estadística del rendimiento entre parejas de clasificadores Bayesianos no se encontró diferencia significativa entre el mejor clasificador, *semi naive Bayes*, y cada uno

de los clasificadores empleados en este estudio (naive Bayes, selective naive Bayes, *k*DB, TAN, FSNB, FTAN y *Fk*DB).

		naive Bayes	semi naive Bayes	selective NB	TAN	<i>k</i> DB
		real	0	1	0	1
predicha	0	36	11	41	14	42
	1	9	7	4	4	3
Precisión		68.25		71.43		68.25
					63.49	60.32

Tabla 8.1. Resultados de la selección de un conjunto de embriones en tratamientos de fertilización in-vitro, expresados en términos de la matriz de confusión y porcentaje de casos bien clasificados, por medio de clasificadores Bayesianos.

Clasificador	Precisión	Sensibilidad	Especificidad	Valor predictivo Implantación	Valor predictivo No-implantación
naive Bayes	68.25	38.89	80.00	43.75	76.59
semi naive Bayes	71.43	22.22	91.11	50.00	74.54
selective naive Bayes	68.25	5.55	93.33	25.00	71.19
TAN	63.49	5.55	86.67	14.29	69.64
<i>k</i> DB	60.32	5.55	82.22	11.11	68.52
FSNB	68.25	11.11	91.11	33.33	71.92
FTAN	68.25	11.11	91.11	33.33	71.92
<i>Fk</i> DB	63.49	0.00	88.89	0.00	68.96

Tabla 8.2. Resultados de la clasificación de un conjunto de tres embriones por medio de diferentes clasificadores Bayesianos en términos de la sensibilidad, la especificidad, el valor de predicción de implantación y del valor predictivo de no-implantación.

8.2.1. Selección de un subconjunto de características (FSS)

Otro aspecto importante en este estudio fue la selección de un subconjunto de características por medio del proceso FSS descrito en la Sección 2.4 para inducir los modelos semi naive Bayes, selective naive Bayes, FSNB, FTAN y *Fk*DB. La Tabla 8.3 muestra el subconjunto de variables predictoras más relevantes. Cabe mencionar que el proceso de selección del subconjunto de variables para cada clasificador se realizó incluyendo todas las variables predictoras listadas en la Sección 6.4.

El subconjunto de variables seleccionado se corresponde con la opinión de los embriólogos, ya que ellos se concentran principalmente en identificar dos características para evaluar el potencial de un embrión para su posible implantación, las cuales son *el tamaño de los blastómeros* y *el grado de fragmentación*. Estas características morfológicas de los embriones fueron observadas durante el proceso de selección de embriones por los embriólogos de la unidad de IVF de la Clínica del Pilar. Además, las características del tamaño de los blastómeros y el grado de fragmentación juegan un papel importante en la habilidad de un embrión para progresar a un embarazo clínico [Alikanin y col.,

Lista de variables	semi NB	selective NB	FSNB	FTAN	FkDB
Embrión1 tamaño de blastómeros	✓		✓	✓	✓
Embrión1 fragmentación	✓				
Embrión1 grosor de la zona pelúcida			✓	✓	
Embrión1 multinucleado	✓		✓	✓	✓
Embrión2 grosor de la zona pelúcida	✓		✓	✓	✓
Embrión2 multinucleado			✓	✓	✓
Embrión3 tamaño de blastómeros	✓		✓	✓	✓
Embrión3 fragmentation		✓			
Embrión3 grosor de la zona pelúcida			✓	✓	✓
Embrión3 multinucleado			✓	✓	✓
Calidad del semen	✓				
Infertilidad primaria			✓	✓	✓
Infertilidad secundaria			✓	✓	✓
Número de embriones transferidos			✓	✓	✓

Tabla 8.3. Lista de variables incluidas en la inducción de los clasificadores Bayesianos con una base de datos de 63 casos.

1999, Sakkas y col., 2001]. El *tamaño de los blastómeros*, es una característica relacionada a la tasa de crecimiento del embrión (*cleavage*).

Otras variables predictoras seleccionadas para el aprendizaje de los clasificadores Bayesianos son: el hecho de si el embrión es *multinucleado* o no, lo cual también está mencionado en la literatura [Ziebe y col., 1997, Alikanin y col., 1999, Shen y col., 2002]; la variable predictoría *grosor de la zona pelúcida* coincide con la investigación presentada en [Gabrielsen y col., 2000, Gabrielsen y col., 2001], en la cual se prueba que el *grosor de la zona pelúcida* tiene relación con la implantación del embrión (embarazo) como resultado de los tratamientos de fertilización in-vitro.

Nótese que otras características clínicas directamente relacionadas con el éxito del tratamiento de IVF además de la calidad de cada uno de los embriones también fueron incluidas en la inducción de los clasificadores Bayesianos, como son el *número de embriones transferidos*, la *infertilidad* siendo de tipo primaria o secundaria, y la *calidad del esperma* todas ellas de acuerdo al criterio presentado en [Baczkowski y col., 2004] para mejorar la tasa de éxito de los tratamientos de reproducción asistida.

8.2.2. Simulación de la experiencia por medio del clasificador naive Bayes

En otro experimento se planteó considerar todas las variables obtenidas para la inducción de los clasificadores Bayesianos listados en la Tabla 8.3. Se construyó un modelo naive Bayes para simular el conocimiento del experto incluyendo todas las variables incluidas en la Tabla 8.3 con el propósito de

explorar si esta información en su conjunto mejora la precisión de clasificación por medio del clasificador naive Bayes.

La Tabla 8.3 muestra los resultados obtenidos por el clasificador naive Bayes diseñado manualmente (véase la Figura 8.2). Se obtuvo un porcentaje de casos bien clasificados del 68.25 % con 18 falsos negativos y 2 falsos positivos. Nótese que ningún caso positivo fue identificado, por lo que se considera que el conjunto de variables no inducen un buen clasificador naive Bayes y por lo tanto el modelo diseñado manualmente no es viable para su aplicación en el problema real objetivo de este trabajo que es la selección embrionaria en tratamientos de fertilización in-vitro.

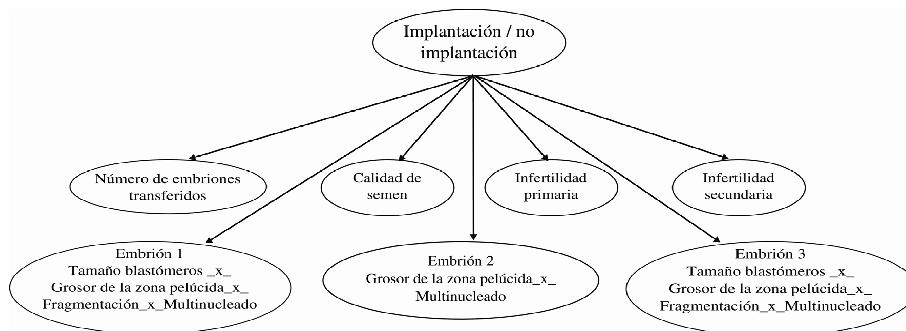


Figura 8.2. Estructura del clasificador naive Bayes construido a partir de las variables seleccionadas por medio del proceso de selección de variables FSS.

En base a los resultados obtenidos en este estudio se consideró adecuado la aplicación de los clasificadores Bayesianos para la selección de un conjunto de tres embriones en tratamientos de fertilización in-vitro. Se encontró una lista de variables predictoras relevantes y en concordancia con respecto a la información considerada por el experto en el proceso de la selección embrionaria. Se descartó la posibilidad de construir manualmente una red Bayesiana en base a las variables encontradas por el proceso de selección aplicado en la inducción de los clasificadores Bayesianos aplicados en este estudio. Finalmente se encontró un modelo del clasificador semi naive Bayes con dos nodos incluyendo las variables más relevantes y directamente relacionadas, lo cual permitió el planteamiento de otro experimento relacionado con el enriquecimiento de la información en base a las características del embrión a partir de la imagen con el propósito de mejorar la tarea de clasificación en base a los clasificadores Bayesianos. Este estudio se publicó en [Morales y col., 2008c] y [Morales y col., 2007].

Mejora de la clasificación en la selección embrionaria en base a información extraída de la imagen del embrión

Introducción

En base a los resultados reportados en el Capítulo 8, en el que se observa la viabilidad de los clasificadores Bayesianos para predecir el éxito del tratamiento de reproducción asistida en base a la selección de un conjunto de un grupo de embriones (en particular el análisis de la lista de variables producto del proceso FSS que se muestran en la Tabla 8.3 y su relación con la literatura en el campo de la biología) se planteó el último estudio considerando que la medición de la variación del grosor de la zona pelúcida está directamente relacionada con el éxito del tratamiento de reproducción asistida en base a la capacidad de implantación del embrión [Gabrielsen y col., 2000, Gabrielsen y col., 2001].

Se propuso enriquecer la información proporcionada por la clínica del Pilar, con la medición de la variación del grosor de la zona pelúcida para sustituir la medida asignada por el experto (delgada, gruesa y muy gruesa) así como otras características extraídas de la imagen para caracterizar la morfología del embrión como los momentos centrales de segundo orden, relacionados con el estudio presentado en el Capítulo 7. Todo ello para comprobar la hipótesis de que al enriquecer la información del mejor embrión se mejora el proceso de FSS y por ende la clasificación de los clasificadores Bayesianos inducidos con el subconjunto de variables producto del mismo.

Para la realización de los estudios presentados en este capítulo se construyeron dos bases de datos diferentes a las usadas en los estudios presentados en los Capítulos 6, 7 y 8.

Dado que la imagen del embrión es una imagen sencilla con formas definidas y contornos difusos, en esta tesis se propone la aplicación de contornos activos para la segmentación de la zona pelúcida para medir la variación del grosor de la membrana. Este método ha sido ampliamente utilizado en la segmentación de imágenes médicas, debido a su capacidad de segmentar contornos difusos y con baja intensidad. El Capítulo 5 describe el proceso de

segmentación así como el método de contornos activos y los resultados obtenidos para la validación de su aplicación.

Se midieron otras características con relación a la textura de la imagen del embrión con el propósito de caracterizar su morfología a partir de la imagen. Se calculó la energía, entropía, momento inverso, homogeneidad, forma, inercia, entre otras medidas extraídas de la imagen del embrión. Dichas medidas son introducidas en el Apéndice B.

Se anexó a las características extraídas de la imagen, las características morfológicas evaluadas por el embriólogo asignándole un tipo en base a una escala de clasificación de embriones, así como las variables clínicas extraídas de los expedientes clínicos utilizando esta información como entrada del algoritmo. Se cumplió el objetivo de verificar que la información extraída de la imagen del embrión mejora la tarea de clasificación.

9.1. Selección embrionaria en base a características morfológicas evaluadas por el experto y los datos clínicos del tratamiento de fertilización in-vitro

En este estudio retrospectivo se integró la información recabada del banco de imágenes de embriones y de sus respectivas historias clínicas de la Unidad de Reproducción de la Clínica del Pilar, Donostia-San Sebastián.

Para llevar a cabo la comprobación de la hipótesis de que la caracterización morfológica del embrión a partir de la imagen mejora la tarea de clasificación de un conjunto de embriones transferidos en tratamientos de fertilización in-vitro, se diseñaron dos estudios. Los resultados del primero fueron utilizados como control para compararlos con el comportamiento de los clasificadores Bayesianos inducidos a partir de la información extendida.

9.1.1. Diseño experimental

Para la realización del *primer* estudio se cuenta con una base de datos con 81 casos seleccionados de un total de 102 ciclos de tratamientos de fertilización in-vitro de pacientes con un rango de edad de 27-44 años (con una media de 35,64 años) realizados en la Clínica del Pilar desde mayo de 2006 a enero de 2008. Se registraron 36 tratamientos exitosos al lograrse la implantación de alguno de los embriones transferidos y 45 casos en los que no se produjo la implantación.

Se consideraron las mismas variables predictoras listadas en la Tabla 6.4 correspondientes a los datos morfológicos de los embriones provistos por el biólogo, los datos clínicos y del tratamiento de fertilización in-vitro.

Se aplicaron los clasificadores *naive Bayes*, *selective naive Bayes*, *semi naive Bayes*, *TAN* y *kDB*, los cuales fueron inducidos por medio de una validación-cruzada con 10 subconjuntos estratificados. Este estudio fue considerado

como referencia para comparar los resultados obtenidos en el *segundo* estudio realizado en este capítulo. Finalmente se compararon los pares de clasificadores del primer y segundo estudio por medio de la prueba de Mann-Whitney.

9.1.2. Resultados experimentales

La Tabla 9.1 muestra la precisión de los clasificadores Bayesianos validados con diez subconjuntos estratificados de validación cruzada. El modelo TAN obtuvo la mayor precisión con un porcentaje de casos bien clasificados de 74.56 ± 2.26 ; le siguió el modelo *kDB* con una precisión del 73.70 ± 3.24 mientras que los modelos semi naive Bayes y naive Bayes presentaron una precisión del 68.27 ± 2.47 y 67.16 ± 4.40 respectivamente. Finalmente el modelo selective naive Bayes obtuvo la precisión más baja con un 64.44 ± 5.57 de casos bien clasificados.

<i>Clasificador</i>	<i>Precisión±SD</i>
<i>naive Bayes</i>	67.16 ± 4.40
<i>selective naive Bayes</i>	64.44 ± 5.57
<i>semi naive Bayes</i>	68.27 ± 2.47
<i>TAN</i>	74.56 ± 2.26
<i>kDB</i>	73.70 ± 3.24

Tabla 9.1. Resultados de cinco clasificadores Bayesianos aplicados a una base de datos con 81 casos con la información clínica y morfológica de los embriones (determinadas por el experto) en tratamientos de fecundación in-vitro.

La Tabla 9.2 muestra las matrices de confusión de los clasificadores naive Bayes, selective naive Bayes, semi naive Bayes, TAN y *kDB* que obtuvieron el porcentaje de casos bien clasificados más alto de los 10-conjuntos de prueba. Los clasificadores TAN y *kDB* obtuvieron la mayor precisión con un porcentaje de casos bien clasificados del 79.01 % mientras que el modelo naive Bayes obtuvo el porcentaje más bajo con un 71.60 %.

		naive Bayes	semi naive Bayes	selective NB	TAN	<i>kDB</i>
real	predicha	0	1	0	1	0
		32	10	36	13	38
0	1	13	26	9	23	7
Precisión		71.60		72.83		76.54
					79.01	79.01

Tabla 9.2. Matrices de confusión de los clasificadores Bayesianos aplicados a una base de datos de 81 casos con un conjunto de tres embriones, sus características morfológicas y datos clínicos.

A continuación en la Tabla 9.3 se reportan los resultados en términos de los valores de sensibilidad, especificidad, valor de predicción verdadero y

valor de predicción negativo para cada uno de los clasificadores Bayesianos aplicados en este estudio, ya que este tipo de resultados resultan más familiares tanto a médicos como embriólogos, para comprender el comportamiento de los clasificadores Bayesianos.

Clasificador	Precisión	Sensibilidad	Especificidad	Valor predictivo Implantación	Valor predictivo No-implantación
naive Bayes	71.60	72.22	71.11	66.66	76.19
selective naive Bayes	72.83	63.88	80.00	71.87	73.46
semi naive Bayes	76.54	66.66	84.44	77.41	76.00
TAN	79.01	75.00	82.22	77.14	80.43
<i>k</i> DB	79.01	69.44	86.66	80.64	78.00

Tabla 9.3. Resultados control de la clasificación de un conjunto de tres embriones por medio de clasificadores Bayesianos.

En la Tabla 9.3 se puede observar que los clasificadores TAN y *k*DB obtuvieron una precisión del 79.01 %, sin embargo el modelo TAN obtuvo una sensibilidad del 75.00 % y una especificidad del 82.22 % con un valor predictivo de embriones implantados del 77.14 % y un 80.43 % de predecir los casos no-implantados. El comportamiento difiere del clasificador *k*DB ya que este modelo obtuvo una sensibilidad del 69.44 % y una especificidad del 86.66 % siendo sus valores predictivos del 80.64 % y 78.00 % para los casos positivos y negativos respectivamente.

9.2. Mejora de la clasificación en base a la información extraída de la imagen del embrión, características morfológicas evaluadas por el experto y datos clínicos

9.2.1. Base de datos

La *segunda* base de datos aplicada en este capítulo se construyó incluyendo 81 casos de la base de datos aplicada en el primer estudio, de los cuales 36 corresponden a los tratamientos de FIV con éxito (implantación) y los restantes 45 casos donde no se logró la implantación. Se anexó a cada caso la información morfológica del embrión extraída de la imagen, así como la medida de la variación del grosor de la zona pelúcida.

La Tabla 9.4 muestra las variables relacionadas con la imagen del embrión, las cuales fueron incluidas en la base de datos complementando la información de las variables morfológicas del embrión determinadas por el experto así como las variables clínicas de la paciente. En conjunto conformaron un nuevo vector de características al cual se le asignó el valor de la clase real.

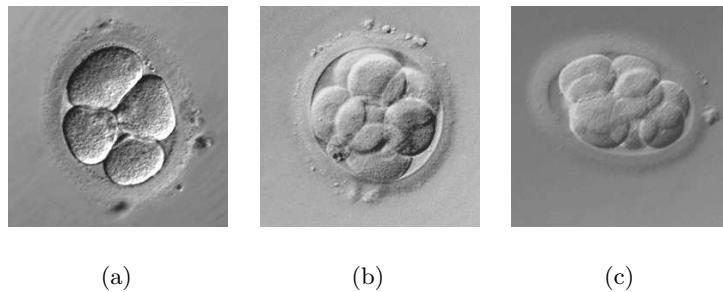


Figura 9.1. Ejemplo de imágenes de embriones adquiridas en la Clínica del Pilar, Donostia-San Sebastián. Las Figuras (a) muestra un embrión en el segundo día después de la fertilización con cuatro blastómeros. La Figura (b) y (c) muestran dos embriones en el tercer día con ocho blastómeros.

Todas las imágenes fueron adquiridas con una cámara Nikon a color adaptada a un microscopio invertido con contraste de Hoffman (HMC). Las observaciones se realizaron con un aumento de 200x y el campo de iluminación se mantuvo constante en todas las observaciones y adquisiciones de las imágenes de los embriones. La Figura 9.1 muestra ejemplos de imágenes de embrones adquiridas en la Unidad de FIV de la Clínica del Pilar.

Las imágenes de los embrones incluidos en esta base de datos corresponden a los mejores embrones de los conjuntos formados por tres embrones. Todos los embrones tienen entre 4 y 5 blastómeros el segundo día después de la fertilización y de 6 a 8 blastómeros el tercer día, todos ellos sin presentar fragmentación.

Variable	Rango de valores
Contraste	Numérico
Energía	Numérico
Entropía	Numérico
Homogeneidad	Numérico
Inercia	Numérico
Inverso	Numérico
Prominencia	Numérico
Forma	Numérico
Variación del grosor de la zona pelúcida	Numérico

Tabla 9.4. Lista de variables extraídas de la imagen del mejor embrión transferido de un conjunto de tres. Imágenes incluidas en la base de datos obtenida de la Clínica del Pilar, San Sebastián.

9.2.2. Diseño experimental

Se aplicaron los modelos clásicos de clasificadores Bayesianos: naive Bayes, selective naive Bayes, semi naive Bayes, TAN y *k*DB. Por medio de la validación-cruzada con diez subconjuntos estratificados se indujo cada uno de los clasificadores Bayesianos. Finalmente se compararon los pares de clasificadores, inducidos a partir de los datos provenientes de la imagen, morfológicos y clínicos, por medio de la prueba Kruskal-Wallis, la cual se introdujo en la Sección 2.3.3.

9.2.3. Resultados experimentales

En este *segundo* estudio se aplicaron los clasificadores Bayesianos naive Bayes, selective naive Bayes, semi naive Bayes, TAN y *k*DB. La Tabla 9.5 muestra los resultados en términos del porcentaje de casos bien clasificados. Siendo los modelos selective naive Bayes y semi naive Bayes los que obtuvieron la mayor precisión con un 80.00 ± 0.97 y 80.00 ± 2.71 respectivamente, muy de cerca, el modelo *k*DB obtuvo una precisión del 79.01 ± 2.72 y finalmente los modelos TAN y naive Bayes obtuvieron la menor precisión con un 74.56 ± 2.67 y un 73.08 ± 5.54 porcentaje de casos bien clasificados respectivamente.

<i>Clasificador</i>	<i>Precisión</i> $\pm SD$
naive Bayes	73.08 ± 5.54
selective naive Bayes	80.00 ± 0.97
semi naive Bayes	80.00 ± 2.71
TAN	74.56 ± 2.67
<i>k</i> DB	79.01 ± 2.72

Tabla 9.5. Resultados de cinco clasificadores Bayesianos aplicados a una base de datos con variables extraídas de la imagen del mejor embrión de un conjunto de tres conjuntamente con las variables clínicas y morfológicas de los embriones determinadas por el experto.

	naive Bayes	semi naive Bayes	selective NB	TAN	<i>k</i> DB
real	0	1	0	0	1
predicha	0	34	6	35	5
0	34	6	35	5	37
1	11	30	10	31	9
Precisión	71.60		79.01	81.14	79.01
					82.71

Tabla 9.6. Matrices de confusión de los clasificadores Bayesianos aplicados a un conjunto de tres embriones con información adicional extraída de la imagen, características morfológicas y datos clínicos.

En términos del valor predictivo de implantación, valor predictivo de no-implantación así como de la sensibilidad y especificidad, se muestran los resultados obtenidos por los clasificadores Bayesianos en la Tabla 9.7.

Clasificador	Precisión	Sensibilidad	Especificidad	Valor predictivo Implantación	Valor predictivo No-implantación
naive Bayes	79.01	83.33	75.55	73.17	85
selective naive Bayes	81.48	86.11	77.77	75.60	87.50
semi naive Bayes	82.71	86.11	80.00	77.50	87.80
TAN	79.01	75.00	82.22	77.14	80.43
kDB	82.71	80.55	84.44	80.55	84.44

Tabla 9.7. Sensibilidad, especificidad, valor predictivo positivo y negativo de la clasificación de un conjunto de tres embriones por medio de clasificadores Bayesianos con información adicional de extraída de la imagen, características morfológicas y datos clínicos.

Nótese que el clasificador semi naive Bayes es el mejor clasificador con una sensibilidad del 86.11% y una especificidad del 77.50% con un valor predictivo de los casos de embriones implantados del 80.00% y con un 87.80% para predecir los no implantados, con un total de casos bien clasificados del 82.71%.

La Tabla 9.8 muestra las variables predictoras seleccionadas en el proceso de la construcción de los clasificadores por medio del proceso de selección de un subconjunto de variables - FSS -. La *homogeneidad* de la imagen mide la relación de variación de la intensidad de los niveles de gris, la cual presenta un valor alto cuando la intensidad de la matriz de muestreo (ventana) de la imagen es homogénea. La *entropía* de la imagen integra la diversidad de los diferentes niveles de grises que se presentan en la imagen. Tiende a tener valores cercanos a 0 o 1. En el caso que la entropía sea 0, significa que no hay variación en el contraste de la imagen presentándose valores de entropía altos cuando los valores de la matriz de co-ocurrencia, tienden a ser iguales, contrario a los valores altos de la homogeneidad que se concentran en la diagonal de la matriz de co-ocurrencia. La medida de *inercia* en la imagen está definida por los *momentos centrales de segundo orden* los cuales son invariantes a la traslación, escalamiento y rotación. Estas propiedades permiten caracterizar las estructuras celulares del embrión por su forma. Esta variable tiene relación con los momentos centrales aplicados en el Capítulo 7, para la selección del mejor embrión a partir de la imagen.

Se encontró que la medida de la variación del grosor de la zona pelúcida está directamente relacionada con la predicción del éxito del tratamiento de fertilización in-vitro como se afirma en [Cohen y col., 1989, Palmstierna y col., 1998, Gabrielsen y col., 2000, Gabrielsen y col., 2001]. De esta forma se corrabora la suposición del embriólogo en que el embrión Tipo I o II en la escala de [Veek, 1999] tiene mayor probabilidad de implantarse.

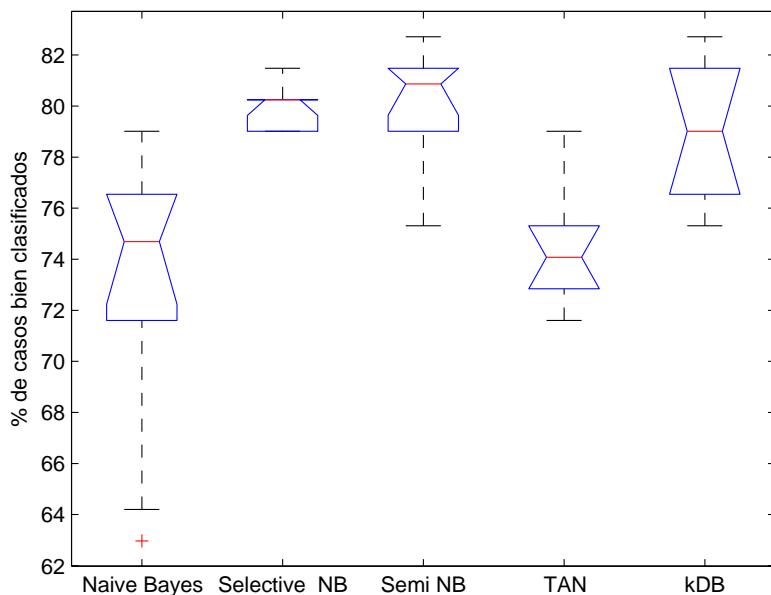


Figura 9.2. Gráfica del comportamiento de los clasificadores Bayesianos (naive Bayes, selective naive Bayes, semi naive Bayes, TAN y *kDB*) aplicando la prueba Kruskal-Wallis. El clasificador naive Bayes presentó un comportamiento significativamente diferente al resto del grupo de clasificadores Bayesianos, presentando la menor precisión.

Con el propósito de comparar el comportamiento de los cinco clasificadores Bayesianos obtenidos en este estudio, en el que se consideraron las características de la imagen, morfológicos y clínicos, se aplicó la prueba no paramétrica de Kruskal-Wallis. En la Figura 9.2 se muestra de forma gráfica el comportamiento de los clasificadores naive Bayes, selective naive Bayes, semi naive Bayes, TAN y *kDB*. En la misma se observa que el clasificador naive Bayes tiene una diferencia significativa con respecto a los otros clasificadores Bayesianos con un valor de $p = 3,1326 \times 10^{-6}$.

Se aplicó la prueba no paramétrica de Mann-Whitney para comparar el comportamiento de cada modelo obtenido a partir de los datos clínicos y morfológicos del embrión cuyos resultados se resumen en la Tabla 9.1 con el correspondiente modelo obtenido con los datos provenientes de la imagen, morfológicos y clínicos cuyos resultados se presentaron en la Tabla 9.6. Se encontró que existe una diferencia significativa entre los pares de clasificadores provenientes de distintas bases de datos, por lo que se comprueba que la

Lista de variables	selective NB	semi NB
Embrión1 inercia	✓	✓
Embrión1 entropía		✓
Embrión1 homogeneidad		✓
Embrión1 variación de VGZP	✓	✓
Embrión1 tamaño de blastómeros		✓
Embrión3 tamaño de blastómeros	✓	✓
Embrión3 número de células	✓	
Embrión3 multinucleado		✓
Calidad de la transferencia		✓

Tabla 9.8. Lista de variables predictoras obtenidas por medio del proceso de la selección de un subconjunto de variables (FSS) para la inducción de diferentes clasificadores Bayesianos en base a información extraída de la imagen, características morfológicas del embrión y datos clínicos.

información enriquecida con las variables extraídas de la imagen mejora la clasificación.

Las variables seleccionadas del tercer embrión *tamaño de blastómeros*, *número de células* y si es multinucleado se corresponden con las características evaluadas para determinar su Tipo. En base a lo expuesto anteriormente se puede concluir que la caracterización morfológica del embrión a partir de la imagen conjuntamente con la medida de la variación del grosor de la zona pelúcida mejoran la clasificación del conjunto de embriones transferidos en tratamientos de fertilización in-vitro.

Parte V

Conclusiones y trabajo futuro

Conclusiones y trabajo futuro

Esta tesis constituye el primer acercamiento desde la clasificación supervisada con clasificadores Bayesianos al problema de la selección embrionaria, concretamente en la predicción del éxito (implantación) del tratamiento de reproducción humana asistida en base a los datos clínicos y del propio tratamiento, así como la información de la caracterización morfológica del embrión realizada por el biólogo y las medidas extraídas de la imagen.

10.1. Conclusiones

Como se mencionó en los objetivos considerados para la realización de esta tesis, se han desarrollado, adaptado e inducido diversos paradigmas de clasificación supervisada en especial los clasificadores Bayesianos, los cuales fueron aplicados al problema biomédico de la infertilidad, en concreto a la selección embrionaria en tratamientos de reproducción asistida.

Se propuso un nuevo multi-clasificador *Gaussian-stacking* basado en el esquema de pila, en el cual se induce el meta-clasificador con redes condicionales Gaussianas, como un nuevo método para combinar diferentes paradigmas de clasificación supervisada.

En base al multi-clasificador Gaussian-stacking se superó la necesidad de discretizar las distribuciones de probabilidad de la clase, resultado de la clasificación previa de los clasificadores-base, proceso hasta ahora realizado en las diferentes adaptaciones realizadas al esquema *staked* para inducción del meta-clasificador.

Se comprobó que la clasificación en base a la combinación de la predicción de los diversos paradigmas de clasificación es más precisa que la realizada por un único clasificador [Dietterich, 1998].

En esta tesis se adaptaron diferentes clasificadores Bayesianos sensibles al coste del error con los cuales se revolvió el problema de clases no balanceadas.

El planteamiento del problema de la selección embrionaria como un problema de clasificación supervisada propuesto en esta tesis se aplicó en los

estudios presentados en los Capítulos 6, 8 y 9 obteniéndose resultados favorables avalados por [Morales y col., 2008c, Morales y col., 2008b, Morales y col., 2009, Morales y col., 2007]. Con lo que se concluye que la aproximación al problema de la selección de un conjunto de embriones para su transferencia en tratamientos de reproducción asistida es correcto.

Se corroboró la interdependencia entre los momentos centrales en base a los clasificadores TAN y *kDB* debido a que la estructura final de estos clasificadores es capaz de representar las dependencias condicionales entre variables, como es el caso de los momentos centrales. Y en base al comportamiento obtenido por los clasificadores Bayesianos sensibles al coste del error se puede concluir que los momentos centrales son una técnica adecuada para la caracterización de la morfología del embrión y su aplicabilidad para enriquecer la información del problema de la selección embrionaria.

Analizando los resultados obtenidos por el multi-clasificador *Gaussian-stacking* en la selección embrionaria de un conjunto de tres embriones en tratamientos de fertilización in-vitro, podemos concluir que el multi-clasificador *Gaussian-stacking* puede ser usado satisfactoriamente como un método válido para problemas médicos complejos.

Se demostró que los clasificadores Bayesianos sensibles al coste son modelos apropiados para la tarea de la *selección del mejor embrión* para su transferencia en tratamientos de reproducción asistida.

En esta tesis se aportó una lista de variables relacionadas con la selección embrionaria resultado del proceso de selección de un subconjunto de variables predictoras para la inducción de los clasificadores Bayesianos y se verificó que las variables predictoras seleccionadas para la inducción de los clasificadores Bayesianos son acorde con el criterio de evaluación y selección embrionaria hecha por el biólogo con lo cual podemos concluir que la aproximación al problema de la selección embrionaria por medio de los clasificadores Bayesianos es correcta.

A partir de la construcción del clasificador naive Bayes de forma manual considerando la lista de variables seleccionadas en los procesos de FSS aplicados para la inducción de algunos clasificadores Bayesianos se concluye que no es confiable para su aplicación en el problema de la selección embrionaria.

Se sugiere una lista de variables acordes con la literatura [Ziebe y col., 1997, Alikanin y col., 1999, Gabrielsen y col., 2001] para considerarse en los modelos de clasificación supervisada, las cuales corroboran la importancia de la morfología del embrión para su selección entre ellas la medida propuesta en esta tesis para mejorar la tarea de clasificación, la variación del grosor de la zona pelúcida.

Aportaciones de la tesis en el problema de la selección embrionaria en tratamientos de reproducción asistida.

En la práctica médica, en países donde la legislación permite la transferencia de un máximo de 3 embriones por ciclo del tratamiento de reproducción asistida, como en la Clínica del Pilar, el modelo de clasificación supervisada propuesto en esta tesis para el estudio de la problemática de la selección embrionaria se puede aplicar también a un conjunto de dos o más embriones transferidos.

Por medio de los clasificadores Bayesianos se puede disminuir la incertidumbre inherente al proceso de la selección de un conjunto de embriones para ser transferidos en tratamientos de reproducción asistida.

A partir de la propuesta de enriquecer la información del tratamiento de fertilización in-vitro y la caracterización del embrión realizada por el experto, con algunas medidas extraídas de la imagen del mejor embrión entre ellas la medición del grosor de la membrana de la zona pelúcida se logró un incremento significativo en la precisión de los clasificadores Bayesianos.

Se comprobó que la medida de la variación del grosor de la zona pelúcida es una característica potencialmente predictiva y está directamente relacionada con la predicción del éxito del tratamiento de fertilización in-vitro como se afirma en [Cohen y col., 1989, Palmstierna y col., 1998, Gabrielsen y col., 2000, Gabrielsen y col., 2001]. Por otra parte su aplicación en la clasificación supervisada mejora los resultados de clasificación.

Considerando la caracterización del mejor embrión por medio de la imagen, la información del experto y del tratamiento así como la información de los dos embriones del conjunto transferido, se corrobora la suposición del embriólogo de que el embrión Tipo I o II en la escala de [Veek, 1999] tiene mayor probabilidad de implantarse. Considerando las variables predictoras relacionadas con las características morfológicas del embrión más viable así como las del embrión menos viable (en el conjunto transferido) se relacionan con la facilitación del medio en el útero materno cuando se transfiere más de un embrión, pronunciándose la implantación de alguno de los embriones transferidos [Matorras y col., 2005].

Finalmente teniendo en cuenta los resultados obtenidos en [Morales y col., 2008c, Morales y col., 2008b, Morales y col., 2009, Morales y col., 2007] con la aportación de diferentes adaptaciones de los clasificadores Bayesianos al problema de la selección embrionaria, se puede concluir que el desarrollo de un sistema de apoyo a la decisión del embriólogo en el proceso de la selección embrionaria es completamente viable, así como su posible aplicación en la formación de nuevos embriólogos.

10.2. Trabajo futuro

Considerando los resultados presentados en esta tesis y la mejora en la precisión de los clasificadores Bayesianos a partir de la información proveniente de la imagen del mejor embrión transferido se plantean los siguientes trabajos por desarrollar:

- La adquisición de nuevos datos que incluyan características morfológicas extraídas de las imágenes del *conjunto de embriones transferidos* e información clínica del paciente, ya que incrementando el número de variables predictoras ha mejorado el proceso de selección de variables FSS y a su vez la precisión y el comportamiento de los clasificadores Bayesianos.
- Aplicar el multi-clasificador *Gaussian-stacking* a la información extraída del conjunto de imágenes de los embriones transferidos.
- Explorar la aplicación del aprendizaje en base a una matriz de costos de los clasificadores Bayesianos en la predicción de un embarazo múltiple.
- Adaptar el multi-clasificador con aprendizaje sensible al coste del error y su aplicación en problemas del dominio real.
- Otro trabajo podría ser el desarrollo de una herramienta para homologar el criterio del experto a través del apoyo en la formación y entrenamiento del embriólogo en la tarea de la selección embrionaria.
- Finalmente se podría plantear una posible solución al problema de la selección de un único embrión para ser trasferido en tratamientos de fertilización in-vitro. Este problema representa el gran reto de las clínicas de reproducción asistida en el mundo, disminuyendo el riesgo que representa un embarazo múltiple para la paciente y la implantación de un embrión sano.

Parte VI

Apéndices

Introducción

Como propuesta para mejorar la tarea de clasificación supervisada por medio de los clasificadores Bayesianos fue necesaria la aplicación de algunas técnicas del análisis de la imagen con el propósito de extraer información relacionada con la morfología del embrión a partir de la imagen, a través de diferentes medidas relacionadas con la textura y la forma de las estructuras celulares del embrión.

La técnica de los contornos activos (Sección 5.6) se aplicó para segmentar la membrana de la zona pelúcida y mediante la extracción de los contornos de sus paredes, se midió la *variación del grosor* de la zona pelúcida y se aplicó en el Capítulo 9. El Apéndice A describe una serie de procesos necesarios para la segmentación automática por medio de los contornos activos.

Se extrajo la información relacionada con la textura de la imagen del embrión y se midió la homogeneidad, entropía, energía, contraste, momento de diferencia inverso, prominencia y forma, las cuales ha sido definidas en el Apéndice B. Estas medidas se incorporaron en la experimentación del Capítulo 9.

En esta tesis, en el Capítulo 7 se aplicaron los momentos centrales como descriptores de forma, para extraer las características morfológicas de los embriones. El Apéndice C describe los momentos estadísticos y se definen los momentos centrales.

A

Técnicas del pre-proceso de la imagen

A.1. Introducción

Con el propósito de describir los procesos aplicados a la imagen para mejorar su visualización y facilitar su análisis en la Figura A.1 se pueden observar de forma esquemática los diferentes métodos aplicados así como las técnicas utilizadas para la extracción de características de la imagen del embrión. En general se pueden considerar las siguientes etapas anteriores a la segmentación automática de los contornos activos (*snakes*).

- Adquisición de la imagen. En primer lugar se adquieren las imágenes por el experto, en nuestro caso, la imagen a color del embrión se transforma en una imagen en escala de grises para posteriormente pasar por una etapa de preprocessamiento.
- Preproceso de la imagen. Consiste en aplicar una serie de filtros de restauración, para obtener una buena visualización de la imagen que se está analizando.
- Mapa de bordes. Se aplican filtros de gradiente para extraer las características de forma de los objetos contenidos en la imagen de forma automática siendo esta la etapa anterior al análisis de la imagen (segmentación).
- Inicialización de los contornos activos.
- Segmentación de la zona pelúcida por medio de *snakes*.

Primeramente se adquieren las imágenes por el experto, en nuestro caso, la imagen a color del embrión se transforma en una imagen en escala de grises para posteriormente pasar por una etapa de preprocessamiento que en general suele consistir en una serie de filtros de restauración, para obtener una buena visualización de la imagen que se está analizando. Posteriormente se aplican filtros de gradiente para finalmente iniciar la segmentación con la inicialización del *snake*. A continuación se describen las etapas mencionadas anteriormente.

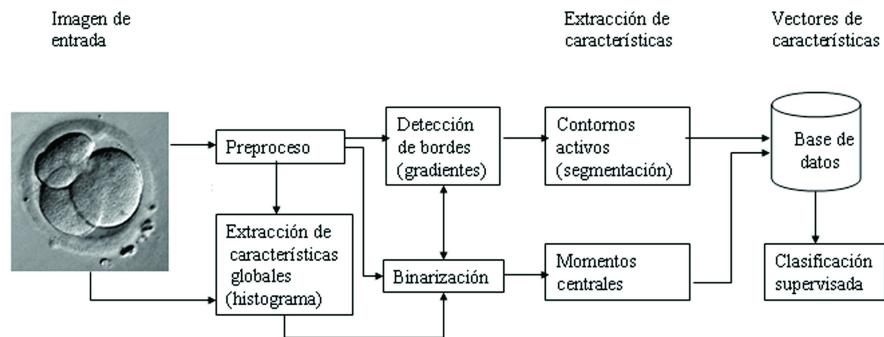


Figura A.1. Esquema del proceso de extracción automática de las características morfológicas del embrión.

A.2. Pre-proceso de la imagen del embrión

Las imágenes de los embriones generalmente son adquiridas en color ya que la observación in-vivo a través del microscopio invertido permite la observación de sus características naturales, entre ellas el color de las células.

Como pre-proceso de la imagen del embrión se aplicaron diversas técnicas para facilitar su análisis posterior. A continuación se describirá la secuencia de cambios realizados a la imagen del embrión y se mencionara brevemente la técnica aplicada.

En este trabajo se transformaron las imágenes de los embriones en imágenes en escala de grises para facilitar su análisis. Como se mencionará en la Sección B.2, una de las características de la imagen del embrión es el contorno difuso de la membrana de la zona pelúcida. En la Figura A.2 se puede observar como varía la intensidad de los niveles de gris a través de la zona pelúcida, el contorno así como en el fondo de la imagen.

Como pre-proceso, se combinó un filtro de convolución Gaussiano paso alto, con el proceso de histéresis propuesto por [Klaus-Ruediger, 1994]. Con estos procesos se aumentaron los niveles de contraste para incrementar la claridad en la visualización de la zona pelúcida en la imagen del embrión.

El filtro Gaussiano se aplicó para el suavizado de la imagen para posteriormente mejorarla por medio de la técnica de histéresis diferencial, la cual realiza comparaciones entre la intensidad de un píxel en la imagen original y su homólogo en la imagen suavizada, detectando así las variaciones entre los niveles de gris. Este método remueve las pequeñas variaciones en los niveles de gris de la imagen mejorada. Estos procesos se encuentran ampliamente descritos en [Gonzales y Woods, 2002].

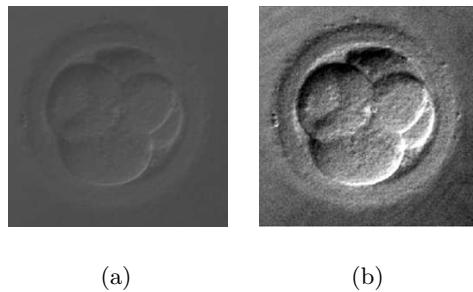


Figura A.2. (a) Imagen original en escala de grises (b) Imagen transformada del embrión para mejorar su análisis y visualización tras el proceso.

A.3. Generación del mapa de bordes de la imagen

Una vez concluida la etapa de mejora de la imagen se generó un mapa de bordes. A continuación se describe el proceso de su construcción.

El primer paso fue aplicar el algoritmo detector de bordes propuesto por [Canny, 1986]. Este algoritmo detecta los máximos locales de los cambios de intensidad en el nivel de gris de un píxel a otro y se caracteriza por utilizar dos valores de corte o umbrales. En este trabajo se aplicó el umbral más bajo para detectar los puntos de los bordes menos intensos. Cabe resaltar que este método genera un mapa con varios puntos pertenecientes a los bordes de los objetos incluidos en la imagen. Sin embargo no hay continuidad entre los puntos del contorno de los objetos por lo que resulta necesario un algoritmo ya sea para unir los puntos o un método como los contornos activos que extraen el contorno completo del objeto a segmentar.

A continuación en un segundo paso se calculó el área de los píxeles conectados entre si considerándose ocho-píxeles vecinos y se removieron los artefactos que contenían un número de píxeles igual o menor del área calculada. De esta forma se removieron los artefactos de la imagen como los restos de células de granulosa.

Posteriormente, en un tercer paso se aplicó un filtro derivativo Gaussiano al mapa de bordes, para calcular el mapa de gradientes.

Finalmente, se aplicó el método de contornos activos para obtener los contornos de las paredes de la membrana de la zona pelúcida. Es necesario inicializar el *snake* lo más cerca posible del contorno a segmentar.

B

Técnicas de extracción automática de características de textura de la imagen del embrión

B.1. Introducción

En este apéndice se introducen brevemente algunas notaciones sobre funciones para el tratamiento y análisis de la imagen en general, así como las medidas y transformaciones basadas en el histograma de la imagen, necesarias para la introducción de las técnicas aplicadas en esta tesis para la extracción de *información de textura* de las imágenes de los embriones.

En el Capítulo 9 se aplican las medidas relacionadas con la textura de la imagen del embrión además de la medición del grosor de la zona pelúcida así como la caracterización morfológica del embrión realizada por el experto, la información clínica y del tratamiento, aplicadas en la selección embrionaria de un conjunto de tres embriones por medio de clasificadores Bayesianos. La información en conjunto ha sido aplicada a los clasificadores Bayesianos clásicos y por medio del proceso de selección de un subconjunto de variables se ha mejorado el proceso de la clasificación supervisada de la selección embrionaria en tratamientos de reproducción asistida.

B.2. Extracción de medidas de la imagen

La imagen se representa mediante una matriz $N \times M$ y su elemento fundamental es el píxel. La información básica que obtiene de cada objeto contenido en la imagen son las coordenadas de todos los píxeles contenidos en el objeto de estudio. A cada píxel le corresponde un valor del nivel de gris presente en la imagen.

Sea una imagen definida por la función $f(x, y)$, donde (x, y) son las coordenadas de un punto y $f(x, y)$, el valor de ese punto; en el caso de las imágenes binarias, este valor será 0 si el punto es distinto de negro y 1 si es negro; en el caso de imágenes en escala de grises este valor dependerá del nivel de gris en ese punto.

El histograma de una imagen $f(x, y)$ cuantifica del número de pixels que hay con un determinado nivel de gris. El histograma de una imagen digital con 256 niveles de gris $l = 0, \dots, 255$, presenta valores en el intervalo $[0, 255]$. El histograma es una función discreta definida como $h(l_i) = n_i$, donde l_i es el i -ésimo nivel de gris y n_i es el número de píxeles en la imagen con el nivel de gris l_i . Una práctica común, es la normalización del histograma dividiendo cada uno de sus valores entre el número de píxeles en la imagen, denotado por n . Por lo tanto el histograma normalizado está definido como:

$$p(l_i) = \frac{n_i}{n}$$

con $i = 0, 1, \dots, l$, donde $p(l_i)$ es un estimador de la probabilidad de la ocurrencia del nivel de gris l_i y donde la suma de todas las componentes del histograma normalizado es igual a uno. Los histogramas son la base de varias transformaciones y cálculos para numerosos métodos de procesamiento de la imagen en el dominio espacial.

Por medio de la gráfica de un histograma se visualiza la distribución de probabilidad de la ocurrencia del nivel de gris l_i . El eje x corresponde a cada valor del nivel de gris y el eje y se corresponde a los valores $h(l_i) = n_i$ (en el caso de valores normalizados a los valores de $p(l_i)$).

Existen varias transformaciones que se pueden realizar en base al histograma, como por ejemplo la variación del contraste para mejorar la visualización de la imagen, pero la utilidad general mayor es la binarización de imágenes, para extraer el valor umbral, de la escala de los niveles de grises presentes en la imagen ya sea como porcentaje o como un mínimo local. Para el estudio de la función de histograma se suele transformar con una función de suavizado en la que para cada nivel de gris, se considera la media de su valor en el histograma y el de los niveles de gris contiguos. Un ejemplo de ello son los filtros Gaussianos.

En base al análisis estadístico de segundo orden es posible extraer información de la textura de los objetos contenidos en una imagen. La textura es una propiedad asociada a las superficies tales como suavizado, rugosidad, granularidad y regularidad. El aspecto de un objeto determina su textura, la cual puede modelarse matemáticamente por medio del método de co-ocurrencia [Haralick y col., 1973] como un conjunto de elementos (patrones) que presentan una determinada relación espacial que se mantiene estable a lo largo de una superficie. La cuantificación de la textura está determinada por la búsqueda de esos elementos estructurales y de la estimación de sus relaciones espaciales, las cuales son almacenadas y ordenadas en una matriz de co-ocurrencia \mathbf{P} . Las matrices de co-ocurrencia son, esencialmente, histogramas de segundo orden en las que se registran el número de elementos $P(i, j)$ con sus respectivos valores de niveles de gris en posiciones espaciales concretas. Para G niveles de gris en la imagen, la matriz $\mathbf{P} = G \times G$.

$$P(i, j, D, \theta) = P(I(l, m) = i, I(l + D \cos(\theta), m + D \sin(\theta)) = j) \quad (\text{B.1})$$

donde $I(l, m)$ es el píxel (l, m) en la imagen, D es la distancia entre píxeles y θ es el ángulo. En la literatura se encuentran trabajos que sugieren que los pares de vecinos a una distancia D en la orientación $\theta=\{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$ conforman el conjunto de datos mínimo para describir las medidas estadísticas de segundo orden de la textura.

Una vez calculada la matriz de co-ocurrencia se pueden extraer varias medidas como *descriptores de textura* por ejemplo la *energía*, *entropía*, *máxima probabilidad*, *contraste*, *momento de diferencia inverso* (MDI), *correlación*, *homogeneidad*, *contraste*, etc. Con el propósito de definir estas funciones sea $P(i, j)$ el elemento (i, j) de la matriz de co-ocurrencia normalizada.

$$\begin{aligned} P_x(i) &= \sum_j P(i, j) \\ P_y(j) &= \sum_i P(i, j) \\ \mu_x &= \sum_i i \sum_j P(i, j) = \sum_i i P_x(i) = E\{i\} \\ \mu_y &= \sum_j j \sum_i P(i, j) = \sum_j j P_y(j) = E\{j\} \end{aligned}$$

A continuación se introducen las medidas de textura calculadas para la caracterización morfológica del embrión a partir de la imagen por medio de la matriz de co-ocurrencia.

$$\begin{aligned} E &= \sum_{i,j} P(i, j)^2 \\ En &= - \sum_{i,j} P(i, j) \log P(i, j) \\ MDI &= \sum_{i,j} \frac{1}{1 - (i - j)^2} P(i, j) \\ Forma &= \sum_{i,j} (i + j - \mu_x - \mu_y)^3 P(i, j) \\ Inercia &= \sum_{i,j} (i - j)^2 P(i, j) \\ Prominencia &= \sum_{i,j} (i + j - \mu_x - \mu_y)^4 P(i, j) \end{aligned}$$

donde E es la medida de energía y En la de entropía.

C

Caracterización morfológica del embrión por medio de momentos centrales

C.1. Introducción

En la literatura se encuentran diversas técnicas de análisis de la imagen, las cuales son aplicadas para la descripción y caracterización de los objetos presentes en la misma. Resulta importante, por tanto, disponer de descripciones capaces de caracterizar, con precisión, los objetos presentes en la imagen ya sea por medio de textura, forma, regiones, contornos, etc. No es objetivo de esta tesis profundizar en las técnicas de análisis de la imagen, para ello se puede consultar el libro [Gonzales y Woods, 2002].

En esta tesis en especial, se aplicaron los momentos como descriptores de la forma de los objetos contenidos en una imagen. En el Capítulo 7 se aplicaron los momentos centrales para la caracterización morfológica del embrión a partir de la imagen.

Los momentos de un objeto, son una serie de parámetros estadísticos obtenidos a partir de las coordenadas de los puntos del objeto. Los momentos centrales normalizados en particular se caracterizan matemáticamente por sus propiedades de invarianza con respecto a la traslación, escalamiento y rotación. A continuación se describe el marco teórico de los momentos centrales y su definición.

C.2. Momentos centrales

La forma de los lados de un contorno se puede describir cuantitativamente utilizando momentos, a partir de una función unidimensional $g(x)$ de una variable x arbitraria. Tratando la amplitud de g como una variable aleatoria v formando un histograma $p(v_i)$ con $i = 1, 2, \dots, k$ siendo k el número de incrementos de amplitud. Entonces, el momento n -ésimo de v respecto de su media está definido de la siguiente manera:

$$\mu_n(v) = \sum_{i=0}^{L-1} (v_i - m)^n p(v_i)$$

La geometría de una región plana se basa en el tamaño, la posición, la orientación y la forma. Todas estas medidas están relacionadas con la familia de momentos geométricos los cuales se clasifican en momentos simples, momentos centrales y momentos centrales normalizados.

Los momentos centrales, los cuales fueron introducidos por [Hu, 1962], son invariantes a la traslación, rotación y escalamiento de un patrón representado en la imagen. Estas propiedades han permitido su amplia utilización en la investigación Biomédica. Los momentos centrales han sido aplicados para extraer las características de forma a partir de imágenes celulares como en el trabajo presentado por [Albregtsen y col., 2006]. En el dominio de las imágenes médicas encontramos la aplicación de los momentos centrales en [Ruggeri y Pajaro, 2002, Mangin y col., 2004].

Momentos simples

Principalmente los momentos se emplean para describir la forma de los objetos en la imagen incluyendo todos los píxeles de la misma y no solo los píxeles de los bordes de los objetos. Si tenemos un objeto en una región \mathbb{R} que viene dado por los puntos en los que $f(x, y) > 0$, definimos el momento de orden p, q con $p, q \in \{1, \dots, n\}$ como:

$$m_{p,q} = \int_{x=1}^N \int_{y=1}^M x^p y^q f(x, y) dx dy$$

En el caso discreto, el momento de orden (p, q) se calcula como:

$$m_{p,q} = \sum_{x=1}^N \sum_{y=1}^M x^p y^q f(x, y)$$

Momentos de orden cero

El momento simple de orden cero $\mu_{0,0}$ representa el área de la figura en imágenes binarias y la superficie en imágenes en escala de grises. Es la suma de los valores de todos los píxeles. Para ello nos basamos en la fórmula de los momentos simples.

$$\mu_{0,0} = \sum_x \sum_y f(x, y)$$

Momentos de primer orden

Los momentos simples de primer orden, $\mu_{1,0}$, $\mu_{0,1}$ se utilizan principalmente para calcular el centro de masas de un objeto.

$$\begin{aligned}\mu_{1,0} &= \sum_x \sum_y x f(x, y) \\ \mu_{0,1} &= \sum_x \sum_y y f(x, y)\end{aligned}$$

Momentos centrales

Los momentos centrales se utilizan para describir un objeto en la imagen independientemente de su situación con respecto a los ejes de coordenadas. A partir de los momentos de orden cero y primer orden de la imagen se calcula el centro de masa ó centroide de un objeto.

El centro de masa de un objeto viene determinado por las coordenadas (x, y) , de forma que el área del objeto que queda a la derecha e izquierda del punto x es la misma, al igual que el área que queda por encima y por debajo del punto y . A partir de este valor, se pueden definir unos nuevos momentos, relativos a este punto, denominados momentos centrales. A partir de las coordenadas del centro de gravedad de un objeto:

$$\begin{aligned}\bar{x} &= \frac{\mu_{1,0}}{\mu_{0,0}} \\ \bar{y} &= \frac{\mu_{0,1}}{\mu_{0,0}}\end{aligned}$$

los momentos centrales se definen en el dominio continuo como:

$$\mu_{p,q} = \int_x^N \int_{y=1}^M (x - \bar{x})^p (y - \bar{y})^q f(x, y) dx dy \quad (\text{C.1})$$

y en el discreto:

$$\mu_{p,q} = \sum_{x=1}^N \sum_{y=1}^M (x - \bar{x})^p (y - \bar{y})^q f(x, y) \quad (\text{C.2})$$

Los momentos centrales de primer orden, $\mu_{1,0}$ y $\mu_{0,1}$, presentan un valor de cero al igual que los momentos centrales normalizados de primer orden, $\eta_{1,0}$ y $\eta_{0,1}$.

A partir de los momentos centrales de primer orden del objeto, se pueden definir unos nuevos ejes de coordenadas, denominados ejes principales del objeto definidos como las coordenadas x', y' a un ángulo θ , de los ejes x, y con

$$\tan 2\theta = \frac{2\mu_{1,1}}{\mu_{2,0} - \mu_{0,2}}$$

Respecto a los ejes principales del objeto x', y' , los *momentos centrales normalizados* son invariantes a la traslación, escalado y rotación. Es decir, se pueden clasificar los distintos objetos simplemente por su forma. A continuación se describen los momentos centrales normalizados de segundo y tercer orden.

Momentos centrales normalizados (momentos invariantes)

A partir de los momentos centrales normalizados de segundo y tercer orden comienza el análisis de las imágenes mediante el reconocimiento de formas. Es posible extraer siete parámetros denominados *momentos invariantes* [Hu, 1962] como una forma de describir matemáticamente la imagen. Al normalizar los momentos centrales se obtiene una descripción independiente del tamaño mediante los *momentos centrales normalizados* definidos como:

$$\eta_{p,q} = \frac{\mu_{p,q}}{\mu_{0,0}^{\alpha}}$$

con $\alpha = \frac{p+q}{2} + 1$.

A continuación se describirán brevemente las propiedades de escalamiento y rotación, ya que los momentos centrales invariantes a la traslación los hemos definido anteriormente.

En el escalamiento $x' = x, y' = y$ de la imagen $f(x, y)$, los momentos centrales $\mu_{p,q}$ de $f(x, y)$ cambian a:

$$\mu'_{p,q} = \frac{\mu_{p,q}}{\alpha^{p+q+z}}$$

Los momentos centrales normalizados también son invariantes a los cambios de escala y en base a estos se calculan los momentos centrales normalizados de mayor orden. Están definidos como:

$$\eta_{p,q} = \frac{\mu'_{p,q}}{\mu'_{0,0}^{\gamma}}$$

con $\gamma = (p + q + 2)/2$.

Rotación y reflexión.

Los momentos centrales forman las componentes de la matriz de rotación MR definida a continuación:

$$\begin{vmatrix} \mu_{0,2} & -\mu_{1,1} \\ -\mu_{1,1} & \mu_{2,0} \end{vmatrix}$$

Tabla C.1. Matriz de rotación.

Partiendo de las componentes de la matriz MR se obtienen tanto el ángulo de rotación como la excentricidad.

Ángulo de rotación

El ángulo de rotación se define como el ángulo que existe entre el eje de abscisas y el eje alrededor del cuál el objeto puede rotar con mínima inercia.

Excentricidad

La excentricidad es el grado de desviación de una sección cónica con respecto a una circunferencia. A medida que aumenta la excentricidad disminuye la curvatura. En nuestro caso podemos definir la excentricidad en términos de una elipse, que rodease al objeto para que quedase fuera un área del objeto igual al área que quedase dentro de una elipse. La excentricidad queda determinada por la distancia entre los dos focos, la cual tiene valores en un rango $[0,1]$.

Momentos centrales de segundo orden

La densidad del objeto, en términos de la resolución, se multiplica por las distancias al cuadrado desde el centro de masas o centroide.

A continuación se describe la interpretación del valor que presentan los momentos centrales normalizados de segundo orden.

$\mu_{2,0}$: Su valor aumentará cuanto mayor sea la componente horizontal de una figura.

$\mu_{0,2}$: Su valor aumentará cuando mayor sea la componente vertical de una figura.

$\mu_{1,1}$: Su valor puede ser positivo o negativo dependiendo de donde se encuentre la componente vertical; si la componente vertical se encuentra en el segundo y cuarto cuadrante, entonces será negativo, si por el contrario está en el primer y tercer cuadrante entonces será positivo. En el caso de

una imagen simétrica respecto a los ejes, el valor de $\mu_{1,1}$ será igual a cero.

Momentos centrales de tercer orden

Los *momentos centrales de tercer orden* al igual que los momentos centrales normalizados de orden n , conservan las características de invarianza con respecto a la traslación, escalamiento y rotación. En el caso de los momentos centrales de tercer orden quedan definidos de la siguiente manera:

$$h_1 = \eta_{2,0} + \eta_{0,2}$$

$$h_2 = (\eta_{2,0} - \eta_{0,2})^2 + 4\eta_{1,1}^2$$

$$h_3 = (\eta_{3,0} - 3\eta_{1,2})^2 + (3\eta_{2,1} - \eta_{0,3})^2$$

$$h_4 = (\eta_{3,0} + \eta_{1,2})^2 + (\eta_{0,3} + \eta_{2,1})^2$$

$$h_5 = (\eta_{3,0} - 3\eta_{1,2})(\eta_{3,0} + \eta_{1,2})[(\eta_{3,0} + \eta_{1,2})^2 - 3(\eta_{0,3} + \eta_{2,1})^2] + \\ (3\eta_{2,1} - \eta_{0,3})(\eta_{3,0} + \eta_{1,2})[3(\eta_{3,0} + \eta_{1,2})^2 - (\eta_{0,3} + \eta_{2,1})^2]$$

$$h_6 = (\eta_{2,0} - \eta_{0,2})[(\eta_{3,0} + \eta_{1,2})^2 - (\eta_{0,3} + \eta_{2,1})^2] + 4\eta_{1,1}(\eta_{3,0} + \eta_{1,2})(\eta_{0,3} + \eta_{2,1})$$

$$h_7 = (3\eta_{2,1} - \eta_{0,3})(\eta_{3,0} + \eta_{1,2})[(\eta_{3,0} + \eta_{1,2})^2 - 3(\eta_{0,3} + \eta_{2,1})^2] - \\ (\eta_{3,0} - 3\eta_{1,2})(\eta_{0,3} + \eta_{2,1})[3(\eta_{3,0} + \eta_{1,2})^2 - (\eta_{0,3} + \eta_{2,1})^2]$$

En el Capítulo 7 se aplicaron las características morfológicas de los embriones extraídas de las imágenes a partir de seis momentos centrales invariantes.

Referencias

- [Aha y Kibler, 1991] Aha, D. y Kibler, D. (1991). Instance-based learning algorithms. *Machine Learning*, 6:37–66.
- [Albregtsen y col., 2006] Albregtsen, F., Schulerud, H., y Yang, L. (2006). Texture classification of mouse liver cell nuclei using invariant moments of consistent regions. En *Computer Analysis of Images and Patterns*, pp 496–502.
- [Alikanin y col., 1999] Alikanin, M., Cohen, J., Tomkin, G., Garrisi, J., Mack, C., y Scott, R.T. (1999). Human embryo fragmentation in vitro and its implications for pregnancy and implantation. *Fertility and Sterility*, 71(5):836–842.
- [Andersen y col., 2008] Andersen, A.N., Goossens, V., Ferraretti, A.P., Bhattacharya, S., Felberbaum, R., de Mouzon, J., Nygren, K.G., y IVF-Monitoring (EIM) Consortium, The European (2008). Assisted reproductive technology in Europe, 2004: Results generated from european registers by ESHRE. *Human Reproduction*, 23(4):756–771.
- [Anderson, 1958] Anderson, F.W. (1958). *An Introduction to Multivariate Statistical Analysis*. John Wiley and Sons.
- [Asuncion y Newman, 2007] Asuncion, A. y Newman, D.J. (2007). UCI machine learning repository. <http://www.ics.uci.edu/~mlearn/MLRepository.html>.
- [Baczkowski y col., 2004] Baczkowski, T., Kurzawa, R., y Glabowski, W. (2004). Methods of embryo scoring in in vitro fertilization. *Reproductive Biology*, 4(1):5–22.
- [Bamber, 1975] Bamber, D. (1975). The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *Journal of Mathematical Psychology*, 12:387–415.
- [Ben-Bassat, 1982] Ben-Bassat, M. (1982). Pattern recognition and reduction of dimensionality. En *Handbook of Statistics II*, pp 773–791. North-Holland.
- [Blanco y col., 2005] Blanco, R., Inza, I., Merino, M., Quiroga, J., y Larrañaga, P. (2005). Feature selection in Bayesian classifiers for the prognosis of survival of cirrhotic patients treated with TIPS. *Biomedical Informatics*, 38:376–388.
- [Blum y Langley, 1997] Blum, A.L. y Langley, P. (1997). Selection of relevant features and examples in machine learning. *Artificial Intelligence*, 97:245–271.
- [Boscolo y col., 2002] Boscolo, R., Brown, M.S., y McNitt-Gray, M.F. (2002). Medical image segmentation with knowledge-guided robust active contours. *RadioGraphics*, 22(2):437–448.

- [Bottcher, 2004] Bottcher, S.G. (2004). *Learning Bayesian Networks with Mixed Variables*. PhD thesis, Aalborg University.
- [Breiman, 1996a] Breiman, L. (1996a). Bagging predictors. *Machine Learning*, 24:123–140.
- [Breiman, 1996b] Breiman, L. (1996b). Stacked regression. *Machine Learning*, 24:49–64.
- [Breiman y col., 1984] Breiman, L., Friedman, J.H., Olshen, R.A., y Stone, C.J. (1984). *Classification and Regression Trees*. Wadsworth and Brooks.
- [Canny, 1986] Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698.
- [Carlson, 2005] Carlson, B.M. (2005). *Embriología Humana y Biología del Desarrollo*. Elsevier Es., 3 ed.
- [Chow y Liu, 1968] Chow, C. y Liu, C. (1968). Approximating discrete probability distributions with dependence trees. *IEEE Transactions on Information Theory*, 14:462–467.
- [Cleary y Trigg, 1995] Cleary, J.G. y Trigg, L.E. (1995). K*: An instance-based learner using an entropic distance measure. *Proceedings of the 12th International Conference on Machine Learning*, pp 108–114.
- [Cohen, 1960] Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20:37–46.
- [Cohen y col., 1989] Cohen, J., Inge, K.L., y Suzman, K. (1989). Videocinematography of fresh and cryopreserved embryos: A retrospective analysis of embryonic morphology and implantation. *Fertility and Sterility*, 51:820–827.
- [Cover y Thomas, 1991] Cover, T.M. y Thomas, J.A. (1991). *Elements of Information Theory*. John Wiley and Sons.
- [Dawid, 1976] Dawid, A. P. (1976). Properties of diagnostic data distributions. *Biometrics*, 32:647–658.
- [Debuse y Rayward-Smith, 1999] Debuse, J.C.W. y Rayward-Smith, V.J. (1999). Feature subset selection within a simulated annealing algorithm. *Journal of Intelligent Information Systems*, 9(1):57–81.
- [della Ragione y col., 2007] della Ragione, T., Verheyen, G., Papanikolaou, E.G., Van Landuyt, L., Devroey, P., y Van Steirteghem, A. (2007). Developmental stage on day-5 and fragmentation rate on day-3 can influence the implantation potential of top-quality blastocysts in IVF cycles with single embryo transfer. *Reproductive Biology and Endocrinology*, 5(1).
- [Dempster, 1972] Dempster, A.P. (1972). Covariance selection. *Biometrika*, 32:95–108.
- [Demšar, 2006] Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, 7:1–30.
- [Dietterich, 1998] Dietterich, T.G. (1998). Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Computation*, 10(7):1895–1924.
- [Domingos y Pazzani, 1997] Domingos, P. y Pazzani, M. (1997). On the optimality of the simple Bayesian classifier under zero-one loss. *Machine Learning*, 29(2–3):103–130.
- [Dougherty y col., 1995] Dougherty, K., Kohavi, R., y Sahami, M. (1995). Supervised and unsupervised discretization of continuous features. In *Proceedings of the Twelfth International Conference on Machine Learning*, pp 194–202.
- [Duda y Hart, 1973] Duda, R. y Hart, P. (1973). *Pattern Classification and Scene Analysis*. John Wiley and Sons, New York.

- [Duda y col., 2001] Duda, R., Hart, P., y Stork, D.G. (2001). *Pattern Classification*. Wiley Interscience.
- [Džeroski y Ženko, 2004] Džeroski, S. y Ženko, B. (2004). Is combining classifiers better than selecting the best one? *Machine Learning*, 54(3):255–273.
- [Edwards y Steptoe, 1980] Edwards, R. y Steptoe, P. (1980). *A Matter of Life*, volume 40 of *The Story of a Medical Breakthrough*. Hutchinson, London.
- [Edwards y col., 1980] Edwards, RG., Steptoe, PC., y Purdy, J.M. (1980). Establishing full term human pregnancies using cleaving embryos grown in vitro. *British Journal of Obstetrics and Gynecology*, 87(8085):737–756.
- [Efron, 1983] Efron, B. (1983). Estimating the error rate of a prediction rule: Improvement on cross-validation. *Journal of the American Statistical Association*, 78(282):316–330.
- [Efron y Tibshirani, 1993] Efron, B. y Tibshirani, R.J. (1993). *An Introduction to the Bootstrap*. Monograph on Statistics and Applied Probability. Chapman and Hall.
- [Elvira Consortium, 2002] Elvira Consortium (2002). Elvira: An environment for creating and using probabilistic graphical models. En *Proceedings of the 1st European Workshop on Probabilistic Graphical Models*, pp 222–230, Cuenca (Spain).
- [Fisch y col., 2001] Fisch, J.D., Rodriguez, H., Ross, R., Overby, G., y Sher, G. (2001). The graduated embryo score (GES) predicts blastocyst formation and pregnancy rate from cleavage-stage embryos. *Human Reproduction*, 16(9):1970–1975.
- [Freund y Schapire, 1996] Freund, Y. y Schapire, R.E. (1996). Experiments with a new boosting algorithm. En *Proceedings of the Thirteenth International Conference on Machine Learning (ICML)*, pp 148–156, Bari, Italy.
- [Friedman, 1937] Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association*, 32(200):675–701.
- [Friedman y col., 1997] Friedman, N., Geiger, D., y Goldsmidt, M. (1997). Bayesian network classifiers. *Machine Learning*, 29(2):131–163.
- [Gabrielsen y col., 2000] Gabrielsen, A., Bhatnager, P.R., Petersen, K., y Lindenberg, S. (2000). Influence of zona pellucida thickness of human embryos on clinical pregnancy outcome following in vitro fertilization treatment. *Journal of Assisted Reproduction and Genetics*, 17(6):323–328.
- [Gabrielsen y col., 2001] Gabrielsen, A., Lindenberg, S., y Petersen, K. (2001). The impact of the zona pellucida thickness variation of human embryos on pregnancy outcome in relation to suboptimal embryo development. A prospective randomized controlled study. *Human Reproduction*, 16(10):2166–2170.
- [Gama y col., 2000] Gama, J., Brazdil, P., y Valdes-Perez, R. (2000). Cascade generalization. *Machine Learning*, 41(3):315–343.
- [Geiger y Heckerman, 1994] Geiger, D. y Heckerman, D. (1994). Learning Gaussian networks. Technical Report MST-TR-94-10, Microsoft Advanced Technology Division, Microsoft Corporation.
- [Giorgietti y col., 2007] Giorgietti, C., Hans, E., Terriou, P., Salzmann, J., Barry, B., Chabert-Orsini, V., Chincole, J.M., Franquebalme, J.P., Glowaczower, E., , M-C. Sitri, Thibault, M-C., y Roulier, R. (2007). Early cleavage: An additional predictor of high implantation rate following elective single embryo transfer. *Reproductive BioMedicine Online*, 14(1):85–91.

- [Giorgotti y col., 1995] Giorgotti, C., Terriou, P., Auquier, P., Hans, E., Spach, J.L., Salzmann, J., y Roulier, R. (1995). Embryo score to predict implantation after in-vitro fertilization: Based on 957 single embryo transfers. *Human Reproduction*, 10:2427–2431.
- [Gonzales y Woods, 2002] Gonzales, R.C. y Woods, R.E. (2002). *Digital Image Processing*. Prentice Hall, second ed.
- [Hand y Yu, 2001] Hand, D.J. y Yu, K. (2001). Idiot's Bayes - not so stupid after all? *International Statistical Review*, 69(3):385–398.
- [Hanley y McNeil, 1982] Hanley, J.A. y McNeil, B.J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143(1):29–36.
- [Haralick y col., 1973] Haralick, R., Shanmugam, K., y Dinstein, I. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, 3:610–621.
- [Heckerman y col., 1995] Heckerman, D., Geiger, D., y Chickering, D. M. (1995). Learning Bayesian networks: The combination of knowledge and statistical data. *Machine Learning*, 20:197–243.
- [HFEA, 2005] HFEA (2005). IVF patient numbers and success rates continue to rise. HFEA issues latest find a clinic guide with 2005 success figures. <http://www.hfea.gov.uk/en/1626.html>.
- [Hill y col., 1989] Hill, G.A., Freeman, M., Bastias, M.C., Rogers, B.J., Herbert, C.M. III, Osteens, K.G., y Wentz, A.C. (1989). The influence of oocyte maturity and embryo quality on pregnancy rate in a program for in vitro fertilization-embryo transfer. *Fertility and Sterility*, 52:801–806.
- [Hu, 1962] Hu, M.K. (1962). Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory*, 8:179–187.
- [Inza y col., 2001] Inza, I., Larrañaga, P., y Sierra, B. (2001). Feature subset selection by Bayesian networks: A comparison with genetic and sequential algorithms. *International Journal of Approximate Reasoning*, 27(2):143–164.
- [John y col., 1994] John, G.H., Kohavi, R., y Pfleger, K. (1994). Irrelevant features and the subset selection problem. En *Proceedings of the 11th International Conference on Machine Learning*, pp 121–129.
- [Jurisica y col., 1998] Jurisica, I., Mylopoulos, J., Glasgow, J., Shapiro, H., y Casper, R. (1998). Case-based reasoning in IVF: Prediction and knowledge mining. *Artificial Intelligence in Medicine*, 12(1):1–24.
- [Karlsson y col., 2004] Karlsson, A., Overgaard, N.C., y Heyden, A. (2004). Automatic segmentation of zona pellucida in HMC images of human embryos. En *Proceedings of the 17th International Conference on Pattern Recognition, ICPR'04*, volume 3, pp 518–521.
- [Karlsson y col., 2005] Karlsson, A., Overgaard, N.C., y Heyden, A. (2005). A two-step area based method for automatic tight segmentation of zona pellucida in HMC images of human embryos. En *Lecture Notes in Computer Science 3459*, pp 503–514.
- [Kass y col., 1987] Kass, M., Witkin, A., y Terzopoulos, D. (1987). Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331.
- [Keenan y col., 2007] Keenan, J., Finger, R., Check, J.H., Daly, D., Dodds, W., y Stoddart, R. (2007). Favorable pregnancy, delivery, and implantation rates experienced in embryo donation programs in the United States. *Fertility and Sterility*.

- [Keogh y Pazzani, 1999] Keogh, E.J. y Pazzani, M.J. (1999). Learning augmented Bayesian classifiers: A comparison of distributions-based and classification-based approaches. En *Uncertainty 99: The 7th International Workshop on Artificial Intelligence and Statistics*, pp 225–230.
- [Kittler, 1978] Kittler, J. (1978). Feature set search algorithms. *Pattern Recognition and Signal Processing*, pp 41–60.
- [Klaus-Ruediger, 1994] Klaus-Ruediger, P. (1994). Digital differential hysteresis image processing displays what the microscope acquires but the eye can't see. En Bailey, G.W. y Garratt-Reed, A.J., editors, *Fifty Second Annual Meeting Microscopy Society America*, pp 416–417. San Francisco Press, Inc.
- [Kohavi, 1995a] Kohavi, R. (1995a). The power of decision tables. En *Proceedings of 8th European Conference on Machine Learning*, volume 912, pp 174–189. Springer-Verlag.
- [Kohavi, 1995b] Kohavi, R. (1995b). *Wrapper for Performance Enhancement and Oblivious Decision Graphs*. PhD thesis, Stanford University.
- [Kohavi, 1996] Kohavi, R. (1996). Scaling up the accuracy of naive-Bayes classifiers: A decision-tree hybrid. En *In Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, pp 114–119.
- [Kohavi y John, 1997] Kohavi, R. y John, G. (1997). Wrappers for feature subset selection. *Artificial Intelligence*, 97(1-2):273–324.
- [Kononenko, 1991] Kononenko, I. (1991). Semi-naïve Bayesian classifiers. En *Proceedings of the 6th European Working Session on Learning*, pp 206–219.
- [Kruskal y Wallis, 1952] Kruskal, W.H. y Wallis, W.A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 47(260):583–621.
- [Landis y Koch, 1977] Landis, J. R. y Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174.
- [Landwehr y col., 2003] Landwehr, N., Hall, M., y Frank, E. (2003). Logistic model trees. En *Machine Learning*, pp 241–252. Springer-Verlag.
- [Langley y Sage, 1994] Langley, P. y Sage, S. (1994). Induction of selective Bayesian classifiers. En *Proceedings of the 10th Conference on Uncertainty in Artificial Intelligence*, pp 399–406, Seattle, WA.
- [Larrañaga, 2003] Larrañaga, P. (2003). *Clasificación Supervisada via Modelos Gráficos Probabilísticos*. Trabajo de investigación presentado para la prueba de habilitación para Catedráticos de Universidad en el área de Ciencia de la Computación e Inteligencia Artificial.
- [Lasko y col., 2005] Lasko, T.A., Bhagwat, J.G., Zou, K.H., y Ohno-Machado, L. (2005). The use of receiver operating characteristic curves in biomedical informatics. *Journal of Biomedical Informatics*, 38:404–415.
- [Lauritzen, 1996] Lauritzen, S.L. (1996). *Graphical Models*. Oxford University.
- [Le Cessie y Van Houwelingen, 1992] Le Cessie, S. y Van Houwelingen, J.C. (1992). Ridge estimators in logistic regression. *Applied Statistics*, 41(1):191–201.
- [Lewis, 1998] Lewis, D. (1998). Naive (Bayes) at forty: The independence assumption in information retrieval. En *Proceedings of 10th European Conference on Machine Learning*, pp 4–15. Springer-Verlag.
- [Littlestone y Warmuth, 1989] Littlestone, N. y Warmuth, M. (1989). The weighted majority algorithm. Technical Report UCSC-CRL-89-16, Computer Research Laboratory University of California, Santa Cruz, C.A.
- [Liu y Motoda, 1998] Liu, H. y Motoda, H. (1998). *Feature Selection for Knowledge Discovery and Data Mining*. Kluwer Academic Publishers.

- [Maloof, 2003] Maloof, M.A. (2003). Learning when data sets are imbalanced and when costs are unequal and unknown. En *Workshop on Learning from Imbalanced Sets II*, Washington D.C.
- [Mangin y col., 2004] Mangin, J.F., Poupon, F., Duchesnay, E., Rivière, D., Cachia, A., Collins, D.L., Evans, A.C., y Régis, J. (2004). Brain morphometry using 3D moment invariants. *Medical Image Analysis*, 8(3):187–196.
- [Mann y Whitney, 1947] Mann, H. y Whitney, D. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics*, 18:50–60.
- [Manna y col., 2004] Manna, C., Patrizi, G., Rahman, A., y Sallam, H. (2004). Experimental results on the recognition of embryos in human assisted reproduction. *Reproductive BioMedicine Online*, 8(4):460–469.
- [Matorras y col., 2005] Matorras, R., Matorras, F., Mendoza, R., Rodríguez, M., Remohí, J., Rodríguez-Escudero, F.J., y Simón, C. (2005). The implantation of every embryo facilitates the chances of the remaining embryos to implant in an IVF programme: A mathematical model to predict pregnancy and multiple pregnancy rates. *Human Reproduction*, 20(10):2923–2931.
- [McInerney y Terzopoulos, 1996] McInerney, T. y Terzopoulos, D. (1996). Deformable models in medical images analysis: A survey. *Medical Image Analysis*, 1(2):91–108.
- [McKay, 2003] McKay, D. (2003). *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press.
- [Merz, 1999] Merz, C.J. (1999). Using correspondence analysis to combine classifiers. *Machine Learning*, 36:33–58.
- [Mills, 1992] Mills, C.L. (1992). *Factors affecting embryological parameters and embryo selection for IVF-ET*. En *A Textbook of In Vitro Fertilization and Assisted Reproduction*, pp 187–204. The Parthenon Group.
- [Minsky, 1961] Minsky, M. (1961). Steps toward artificial intelligence. *Transactions on Institute of Radio Engineers*, 49:8–30.
- [Morales y col., 2008a] Morales, D.A., Bengoetxea, E., y Larrañaga, P. (2008a). Automatic segmentation of zona pellucida in human embryo images applying an active contour model. En Hoey, J. y McKenna, S., editors, *Medical Image Understanding and Analysis*, pp 209–213, Dundee, Scotland.
- [Morales y col., 2008b] Morales, D.A., Bengoetxea, E., y Larrañaga, P. (2008b). Selection of human embryos for transfer by Bayesian classifiers. *Computers in Biology and Medicine*, 38(11-12):1177–1186.
- [Morales y col., 2009] Morales, D.A., Bengoetxea, E., y Larrañaga, P. (2009). *Data Mining and Medical Knowledge Management: Cases and Applications*, capítulo Gaussian-stacking multiclassifiers for human embryo selection, pp 307–331. IGI Global Inc.
- [Morales y col., 2007] Morales, D.A., Bengoetxea, E., Larrañaga, P., García, M., Franco, Y., Fresnada, M., y Merino, M. (2007). Selection of human embryos in IVF by Bayesian classifiers. En *CEDI 2007, II Simposio de Inteligencia Computacional*, pp 251–260, Spain.
- [Morales y col., 2008c] Morales, D.A., Bengoetxea, E., Larrañaga, P., García, M., Franco, Y., Fresnada, M., y Merino, M. (2008c). Bayesian classification for the selection of in-vitro human embryos using morphological and clinical data. *Computer Methods and Programs in Biomedicine*, 90(2):104–116.
- [Palmstierna y col., 1998] Palmstierna, M., Murkes, D., Csemiczky, G., Andersson, O., y Wrambsy, H. (1998). Zona pellucida thickness variation and occurrence

- of visible mononucleated blastomeres in preembryos are associated with a high pregnancy rate in IVF treatment. *Journal of Assisted Reproduction and Genetics*, 15(2):70–75.
- [Patrizi y col., 2004] Patrizi, G., Manna, C., Moscatelli, C., y Nieddu, L. (2004). Pattern recognition methods in human-assisted reproduction. *International Transactions in Operational Research*, 11(4):365–379.
- [Pazzani, 1997] Pazzani, M.J. (1997). Searching for dependencies in Bayesian classifiers. En Fisher, D. y Lenz, H., editors, *Artificial Intelligence and Statistics IV, Lecture Notes in Statistics*, pp 239–248, New York, NY. Springer–Verlag.
- [Pearl, 1988] Pearl, J. (1988). *Probabilistic Reasoning in Intelligence Systems*. Morgan Kaufmann, Los Altos CA.
- [Pérez y col., 2006] Pérez, A., Larrañaga, P., y Inza, I. (2006). Supervised classification with conditional Gaussian networks: Increasing the structure complexity from naive Bayes. *International Journal of Approximate Reasoning*, 43:1–25.
- [Pham y col., 2000] Pham, D.L., Chenyang, X., y Jerry, L.P. (2000). A survey of current methods in medical image segmentation. *Annual Review of Biomedical Engineering*, 2:315–337.
- [Provost y Fawcett, 2001] Provost, F. y Fawcett, T. (2001). Robust classification for imprecise environments. *Machine Learning*, 42:203–231.
- [Provost y col., 1998] Provost, F., Fawcett, T., y Kohavi, R. (1998). The case against accuracy estimation for comparing induction algorithms. En *Proceedings of the Fifteenth International Conference on Machine Learning*, pp 445–453, San Francisco. CA: Morgan Kaufmann.
- [Puissant y col., 1987] Puissant, F., Rysselberge, M. Van, Barlow, P., Deweze, J., y Leroy, F. (1987). Embryo scoring as a prognostic tool in IVF treatment. *Human Reproduction*, 2:705–708.
- [Quinlan, 1986] Quinlan, J.R. (1986). Induction of decision trees. *Machine Learning*, 1(1):81–106.
- [Rubinstein y Hastie, 1997] Rubinstein, Y. D. y Hastie, T. (1997). Discriminative vs. informative learning. En *Proceedings of the Third International Conference on Knowledge Discovery and Data Mining*, pp 49–53.
- [Ruggeri y Pajaro, 2002] Ruggeri, A. y Pajaro, S. (2002). Automatic recognition of cell layers in corneal confocal microscopy images. *Computer Methods and Programs in Biomedicine*, 68(1):25–35.
- [Sahami, 1996] Sahami, M. (1996). Learning limited dependence Bayesian classifiers. En *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining*, pp 335–338.
- [Saith y col., 1998] Saith, R., Srinivasan, A., Michie, D., y Sargent, I. (1998). Relationships between the developmental potential of human in-vitro fertilization embryos and features describing the embryo, oocyte and follicle. *Human Reproduction Update*, 4(2):121–134.
- [Sakkas y col., 2001] Sakkas, D., Percival, G., DÁrcy, Y., Sharif, K., y Afnan, M. (2001). Assessment of early cleaving in vitro fertilized human embryos at the 2-cell stage before transfer improves embryo selection. *Fertility and Sterility*, 76(6):1150–1156.
- [Schulman y col., 1993] Schulman, A., Ben-Num, I., Gethler, Y., Kaneti, H., Shilon, M., y Beyth, Y. (1993). Relationship between embryo morphology and implantation rate after in vitro fertilization treatment in conception cycles. *Fertility and Sterility*, 60:123–126.

- [Schwarz, 1978] Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 7(2):461–464.
- [Scott, 2003] Scott, L. (2003). Pronuclear scoring as a predictor of embryo development. *Reproductive BioMedicine Online*, 6(2):201–214.
- [Scott y col., 2000] Scott, L., Alvero, R., y Leondires, M. (2000). The morphology of human pronuclear embryos is positively related to blastocyst development and implantation. *Human Reproduction*, 15(11):2394–2403.
- [Seewald, 2003] Seewald, A. (2003). *Towards Understanding Stacking*. PhD thesis, Department of Informatics, Technical University of Wien.
- [Shannon, 1948] Shannon, C.E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423.
- [Shen y col., 2002] Shen, S., Wong, C., Ho, K., Telles, T.L., Fujimoto, V.Y., y Cedars, M.I. (2002). The morphology of 2 pronuclear (2PN) embryos is related to the quality of day 3 embryos. *Fertility and Sterility*, 78:S52–S53.
- [Sierra y col., 2001] Sierra, B., Serrano, N., naga, P. Larra Plasencia, E.J., Inza, I., Jiénez, J.J., Revuelta, P., y Mora, M.L. (2001). Using Bayesian networks in the construction of a bi-level multi-classifier. A case study using intensive care unit patients data. *Artificial Intelligence in Medicine*, 22(3):233–248.
- [Smith y Whittaker, 1998] Smith, P.W.F. y Whittaker, J. (1998). Edge exclusion tests for graphical Gaussian models. En *Learning in Graphical Models*, pp 555–574. Kluwer Academic Publishers.
- [Spackman, 1989] Spackman, K.A. (1989). Signal detection theory: Valuable tools for evaluating inductive learning. En *Proceedings of the 6th International Workshop on Machine Learning*, pp 1285–1293.
- [Spearman, 1904] Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology*, pp 72–101.
- [Steptoe y Edwards, 1978] Steptoe, PC. y Edwards, RG. (1978). Birth after reimplantation of a human embryo. *Lancet*, 2(8085):366.
- [Stone, 1974] Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society Series B*, 36:111–147.
- [Terriou y col., 2001] Terriou, P., Sapin, C., Giorgetti, C., Hans, E., Spach, J.L., y Roulier, R. (2001). Embryo score is a better predictor of pregnancy than the number of transferred embryos or female age. *Fertility and Sterility*, 75(3):525–531.
- [Tesarik y Greco, 1999] Tesarik, J. y Greco, E. (1999). The probability of abnormal preimplantation development can be predicted by a single static observation on pronuclear stage morphology. *Human Reproduction*, 14(5):1318–1323.
- [Ting y Witten, 1999] Ting, K. y Witten, M. (1999). Issues in stacked generalization. *Journal of Artificial Intelligence Research*, 10:271–289.
- [Todorovski y Džeroski, 2002] Todorovski, L. y Džeroski, S. (2002). Combining classifiers with meta decision trees. *Machine Learning*, 50(3):223–249.
- [Trimarchi y col., 2003] Trimarchi, J.R., Goodside, J., Passmore, L., Silberstein, T., Hamel, L., y Gonzalez, L. (2003). Comparing data mining and logistic regression for predicting IVF outcome. *Fertility and Sterility*, 80(3):100–100.
- [Van Peperstraten y col., 2008] Van Peperstraten, A.M., Kreuwel, I.A., Hermens, R.P., Nelen, W.L., Van Dop, P.A., Grol, R.P., y Kremer, J.A. (2008). Determinants of the choice for single or double embryo transfer in twin prone couples. *Acta Obstetricia et Gynecologica Scandinavica*, 87(2):226–231.
- [Van Royen y col., 1999] Van Royen, E., Mangelschots, K., DeNeubourg, D., Valkenburg, M., Van de Meerssche, M., Ryckaert, G., Eestermans, W., y Gerris, J.

- (1999). Characterization of a top quality embryo, a step towards single-embryo transfer. *Human Reproduction*, 14(9):2345–2349.
- [Vapnik, 1995] Vapnik, V. (1995). *The Nature of Statistical Learning Theory*. Springer-Verlag.
- [Veek, 1999] Veek, L.L. (1999). *An Atlas of Human Gametes and Conspectuses: An Illustrated Reference for Assisted Reproductive Technology*. Parthenon Publishing.
- [Verberg y col., 2007] Verberg, M.F., Eijkemans, M.J., Macklon, N.S., Heijnen, E.M., Fauser, B.C., y Broekmans, F.J. (2007). Predictors of ongoing pregnancy after single-embryo transfer following mild ovarian stimulation for IVF. *Fertility and Sterility*, 89(5):1159–1165.
- [Wilcoxon, 1945] Wilcoxon, F. (1945). Individual comparison by ranking methods. *Biometrics*, 1:80–83.
- [Windt y col., 2004] Windt, M.L., Kruger, T.F., Coetze, K., y Lombard, C.J. (2004). Comparative analysis of pregnancy rates after the transfer of early dividing embryos versus slower dividing embryos. *Human Reproduction*, 19(5):1155–1162.
- [Withey y Koles, 2007] Withey, D.J. y Koles, Z.J. (2007). Medical image segmentation: Methods and software. En *Noninvasive Functional Source Imaging of the Brain and Heart and the International Conference on Functional Biomedical Imaging*, pp 140–143.
- [Witten y Frank, 2005] Witten, I.H. y Frank, E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. Morgan Kaufmann Publishers.
- [Wolpert, 1992] Wolpert, D.H. (1992). Stacked generalization. *Neural Networks*, 5:241–259.
- [Xu y col., 2000] Xu, C., Pham, D.L., y Prince, J.L. (2000). Image segmentation using deformable models. En Fitzpatrick, J.M. y Sonka, M., editors, *Handbook on Medical Imaging*, volume 2, pp 129–174. SPIE Press.
- [Xu y Prince, 1998] Xu, C. y Prince, J.L. (1998). Snakes, shapes, and gradient vector flow. *IEEE Transactions on Image Processing*, 7(3):359–369.
- [Xu y col., 1992] Xu, L., Krzyzak, A., y Suen, C. (1992). Methods of combining multiple classifiers and their applications to handwriting recognition. *IEEE Transactions on Systems, Man and Cybernetics*, 22:418–435.
- [Yang y Honavar, 1998] Yang, J. y Honavar, V. (1998). Feature subset selection using a genetic algorithm. *IEEE Intelligent Systems*, 13(2):44–49.
- [Zheng y Webb, 1998] Zheng, Z. y Webb, G.I. (1998). Lazy Bayesian rules. Technical Report TR C98/17, Deakin University, Geelong, Victoria, Australia.
- [Ziebe y col., 1997] Ziebe, S., Petersen, K., Lindenberg, S., Andersen, A.G., Gabrielsen, A., y Andersen, A.N. (1997). Embryo morphology or cleavage stage: How to select the best embryos for transfer after in-vitro fertilization. *Human Reproduction*, 12(7):1545–1549.
- [Zollner y col., 2002] Zollner, U., Zollner, K.P., Hartl, G., Dietl, J., y Steck, T. (2002). The use of a detailed zygote score after IVF/ICSI to obtain good quality blastocysts: The German experience. *Human Reproduction*, 17(5):1327–1333.

