Análisis Multivariado y Reconocimiento de Patrones

Métodos y Aplicaciones Médicas

Generación y Extracción/Selección de Características en la Detección Automática de Isquemia y Deficiencias Valvulares sobre Registros de la Actividad Cardíaca

Edilson Delgado Trejos

Departamento de Ingeniería Eléctrica, Electrónica y Computación
Universidad Nacional de Colombia Sede Manizales

Generación y extracción/selección de características en la detección automática de isquemia y deficiencias valvulares sobre registros de la actividad cardíaca

Edilson Delgado Trejos

Tesis doctoral

Departamento de Ingeniería Eléctrica, Electrónica y Computación

Universidad Nacional de Colombia

Director:
Prof. Germán Castellanos

Memoria presentada para optar al título de
Doctor en Ingeniería
Línea de investigación en Automática
©6 de Junio de 2008

Grupo de Control y Procesamiento Digital de Señales
Manizales 2008

Measurement and Feature Extraction/Selection in the Automatic Detection of Ischemia and Valvular Pathologies on Cardiac Signals

Edilson Delgado Trejos

Doctoral thesis

Department of Electrical, Electronic and Computer Engineering

National University of Colombia

Supervisor:
Prof. Germán Castellanos

This document is presented
in partial fulfillment of the requirements for the
Doctorate Degree in Engineering Sciences
©June 6 of 2008

Control and Digital Signal Processing Group
Manizales 2008

A Jesús por ser mi guía en el camino.
A la memoria de mi abuela: Idalí Ladino.

ÍNDICE GENERAL

Índice general I

Índice de figuras V

Índice de tablas VII

Agradecimientos XI

Notaciones XIII

Resumen (+*Abstract*) XV

Introducción **1**
Pertinencia . . . 1
Antecedentes . . . 3
Justificación . . . 3
Planteamiento del problema . . . 4
Objetivos . . . 5
Organización del manuscrito . . . 6

1. Registro de la actividad cardíaca y diagnóstico asistido **9**
1.1. Señales electrocardiográficas . . . 9
1.1.1. Morfología básica . . . 10
1.1.2. Derivaciones . . . 11
1.1.3. Descripción de las componentes . . . 13
1.1.4. Modos de registro . . . 16
1.1.5. Variaciones del ECG debidas a enfermedades isquémicas . . . 16
1.2. Señales fonocardiográficas . . . 19
1.2.1. Origen fisiológico . . . 20
1.2.2. Descripción de las componentes . . . 21
1.2.3. Focos de auscultación . . . 22
1.2.4. Variaciones del FCG debidas a soplos cardíacos . . . 23
1.3. Estado del arte en el diagnóstico asistido de patologías . . . 27
1.3.1. El diagnóstico asistido y sus aplicaciones . . . 27
1.3.2. Consideraciones para el diagnóstico asistido . . . 29
1.3.3. Identificación automática en el diagnóstico asistido . . . 32

Bibliografía **39**

2. Modelos de representación para las señales cardíacas **41**
2.1. Medidas de diagnóstico: análisis temporal y análisis acústico . . . 41
2.2. Análisis espectro-temporal . . . 45
2.2.1. El espectro de Fourier . . . 46
2.2.2. Densidad espectral de potencia . . . 47
2.2.3. Transformada de Fourier de tiempo corto (STFT) . . . 48
2.2.4. Transformada de Gabor . . . 49
2.2.5. Análisis multiresolución y transformada wavelet . . . 49
2.3. Análisis de complejidad . . . 50
2.3.1. Determinación de la no linealidad determinista . . . 51
2.3.2. Reconstrucción del espacio de estados . . . 52
2.3.3. Medidas de complejidad . . . 55
2.4. Estado del arte en el análisis de señales cardíacas . . . 60
2.4.1. Filtración . . . 60
2.4.2. Segmentación . . . 68
2.4.3. Caracterización y reconocimiento . . . 73

Bibliografía **79**

3. Análisis de relevancia en estructuras multivariadas **85**
3.1. Modelo general de sistemas gaussianos . . . 85
3.2. Relevancia como medida de representación . . . 89
3.2.1. Definición generalizada de la relevancia . . . 89
3.2.2. Función de evaluación en el análisis de relevancia . . . 93
3.3. Reducción de dimensiones y niveles de operación . . . 94
3.3.1. Nivel 1: Proyección . . . 94
3.3.2. Nivel 2: Interpretación . . . 101
3.3.3. Nivel 3: Visualización . . . 117
3.4. Estado del arte en la reducción de dimensiones . . . 120

Bibliografía **125**

4. Marco experimental **129**
4.1. Esquema metodológico . . . 129
4.2. Bases de datos . . . 129
4.2.1. Señales electrocardiográficas BD-UNAL-GCPDS-ECG . . . 129
4.2.2. Señales electrocardiográficas BD-MIT E-STT . . . 130
4.2.3. Señales fonocardiográficas BD-UNAL-GCPDS-FCG . . . 131
4.2.4. Conjuntos de datos artificiales . . . 132
4.3. Preproceso y adecuación de señales . . . 132
4.3.1. Reducción de perturbaciones . . . 132
4.3.2. Detección de eventos . . . 139
4.4. Extracción de las medidas de representación . . . 148
4.4.1. Medidas de diagnóstico . . . 148
4.4.2. Análisis espectro-temporal . . . 149
4.4.3. Análisis de complejidad . . . 154
4.5. Preproceso experimental de datos . . . 156
4.5.1. Normalización . . . 156
4.5.2. Remoción de valores atípicos . . . 156
4.5.3. Verificación de normalidad . . . 158
4.5.4. Transformación de distribuciones . . . 160
4.6. Modelos desarrollados para reducir dimensiones . . . 161
4.6.1. Nivel 1 . . . 161
4.6.2. Nivel 2 . . . 163

4.6.3. Nivel 3 166

Bibliografía **171**

5. Resultados y discusión **175**
5.1. Etapa de preprocesamiento 175
5.1.1. Filtración 175
5.1.2. Segmentación 179
5.2. Generación del espacio de características 181
5.3. Reducción de dimensiones: Nivel 1 183
5.4. Reducción de dimensiones: Nivel 2 184
5.5. Reducción de dimensiones: Nivel 3 187
5.5.1. Reducción de dimensiones mediante cargas agrupadas 187
5.5.2. Reducción de dimensiones para el reconocimiento de múltiples formas 190
5.6. Discusión académica de los resultados 191

Bibliografía **195**

6. Conclusiones **197**
6.1. Conclusiones generales de la tesis 197
6.2. Trabajo Futuro 200
6.3. Contribuciones científicas 200
6.4. Significado práctico de la tesis 201

A. Algoritmos genéticos **203**
A.1. El algoritmo genético simple 203
A.2. Fundamentos teóricos de los algoritmos genéticos 206
A.2.1. Teoría de esquemas 207
A.2.2. Caminos reales (*royal roads*) 208
A.3. Mejoramientos al modelo básico 210
A.3.1. Métodos de escalamiento de la calidad 210
A.3.2. Métodos de ordenamiento (*Ranking*) 211
A.3.3. Nichos y especies 212
A.4. Sintonización de los parámetros de control 213

Bibliografía **215**

B. Predicción en series de tiempo caóticas **217**

Bibliografía **219**

Índice alfabético **221**

ÍNDICE DE FIGURAS

1.1. Electrofisiología del corazón . . . 10
1.2. Derivaciones bipolares y el triángulo de Einthoven . . . 12
1.3. Derivaciones unipolares de las extremidades . . . 13
1.4. Derivaciones unipolares precordiales y sus registros típicos . . . 14
1.5. Ciclo cardíaco . . . 21
1.6. Componentes de los sonidos cardíacos y su auscultación . . . 22
1.7. Variaciones del FCG debidas a deficiencias valvulares . . . 27
1.8. Exigencias para la calidad del entrenamiento . . . 33
1.9. Diagrama de bloques correspondiente al reconocimiento de patrones . . . 34
1.10. Rango de solución . . . 36
1.11. Capacidad de generalización . . . 37

2.1. Diferentes métodos de envolvente . . . 43
2.2. Banco de filtros para obtener los MFCC . . . 44
2.3. Mapa plano de una región . . . 53
2.4. Estimación de D_2 para una señal FCG . . . 57
2.5. Divergencia de trayectorias . . . 58

3.1. Modelo general de un sistema dinámico lineal . . . 86
3.2. Análisis de relevancia sobre el espacio de representación . . . 89
3.3. Interpretación gráfica de la relevancia generalizada . . . 90
3.4. Capas o niveles en la reducción de dimensiones . . . 95
3.5. Diagrama general de la selección de características . . . 102
3.6. Diagrama de flujo del algoritmo ID3 . . . 111
3.7. Taxonomía de los métodos de proyección . . . 117
3.8. Diagrama de bloques del sistema de análisis y visualización de datos multivariados . . . 119

4.1. Esquema general de procedimiento . . . 130
4.2. Perturbaciones frecuentes en las señales ECG . . . 133
4.3. Estadística de tiempos de recurrencia sobre una señal FCG normal . . . 143
4.4. Remoción de soplo sistólico y detección de $S2$ . . . 145
4.5. Envelograma de Shannon de una señal FCG . . . 146
4.6. Segmentación intra-latido de señales FCG . . . 147
4.7. Promedio de los espectrogramas obtenidos usando STFT . . . 152
4.8. Promedio de los espectrogramas obtenidos usando transformada Gabor . . . 152
4.9. Suma de los detalles 1 y 2 de la descomposición WT–$db4$ . . . 153

4.10. Función de evaluación 154
4.11. Metodología de optimización para los parámetros de control 155
4.12. Distribuciones con kurtosis mayores y menores que 3 comparadas a la gaussiana 157
4.13. Intervalos de longitud equiprobable 159
4.14. Reducción de dimensiones basada en el peso de las características 164
4.15. Modelo híbrido propuesto para la selección de características 164
4.16. Esquema del procedimiento de cargas agrupadas 166
4.17. Esquema del procedimiento para formas agrupadas 168

5.1. Visualización en las distintas etapas del procedimiento 188
5.2. Interpretación visual de los grupos discriminates de características 189
5.3. Visualización de las diferentes formas de onda relacionadas a episodios isquémicos 191

A.1. Proceso de selección y recombinación en una generación del GA 205
A.2. Recombinación de un solo punto 206
A.3. Cadena óptima, dividida en ocho bloques constructores 209

ÍNDICE DE TABLAS

1.1. Eventos eléctricos en el corazón 11
1.2. Descripción de las 12 derivaciones del ECG 14
1.3. Derivaciones y polaridad de las ondas 14
1.4. Duración del intervalo QT 16
1.5. Variaciones del ECG debidas a enfermedades isquémicas en diversas localizaciones 19
1.6. Parámetros en frecuencia de los sonidos cardíacos 22
1.7. Escala de Levine-Harvey 24
1.8. Freeman-Levine simplificada 24
1.9. Publicación de trabajos en medios de divulgación científica 29
1.10. Cantidad de trabajos por área de procesamiento 35

2.1. Medidas de diagnóstico sobre señales ECG 42
2.2. Comportamiento de una señal según el exponente de Hurst (H) 59
2.3. Resultados de la segmentación basada en el envelograma 72
2.4. Resultados de segmentación usando descomposición y reconstrucción wavelet 72
2.5. Raíz cuadrada del error cuadrático medio normalizado 77
2.6. Primer pico de frecuencia dominante 77
2.7. Segundo pico de frecuencia dominante 77

3.1. Comparación de diferentes medidas de relevancia 93
3.2. Conjunto de observaciones con distribución normal 114
3.3. Comparación de algoritmos de selección de características 123

4.1. Respuesta en frecuencia a diferentes escalas 139
4.2. Medidas de diagnóstico extraídas de las señales FCG 150
4.3. Representación espectro-temporal de las señales ECG 151

5.1. Distorsión filtro IE $SNR_i = -6dB$ 176
5.2. Distorsión filtro IE $SNR_i = -3dB$ 176
5.3. Distorsión filtro IE $SNR_i = 0dB$ 176
5.4. Distorsión filtro ASIC $SNR_i = -6dB$ 177
5.5. Distorsión filtro ASIC $SNR_i = -3dB$ 177
5.6. Distorsión filtro ASIC $SNR_i = 0dB$ 177
5.7. Filtración de señales perturbadas con 2 armónicos y SNR=-6 dB 178
5.8. Filtración de línea base usando la transformada wavelet 178
5.9. Detección QRS para la base de datos BD-MIT E-STT 179

5.10. Comparación con otros algoritmos de segmentación QRS 180
5.11. Resultados de la segmentación intra-latido en señales FCG 180
5.12. Comparación con otros algoritmos de segmentación FCG 180
5.13. Resultados de clasificación para el ensamble $\mathbf{X}_{ECG}$ 181
5.14. Resultados de clasificación para el ensamble $\mathbf{X}_{STT}$ 182
5.15. Resultados de clasificación para el ensamble $\mathbf{X}_{FCG}$ 183
5.16. Resultados obtenidos con PCA e ICA para los datos artificiales 183
5.17. Resultados obtenidos con PCA e ICA para los ensambles de características 184
5.18. Resultados obtenidos con PPCA y kernel–ICA 184
5.19. Porcentaje de error de clasificación 185
5.20. Porcentaje de reducción de dimensiones 185
5.21. Porcentaje de error de verificación 185
5.22. Resultado de clasificación por familia wavelet del ensamble $\mathbf{X}_{ECG}$ 186
5.23. Resultado de clasificación por familia wavelet del ensamble $\mathbf{X}_{STT}$ 186
5.24. Comparación con otros algoritmos de detección de soplos 187
5.25. Matriz de confusión 188
5.26. Comparación con otros algoritmos de detección de isquemia 190
5.27. Matriz de confusión relacionada con la agrupación de latidos 190

LISTA DE ALGORITMOS

2.1. Cálculo de los valores sustitutos usando la transformada de Fourier 52

3.1. Clasificador basado en los k vecinos más cercanos 108
3.2. Algoritmo SFFS 110
3.3. Algoritmo genético simple 113
3.4. Algoritmo Isodata 116

4.1. Algoritmo genético de bajo nivel 155
4.2. Detección multivariada de datos atípicos 158
4.3. Visualización mediante cargas agrupadas 167
4.4. Visualización de formas agrupadas 169

Agradecimientos

Durante el desarrollo de este trabajo, he sentido el apoyo de un conjunto de personas que no solo me han brindado su cariño sino su talento para hacer que mi vida sea continuamente construida en conocimiento y espíritu... a todos ellos: ¡Gracias!

Quiero agradecer al director de esta tesis, el profesor Germán Castellanos, ya que fue su orientación la que permitió encontrar solución a los problemas que se fueron presentando y así lograr los resultados que se exponen en este trabajo. Quiero expresar mi agradecimiento a todos los que han sido y son integrantes del Grupo de Control y Procesamiento Digital de Señales, ya que esta tesis doctoral es el resultado de un enorme trabajo en equipo. Especialmente quiero agradecer a Victoria Eugenia Montes, Johanna Carvajal, Diana Marín, Lolita González, María Constanza Torres, Olga Beatriz Naranjo, María Andrea Lorza, Carolina Ospina, Carolina Acosta, Gustavo Guarín, Juan Camilo Arias, Nicolás Ramírez, David Echeverri, Oscar Henao, Mario Ladino, Julián Felipe Ramírez, Manuel Arango, Oscar Arellano, Carlos Lino Rengifo, John Eder Cuitiva, Ricardo Grisales, Jose Fernando Marín, David Jerónimo Giraldo, Luis Felipe Giraldo, Milton Orlando Sarria, David Márquez, Haendel Gabriel Guerrero, Marcelo Castaño, Alejandro Giraldo, Pablo Daniel Perugache, Jose Fernando Restrepo, Andrés Felipe Quiceno, Luis David Avendaño, Fernando Martínez, Jorge Jaramillo, Julián David Arias, Genaro Daza, Luis Gonzalo Sánchez, José Luis Rodríguez, Alexander Cerquera, Ricardo Henao, Juan Diego Pulgarín y los profesores: Juan Carlos Riaño, Julio Fernando Suárez, Gerard Olivar, Julio César García, Franklin Sepúlveda, Mauricio Orozco, Álvaro Orozco, Jorge Rivera, Rafael Gutiérrez, Ana María Matijasevic, Guillermo Agudelo, Gustavo Reyes, Óscar Jaramillo y Óscar Castaño, cada uno de ellos aportaron valiosa parte de su tiempo y experiencia para que esta tesis doctoral alcanzara un buen nivel en los resultados junto a un alto grado de discusión.

No se puede pasar por alto el enorme apoyo recibido por parte del Instituto Colombiano para el Desarrollo de la Ciencia y la Tecnología Francisco José de Caldas – COLCIENCIAS con su financiación mediante el programa de Apoyo a la Comunidad Científica a través de los Programas de Doctorados Nacionales, también destaco el respaldo recibido de la Di-

rección de Investigaciones de Manizales (DIMA) por su financiación a través del programa permanente de apoyo económico para la investigación en las Maestrías y los Doctorados de la Sede y el apoyo invaluable de la Vicerrectoría de Investigaciones de la Universidad Nacional de Colombia en cada una de las etapas de realización de esta tesis.

Quiero de una manera especial agradecer a las instituciones, hospitales y grupos de investigación que colaboraron de alguna manera en la realización de este trabajo, en forma particular, al Centro de Investigaciones en Ingeniería Biomédica (CREB) de la Universidad Politécnica de Cataluña, al Grupo de Informática Industrial, Comunicaciones y Automática (GIICA) de la Universidad Politécnica de Valencia, al Instituto de Investigación en Ingeniería de Aragón (I3A) en la Universidad de Zaragoza, al Grupo de Investigación en Algoritmos y Arquitecturas para Procesamiento de Señales Avanzado de la Universidad de Granada, al Grupo de Procesado de Señales Biológicas (GPSB) de la Universidad de las Palmas de Gran Canaria, a Telesalud de la Universidad de Caldas, a la Fundación Cardiovascular de Colombia, al Hospital Santa Sofía, a la Clínica del Seguro Social de Manizales y a los Centros de Salud ubicados en los principales municipios del Departamento de Caldas.

A nivel internacional quiero destacar la contribución de los profesores e investigadores: Peré Caminal, Montserrat Vallverdú, Raymon Jané, Beatriz Giraldo, Pedro Gomis, Pablo Laguna, David Cuesta, Carlos G. Puntonet, Manuel Rodríguez, Miguel Ángel Ferrer, Jesús B. Alonso, Carlos Manuel Travieso, Alexandre Perera, Fernando Rojas y Francisco Jiménez, a todos ellos les agradezco la oportunidad que me brindaron para realizar una pasantía de tanto provecho en España y el tiempo que invirtieron en analizar y discutir asuntos importantes de la tesis. A los profesores Juan Ignacio Godino, Miguel Ángel Carreira-Perpiñán, Sakari Lukkarinen y Leif Sörnmo les agradezco las valiosas observaciones que hicieron relacionadas a algunos esquemas metodológicos propuestos en este trabajo.

Fuera de las aulas de la academia, quiero agradecer a mis padres, Gerardo Delgado y María del Socorro Trejos, como también a mis hermanas Gisleny y Francy, por estar conmigo en cada momento ofreciéndome su apoyo incondicional. A Los Primos (Abner, Nacho, Caliche y Esteban) por su entusiasmo y permitirme desahogar en las partituras de su música. A mi novia Sarah Röthlisberger por su paciencia, sus contribuciones relacionadas a toda la divulgación que se hizo de esta tesis en inglés, por sus acertadas apreciaciones y la motivación que siempre recibí con la inspiración fruto del aliento de su amor... te amo. Finalmente, quiero a agradecer con toda mi alma y corazón al dueño y dador de todo, a Dios, por permitirme llegar a esta etapa del camino.

> *"El Señor te pastoreará siempre y en las sequías saciará tu alma, y dará vigor a tus huesos; y serás como huerto de riego, y como manantial de aguas, cuyas aguas nunca faltan".* La Biblia – Isaías 58:11.

Edilson Delgado Trejos, Junio de 2008.

Notaciones

Notación	*Significado*
$x(t)$, $s(t)$	Señal de tiempo continuo
$x[k]$, $s[k]$	Señal discretizada (con argumento normalizado)
$\{x_k : k = 1, \ldots, N\}$	Conjunto o sucesión de N valores puntuales x_k
$d(x_m, x_n)$	Distancia métrica entre los elementos x_m y x_n
$\boldsymbol{\xi}$; $\boldsymbol{\zeta}$	Vector de características; Vector de componentes
$\mathbf{X}$	Matriz de características
y; $\mathbf{a}$; $\mathbf{Z}$	Escalar; Vector; Matriz
rank$\{\mathbf{X}\}$	Rango de $\mathbf{X}$
trace$\{\mathbf{X}\}$	Traza de $\mathbf{X}$
det$\{\mathbf{X}\}$	Determinante de $\mathbf{X}$
$\mathbf{X}^{-1}$; $\mathbf{X}^{\top}$; $\mathbf{X}^{\mathsf{H}}$	Matriz inversa; transpuesta, hessiana
$\mathfrak{X}$	Espacio métrico funcional
$\mathbb{N}, \mathbb{Z}, \mathbb{R}, \mathbb{C}$	Dominio de los naturales, enteros, reales y complejos
$\mathbb{R}^n$	Espacio euclídeo con dimensión n
D_q	Dimensión fractal generalizada
$P(\Omega)$	Función de distribución de probabilidad sobre Ω
$\mathscr{N}(\mu_\xi, \sigma_\xi)$	Distribución normal con momentos μ y σ de ξ
$E\{\xi_i\}$	Valor esperado - promedio de ensamble de la variable ξ_i
$R_{\xi,\eta}(\tau)$	Función de correlación
$S_\xi(\omega)$	Densidad espectral potencia de ξ
$\boldsymbol{\Pi} = \{\pi_{mn}\}$	Matriz probabilidad de transición de estados
$\boldsymbol{P}(\boldsymbol{\nu}\|\boldsymbol{\vartheta})$	Matriz probabilidad condicional de aparición de estados
$\boldsymbol{p}_{\theta 1} = \{P(\theta_1 = \vartheta_i)\}$	Vector probabilidad inicial de aparición
$\mu(\boldsymbol{\xi})$	Medida de relevancia sobre el conjunto $\boldsymbol{\xi}$
$\mathcal{L}$	Operador matemático

Siglas frecuentes

Notación	*Significado*
ECG	Señal electrocardiográfica
FCG	Señal fonocardiográfica
RR	Intervalo que determina la frecuencia cardíaca
AV	Auriculoventricular
P, QRS, ST, T	Ondas del ECG
IAM	Infarto agudo de miocardio
$S1$, $S2$	Sonidos cardíacos
STFT	Transformada de Fourier de tiempo corto
WT	Transformada wavelet
PCA	Análisis de Componentes Principales
ICA	Análisis de Componentes Independientes
MDS	Escalamiento multidimensional
FA	Análisis factorial
ANOVA – MANOVA	Análisis de varianza univariada – multivariada
AMI	Información mutua promedio
GA	Algoritmos genéticos

Resumen

La detección automática de patologías cardíacas depende fuertemente de la calidad de representación de los fenómenos fisiológicos subyacentes, y esto ha motivado la realización de estudios para encontrar la representación efectiva de los datos. La representación de la dinámica del corazón está relacionada con las propiedades temporales, morfológicas, espectrales y de complejidad de las señales cardíacas. Como el comportamiento cardíaco posee alta variabilidad intra-clase entonces la tarea de encontrar un espacio de representación efectivo es muy compleja, y desafortunadamente, muchas de las características estimadas son irrelevantes al concepto de clasificación. En este estudio, se presenta un nuevo método de entrenamiento para obtener la representación efectiva de datos involucrados en la dinámica cardíaca, con el fin de desarrollar un sistema de detección automática de isquemia y deficiencias valvulares sobre registros de la actividad cardíaca. Se expone una comparación de diferentes modelos de representación basados en medidas de diagnóstico y coeficientes de representación abstracta. Se presenta un método de reducción de dimensiones basado en un esquema nuevo compuesto por tres niveles de aplicación (proyección, interpretación y visualización), el cual ayuda a encontrar la interpretación del espacio inicial de entrenamiento, indicando cuáles características son las que realmente aportan a la detección y también conecta esas características con la representación fisiológica de las señales cardíacas (electrocardiograma – ECG y fonocardiograma – FCG). Varias bases de datos fueron usadas para evaluar los modelos propuestos de representación, encontrando que el análisis de complejidad permitió la conformación del espacio de representación más discriminante. También fueron evaluadas diferentes esquemas metodológicos de reducción de dimensiones, obteniendo resultados del 99.5 % en la detección de eventos isquémicos y del 97.8 % para la detección de soplos cardíacos (clasificador 5-NN – validación cruzada 15×2). Para incrementar la capacidad de generalización en la etapa de clasificación fue usado el modelo de complejidad de Rademacher y comparado con técnicas convencionales de entrenamiento, resultando que la penalización de Rademacher suma capacidad de generalización al clasificador, lo cual es una restricción necesaria debida a la alta variabilidad intra-clase de las señales cardíacas. Esta componente de incertidumbre incluida en la reducción de dimensiones permite hallar una representación efectiva de los datos.

Abstract

The automatic detection of cardiac pathologies strongly depends on the representation quality of subjacent physiological phenomena, and this has been the motivation behind several studies to seek effective data representation. Representation of heart dynamics is mostly related to timing, morphology, spectral and complexity properties of cardiac signals. It is important to note that because cardiac behavior has high within-class variability then the task of finding an effective representation space is very complex, and unfortunately, many of the candidate features are irrelevant to the target concept. In this study, a new training method for obtaining the effective data representation involved in cardiac dynamics is presented, with the aim of achieving an automatic detection system of ischemia and valvular pathologies on cardiac signals. A comparison of different representation models based on clinical measures and abstract coefficients was carried out. A dimensionality reduction method based on a new scheme founded in three applicability layers (projection, interpretation and visualization) is presented, which helps to find interpretation of the initial space, locating the features that increase the detection capability of cardiac pathologies and also connecting these features with the physiologic representation of the cardiac signals (electrocardiogram – ECG and phonocardiogram – PCG). Several databases were used for evaluating the proposed representation models, finding that complexity analysis made the representation space more discriminant. Also, different dimensionality reduction methodologies were evaluated obtaining results over 99.5 % of classification accuracy for detecting ischemia events and 97.8 % for episodes with cardiac murmurs, using a 5-NN classifier and 15×2 cross-validation. For increasing the generalization capability in the classification stage, the Rademacher complexity model was used and compared with conventional training techniques, as a result that the Rademacher penalty added generalization capability to the classifier, which is a necessary constraint due to the high within-class variability of cardiac signals. This uncertainty included in the dimensionality reduction procedure allowed effective data representation.

Introducción

El *entrenamiento* de sistemas automáticos usados en la detección de patologías a partir de registros de la actividad cardíaca, lleva consigo el problema de estimar y seleccionar de manera efectiva un subconjunto de variables que contengan la información suficiente contenida en la señal para su adecuada representación. Aunque los sistemas modernos de cómputo, debido a su alto rendimiento, permiten analizar cantidades grandes de datos en espacios de dimensiones altas, cuando el número de características se incrementa, resultan inconvenientes que requieren ser tratados para asegurar el buen rendimiento del sistema automático [10]. El aumento en el número de dimensiones del espacio de características es originado, entre otras cosas, por las exigencias de *calidad* y *precisión* en la representación de los estados funcionales, teniendo en cuenta que la complejidad crece exponencialmente a como se adicionan dimensiones; de acuerdo a esto, el problema se relaciona con el aumento significativo del tiempo de ejecución de los algoritmos y el exagerado costo en los estudios relacionados a la detección automática de isquemia y deficiencias valvulares.

Pertinencia de la tesis

La realización de un modelo automático de asesoría en línea, basado en la señales electrocardiográficas y fonocardiográficas, genera un impacto real en el servicio médico, al establecerse el análisis objetivo, evaluación e interpretación de los registros. Desde esta perspectiva, se hacen necesarios mecanismos, modelos e instrumentos que puedan potenciar la precisión de las observaciones en las señales propias del cuerpo humano, así mismo que permitan almacenarse para luego poder tomar de forma objetiva y colaborativa mejores decisiones clínicas. Actualmente, existen diferentes dispositivos y sistemas que ayudan en el estudio de las disfunciones cardíacas, brindando una visión confiable para el diagnóstico de dichas patologías. Sin embargo, en todos ellos el análisis de factores de influencia externos (estado funcional de otros órganos, hábitos del paciente, factores congénitos, etc.) no se lleva a cabo, limitando la efectividad de los métodos de prevención, sumado todo ello a los costos de adquisición considerablemente altos de esta tecnología.

La *generación de características*, orientado a la detección de patologías en señales biomédicas, consiste en calcular medidas relacionadas con los procedimientos que son requeridos

para el diagnóstico médico. De esta forma, la caracterización genera una representación compuesta por atributos con utilidad fisiológica y clínica [34]. Sin embargo, el procesamiento de señales ofrece otras alternativas de representación mediante la estimación de coeficientes o índices matemáticos que aunque no son medidas con sentido clínico, pueden brindar de manera consistente la separabilidad entre patrones correspondientes a distintas clases. La integración del conjunto de características, con y sin sentido clínico, conforma el espacio completo de representación. Es importante resaltar, que el aumento del número de características de representación no tiene una relación directamente proporcional con el buen desempeño de un *sistema de reconocimiento* automático [43].

La detección de patologías, mediante el procesamiento digital de *bioseñales*, involucra el *entrenamiento* de los sistemas como una de las etapas fundamentales del reconocimiento de patrones, que puede ser mejorado si se realiza una extracción y selección de características adecuada. En la práctica, los procedimientos de entrenamiento requieren de un alto número de características (i.e., una alta dimensión del espacio inicial de entrenamiento) que conduce a un tiempo de procesamiento significativo, alto costo en la recolección de las señales biomédicas, y el conocido fenómeno de la maldición de la dimensión [6]. Como resultado, el rendimiento del sistema de entrenamiento decrece. En este sentido, reducir el número de dimensiones del espacio de representación, a través de un método basado en la *extracción* y/o *selección de características*, debe tomarse en cuenta, en orden de distinguir aquellas variables que ofrezcan mayor capacidad discriminante, sin afectar negativamente la precisión de clasificación [16]. En muchas aplicaciones reales, reducir la dimensión de representación es un paso esencial antes de cualquier análisis que pueda ser desarrollado. El criterio general para reducir la dimensión es preservar al máximo la información relevante de los datos originales, de acuerdo a algún criterio de optimización [9].

En este trabajo, se analizan estrategias basadas en la combinación y selección de características, con el fin de explotar lo mejor de cada una de ellas, además de la sinergia potencial. De otro modo, se propone la búsqueda de un subconjunto de variables tan reducido como sea posible, incluso compuesto por solo una característica dotada de sensibilidad y robustez suficiente para los propósitos de reconocimiento. En este caso, el inconveniente consiste en cómo hallar ese espacio reducido de representación o cómo construirlo con base a los conceptos nuevos de procesamiento de información (representación efectiva) y una profunda comprensión de la fisiología involucrada. De acuerdo a esto, el estudio se concentra en la representación efectiva de clases particulares con fuentes de información específicas. Por lo tanto, se desarrolla un método basado en técnicas de representación de datos multidimensionales para reducir la dimensión del espacio inicial de características, como también, rutinas que llevan a cabo la transformación mediante un mapeo lineal o no lineal de los datos a espacios con dimensión mucho menor, considerando funciones de evaluación que involucran algún concepto de relevancia asociada a una medida (estadística, geométrica o de información). De esta forma, se logran descartar las variables que no ofrecen un valor adecuado de representación o que son combinación lineal de las otras, con el objetivo de hacer más preciso y eficiente el *sistema de detección*, sumando a ello, la disminución en la complejidad computacional requerida para el entrenamiento y posterior ejecución del sistema.

La tesis fue realizada entre los años 2004 y 2007 en el Grupo de Control y Procesamiento Digital de Señales de la Universidad Nacional de Colombia sede Manizales. Algunos desa-

rrollos experimentales fueron realizados de forma parcial durante la pasantía en 5 grupos de investigación en España: el *Centro de Investigaciones en Ingeniería Biomédica* de la Universidad Politécnica de Cataluña, el *Grupo de Informática Industrial, Comunicaciones y Automática* en la Universidad Politécnica de Valencia, el *Instituto de Investigación en Ingeniería de Aragón* en la Universidad de Zaragoza, el *Grupo de Investigación en Algoritmos y Arquitecturas para Procesamiento de Señales Avanzado* de la Universidad de Granada y el *Grupo de Procesado de Señales Biológicas* de la Universidad de Las Palmas de Gran Canaria. Es importante anotar que el desarrollo de este trabajo tuvo el apoyo financiero de Colciencias y la Universidad Nacional de Colombia.

Antecedentes

El Grupo de Control y Procesamiento Digital de Señales (GC&PDS) adscrito al Departamento de Ingeniería Eléctrica, Electrónica y Computación de la Universidad Nacional Sede Manizales y registrado ante Colciencias, en conjunto con el Hospital de Caldas y el Hospital Santa Sofía de la ciudad de Manizales, han desarrollado una serie de trabajos relacionados con los métodos de análisis de la actividad biológica, que incluye el desarrollo de sensores de signos vitales (ECG clínico de 12 derivaciones, monitoreo cardíaco por radio, etc.), el proceso de bioseñales (reducción de perturbaciones, segmentación, caracterización y reconocimiento automático de señales ECG, reconocimiento de patologías de emisión vocal, etc.), e imágenes médicas (caracterización y reconocimiento de células cancerosas, caracterización biométrica de pacientes infantiles, etc). En el área de procesamiento digital de señales cardíacas, para el mes de Octubre de 2004 se finalizó el proyecto de investigación titulado *"Análisis Automatizado de Señales Cardíacas"*, en Agosto de 2005 se entregó a la Vicerrectoría de Investigaciones de la Universidad de Caldas el proyecto titulado *"Desarrollo de Aplicativos de Proceso de Bio-señales para Servicios de Telemedicina en el Departamento de Caldas"* y en Septiembre de 2007 se entregó a Colciencias el proyecto titulado *"Auscultación y Registro Electrocardiográfico sobre la Web para Apoyo a la Teleconsulta Médica"*. Actualmente se está desarrollando un proyecto financiado por la Universidad Nacional de Colombia titulado *"Técnicas de computación de alto rendimiento en la interpretación automatizada de imágenes médicas y bioseñales"*.

Justificación

Para el caso de Manizales, la mortalidad, por enfermedades cardíacas, ocupa el primer lugar con una tasa de 156,4 x 100.000 habitantes, muy por encima de la ocasionada por la violencia. En el año 2001 el Hospital de Caldas internó 215 pacientes por enfermedades isquémicas. Cerca del 50 % de pacientes fallecen antes de su ingreso al hospital, mientras, la mortalidad hospitalaria está alrededor del 15 % [11]. En los centros de atención primaria del departamento de Caldas, los médicos realizan la percepción de normalidad y anormalidad de los *sistemas orgánicos* a través de instrumentos básicos (fonendoscopio, tensiómetro), cuya interpretación es valorada y analizada subjetivamente, lo cual puede condicionar a un mayor riesgo de error. Frecuentemente el médico se ve enfrentado a múltiples limitaciones: la falta de habilidad y destreza para identificar plenamente la verdadera *calidad* de un sonido en cuanto a tono y timbre, la experiencia requerida para familiarizarse con los sonidos normales y anormales emitidos por el organismo y también la falta de instrumentos de alta calidad y precisión. La realización de un modelo automático de asesoría en línea, basado en

la señales electrocardiográficas y fonocardiográficas, genera un impacto real en el servicio médico, al establecerse un análisis más objetivo para la evaluación e interpretación de los registros. En la práctica médica, se han establecido las juntas de decisiones con el fin de compartir, analizar e interpretar lo percibido desde varios puntos de vista, las cuales en la mayoría de las veces están fundamentadas en la percepción de una sola persona. Desde esta manera, se hacen necesarios mecanismos, modelos e instrumentos que puedan potenciar la precisión de las observaciones en las señales propias del cuerpo humano, como también que permitan almacenarse para luego poder tomar de forma objetiva y colaborativa mejores decisiones clínicas. Actualmente, existen diferentes dispositivos y sistemas que ayudan en el estudio de las disfunciones cardíacas, brindando una visión confiable para el diagnóstico de dichas patologías. Sin embargo, en todos ellos el análisis de factores de influencia externa (estado funcional de otros órganos, hábitos del paciente, factores congénitos, etc.) no se lleva a cabo, limitando la efectividad de los métodos de prevención, sumado a ello los altos costos de adquisición de esta tecnología.

Planteamiento del problema

Las técnicas de *análisis multivariado* han sido ampliamente aplicadas en la reducción de dimensiones, particularmente, el análisis de componente principales (PCA), entendida como una representación ortogonal de los datos que permite encontrar un subespacio de menor dimensión basado en la varianza, pero este procedimiento no incluye información acerca de la separabilidad de las clases, así, PCA no es la técnica apropiada cuando la dirección de máxima varianza no corresponde a la dirección de máxima separabilidad [24]. Es importante considerar también, que PCA puede fallar en la preservación de la estructura local si los datos residen en alguna variedad no lineal, lo cual es irremediable en muchas aplicaciones reales [5]. Por otro lado, el análisis de separabilidad entre subconjuntos de características puede ser desarrollado a través del análisis multivariado de la varianza (MANOVA) o múltiples análisis de varianza univariada (ANOVA); sin embargo, ninguna de estas técnicas toman en cuenta la dependencia estadística que puede existir entre las características. De acuerdo a esto, el funcionamiento conjunto de PCA y ANOVA puede ser tomado en cuenta aunque el costo de cómputo tiende a ser elevado [27]. Como resultado, se tiene que tanto las técnicas univariadas como las multivariadas tienen fuertes restricciones en la reducción de espacios con alta dimensión.

La selección de características, como un procedimiento iterativo, puede ser desarrollada mediante algoritmos de búsqueda completa o heurística [10]. Los de búsqueda *completa* aseguran encontrar el subconjunto óptimo de características respecto a una función de evaluación, después de realizar un procedimiento exhaustivo sobre todo el espacio inicial de entrenamiento. De otra forma, los algoritmos de búsqueda *heurística* encuentran subconjuntos sub-óptimos respecto a una función de evaluación a través de criterios de búsqueda usados para reducir la complejidad computational, sin disminuir el rendimiento del sistema; estos métodos requieren de una condición de parada para prevenir que la búsqueda de subconjuntos se vuelva exhaustiva. A menudo, la implementación de técnicas de búsqueda heurística están basadas en métodos estadísticos. En [31], se muestra la bondad de un subconjunto de características usando como función de evaluación la precisión de clasificación; donde los métodos de búsqueda flotante secuencial (heurístico), son adecuados para espacios iniciales de entrenamiento de pequeña y mediana escala.

El entrenamiento de sistemas inteligentes para la detección de estados funcionales en aplicaciones médicas no es un tema nuevo en la literatura, en [8] se presenta la revisión de diferentes procedimientos para el reconocimiento de patrones sobre bioseñales y sistemas de interpretación. Este estudio reporta la viabilidad de desarrollar técnicas de clasificación de patologías siempre y cuando la caracterización de los registros dependan fuertemente de la información previa que se tenga del fenómeno fisiológico, además del conocimiento a priori que se tenga de las distribuciones de probabilidad de las clases a identificar. En [26] se resumen y comparan un conjunto amplio de métodos bien conocidos para las varias etapas involucradas en el reconocimiento de patrones. Este estudio muestra que el entrenamiento de *sistemas inteligentes* es altamente cambiante y proliferante, por lo cual es posible que un problema pueda ser solucionado por un conjunto diferentes de procedimientos matemáticos, estadísticos y de lógica simbólica. En [30] se discute acerca de las diferentes formas en que se puede conseguir algún nivel de comprensión de las estructuras complejas que se generan en los espacios de alta dimensión, en particular, las facultades que tiene la representación visual en la interpretación de dinámicas ocultas del fenómeno fisiológico subyacente. El inconveniente radica en que la visualización de espacios de representación requiere una reducción de dimensiones fuertemente restringida y es difícil definir el direccionamiento del mapeo de reducción sin que se pierda información relevante respecto a la dinámica y variabilidad fisiológica.

En general, un número limitado de características simplifica la representación tanto del patrón como del clasificador, resultando una extracción y análisis menos denso y extenuante, al igual que un clasificador más rápido con mínimos requerimientos de memoria. A este *sistema de representación*, se le conoce con el nombre de *representación efectiva*. Sin embargo, una reducción exagerada en el número de características podría llevar a una pérdida en el poder discriminante, deteriorando la precisión del *sistema de reconocimiento*. Cuando se posee un hiperespacio de análisis en el cual el agrupamiento de clases no se presenta en forma de distribución gaussiana, sino que se presentan conjuntos con disposiciones de separabilidad no lineal, las medidas de distancia y dependencia lineal ya no son efectivas, y se hace necesario hacer el análisis basado en herramientas más complejas. Por lo tanto, la representación efectiva requiere de modelos de representación de naturaleza consistente, discriminante y no lineal, además de un esquema de reducción de espacios representativos que incluya una medida de relevancia para preservar la estructura discriminante y representativa de los datos, de manera que la efectividad en la detección de eventos no se atribuya exclusivamente al clasificador. Así, esta tesis doctoral busca establecer un método de entrenamiento que permita a un conjunto de objetos embeberse en un espacio de representación dotado de alta capacidad de interpretación del fenómeno fisiológico subyacente, mínima complejidad computacional y máxima capacidad de generalización para la evaluación de nuevos ejemplos.

Objetivos de la tesis

Objetivo general

Desarrollar un método de entrenamiento que mejore la tasa de acierto con un baja complejidad computacional en la identificación automática de *isquemia* y *deficiencias valvulares*, realizada sobre el estudio de los registros de comportamiento eléctrico y acústico

del corazón. El método de entrenamiento incluye el estudio y la generación de parámetros para la extracción de información de las clases a identificar, así como el desarrollo de un esquema metodológico de reducción de dimensiones basado en el análisis de relevancia sobre estructuras multivariadas, asociando una función de evaluación y alguna medida de representación.

Objetivos específicos

1. Analizar y comparar las diferentes técnicas de extracción de medidas de representación sobre señales electrocardiográficas y fonocardiográficas, que correspondan tanto a medidas de diagnóstico como parámetros de representación abstracta, estimadas mediante el análisis espectral y técnicas de dinámica no lineal, las cuales se orienten a una precisión adecuada de clasificación de los estados de normalidad y anormalidad cardíaca, para los casos patológicos de *isquemia* y *deficiencias valvulares*.

2. Desarrollar un esquema basado en medidas de relevancia, considerando las diferentes métricas de representación (en particular, geométrica, estadística y de información), que permita evaluar la capacidad representativa de las diferentes características de la dinámica cardíaca obtenida mediante registros de las señales electrocardiográficas y fonocardiográficas en la identificación de los casos patológicos de *isquemia* y *deficiencias valvulares*.

3. Desarrollar un método para reducir la dimensión de los espacios iniciales de entrenamiento correspondiente a las diferentes características de la dinámica cardíaca (señales electrocardiográficas y fonocardiográficas), que sobre las medidas propuestas de relevancia, disminuya significativamente el costo computacional para un rendimiento de clasificación dado, durante el entrenamiento de los sistemas de identificación automatizada de los casos patológicos de *isquemia* y *deficiencias valvulares*.

4. Desarrollar un método de selección efectiva de características, correspondientes a los diferentes parámetros de diagnóstico de la dinámica cardíaca, que brinde la interpretación y representación visual adecuada de los espacios solución necesaria en el diagnóstico asistido de los casos de patología de *isquemia* y *deficiencias valvulares*.

5. Evaluar la capacidad de representación de las variables de salida del modelo desarrollado mediante la precisión de clasificadores basados en algún principio discriminante.

Organización del manuscrito

El manuscrito de esta tesis inicia en el Capítulo 1 con los preliminares fisiológicos de las señales cardíacas, como también los principales requerimientos que incurren en los *sistemas automáticos* de identificación de patologías sobre registros que contienen la información clínica del *sistema eléctrico* y *mecánico* del corazón. En el Capítulo 2, se presentan las formas en que pueden ser estimadas las medidas de representación sobre los segmentos del registro, con el fin de transformar la información clínica en un ensamble multivariado de características a través de diversos modelos correspondientes al procesamiento digital de señales. En el Capítulo 3, se presenta una discusión acerca del modelo general de *sistemas*

gaussianos, con el fin de plantear de manera formal el problema de reducción de dimensiones y a partir del problema generalizado, orientarlo a diferentes aspectos: disminuir la complejidad en la representación, mejorar los resultados de clasificación, remover la información redundante e irrelevante y descubrir la estructura esencial de los datos para obtener una representación gráfica; todo esto basado en el denominado *análisis de relevancia*, el cual se encarga de dirigir las proyecciones en el espacio de representación. En el Capítulo 4, se presenta el marco experimental del trabajo, de forma que se exponen los modelos propuestos para el tratamiento digital de las señales ECG y FCG, los procedimientos de reducción de dimensiones de acuerdo al esquema propuesto y la detección de patrones funcionales relacionados a eventos isquémicos y deficiencias valvulares. Finalmente, en los Capítulos 5 y 6 se presentan tanto los resultados y la discusión como las conclusiones finales del trabajo.

CAPÍTULO 1

Registro de la actividad cardíaca y diagnóstico asistido

Da ubi consistam, et terram caelumque movebo.

(Arquímedes)

En el presente capítulo, se analizan los preliminares fisiológicos de las bioseñales, así como los principales requerimientos en los *sistemas automáticos* de identificación de patologías sobre registros que contienen la información discreta de las bioseñales.

Definición 1.1

Una *bioseñal*, a veces también denominada *señal biomédica*, corresponde al término que incluye a toda clase de variables, que pueden ser medidas, monitoreadas y además reflejan un estado biológico funcional. Típicamente, las variables son traducidas a señales eléctricas, entre otras razones, porque éstas son relativamente simples de procesar por medio de equipos electrónicos con todas sus ventajas conocidas: confiabilidad, flexibilidad de análisis, precisión, bajo costo energético, etc.

La bioseñal se toma a partir de corrientes producidas por cambios en los *potenciales eléctricos* de algún tejido, órgano o *sistema celular*. Entre las bioseñales más comúnmente analizadas, están: el electroencefalograma, el magnetoencefalograma, el electrocardiograma, el miograma, y similares. De otra parte, las bioseñales también se refieren a variables con naturaleza no eléctrica que pueden ser medidas sobre un *sistema biológico*, por ejemplo, las señales acústicas (voz, respiración, fonocardiograma, ruidos pulmonares), los movimientos oculares, las posturas faciales, etc.

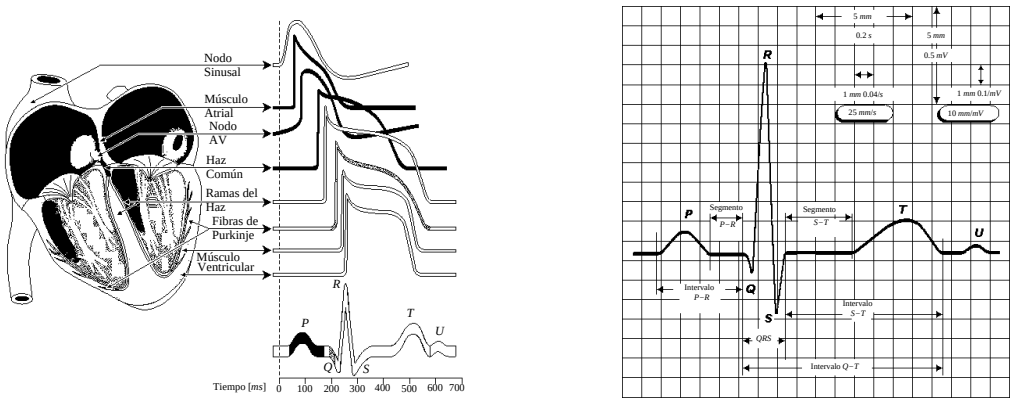
1.1. Señales electrocardiográficas

El estudio de la señal electrocardiográfica (ECG) se basa en el análisis del sistema de producción y conducción del estímulo *eléctrico* cardíaco, capturado mediante electrodos de

superficie, siendo una de las pruebas médicas más frecuentes en la exploración de la actividad cardíaca para el diagnóstico de diversas formas de arritmia, defectos de conducción, infartos, hipertrofias y otras anomalías [2].

1.1.1. Morfología básica

La actividad eléctrica del miocardio es recogida en un registro eléctrico (señal ECG), como se observa en la Figura 1.1(a). Una señal ECG normal sigue un trazo similar al mostrado en la Figura 1.1(b). Desde el lado interno de la pared ventricular, los diversos sitios de activación causan la formación de un frente de onda, el cual se propaga a través de la masa ventricular hacia la pared externa del músculo cardíaco, de forma que la activación es realizada célula a célula. Después de que cada región ventricular se ha despolarizado, ocurre la repolarización, la cual no es un fenómeno de propagación. Debido a que la duración de la acción del impulso es mucho más corta en el epicardio (la parte más externa del miocardio) que en el endocardio (la parte interna del miocardio), la actividad eléctrica aparece como si se propagara del epicardio hacia el endocardio [21].



(a) Origen de la actividad eléctrica

(b) Prototipo de registro normal de ECG.

Figura 1.1. Electrofisiología del corazón

Debido a que la frecuencia intrínseca del nodo sinusal es mayor que la del nodo auriculoventricular (AV), su valor establece la frecuencia de activación de todo el corazón. Si la conexión entre la aurícula y el nodo AV falla, el nodo AV adopta su frecuencia intrínseca como frecuencia de activación. Si el sistema de conducción falla en el haz de His, los ventrículos se contraen a la mayor frecuencia intrínseca de la región [28].

Los eventos *eléctricos* involucrados en la dinámica cardíaca se resumen en la Tabla 1.1. El registro ECG normal está constituido por la onda P, el complejo QRS y la onda T. El complejo QRS, en realidad incluye tres ondas separadas, la Q, la R y la S; todas ellas causadas por el paso del impulso cardíaco a través de los ventrículos. En el ECG las ondas Q y S suelen ser mucho menos prominentes que la onda P e incluso a veces pueden faltar. La onda P depende de corrientes eléctricas generadas, cuando las aurículas se despolarizan antes de la contracción, y el complejo QRS, que es producido por corrientes

Localización	*Evento*	*Tiempo* $[ms]$	*ECG*[a]	*Velocidad* $[m/s]$[b]	*Frecuencia* $[1/min]$
Nodo sinusal	Gen. del impulso	0		0.05	$70-80$
Aurícula derecha	Despolarización	5	P	$0.8-1.0$	
Aurícula izquierda	Despolarización	85	P	$0.8-1.0$	
Nodo AV	Llegada impulso	50	PQ	$0.02-0.05$	
Nodo AV	Salida impulso	125	PQ		
Haz de His	Activado	130		$1.0-1.5$	$20-40$
Ramas del haz	Activado	145		$1.0-1.5$	$20-40$
Fibras de purkinje	Activado	150		$3.0-3.5$	$20-40$
Endocardio					
Septo	Despolarización	175	QRS	0.3 (axial)	$20-40$
Ventrículo izquierdo	Despolarización	190	QRS	–	$20-40$
Epicardio					
Ventrículo izquierdo	Despolarización	225	QRS	0.8	$20-40$
Ventrículo derecho	Despolarización	250	QRS	(transv.)	$20-40$
Epicardio					
Ventrículo izquierdo	Repolarización	400			
Ventrículo derecho	Repolarización				
Endocardio					
Ventrículo izquierdo	Repolarización	600	T	0.5	

Tabla 1.1. Eventos eléctricos en el corazón.[a]Terminología en lectura de ECG. [b]Velocidad de conducción

nacidas cuando los ventrículos se despolarizan antes de contraerse. Por lo tanto, la onda P como los componentes del complejo QRS corresponden a las ondas de despolarización. La onda T, que es causada por las corrientes nacidas cuando los ventrículos se recuperan del estado de despolarización, se conoce como la onda de repolarización. En esencia, el ECG está formado por las ondas de despolarización y repolarización (ver Figura 1.1(a)). Los complejos QRS, que están presentes en la mayoría de los latidos, se asocian con la actividad eléctrica ventricular y contienen información clínica importante, por lo que su relación señal a ruido es la más alta de todas las ondas presentes en la señal electrocardiográfica [21].

1.1.2. Derivaciones

Las derivaciones proporcionan una vista particular de la actividad eléctrica del corazón entre dos puntos, además, ayuda a precisar el sitio anatómico y el grado de alteración de alguna patología. Por cuanto el corazón es un órgano tridimensional, y dado que una imagen tridimensional no es registrable en una superficie plana, es necesario proyectar las fuerzas eléctricas cardíacas sobre dos planos: frontal y horizontal, con el fin de conocer si dichas fuerzas se dirigen de arriba-abajo, derecha-izquierda, delante-atrás. Para ello es suficiente que se registren, mediante diferentes derivaciones, las componentes eléctricas cardíacas proyectadas sobre el plano frontal (arriba-abajo), y sobre el plano horizontal (derecha-izquierda). De acuerdo a la ubicación de los electrodos de registro se detecta el

flujo de corriente de despolarización y repolarización de la estructura tridimensional del miocardio. De manera que para obtener una imagen completa y correcta de la forma en que se transmite la excitación eléctrica en el corazón, se requiere información de electrodos situados en distintas posiciones, tomando las siguientes consideraciones: si la corriente eléctrica se acerca al electrodo aparece una onda con polaridad positiva, y si se aleja, una onda con polaridad negativa [2]. La disposición específica de los electrodos recibe el nombre de *derivación*. La forma de onda en cada derivación varía debido a que se registra la actividad cardíaca desde diferentes ángulos. Se conocen más de 40 derivaciones distintas, pero habitualmente sólo 12 son las más utilizadas. Entre las derivaciones del plano frontal se encuentran, las bipolares y las aumentadas; y del plano vertical, están las precordiales. Cabe anotar que en cada derivación se obtiene la actividad *eléctrica* del mismo *sistema cardíaco*.

– *Bipolares*. Las derivaciones estándar bipolares (I, II y III) son las derivaciones elegidas originalmente por Einthoven para registrar los potenciales eléctricos en el plano frontal del cuerpo [20]. Como se observa en la Figura 1.2, se colocan electrodos en el brazo izquierdo (LA), brazo derecho (RA) y pierna izquierda (LL). La pierna derecha funciona como tierra y no realiza ninguna función en la producción del trazo ECG. Las derivaciones bipolares muestran diferencias de potencial entre dos electrodos seleccionados, así: la derivación I es la diferencia de potencial entre el brazo izquierdo y el brazo derecho (LA-RA); la derivación II es la diferencia de potencial entre la pierna izquierda y el brazo derecho (LL-RA); y la derivación III es la diferencia de potencial entre la pierna izquierda y el brazo izquierdo. La relación entre las derivaciones se expresa algebraicamente por la ecuación de Einthoven: DII = DI + DIII.

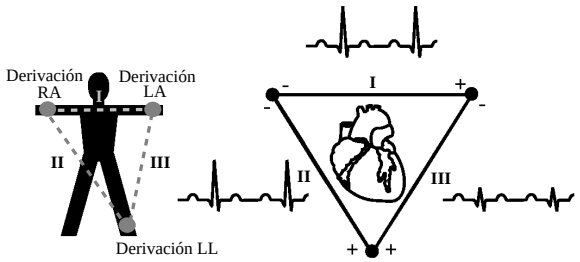


Figura 1.2. Derivaciones bipolares y el triángulo de Einthoven

– *Aumentadas*. Las derivaciones bipolares de las extremidades tienen el inconveniente de que registran sólo diferencias de potencial eléctrico y no el potencial real neto en un punto de la superficie del cuerpo. Para separar las derivaciones bipolares en sus dos componentes, se emplean las derivaciones unipolares (VR, VL y VF). En estas derivaciones no se coloca el electrodo positivo en un miembro y el negativo en otro como en el caso anterior, sino que se coloca el electrodo positivo en uno de los miembros y se compara contra la sumatoria de los otros dos miembros conectados al

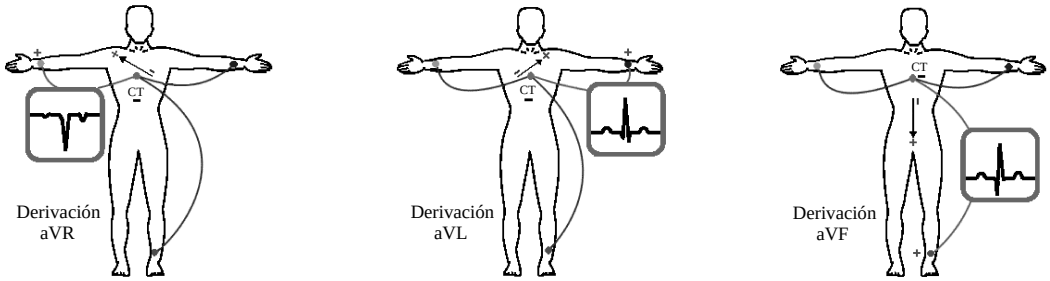


Figura 1.3. Derivaciones unipolares de las extremidades

polo negativo (central terminal – CT). En estas condiciones el electrodo explorador funciona como unipolar, y se conecta al brazo derecho (R), al brazo izquierdo (L) o a la pierna izquierda (F), registrando respectivamente las derivaciones VR, VL y VF, como se observa en la Figura 1.3. La mayoría de los electrocardiógrafos están construidos de manera que los voltajes obtenidos por medio de las derivaciones unipolares experimentan una ganancia en magnitud de 3/2. En estos casos los registros de ECG usan el prefijo *a* (amplificada), así que las derivaciones son llamadas unipolares aumentadas y se designan como aVR, aVL y aVF.

– *Precordiales*. Representan la actividad eléctrica del corazón desde un plano horizontal (anterior-posterior) para complementar la información registrada del plano frontal. Estas derivaciones son particularmente útiles cuando los vectores cardíacos son perpendiculares al plano frontal, ya que la proyección de un vector perpendicular al plano es igual a cero. Estas derivaciones comparan la actividad del punto en que se coloca el electrodo a nivel precordial (electrodo explorador) contra la suma de los tres miembros activos que se reúnen en una central terminal de *Wilson* (CTw). El CTw une mediante resistencias los tres miembros activos, de forma que su potencial combinado se puede considerar cero, por lo que el electrodo negativo es insensible a las variaciones del potencial del músculo cardíaco y las diferencias de potencial entre el positivo y el corazón son registradas. Las derivaciones del plano horizontal que se utilizan en electrocardiografía clínica normalmente son seis ($V1$ a $V6$), cuya disposición se muestra en la Figura 1.4. Así, las derivaciones precordiales se proyectan desde el electrodo positivo a través del nodo AV, hacia la espalda del paciente (en un plano horizontal). De otra parte, existen otras derivaciones que son variantes de las anteriores pero menos conocidas, las cuales corresponden a $V7$, $V8$ y $V9$, encargadas de informar acerca de la parte posterior del corazón, además de las R$V1$ y R$V2$, que se ubican frente al ventrículo derecho.

Las Tablas 1.2 y 1.3 muestran las principales derivaciones, con la posición de los electrodos, la nomenclatura utilizada, y la polaridad de las ondas, respectivamente.

1.1.3. Descripción de las componentes

Las distancias entre las ondas que componen el ECG se denominan *segmentos* o *intervalos*. Un período del ECG, perteneciente a un individuo sano, consta de las ondas: P, el complejo

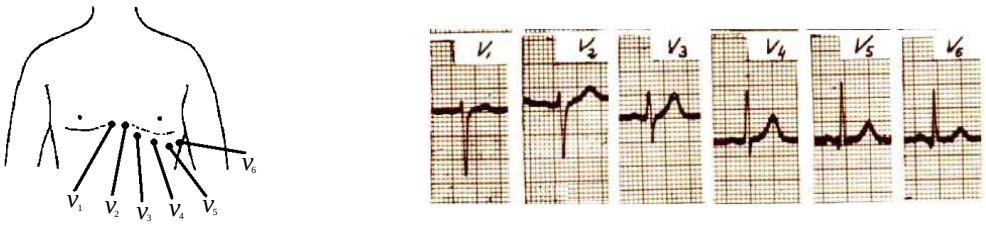


Figura 1.4. Derivaciones unipolares precordiales y sus registros típicos

Derivación	Posición electrodos	Nomenclatura
Bipolar	LL, RA, LA	I=LA-RA II=LL-RA III=LL-LA
Aumentadas	LL, RA, LA	aVR=RA-0.5(LA+LL) aVL=LA-0.5(LL+RA) aVF=LL-0.5(LA+RA)
Precordiales	ν_1, ν_2, ν_3, ν_4, ν_5, ν_6	$V_1 = \nu_1 - \frac{LA+RA+LL}{3}$ $V_2 = \nu_2 - \frac{LA+RA+LL}{3}$ $V_3 = \nu_3 - \frac{LA+RA+LL}{3}$ $V_4 = \nu_4 - \frac{LA+RA+LL}{3}$ $V_5 = \nu_5 - \frac{LA+RA+LL}{3}$ $V_6 = \nu_6 - \frac{LA+RA+LL}{3}$

Tabla 1.2. Descripción de las 12 derivaciones del ECG

Derivación	Onda P	Complejo QRS	Onda T
I	+	+	+
II	+	+	+
III	+/-	+/-	+/-
aVR	-	-	-
aVL	+/-	+/-	+
aVF	+	+	+
$V1$	+/-	-	-
$V2$	+/-	+/-	+/-
$V3$	+	+	+
$V4$	+	+	+
$V5$	+	+	+
$V6$	+	+	+

Tabla 1.3. Derivaciones y polaridad de las ondas

QRS, el segmento ST, la onda T y la onda U, tal como se muestra en la Figura 1.1(b).

Onda *P*. Representa la despolarización de la aurícula. Su duración normal es de 0.1 s y tiene una amplitud menor a 0.25 mV. La forma de onda depende de la derivación en que se registre. Un aumento del voltaje o de la duración de esta onda indica una anomalía auricular. La ausencia de esta onda ocurre en una parada del nodo sinusal, y en el bloqueo sinoauricular (situación en la que sí se despolariza el nodo sinusal, pero no se transmite el

impulso a las células contiguas de la aurícula).

Complejo *QRS*. Representa la despolarización de los ventrículos y se presenta en forma de frentes de activación sinusal continuos y rápidos. Está formado por las ondas Q, R y S. Su duración es de aproximadamente 0.1 s y su amplitud se encuentra entre 1 y 3 mV. Durante este evento tiene lugar la repolarización auricular, llamada onda T_a, la cual queda enmascarada por el complejo QRS y, por lo tanto, esta deflexión no suele observarse en la señal ECG de 12 derivaciones. La onda T_a tiene dirección opuesta a la onda P.

Segmento *ST*. Comprende el intervalo de tiempo desde el retorno del QRS a la línea base hasta la primera deflexión, positiva o negativa, de la onda T. Aunque el mismo valor de duración del ST no tiene un evidente significado clínico, su desviación respecto a la línea base, que se mide generalmente a 0.08 s después del final del complejo QRS, sí tiene evidencia clínica.

Onda *T*. Normalmente, el vector medio de la onda T se orienta de forma similar al vector QRS. Como la despolarización y la repolarización son procesos eléctricos opuestos, esta concordancia entre los vectores de las ondas QRS-T indica que la repolarización debe ocurrir en dirección inversa a la despolarización (es decir, desde el epicardio al endocardio o desde la punta del corazón hacia la base). La onda T representa la repolarización de los ventrículos. La forma de onda normal es asimétrica en sus ramas y redondeada en su vértice, de manera que la pendiente de la rama inicial es más suave que la de la rama terminal.

Onda *U*. Es una pequeña deflexión redondeada (≤ 1 mm) que sigue a la onda T, generalmente, con la misma polaridad. El aumento patológico de la amplitud de la onda U suele ocurrir debido a determinados fármacos (p. ej. quinidina, procainamida, disopiramida) o a una hipopotasemia. Las ondas U más prominentes se asocian a una mayor susceptibilidad a la taquicardia ventricular del tipo *torsades de pointes*. La inversión de la onda U en las derivaciones precordiales es patológica y constituye un signo de isquemia [28].

Intervalo *PR*. Se mide a partir de la primera deflexión hacia arriba de la onda P hasta la primera deflexión del complejo QRS, ya sea positiva (R) o negativa (Q). El intervalo PR varía ligeramente de acuerdo a la edad y al ritmo cardíaco, pero para propósitos prácticos, se puede decir que el rango está entre 0.12 y 0.20 s.

Segmento *PR*. Corresponde a la línea isoeléctrica entre el final de la onda P hasta la primera deflexión del complejo QRS ya sea positiva (R) o negativa (Q). Es importante decir que no es frecuente usar esta medida para la distinción de anormalidades, solo es tomada para definir patrones de normalidad.

Intervalo PP. Corresponde al espacio de tiempo entre el comienzo de la onda P y el comienzo de la siguiente onda P.

Intervalo RR. Corresponde al intervalo de tiempo entre la onda R de un complejo QRS y la onda R del siguiente complejo QRS.

Intervalo *QT*. Se mide a partir de la primera deflexión del complejo QRS, positiva (R) o negativa (Q), hasta el retorno de la onda T a la línea isoeléctrica. El intervalo QT se muestra marcadamente afectado por el ritmo cardíaco y sus límites máximos normales se

muestran en la Tabla 1.4, teniendo en cuenta que con ritmos cardíacos normales entre 60 y 100 lpm, los intervalos QT están entre 0.3 y 0.4 s. Además, el máximo intervalo QT es generalmente 10 % más largo en las mujeres que en los hombres. A menos que el ritmo cardíaco sea muy lento, un intervalo QT que exceda 0.4 a 0.44 s es anormalmente prolongado.

Ritmo	*Máxima duración*	
cardíaco [lpm]	*Mujeres* [s]	*Hombres* [s]
300	0.20	0.19
187	0.25	0.23
136	0.29	0.26
107	0.33	0.30
88	0.36	0.33
75	0.39	0.36
68	0.41	0.38
55	0.46	0.42
41	0.53	0.48
30	0.62	0.57

Tabla 1.4. Duración del intervalo QT [17]

1.1.4. Modos de registro

ECG en reposo. Se utiliza con el paciente en reposo y se usa básicamente para la identificación de trastornos primarios de la conducción cardíaca, arritmias, hipertrofia cardíaca, pericarditis, desequilibrios deselectrolíticos, evaluación de marcapasos, diagnóstico y seguimiento del infarto de miocardio con su localización.

ECG de ejercicio. Busca medir los efectos cardiovasculares que se presenta en esfuerzo físico controlado (ejercicio en bicicleta estática o en banda sinfín). Se usa para identificar las arritmias que se desarrollan durante el ejercicio, detectar coronariopatía asintomática, determinar la capacidad funcional del corazón y el origen del dolor torácico.

ECG ambulatorio o Holter. Consiste en un registro continuo de la actividad eléctrica del corazón, por un lapso de tiempo mayor o igual a 24 horas, mientras el paciente realiza sus actividades rutinarias. Se usa para identificar arritmias, evaluar el dolor torácico, valorar el estado del corazón después de un infarto agudo de miocardio, evaluar la disnea, síncope, lipotimias, palpitaciones y otras manifestaciones cardíacas.

ECG de alta resolución. Se utiliza para detectar señales de pequeña amplitud que no pueden ser observados en el electrocardiograma estándar.

1.1.5. Variaciones del ECG debidas a enfermedades isquémicas

Las enfermedades isquémicas se relacionan con cualquier trastorno en los tejidos del corazón causado por restricciones en las arterias coronarias para el suministro de sangre al músculo cardíaco. Estas enfermedades presentan numerosas etiologías, siendo la más frecuente la aterosclerosis coronaria en la que se presenta endurecimiento de las arterias por la formación de la llamada *placa de ateroma*. Los fenómenos que dan inicio a esta placa no están completamente establecidos, pero se relacionan con la penetración y acumulación subendotelial de colesterol, debido a factores como hipertensión arterial, diabetes, tabaquismo, estrés, entre otros. Una de las principales características de la aterosclerosis es la *estenosis luminal*, la cual afecta principalmente las arterias epicárdicas de la siguiente forma: taponamiento de las arterias epicárdicas, reducción del lumen, y limitación de las demandas de riego coronario. A medida que aumenta el grado de estenosis, se observa una progresiva disminución de la reserva coronaria. Las estenosis mayores del 90 % pueden casi anular la capacidad del transporte de oxígeno, desencadenando la aparición de isquemia miocárdica

que tiene efectos metabólicos, mecánicos y eléctricos sobre la zona afectada. Las manifestaciones clínicas de las enfermedades isquémicas se pueden agrupar en dos clases: síndromes coronarios crónicos (incluyen la isquemia silente y la angina de esfuerzo) y síndromes coronarios agudos, los cuales incluyen la angina inestable, la angina variante o de Prinzmetal y el infarto agudo de miocardio (IAM) [28].

Angina de pecho estable o de esfuerzo

Los cambios morfológicos dependen del tipo de ECG que se analiza. Para el ECG en reposo, las anormalidades más comunes consisten en alteraciones no específicas del segmento ST y la onda T, con o sin evidencia de infarto transmural previo; sin embargo, se debe tener en cuenta que también son muy comunes las alteraciones del segmento ST y la onda T en personas con estado funcional normal. En pacientes con enfermedad coronaria, la ocurrencia de alteraciones en el segmento ST y onda T, puede relacionarse con la severidad de la enfermedad de base, incluyendo el número de vasos comprometidos y la presencia de disfunción ventricular izquierda. En contraste, un ECG normal en reposo es un signo de pronóstico favorable. Ahora, el ECG de esfuerzo pretende aumentar el consumo de oxígeno en el miocardio para hacer más evidente la isquemia cardíaca. El cambio electrocardiográfico más específico para determinar que la prueba es positiva, es la depresión horizontal o descendente del segmento ST de más de 0.1 mV (1 mm) y con una duración mayor de 80 ms. La reproducción de la angina con el ejercicio, o una respuesta hipotensiva, refuerzan la confirmación de la prueba [37].

Angina de pecho inestable

Aunque el diagnóstico de angina inestable es clínico y no requiere de anormalidades en el electrocardiograma, es importante realizarlo en los pacientes, tanto para la admisión como en la evaluación durante los episodios de dolor. Con frecuencia hay cambios en el segmento ST y onda T, con los cuales se incrementa la especificidad del diagnóstico de isquemia miocárdica aguda. El trazado debe realizarse mientras el dolor esté presente. El comportamiento del segmento ST es importante en los síndromes isquémicos agudos para determinar las pautas terapéuticas a seguir. En la angina inestable, la mayoría de las veces, se produce un infradesnivel del segmento ST e inversión de la onda T. Estos cambios son transitorios o más acentuados con el dolor. Los cambios en el complejo ST-T durante el dolor torácico, así sea atípico, con normalización de la onda una vez desaparece el dolor, son indicadores de evidencia isquémica. Los desplazamientos del segmento ST asociado con los síntomas identifican la población de más alto riesgo [32].

Angina variante o de Prinzmetal

La clave diagnóstica se basa en la detección del supradesnivel del segmento ST durante el episodio de dolor torácico. Con la administración de dinitratato de isosorbide sublingual o nitroglicerina hay alivio del dolor y normalización del segmento ST. Esta maniobra terapéutica es de gran utilidad para la diferenciación con el infarto agudo de miocardio (IAM). Durante los episodios de dolor se pueden ver cambios alternantes del segmento ST y la onda T, trastornos de la conducción AV y arritmias ventriculares. Todos esos hallazgos se asocian con un riesgo aumentado de muerte súbita para el paciente. Algunos presentan dolor típico y supradesnivel del segmento ST no sólo en reposo, sino durante o después del

ejercicio, reflejando la presencia de lesiones fijas en las arterias coronarias. El espasmo es más común en la arteria coronaria derecha, presentándose una elevación cóncava del ST en DII, DIII y aVF [32].

Infarto agudo de miocardio (IAM)

El trazo ECG es de gran utilidad para localizar el IAM ya que los cambios del ECG varían de acuerdo a si es en la fase evolutiva aguda o si es en la fase estabilizada crónica.

Fase evolutiva aguda. Los cambios pueden ser: hiperagudos, agudos y subagudos. En los cambios hiperagudos, los signos en el ECG más inmediatos se presentan sobre el segmento ST y la onda T. Puede verse un aumento en la magnitud de la onda T, de manera que las ondas son altas y picudas como signo de isquemia subendocárdica. Esto es más frecuente en infarto de la pared anterior. Otro de los signos en el ECG más tempranos es la pérdida de la concavidad normal del segmento ST. El aumento del tiempo de activación ventricular y de la amplitud de las ondas R son signos de retardo de la lesión aguda en la conducción intramural. En cuanto a los cambios agudos, es característica la forma convexa del segmento ST, terminando con una onda T que tiende a ser negativa, aunque en esta fase puede estar involucrada en la parte terminal del segmento ST y en cierta manera arrastrada hacia arriba. Posterior a los cambios descritos del segmento ST y onda T, aparecen los signos indicativos de inactividad eléctrica de las zonas comprometidas: pérdida de amplitud de las zonas R y el desarrollo de ondas Q anormales. Estos cambios pueden ser vistos de forma temprana (a las dos horas después de comenzar el dolor torácico) o tardar hasta doce horas para su aparición. Finalmente, los cambios subagudos relacionados al IAM registran un supradesnivel del ST estable por unos pocos días, retornando a la línea isoeléctrica más tempranamente en los infartos de cara inferior y persistiendo por más tiempo en aquellos de la cara anterior. Para la segunda semana, casi todos los infartos de la cara inferior tienen el segmento ST en posición isoeléctrica. En el infarto de la cara anterior, la persistencia por más de dos semanas del supradesnivel del segmento ST, puede indicar la presencia de un aneurisma ventricular o compromiso severo de la motilidad de esa zona.

Fase estabilizada crónica. Esta es la fase residual donde son considerados los signos de infarto antiguo. Hay persistencia de ondas Q patológicas haciendo parte de los complejos QS y QR. Las ondas R poseen amplitud más baja comparada con registros de pre-infarto. A veces, la única manifestación del infarto previo puede ser una progresión inadecuada de la onda R en derivaciones precordiales o la aparición de la onda q temprana en las derivaciones $V3$ y $V4$. Las ondas T en un principio son negativas y simétricas en la zona de infarto antiguo, sin embargo con el tiempo tienden a normalizarse.

Localización del IAM. El IAM puede localizarse en alguna de las siguientes áreas:

a) Infarto agudo de cara anterior.

b) Infarto agudo de cara lateral.

c) Infarto agudo de cara inferior.

d) Infarto agudo de cara posterior.

e) Infarto agudo del ventrículo derecho.

La Tabla 1.5 muestra los tipos de cardiopatía de acuerdo a su localización y los datos que permiten su diagnóstico. Es importante tomar en cuenta las siguientes consideraciones: la inclusión de la letra Q indica que existe alguna anormalidad en esta onda, $R\uparrow$ señala una progresión del pico R, $ST\uparrow$ significa un supradesnivel del segmento ST, asimismo $ST\downarrow$ es un infradesnivel del segmento ST, $T\uparrow$ indica que la onda T es positiva, $T\downarrow$ quiere decir que la onda T es invertida y QS significa que se presenta una deflexión QS. Por otro lado, la derivación $V4$R es entendida como la derivación precordial derecha.

Tipo de cardiopatía	**Derivaciones**	**Cambios *QRS***	**Cambios ST-*T***
Angina inestable	aVR Más de 8 derivaciones $V1$ a $V4$		$ST\uparrow$ $ST\downarrow$ $T\downarrow$
Angina variante	DII, DIII y aVF		$ST\uparrow$
Infarto anterior	$V2$, $V3$	Q, pérdida de $R\uparrow$ en derivaciones precordiales	$ST\uparrow$
Infarto lateral	$V5$, $V6$, DI y aVL	Q	$ST\uparrow$
Infarto inferior	DII, DIII y aVF	Q ó QS en mín. dos de las derivaciones	$ST\uparrow$ $T\downarrow$
Infarto posterior	$V1$, $V2$	R en $V1 \leq 40ms$ con $R/S > 1$	$ST\downarrow$ $T\uparrow$
Inf. ventrículo derecho	$V4$R	Q	$ST\uparrow$

Tabla 1.5. Variaciones del ECG debidas a enfermedades isquémicas en diversas localizaciones

También es frecuente considerar los cambios en el ECG debidos a infartos asociados a bloqueos de rama del haz de His. El diagnóstico de IAM en pacientes con bloqueo de rama derecha se sospecha cuando los cambios secundarios de la onda T (opuestos al complejo QRS) en las derivaciones $V1$ a $V3$ son reemplazados por ondas T de polaridad similar con el QRS [32]. De otra parte, el IAM puede causar bloqueo completo de la rama izquierda por oclusión proximal de la arteria descendente anterior o puede asociarse a un bloqueo de rama pre-establecido. Algunos cambios del ECG que pueden ayudar a sospechar la coexistencia de ambas condiciones son: supradesnivel del segmento ST de 1 mm en el sentido de la deflexión principal del complejo QRS, infradesnivel del segmento ST de 1 mm en las derivaciones $V1$ a $V3$ y un supradesnivel del segmento ST de 5 mm contrario a la polaridad del QRS. A veces el bloqueo de rama izquierda puede estar presente y los cambios del segmento ST pueden faltar; en estos casos un trazado antiguo es útil para comparar [38].

1.2. Señales fonocardiográficas

La señal fonocardiográfica (FCG) corresponde a la adquisición de las vibraciones *mecánicas* provenientes del corazón y transmitidas por los tejidos vecinos hacia la pared torácica. En la génesis de estos fenómenos acústicos participan los diferentes eventos mecánicos que se presentan durante el ciclo cardíaco, como son: la contracción muscular, la apertura y cierre de las válvulas y el desplazamiento del flujo sanguíneo. En este sentido, existen discrepancias en cuanto al origen específico de los sonidos cardíacos, aunque se consideran dos teorías principales: la explicación valvular y la explicación hemodinámica. La explicación valvular afirma que los sonidos del corazón son vibraciones variantes en el tiempo, resultantes de la tensión abrupta que se produce sobre las valvas u hojillas de las válvulas cardíacas al final

de su cierre y apertura; adicionalmente, se asume que no se produce ningún sonido en el tiempo transcurrido entre el cierre y la apertura de las mismas. Por otro lado, la explicación hemodinámica, la cual tiene mayor aceptación, atribuye el origen de los sonidos cardíacos a las vibraciones de toda la estructura del corazón, como consecuencia de la aceleración y desaceleración de los fluidos de sangre intracardíacos, seguidos por el cierre y la apertura de las válvulas. Así, el corazón se compara con una bomba llena de fluido, que vibra completamente cuando es estimulada en cualquiera de sus puntos [35].

En general, el FCG representa en forma gráfica el sonido producido por la actividad mecánica del corazón, en el cual pueden observarse los tiempos y las intensidades relativas de los sonidos cardíacos en forma clara y repetida. La fonocardiografía fue desarrollada para mejorar los resultados obtenidos con el estetoscopio convencional. El estado actual de la tecnología electrónica ha abierto nuevas posibilidades en el campo de la instrumentación para la auscultación de alta calidad, que incluye el desarrollo de *sistemas digitales* para la adquisición, registro, almacenamiento, y análisis de los sonidos cardíacos.

1.2.1. Origen fisiológico

El corazón cuenta con cuatro *válvulas*: las válvulas *auriculoventriculares* (AV) y las *sigmoideas*. Las válvulas AV (*Mitral* y *Tricúspide*) impiden que de los ventrículos se regrese la sangre a las aurículas durante la sístole, y las válvulas sigmoideas (*Aórtica* y *Pulmonar*) impiden que se regrese la sangre de las arterias aorta y pulmonar a los ventrículos durante la diástole [19]. En condiciones normales, la sangre fluye de forma continua de las grandes venas a las aurículas; allí tres cuartas partes de la sangre que fluye por ellas pasan directamente a los ventrículos; para el flujo de la otra cuarta parte se hace necesario una contracción auricular que termina cuando los ventrículos se han llenado, y debido a esto, las aurículas son consideradas como bombas auxiliares. Durante la sístole ventricular, se acumulan grandes cantidades de sangre en las aurículas debido a que las válvulas AV permanecen cerradas. Por tanto, en cuanto la sístole ha terminado y las presiones ventriculares caen de nuevo a los valores de diástole, la presión auricular moderadamente elevada abre inmediatamente las válvulas AV y permite que la sangre fluya rápidamente a los ventrículos. Durante el último tercio de la diástole, las aurículas se contraen y con un impulso adicional llenan los ventrículos. Inmediatamente después del comienzo de la contracción del ventrículo, la presión ventricular crece notablemente, provocando el cierre de las válvulas AV. Después de transcurrir un tiempo entre 0.02 a 0.03 s, se eleva la presión a un valor suficiente que permita abrir las válvulas sigmoideas contra las presiones de la aorta y la arteria pulmonar [19]. Así, durante este período, se produce la contracción en los ventrículos, pero no el vaciamiento. Cuando la presión ventricular izquierda se eleva ligeramente por encima de los 80 $mmHg$, las presiones ventriculares impulsan la apertura de las válvulas sigmoideas e inmediatamente comienza la expulsión de la sangre contenida en los ventrículos. Al final de la sístole comienza rápidamente la relajación ventricular, permitiendo que disminuyan las presiones intraventriculares. La presión que se genera en las arterias es elevada e intenta el retroceso de la sangre hacia los ventrículos, sin embargo, las válvulas aórtica y pulmonar se cierran para impedirlo. Las presiones intraventriculares disminuyen hasta tomar los valores normales de diástole. Luego, se abren las válvulas AV para comenzar un nuevo ciclo de bombeo ventricular [25].

1.2.2. Descripción de las componentes

En la dinámica cardíaca, los hechos que ocurren desde el inicio de un latido hasta el inicio del próximo se conocen como *ciclo cardíaco* (ver Figura 1.5). El ciclo cardíaco consta de un período de contracción llamado sístole, seguido de un período de relajación durante el cual el corazón se llena de sangre, llamado diástole. De esta manera, la señal FCG está conformada por ciclos cardíacos sucesivos y generalmente cada ciclo puede dividirse en 4 partes: $S1$, sístole, $S2$ y diástole. Ocasionalmente, también pueden ser escuchados dos sonidos adicionales, conocidos como $S3$ y $S4$.

Primer sonido cardíaco ($S1$). Se atribuye al sonido generado por el cierre de las válvulas mitral (M) y tricúspide (T) en el inicio de la sístole ventricular. El sonido $S1$ está dividido en cuatro componentes: la primera, de baja frecuencia, ocurre cuando se contraen los ventrículos impulsando la sangre hacia las aurículas; la segunda componente es de alta frecuencia y comienza con la tensión abrupta del cierre de las válvulas AV, seguido de la apertura de las válvulas sigmoideas (A y P) cuando el flujo de sangre es expulsado hacia los ventrículos; la tercera componente es generada por la oscilación de la sangre entre la raíz de la aorta y las paredes ventriculares; y la cuarta componente es producida por la turbulencia ocasionada por la expulsión de la sangre a través de la aorta y la arteria pulmonar [14].

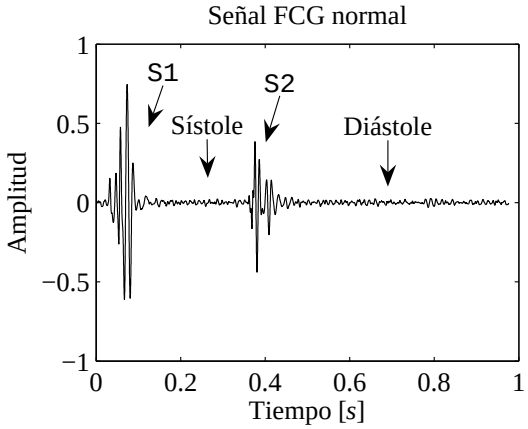


Figura 1.5. Ciclo cardíaco

Segundo sonido cardíaco ($S2$). Ocurre en el fin de la sístole ventricular y el comienzo de la relajación ventricular, constituyendo dos componentes de alta frecuencia: cierre de las válvulas aórtica A y pulmonar P, y apertura de las válvulas mitral (M) y tricúspide (T). El sonido $S2$ usualmente presenta componentes de más alta frecuencia que $S1$, y generalmente su duración es menor que la de $S1$.

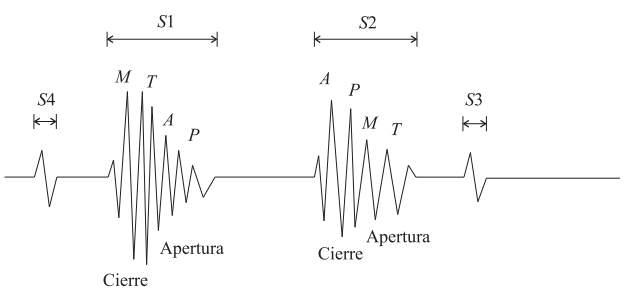
Tercer y cuarto sonido ($S3$ y $S4$). Ocasionalmente se escucha un tercer sonido cardíaco $S3$, que corresponde a la fase de llenado ventricular. Es una vibración acústica de baja frecuencia, y se puede presentar entre 0.12 a 0.16 s después de $S2$ [7]. El cuarto sonido $S4$ es producido en la final de la diástole, y se genera por la contracción de las aurículas desplazando el flujo sanguíneo hacia los ventrículos.

En general, la frecuencia de las vibraciones acústicas de los fenómenos cardíacos está determinada por el mecanismo productor, encontrándose variaciones en el rango de 16 a 1000 Hz. La frecuencia de vibración de cada fenómeno determina a su vez la posibilidad de ser detectada por el oído humano, como también las diferencias en frecuencia. Los sonidos generados por la contracción muscular son de baja frecuencia ($1 - 30$ Hz), mientras que los producidos por la acción de las válvulas y el flujo turbulento de la sangre se consideran de alta frecuencia ($30 - 1000$ Hz).

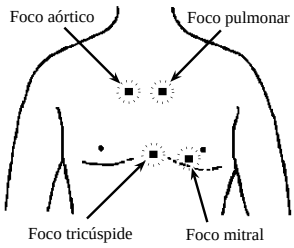
La duración y rango de frecuencias de los diferentes sonidos cardíacos se muestran en la Tabla 1.6, asimismo, en la Figura 1.6(a) se muestran las diversas componentes morfológicas. El tiempo comprendido entre $S1$ y $S2$ se conoce como *período sistólico*, mientras que entre $S2$ y $S1$ se presenta el *período diastólico*.

Sonido cardíaco	*Duración* $[s]$	*Rango de frecuencia* $[Hz]$
$S1$	$0.1 - 0.12$	$20 - 150$
$S2$	$0.08 - 0.14$	$50 - 60$
$S3$	$0.04 - 0.05$	$20 - 50$
$S4$	$0.04 - 0.05$	<25

Tabla 1.6. Parámetros en frecuencia de los sonidos cardíacos



(a) Sonidos cardíacos y sus componentes.



(b) Focos de auscultación.

Figura 1.6. Componentes de los sonidos cardíacos y su auscultación

1.2.3. Focos de auscultación

La máxima intensidad y claridad de los sonidos cardíacos puede auscultarse en zonas determinadas del pecho, llamadas *focos de auscultación*, y corresponden a las siguientes áreas (ver Figura 1.6(b)):

Foco Mitral. Es determinado por el ápex cardíaco y en la mayoría de personas normales está situado en el quinto espacio intercostal izquierdo con línea medio-clavicular. En quienes tienen hipertrofia del ventrículo izquierdo, el ápex, y por ende el foco mitral, se localiza más hacia abajo y hacia afuera de lo normal.

Foco Tricúspide. Se asienta en el quinto espacio intercostal izquierdo al lado del borde esternal.

Foco Pulmonar. Se localiza en el segundo espacio intercostal izquierdo al lado del borde esternal.

Foco Aórtico. Situado en el segundo espacio intercostal derecho al lado del borde esternal del mismo lado.

Frecuentemente, los focos mitral y tricúspide son denominados *focos de la punta*, y los focos pulmonar y aórtico como *focos de la base*. Existen, además, otros dos focos conocidos como *accesorios*, que son:

– *Foco accesorio o de Erb.* Se localiza en el tercer espacio intercostal izquierdo al lado de la línea esternal, inmediatamente por debajo del foco pulmonar.

– *Foco mesocárdico.* Ubicado en el mesocardio, en un punto equidistante entre los focos de la base y los de la punta: cuarto espacio intercostal izquierdo entre las líneas medio-clavicular y medio esternal.

1.2.4. Variaciones del FCG debidas a soplos cardíacos

Normalmente, en el *silencio* de sístole y de diástole no deben existir sonidos adicionales, sin embargo en determinadas condiciones es posible escuchar en estos intervalos ruidos de alta frecuencia, denominados *soplos cardíacos*, en forma de sonidos ásperos o sibilantes que nacen, la mayoría de las veces, por vibraciones anormales del flujo sanguíneo debidas a daños o defectos valvulares, aunque también pueden producirse por defectos en las paredes del corazón. Con respecto a la primera causa, se conocen principalmente dos tipos de deficiencias en el funcionamiento valvular: la estenosis, que consiste en la inadecuada apertura valvular, y la insuficiencia o regurgitación, que se presenta cuando la válvula no se cierra suficientemente, ocasionando un reflujo de sangre en sentido inverso al normal. En ocasiones pueden existir soplos sin daño valvular asociado, como ocurre en las anemias. Es importante decir que no todos los problemas cardíacos producen soplos.

Al momento auscultar y describir los soplos cardíacos deben tomarse en cuenta las siguientes características:

Área de registro de los soplos: Tienden a escucharse mejor sobre el foco correspondiente a la válvula cardíaca donde se producen.

Irradiación: El sonido producido por el soplo generalmente se extiende en el sentido del flujo, ya sea eyectivo, regurgitante o de llenado, a otras áreas donde se ausculta con mayor intensidad. En la irradiación también influye la intensidad, pues cuanto más intenso es un soplo más ampliamente se extiende en todas las direcciones.

Intensidad: La intensidad del soplo se determina por el volumen de sangre responsable de la turbulencia en el flujo sanguíneo y/o del gradiente de tensión que existe entre los dos espacios que separa la válvula, aunque también pueden intervenir, en forma menos importante, circunstancias como la obesidad del paciente, el edema en la pared del tórax, entre otras. La intensidad del soplo no se relaciona siempre en forma directa con la severidad del daño cardíaco, ya que existen soplos de escasa intensidad involucrados en grandes alteraciones endocárdicas. La intensidad del soplo se puede precisar en grados, para lo cual existen dos clasificaciones, una fue introducida por Levine-Harvey en 1933, que los cataloga en 6 grados (Tabla 1.7) y otra, llamada simplificada, que se atribuye a Freeman-Levine, la cual consta de sólo 4 grados (Tabla 1.8). Una característica de intensidad muy conocida es denominada *Frémito* o *Thrill*, en la cual se obtiene una sensación táctil del soplo y sólo aparece en los soplos intensos que producen una corriente sanguínea extremadamente turbulenta al pasar por los orificios valvulares.

De otra parte, la *clasificación de los soplos cardíacos* puede tomarse de acuerdo a las siguientes consideraciones:

Fase del ciclo cardíaco: Según la fase del ciclo cardíaco en el que se produzcan los soplos, estos pueden ser: Sistólicos, Diastólicos y Continuos.

Levine-Harvey	
Grado I	Muy débil, sólo se escucha después que el explorador "se sintonizó"; es probable que no se escuche en todas las posiciones
Grado II	Ligero, pero se escucha de inmediato después de colocar el estetoscopio sobre el pecho
Grado III	Intensidad moderada
Grado IV	Intenso. Tiene frémito.
Grado V	Muy intenso. Puede escucharse cuando el estetoscopio se separa parcialmente del pecho. Tiene frémito.
Grado VI	Se escucha cuando el estetoscopio está separado del pecho. Tiene frémito.

Tabla 1.7. Escala de Levine-Harvey

Freeman-Levine simplificada	
Grado I	Débil, no se oye de primera intención, se debe concentrar la audición para poder detectarlo
Grado II	Moderado, se oye sin dificultad desde el primer momento
Grado III	Fuerte, se ausculta intensamente al colocar el estetoscopio, pero deja de oírse al retirarlo. Tiene frémito.
Grado IV	Muy fuerte, se puede percibir sin el estetoscopio. Tiene frémito.

Tabla 1.8. Freeman-Levine simplificada

Localización temporal en una fase: No siempre los soplos ocupan la totalidad de la sístole o la diástole, y de acuerdo a esto, los soplos pueden llamarse protosistólicos si están únicamente al comienzo de la sístole, mesosistólicos, si están en la mitad y telesistólicos, si están al final; la misma clasificación reciben los soplos diastólicos, sólo que los telediastólicos son más conocidos como presistólicos.

Localización en el tórax: La mayoría de las veces los soplos tienden a tener mayor intensidad en el foco correspondiente a la válvula lesionada. Por ejemplo, es común que las lesiones de la válvula aórtica produzcan soplos perceptibles únicamente en el foco aórtico, o si es audible en varios focos, en este foco es donde se va a escuchar con mayor intensidad; de acuerdo a esta circunstancia, el explorador puede denominar un soplo basado en su localización en el tórax, para este caso, el soplo sería denominado soplo aórtico.

Soplos sistólicos

Por lo general son menos significativos que los soplos diastólicos y pueden dividirse en eyectivos o de eyección, en pansistólicos u holosistólicos y tardíos.

Soplos sistólicos de eyección. Pueden corresponder tanto a soplos inocentes como patológicos. Se escuchan más fácilmente con el estetoscopio de membrana debido a que son dominantes las componentes de alta frecuencia. Es el resultado de un flujo turbulento a través de una válvula semilunar estenótica, alterada y, ocasionalmente, puede ser producido por un flujo aumentado a través de una válvula normal. Comienza después de $S1$, característicamente va aumentando hasta la mitad de la sístole, donde alcanza su máxima intensidad, y luego disminuye progresivamente hasta desaparecer justo antes de $S2$.

Estenosis de la válvula aórtica o pulmonar: este tipo de soplo generalmente es causado por una pérdida en la elasticidad de la válvula, generado por la formación de depósitos calcificados en sus valvas. Por lo tanto, la válvula no abre completamente y como resultado hay una interferencia con el flujo de la sangre. En el fonocardiograma, la estenosis aórtica se caracteriza por tener la forma de un *diamante simétrico* (crescendo - decrescendo), con máxima intensidad hacia la mitad de la sístole. La estenosis de la válvula pulmonar es similar a la estenosis de la válvula aórtica, con la excepción de que la forma del diamante no es simétrico.

Insuficiencia de la válvula mitral o tricúspide: la regurgitación de la sangre dentro de las aurículas ocurre durante la contracción ventricular cuando las valvas no cierran completamente. La forma de este tipo de soplo en el fonocardiograma es *plano* y se extiende desde $S1$ hasta $S2$. La principal causa de la insuficiencia mitral es el prolapso de la válvula mitral y, menos común, la fiebre reumática. Dentro de las causas de insuficiencia de la válvula tricúspide está la fiebre reumática, la carditis infecciosa y la dilatación del ventrículo derecho.

Soplos inocentes: estos soplos son muy comunes, especialmente en los niños, y no tienen un origen patológico. Son causados por estados hiperdinámicos (se incrementa la velocidad del flujo a través de la válvula). También se pueden presentar durante el embarazo normal, y en situaciones patológicas que no están relacionadas al corazón tales como anemia, falla renal crónica y enfermedad hepática.

Soplos pansistólicos. Estos soplos son siempre patológicos y generalmente son ocasionados por insuficiencia de las válvulas auriculoventriculares (AV) o por comunicaciones interventriculares (CIV). Se extienden a lo largo de toda la sístole y, a veces, llegan a sobrepasar el segundo sonido cardíaco. Se producen como consecuencia del escape de sangre del ventrículo a una cámara de presión más baja. La corriente y el soplo continúan uniformes hasta $S2$, porque la diferencia de presión entre las dos cámaras es todavía grande en el momento del cierre de la válvula aórtica, debido a esto, es muy difícil de identificar $S2$ en el área mitral. Es un soplo de frecuencia aspirativa, por lo que se ausculta mejor con el estetoscopio de membrana.

Defecto del tabique interauricular: se debe a que el agujero oval que está permeable durante la gestación, no se cierra completamente después del nacimiento. En esta patología se produce un soplo de tonalidad suave con un retraso del componente pulmonar causando un $S2$ fijo.

Defecto del tabique interventricular: en este tipo de defecto, la sangre oxigenada del ventrículo izquierdo pasa al ventrículo derecho, causando un soplo que a menudo absorbe a $S2$. El componente aórtico de $S2$ se adelanta debido a un corto periodo de eyección, como consecuencia de eso, en la inspiración se incrementa el tiempo de duración del sonido que produce el cierre de las válvulas aórtica y pulmonar.

Soplos sistólicos tardíos. Estos no comienzan inmediatamente después de $S1$, se presentan algo más tarde durante la sístole.

Miocardiopatía hipertrófica obstructiva: suele presentar un soplo de eyección, generalmente largo, pero a veces la obstrucción se produce en la mitad de la sístole o incluso más tarde. Es más intenso en el área precordial y no se irradia a carótidas.

Prolapso de la válvula mitral: el prolapso de la válvula suele producir un clic mesosistólico y, si existe insuficiencia mitral, se asocia a un soplo sistólico tardío que se oye mejor con el paciente en bipedestación que a menudo presenta características piantes.

Soplos diastólicos

Son producidos por turbulencias generadas durante la diástole de uno o ambos ventrículos y son siempre patológicos.

Soplos diastólicos precoces. Se deben a la incompetencia de las válvulas semilunares, aórtica y/o pulmonar, por lo que se oyen al inicio de la diástole, inmediatamente después de $S2$ y tienen un carácter decreciente. Son de alta frecuencia y finalizan antes de $S1$. Son característicos, por tanto, de la *insuficiencia aórtica* (IA) y de la *insuficiencia pulmonar* (IP). Estos dos tipos de patologías son difíciles de distinguir, pero la localización ayuda al diagnóstico. La IA tiene diferentes causas, entre las cuales se destaca la fiebre reumática y la endocarditis infecciosa, además, el soplo aumenta con la espiración forzada. En el fonocardiograma este tipo de soplo se aprecia desde el inicio de la diástole con gran magnitud y disminuye progresivamente (decrescendo). En la mayoría de los casos, la IP obedece a una hipertensión de la arteria pulmonar, en cuyo caso el soplo se halla precedido de un $S2$ muy intenso en el foco pulmonar (soplo de Graham-Steell).

Soplos mesodiastólicos. Son soplos de llenado ventricular debidos a turbulencias a través de las válvulas auriculoventriculares que comienzan ligeramente después en la diástole. Son característicos, por tanto, de la *estenosis mitral* (EM) y de la *estenosis tricúspide* (ET). Este tipo de patologías no permiten el flujo laminar de la sangre al pasar de las aurículas a los ventrículos. La EM es mucho más común que la ET y su principal causa es la fiebre reumática. En el fonocardiograma se aprecia que el soplo inicia desde la mitad hasta la final de la diástole. La duración del soplo mantiene una relación directa con el grado de estenosis, cuanto más largo sea el soplo, más severa es la estenosis.

Soplos telediastólicos o presistólicos. Son debidas a las turbulencias que se generan cuando pasa la sangre desde las aurículas a los ventrículos, coincidiendo con la fase de llenado activo ventricular. Los soplos presistólicos fuertes son diagnósticos de estenosis mitral y/o de tricúspide. Los soplos de intensidad débil pueden darse por el aumento de corriente auriculoventricular, como es el caso del soplo presistólico tricuspídeo de la comunicación interauricular. El *soplo presistólico de la estenosis mitral* o de arrastre presistólico, es un soplo intenso, de baja frecuencia, característicamente creciente y que termina abruptamente con $S1$. Se oye mejor con el estetoscopio de campana apoyado suavemente sobre la piel y el paciente reclinado hacia la izquierda. El *soplo presistólico tricuspídeo* es más precoz, su forma es crescendo-decrescendo, similar a los soplos eyectivos y con un intervalo libre antes de $S1$. Se oye mejor con el estetoscopio de membrana y se acentúa con la inspiración.

Soplos continuos

Ocurren durante la sístole y la diástole. Se deben a que el gradiente de presión que facilita la corriente de sangre desde una zona de alta presión a otra de baja presión, se mantiene

a lo largo de todo el ciclo cardíaco. Son los propios del *ductus arterioso persistente* (o en menos frecuencia, de la fístula arteriovenosa, defecto aorto-pulmonar, etc). El soplo del ductus arterioso persistente o soplo en maquinaria de Gibson, es un soplo de alta frecuencia, intensidad fuerte, la mayoría de las veces con thrill, máximo en el primer o segundo espacio intercostal izquierdo. Comienza tras un intervalo con el primer sonido cardíaco, aumenta gradualmente en intensidad al final de la sístole y después disminuye gradualmente en la segunda mitad de la diástole. Tiene una forma crescendo-decrescendo, con una intensidad máxima alrededor de $S2$. El soplo disminuye durante la inspiración por aumento de la presión arterial pulmonar y se hace más audible con el ejercicio.

En la Figura 1.7, se aprecian los tipos de soplos cardíacos más frecuentes y su ubicación en la señal FCG.

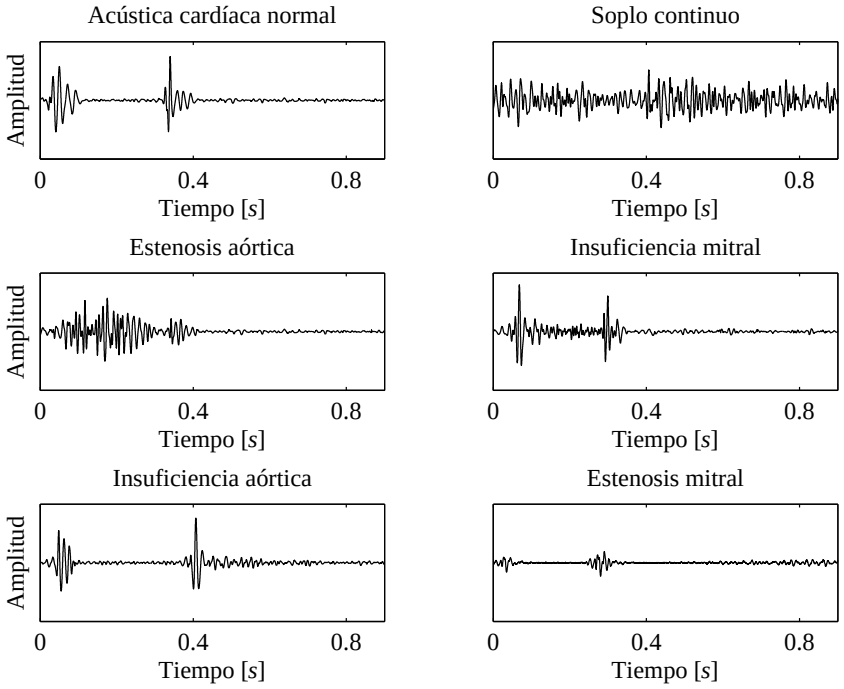


Figura 1.7. Variaciones del FCG debidas a deficiencias valvulares

1.3. Estado del arte en el diagnóstico asistido de patologías

1.3.1. El diagnóstico asistido y sus aplicaciones

Los médicos deben identificar los signos, síntomas y señales orgánicas que sirvan de referentes para determinar los estados de *normalidad* o *anormalidad*, asociada a posibles enfermedades. La identificación se realiza a través de la percepción sensorial, en ocasiones con mediación de instrumentos o dispositivos elementales que magnifican las señales orgánicas o facilitan el juicio sobre el grado de normalidad o anormalidad de los diferentes estados

funcionales del organismo. Este modelo tiene varias limitaciones: alta subjetividad cuando solo se interpreta por un médico, imposibilidad de almacenamiento y réplica para un futuro análisis en caso de requerirse una asesoría de mayor *precisión en el diagnóstico* o en las decisiones terapéuticas, por último, se presenta el sesgo y otras clases de error, a pesar de la destreza generada por el *entrenamiento médico* y la formación del recurso humano. De ahí que, a menudo se requiera de equipos y ayudas de tecnología que faciliten la exploración y auscultación que de manera objetiva mejoren la calidad de las decisiones de tipo clínico; o en otros casos, que provean de la información suficiente al especialista con el objetivo de que sirva de soporte a una adecuada solución del problema.

El avance en las nuevas tecnologías de la electrónica, entre ellas, la instrumentación médica, el procesamiento digital de señales e imágenes, ha mostrado que puede mejorar la *calidad* de la percepción, el análisis, la precisión y la sensibilidad en la toma de las señales emitidas por el *sistema corporal* humano. En este sentido, un particular interés ha surgido alrededor de las técnicas no invasivas de diagnóstico médico, por cuanto, los registros electrónicos de señales biológicas, obtenidos sobre la superficie del cuerpo, reflejan el comportamiento interno del organismo o alguna de sus partes, y pueden ser suficientemente adecuados para proveer la información esencial clínica sin apelar a medidas invasivas. No obstante, el modelo de diagnóstico médico requiere formas objetivas relacionadas a los modelos de proceso, medida en términos de sensibilidad y especificidad, asegurando que las ayudas tecnológicas faciliten la captura, el análisis y el almacenamiento adecuado de las diferentes señales registradas, de tal manera que se permita la posterior interpretación de la información por parte de otros observadores.

Actualmente, los adelantos relacionados con el análisis de señales y la automatización de sistemas, hacen que el modelo de diagnóstico presente cada vez mayor interés en el empleo de los *sistemas automáticos* de procesamiento e identificación de estados funcionales del organismo, los cuales brindan soporte al personal médico especialista que interpreta y toma las decisiones finales sobre el diagnóstico. Como resultado, el diagnóstico asistido, en el cual un sistema de proceso digital ofrece mayor evidencia e información al especialista, permite mejorar la calidad de su veredicto clínico.

Cabe anotar, que desde los años 80 se intentaba la construcción de *sistemas expertos* en el área médica bajo condiciones exploratorias con poca o ninguna experticia clínica, como una de las aplicaciones más importantes de la inteligencia artificial [29]. Las ventajas relacionadas con los altos valores de velocidad, precisión y memoria para el almacenamiento de datos que ofrecen los sistemas modernos de procesamiento digital permite el desarrollo de sistemas de proceso de información sobre una vasta cantidad de datos médicos necesarios para el diagnóstico. De este modo se incrementa la productividad y la eficacia del diagnóstico, tanto terapéutico, como de acciones preventivas por el uso económicamente justificado de computadores para resolver problemas médicos. No obstante, se debe hacer hincapié en que el diagnóstico asistido es siempre de carácter consultivo. Un computador no puede reemplazar el veredicto médico; sin embargo, puede sugerir una serie de decisiones con cierto nivel de confiabilidad, de manera que sean finalmente aceptadas o rechazadas por el médico. El especialista utiliza la respuesta del computador como una segunda opinión, pero es el médico quien toma la decisión final.

El uso de dispositivos digitales se ha convertido en una alternativa efectiva en el proceso de hallar un diagnóstico acertado, el cual gradualmente ha dejado de ser una decisión in-

dividual. Por muchos años el diagnóstico de complicadas enfermedades se ha venido dando no por un médico sino por uno grupo de médicos; cada uno especialista en un campo. En décadas recientes, más y más registros de exámenes complejos han implicado diagnósticos realizados y descifrados por un colectivo de doctores especialmente entrenados (clínicas especializadas). En ese sentido, además de su alta capacidad de proceso de información, los *sistemas de cómputo* brindan la posibilidad de compartir información de forma global. Por lo tanto el uso de *sistemas de proceso digital* en medicina puede resolver dos tareas estrechamente relacionadas: *a)* mayor *precisión en el diagnostico*, y *b)* ampliar la cobertura en el campo aplicativo de un diagnóstico acertado. En general, se puede afirmar que la clase de enfermedades, para las cuales el diagnóstico asistido prueba ser importante, se determinan por dos factores: primero, en enfermedades, donde la *precisión en el diagnóstico* es necesaria para el éxito de la terapia, es decir, diagnósticos que exigen la consideración de muchos signos o síntomas. Como regla, estás son enfermedades que no necesitan cirugía urgente, pero que demandan técnicas operativas complejas. Segundo, en caso de enfermedades agudas donde el diagnóstico exige además de precisión, más importante aun, con urgencia. En estos casos el diagnóstico frecuentemente es hecho antes de obtener los resultados de exámenes adicionales (si éstos son posibles) y concluir con una información mínima para ejecutar procedimientos de acción rápida (una cirugía, por ejemplo).

El número de trabajos escritos presentados relacionados con la investigación y desarrollo de aplicaciones de diagnóstico asistido en *IEEE EMB Society*, *Biosignals* y las publicaciones en *IEEE Transactions on Biomedical Engineering* del 2000 al 2007 están listados en la Tabla 1.9. La mayoría de estas presentaciones han centrado sus estudios en el área de preproceso y extracción/selección de características orientadas al reconocimiento de estados funcionales.

Área de aplicación	*Número de trabajos*
Electrocardiografía	210
Fonocardiografía	73

Tabla 1.9. Publicación de trabajos en medios de divulgación científica

1.3.2. Consideraciones para el diagnóstico asistido

El *diagnóstico clínico*, entre otras, tiene las siguientes particularidades:

Naturaleza subjetiva. Aunque el especialista requiere la mayor información y evidencia posible, el veredicto final es de su potestad, y por ende, la *calidad* del veredicto puede cambiar de especialista a especialista. Los respectivos estudios de concordancia que se consideran aceptables para el dictamen médico de varios especialistas, pueden mostrar valores del orden del $60-70\,\%$, con lo cual generan suficiente incertidumbre en el entrenamiento de los sistemas automáticos e imponen restricciones fuertes sobre su efectividad.

Naturaleza cualitativa. Gran parte de la evidencia e información obtenida de los registros de la actividad del cuerpo humano, los exámenes clínicos, y demás, se ha parametrizado, y a partir de sus valores se han definido franjas cualitativas de juzgamiento, sin embargo, existe un margen amplio subjetivo sobre la escala de calificación final en el veredicto clínico. La generación de características debe acompañarse de una etapa donde se definan valores representativos para cada estado funcional, que cumplan con

las condiciones de consistencia y efectividad impuestas sobre los procedimientos de estimación, a efectos de que los resultados puedan ser reproducibles y transportables. En cuanto a los procedimientos de identificación, estos deben tener una relación estrecha con la evaluación cualitativa de los estados funcionales y no sólo restringirse a la detección.

Naturaleza local. El juzgamiento sobre la normalidad o anormalidad de los estados funcionales del cuerpo humano, además de ser subjetivo, no siempre tiene carácter generalizado y universal. Así por ejemplo, la composición bioquímica, morfológica, antropométrica o biológica de los exámenes clínicos varía de acuerdo a factores como el lugar geográfico, la naturaleza étnica, el origen social, la actividad humana, entre otros, teniendo en cuenta los valores específicos que los diferentes asentamientos humanos han desarrollado en su proceso de adaptación local. Además, la localidad puede ser referida al tiempo, en la medida en que el comportamiento del organismo tiene naturaleza no estacionaria. El organismo, como *sistema complejo*, posee mecanismos dinámicos de compensación que hacen que la medición de las variables fisiológicas cambien con el tiempo.

Trascendencia heurística. El ejercicio de la medicina a lo largo de su historia ha desarrollado de forma empírica una serie de variables de diagnóstico complejas y depuradas, muchas de ellas, basadas en la percepción sensorial directa del médico, que no siempre son fáciles de modelar y parametrizar (por ejemplo, el timbre de los sonidos cardíacos, los cambios morfológicos del registro ECG, etc). Sin embargo, esas variables clínicas de diagnóstico son de gran utilidad, tanto para el descubrimiento de estados de normalidad o anormalidad, como para su tratamiento terapéutico. Este aspecto genera una gran restricción en la generación de nuevas variables en los modelos automáticos de diagnóstico, ya que siempre es preferible que la característica de análisis tenga un sentido físico y sea de fácil comprensión al personal médico.

Además de lo anterior, en el diseño de *sistemas automáticos* de diagnóstico deben tomarse en cuenta los siguientes aspectos:

Modelos irregulares de representación. En la práctica médica, es frecuente que a varios estados de normalidad o anormalidad del cuerpo humano, le correspondan modelos idénticos de representación, esto es, al paciente se le entrega un diagnóstico compuesto por un sólo estado de funcionalidad, cuando en realidad puede presentar varios estados de anormalidad, o el caso contrario, cuando el paciente recibe un diagnóstico que indica la presencia de varias disfuncionalidades y en realidad sólo presenta una de ellas. Este aspecto se hace más complejo debido a la naturaleza no estacionaria del diagnóstico clínico.

Asimetría en la distribución poblacional de las clases. En general, la población no presenta la misma cantidad de personas en cada una de las clases a identificar, y aunque este factor es tenido en cuenta en el diseño de los experimentos relacionados al entrenamiento de sistemas automáticos, existe una gran dificultad en la consecución del número suficiente de personas por clase para el aprendizaje de las máquinas. Como resultado surge la necesidad de aprovechar al máximo la información brindada por cada

observación clínica, o de generar estrategias de registro que impliquen la consecución y extracción de mayor información sobre cada caso. Con el diagnóstico asistido, sin embargo, se requiere que el nivel de desempeño de la respuesta del computador sea alto. Por ejemplo, si la sensibilidad en la detección por computadora de lesiones fuera menor que la sensibilidad promedio de los médicos, sería difícil justificar el uso del diagnóstico asistido.

Orientabilidad del diagnóstico. El sistema de prestación de servicios de salud está dividido en diferentes categorías o niveles, que responden a las diferentes necesidades del servicio. Así por ejemplo, existen los centros de atención primaria, ubicados en zonas remotas o con problemas de acceso, destinados a realizar la percepción básica de normalidad y anormalidad de los *sistemas biológicos* y donde frecuentemente el médico se ve enfrentado a múltiples limitaciones de orden técnico para el registro de las bioseñales. Por su orientación, en estos centros, es preferible el desarrollo de sistemas de identificación automática orientados al tamizaje y/o que alimenten *sistemas centralizados* de prevención de enfermedades. En estos sistemas el énfasis del desarrollo está en la detección con alta precisión de los estados de funcionamiento anormal, los cuales presentan una naturaleza muy diversa frente a los estados normales, es por esto, que se presentan fuertes restricciones en la etapa de adquisición de los registros. Los centros de tercer y cuarto nivel ubicados en las grandes ciudades, que tienen especialistas calificados y presentan menores restricciones en cuanto a la dotación de equipos, exige de los sistemas automáticos un mayor énfasis en la diferenciación con alta precisión entre varios estados de anormalidad con naturaleza muy similar. La solución de ambos requerimientos en un mismo sistema automático hace muy compleja su implementación y manejo, siendo preferible su desarrollo paralelo e integración jerárquica dentro plataformas que incluyan ayudas adicionales, tales como la utilización de las Tecnologías de Informática y Comunicaciones.

Complejidad y alto costo. El registro de las *bioseñales* y la conformación de las respectivas bases de datos son un aspecto básico en el desarrollo de *sistemas automáticos* que sirvan de soporte al diagnóstico clínico. En primer lugar, los altos costos están relacionados con la ubicación, transporte y preparación de los pacientes para la toma y el registro correctos de las bioseñales. Seguidamente, el etiquetado adecuado de los registros y la sintonización de los diversos algoritmos de proceso, agregan una carga considerable al presupuesto de las investigaciones, ya que se requiere personal de alta calificación profesional en el peritazgo de las bases de datos, así como en la asesoría durante el entrenamiento del sistema, sin contar los gastos derivados de la utilización del equipo médico que son asumidos por los hospitales y centros de salud.

La normatividad en el área. El área de la salud, por su impacto e importancia, está suficientemente normativizada, tanto en su ejercicio profesional, como en los demás aspectos concernientes a la prestación de servicios de salud. Así por ejemplo, en Colombia se definen las condiciones de habilitación para las instituciones que prestan servicios de salud en la modalidad de Telemedicina. La Resolución Número 1995 de 1999 establece las normas para el manejo de la historia clínica, incluyendo su forma de presentación electrónica. La transmisión y almacenamiento de bioseñales e imágenes

médicas tiene como *estándar de facto* la norma DICOM, y así sucesivamente.

En este trabajo se presentan diferentes modelos para el desarrollo de *sistemas automáticos*, que a partir de señales cardíacas (ECG y FCG), permiten tomar de forma objetiva y colaborativa mejores decisiones clínicas. La determinación de factores de riesgo de enfermedades cardiovasculares es crítica en nuestro medio Colombiano. Las 10 primeras causas de mortalidad en el departamento de Caldas en el año de 1999 (último registro estadístico encontrado en la respectiva Dirección Territorial de Salud) indican que la mortalidad por enfermedades cardiovasculares y pulmonares ocupan los primeros lugares, superando incluso las muertes causadas debido a la violencia.

El empleo de métodos de control y prevención de las enfermedades cardiovasculares, implica, entre otros, el desarrollo de herramientas más efectivas en el diagnóstico de la funcionalidad cardíaca, en particular se consideran dos formas en los centros de primero y segundo nivel de atención:

a) Análisis de la actividad *eléctrica* del corazón usando registros de electrocardiografía, para lo cual se realiza la cuantificación directa de diferentes parámetros (ritmo, frecuencia, estimación de intervalos, amplitudes), orientado a la identificación de eventos eléctricos relacionados a trastornos isquémicos.

b) Análisis *mecánico* del corazón a través de las señales acústicas obtenidas mediante auscultación, buscando el reconocimiento de anormalidades relacionadas a soplos cardíacos.

1.3.3. Identificación automática en el diagnóstico asistido

El desarrollo del diagnóstico asistido está relacionado con las siguientes preguntas fundamentales: *¿Cómo interpretar los registros corrientes en el tiempo con respecto a las posibles decisiones de diagnóstico?*, *¿Cuáles son las características esenciales?* y, finalmente, *¿Cómo extraer la suficiente información que se encuentra inmersa en las señales obtenidas?*. Estas preguntas son típicas en las tareas concernientes a la automatización de sistemas, particularmente, al entrenamiento de sistemas inteligentes orientado al *reconocimiento de patrones*.

En [1] se discute acerca de exigencias frecuentemente requeridas por la identificación automática relacionadas a la *calidad del entrenamiento* que se necesita para definir la ejecución de procedimientos bajo condiciones necesarias y suficientes en la resolución de tareas específicas concernientes al reconocimiento de patrones (ver Figura 1.8). En este sentido, existe una cota superior que restringe la calidad del entrenamiento dada por el tamaño de la muestra y un límite inferior dado por el conocimiento clínico desde donde los procedimientos diseñados intentan capturar la dinámica fisiológica con el fin de agregar consistencia entre los resultados del sistema automático y la inspección clínica dentro de un rango de calidad. El problema de considerar una *muestra de entrenamiento* de tamaño suficientemente grande tiene su origen en que la adquisición y el etiquetado adecuado de registros patológicos es una tarea compleja y costosa, por lo tanto, las bases de datos disponibles para el entrenamiento usualmente cuentan con un número reducido de registros. Por otro lado, el conocimiento clínico no siempre es de fácil acceso debido a las diferencias que existen entre la estructura del conocimiento médico y la comprensión del diseñador de software.

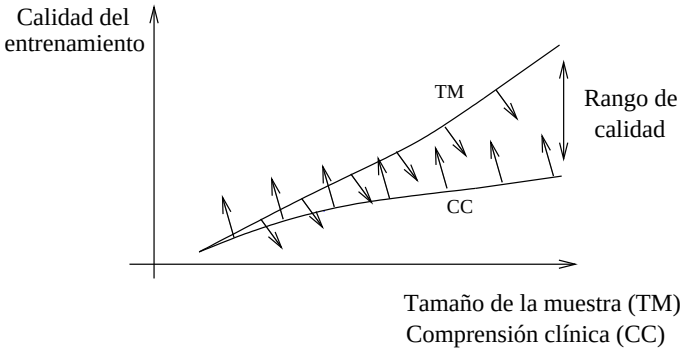


Figura 1.8. Exigencias para la calidad del entrenamiento

Con el ánimo de mejorar la calidad en el entrenamiento se podría pensar en mejorar la *calidad en la representación* aumentando el número de características consideradas en el análisis, sin embargo, en [36] se expone la siguiente relación:

$$\text{Calidad de representación} = \#\text{ de características} \times \frac{|R_o \cap R_e|}{|R_\eta|}$$

donde la calidad de la representación es directamente proporcional al número de características, sin embargo, sujeta a dos restricciones fuertes: la intersección entre la representación dada por las características estimadas R_e y la representación natural de los objetos de análisis R_o debe ser lo más alta posible, y además, las características consideradas en los procedimientos deben minimizar el ruido de representación R_η debido a las condición física o fisiológica de los sujetos (condición atlética, hábitos alimenticios, etc). De esta forma, no siempre un alto número de características significa alta calidad en la representación de un fenómeno fisiológico.

En la terminología del diagnóstico médico, cada uno de los patrones puede ser identificado en términos de síntomas formalizados, mediante un registro electrónico de señales o un conjunto de imágenes del paciente. Mientras, las clases obtenidas representan la variedad de posibles diagnósticos o aseveraciones clínicas. En este sentido, se han planteado soluciones basadas en lógica determinista, aunque son más usuales los enfoques métricos y, en particular, los probabilísticos. Un conjunto ordenado de características relevantes conforman un vector en un espacio de dimensión múltiple, siendo éste un *punto* en ese espacio de representación. La interpretación puede basarse en la proximidad geométrica de las características en el espacio coordenado y, adicionalmente, puede relacionarse con los síntomas significantes de la patología en análisis. De acuerdo a esto, el método de diagnóstico basado en el precedente clínico (completo o parcial) toma como cierto que los casos relacionados con la misma patología generalmente se describen por características de síntomas similares, esto es, en el espacio de características de dimensión múltiple, la similaridad entre casos se determina por la distancia entre sus puntos de representación [33]. Sin embargo, existen problemas sin resolver para casos cuando las características tienen densidad de probabilidad diferente a la Gaussiana o cuando existen relaciones no lineales entre estas. En la mayoría de los casos, se emplean suposiciones sobre distribuciones de probabilidad aproxi-

madas, basados en alguna información adicional. Así, usando estas suposiciones para cada una de la patologías, se pueden establecer relaciones o distancias entre los diversos puntos del espacio con el fin de distinguir entre las clases. También es importante considerar que cuando se tiene un número relativamente pequeño de características, es posible encontrar una completa interpretación entre sus significados. En un caso más general, la distancia en el espacio de características se puede usar en calidad de criterio de diagnóstico [41].

De otra parte, es conocida la información que identifica cada uno de lo síntomas, y por lo tanto, es posible hallar las características que mejor describen el diagnóstico, debido a esto, el espacio de características posee propiedades anisótropas complejas. Una consideración parcial puede ser realizada en el espacio de representación atribuyendo a sus coordenadas diferentes pesos, determinados por cálculos basados en propiedades estadísticas de las características [41]. Esta operación equivale a cambiar la escala en diferentes direcciones del espacio de representación. Algunos algoritmos usan funciones de evaluación basados en el poder discriminante y llevan a cabo transformaciones más complejas de coordenadas.

La toma de decisiones con algoritmos basados en la proximidad en el espacio de características depende de la ubicación del punto con respecto a las áreas que pertenecen a los diferentes estados funcionales. Esto implica que las superficies de decisión deben pasar por áreas donde el número escaso de puntos de representación determinan la frontera entre regiones con densidad mucho mayor. La determinación de fronteras de decisión puede ser realizada usando métodos convencionales de clasificación objetiva [18, 13]. Es importante anotar que el aumento del número de patologías a diagnosticar exige métodos de representación más complejos, lo cual usualmente involucra en las diferentes fases del proceso la combinación entre lógica determinista, enfoques métricos y medidas de información.

En general, el reconocimiento de patrones se puede dividir en una serie de etapas como se muestra en la Figura 1.9. La primera etapa corresponde a la *estimación de características*

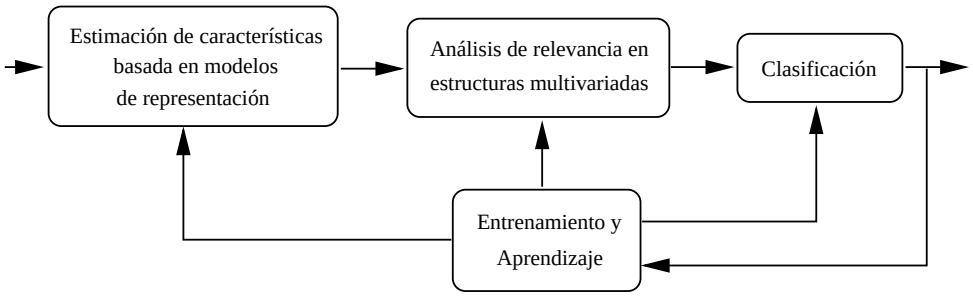


Figura 1.9. Diagrama de bloques correspondiente al reconocimiento de patrones

basada en modelos de representación, donde se extrae la información relacionada con la descripción de los estados funcionales a identificar, y consiste en transformar cada uno de los patrones en cantidades medibles de representación. En la siguiente etapa, denominada *análisis de relevancia en estructuras multivariadas*, se escoge un conjunto reducido de características para representar los patrones, de manera que se preserve la información discriminante de las clases que se requieren identificar en el diagnóstico asistido. Este procedimiento también es llamado reducción de dimensiones y está sujeto a un criterio de relevancia que busca precisión en la identificación, bajos costos de cómputo y reducción en

la complejidad de operaciones. Finalmente, se encuentra la etapa de la *clasificación* que es la encargada de asignar una etiqueta de clase a un patrón específico, usando el conjunto de características seleccionadas para esto. En cada etapa se desarrolla paralelamente el proceso de *entrenamiento* y *aprendizaje*, siendo éste, el procedimiento más importante del reconocimiento de patrones, ya que es donde se determinan y se ajustan los parámetros de funcionamiento.

Estimación de características. Para la estimación de características es importante realizar previamente la reducción de perturbaciones que se encuentran comúnmente en las bioseñales y la segmentación de los eventos que caracterizan los estados funcionales, usando estrategias que abarcan desde filtros lineales hasta el proceso conjunto de tiempo-frecuencia. La magnitud del problema que conlleva el preproceso de la señal, y su importancia, se refleja en la cantidad de trabajos en esta área reportados en la literatura (ver Tabla 1.10), como también el hecho de ser considerada la primera fase de procesamiento de las bioseñales. Los porcentajes por etapa de proceso mostrados en la Tabla 1.10, son obtenidos a partir de las cantidades que fueron expuestas en la Tabla 1.9. Aunque las bioseñales son procesos con

Área de aplicación	Preproceso de la señal	Estimación de características	Reducción de dimensiones	Estrategias de clasificación
ECG	28 %	33 %	29 %	10 %
FCG	37 %	31 %	6 %	26 %

Tabla 1.10. Cantidad de trabajos por área de procesamiento

amplio carácter no estacionario, hasta épocas recientes, el aparato matemático básicamente empleado en el *análisis de bioseñales*, era el de Fourier, particularmente, el proceso en intervalos cortos de tiempo. El problema radica en que las funciones trigonométricas no permiten localizar el análisis en intervalos fijos de tiempo (por ejemplo, durante la aparición de cambios abruptos en diferentes estados funcionales del cuerpo). Debido a esto, la transformada de Fourier no tiene en cuenta que los parámetros periódicos o cuasi-periódicos (amplitud, frecuencia, fase) pueden evolucionar en el tiempo, generando dificultades relacionadas con la longitud insuficiente de la señal con respecto al periodo de análisis, presencia de fluctuaciones con origen no estacionario, pérdida de datos en algunos momentos del tiempo, etc. La insuficiencia de los métodos convencionales de análisis espectral, en la descripción y análisis de las bioseñales no estacionarias, ha hecho que se empleen nuevas formas matemáticas y estadísticas de representación. Particular interés han ganado las wavelets, basadas en una clase de funciones especiales con propiedades de localización, tanto en el espacio físico (tiempo), como en el espacio de Fourier. Cabe anotar, que aunque las funciones wavelet tienen amplio uso en el análisis de bioseñales con estructura no estacionaria, no se excluye el análisis armónico de Fourier [39]. De otro modo, el *sistema de representación* en forma de valores puntuales es una estrategia básica de caracterización de las variables aleatorias y su adecuado empleo influye sobre la efectividad del sistema. Sin embargo, la naturaleza no estacionaria de las bioseñales dificulta que la representación a través de estos valores puntuales sea suficiente, por lo que es necesario la generación de características de aleatoriedad en forma de arreglos (contornos o superficies), buscando que la dinámica de cambio de estas nuevas características revele información importante al concepto de clasificación sobre un

intervalo de análisis, e.g., el tiempo, la frecuencia, las escalas de representación wavelet, etc. En el caso particular del reconocimiento automático de patologías cardíacas, se reportan avances en los casos específicos de electrocardiografía [22, 40] y fonocardiografía [15, 23].

Análisis de relevancia. Existe una restricción fuerte en los métodos básicos de aprendizaje, y consiste en que el número de características o variables descriptivas debe ser mucho menor que el número de registros disponibles para el entrenamiento; a medida que el número de características aumenta, se requiere el incremento exponencial del número de observaciones para mantener la densidad de la muestra dada (*maldición de la dimensión* [3]). Como se puede observar en la Figura 1.10, para un conjunto de registros de entrenamiento dado, hay un número de características por encima del cual, el desempeño de los clasificadores disminuirá en lugar de mejorar [43], lo cual puede estar relacionado con el hecho de que, en los conjuntos de datos de alta dimensión, no todas las variables medidas son relevantes en términos de representación [12]. Así, la Figura 1.10 muestra el compromiso entre los valores aceptables de complejidad computacional (máximo valor permitido C_{max}) y la tasa de acierto en la clasificación (mínimo valor aceptable A_{min}), los cuales determinan un rango de operación adecuado respecto al número de características donde el espacio reducido no compromete la precisión de clasificación del *sistema de detección*. De acuerdo a esto, la reducción de dimensiones realizada sobre espacios de representación puede tomarse como una tarea de optimización asociada a un *criterio de relevancia*, tomando en cuenta que la relevancia es la que define el contexto de la representación reducida. En la práctica, el problema consiste en definir con exactitud el objetivo de la relevancia para que el espacio reducido quede dotado con la información suficiente, de forma que permita una alta capacidad de generalización al momento de evaluar nuevos ejemplos, y de esta manera, se busca que la virtud del *sistema de detección* no dependa exclusivamente de la eficiencia del clasificador. En [4] se presentan diferentes aspectos del análisis de relevancia y se introduce el planteamiento del problema desde la perspectiva de la siguiente definición:

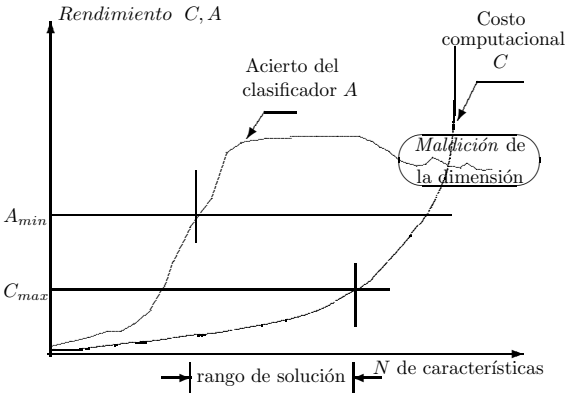


Figura 1.10. Rango de solución

Definición 1.2 (Objetivo de la relevancia)

> Una característica ξ es relevante al objetivo c si existe un par de ejemplos A y B en el espacio de representación, tal que A y B difieren solamente en su asignación a ξ y $c(A) \neq c(B)$.

En general, un número limitado de características simplifica el espacio de representación, convirtiendo el análisis en menos denso y extenuante, de forma que se logra una etapa de clasificación más rápida y con mínimos requerimientos de memoria. Sin embargo, una reducción de dimensiones exagerada podría llevar a una pérdida de representación, disminuyendo la calidad y precisión del *sistema de reconocimiento*.

Clasificación. En el diagnóstico automatizado, esta etapa define a qué clase o estado funcional pertenece un patrón de análisis representado en un espacio de características. En el *análisis de bioseñales*, después que han sido estimadas las características sobre el conjunto de registros de entrenamiento, cada señal es representada por un vector $\mathbf{x} = [x_1, \ldots, x_p]$, el cual es conocido como vector de medidas o vector patrón. Para una clasificación efectiva, se desea que los patrones por clase formen agrupaciones o clusters.

Las características que son comunes a lo largo de los patrones de una clase particular son conocidas como características intra-clase. Las características discriminantes que representan las diferencias entre las distintas clases de patrones son llamadas características inter-clase. El problema de clasificación consiste en generar límites, o fronteras de decisión, que de forma óptima separen los datos de acuerdo a las diferentes clases de patrones existentes en el conjunto de entrenamiento (ver Figura 1.11), y para la resolución de este problema, se han propuesto diversas técnicas supervisadas y no supervisadas de clasificación [35]. El inconveniente se presenta cuando las clases a discriminar tienen distribución de probabilidad no gaussiana y las fronteras de separación son altamente no lineales, puesto que las técnicas convencionales de clasificación para estos casos se vuelven no efectivas. De otra parte, para la evaluación del clasificador, el conjunto de muestras disponible puede dividirse en dos grupos: *conjunto de entrenamiento* y *conjunto de validación*. El método básico de validación es conocido con el nombre de *validación simple* y los pasos para evaluar un clasificador mediante este método, se pueden listar como sigue:

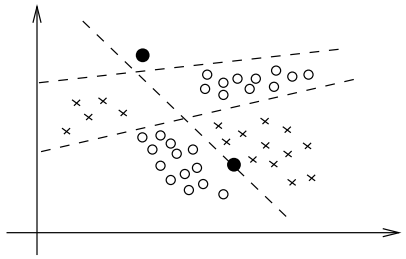


Figura 1.11. Capacidad de generalización

1. Separar aleatoriamente del tamaño total de la muestra N_s, dos conjuntos de tamaño N_t y N_v, donde $N_t + N_v = N_s$.

2. Construir el clasificador usando el conjunto de entrenamiento.

3. Usando el clasificador, se asigna una etiqueta de clase a cada patrón del conjunto de validación. Dado que se conoce previamente la clase correcta para cada patrón evaluado, se puede obtener el número de casos correctamente clasificados N_{CC}.

4. Se calcula la *precisión de clasificación* mediante la expresión: $P = 100 * N_{CC}/N_v$.

La ventaja de este método reside en que no emplea mucho tiempo de proceso, sin embargo, el resultado puede tener una alta variabilidad no deseable, ya que la prueba depende en gran medida de la elección de los datos que forman los conjuntos de entrenamiento y validación. Debido a esto, existe entre otras técnicas la validación cruzada (cv) que es frecuentemente usada cuando el tamaño de la muestra no es suficientemente grande, aunque conlleva a una mayor complejidad computacional [42].

BIBLIOGRAFÍA

[1] E. Alpaydin. *Introduction to Machine Learning*. Massachusetts Institute of Technology, 2004.

[2] J. J. Arango. *Manual de electrocardiografía*. Asociación Colombiana de Facultades de Medicina, 5 edition, 2003.

[3] R. E. Bellman. *Adaptive Control Processes: A Guided Tour*. Princeton University Press, 1961.

[4] A. L. Blum and P. Langley. Selection of relevant features and examples in machine learning. *Artificial Intelligence*, 97(1-2):245–271, 1997.

[5] L. J. Cao, K. S. Chua, W. K. Chong, H. P. Lee, and Q. M. Gu. A comparison of PCA, KPCA and ICA for dimensionality reduction in support vector machine. *Neurocomputing*, 55:321–336, 2003.

[6] M. A. Carreira-Perpiñán. A review of dimension reduction techniques. Technical Report CS-96-09, Department of Computer Science, University of Sheffeld, 1997.

[7] M. Cheitlin and D. Skolow. *Clinical Cardiology*. Prentice Hall international Inc., 1993.

[8] E. J. Ciaccio, S. M. Dunn, and M. Akay. Biosignal pattern recognition and interpretation systems. *IEEE Engineering in Medicine and Biology*, pages 129–135, February/March 1994.

[9] I. Cohen, Q. Tian, X. S. Zhou, and T. S. Huang. Feature selection and dimensionality reduction using principal feature analysis. In *IEEE International Conference on Image Processing*, 2002.

[10] M. Dash and H. Liu. Consistency-based search in feature selection. *Artificial Intelligence*, 151:155–176, 2003.

[11] J. A. del Río. Aspectos del perfil de morbimortalidad. Technical report, Hospital Universitario de Caldas, Manizales, 2002.

[12] D. L. Donoho. High-dimensional data analysis: The curses and blessings of dimensionality. In *Mathematical Challenges of the 21st Century*, 2000.

[13] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. John Wiley Sons, Inc., 2 edition, 2001.

[14] L. Durand and P. Pibarot. Digital signal processing of the phonocardiogram: Review of the most recent advancements. *Critical Reviews in Biomedical Engineering*, 23(3/4):163–219, 1995.

[15] M. El-Segaier, O. Lilja, S. Lukkarinen, L. Sörnmo, R. Sepponen, and E. Pesonen. Computer-based detection and analysis of heart sound and murmur. *Annals of Biomedical Engineering*, 33(7):937–942, 2005.

[16] R. Fried and V. Didelez. Latent variable analysis and partial correlation graphs for multivariate time series. *Statistics & Probability Letters*, 73:287–296, 2005.

[17] H. H. Friedman and M. P. Guinovart. *Diagnostico electrocardiográfico y vectocardiográfico*. Salvat Editores, Barcelona, 1990.

[18] K. Fukunaga. *Introduction to Statistical Pattern Recognition*. Academic Press, 2 edition, 1990.

[19] W. F. Ganong. *Fisiología médica*. Manual Moderno, 20 edition, 2006.

[20] N. Goldschlager and M. J. Goldman. *Principios de Electrocardiografía Clínica*. El Manual Moderno, México, 1992.

[21] A. C. Guyton and J. E. Hall. *Tratado de Fisiología médica*. McGraw-Hill, 9 edition, 1997.

[22] A. I. Hernández, G. Carrault, F. Mora, and A. Bardou. Model-based interpretation of cardiac beats by evolutionary algorithms: signal and model interaction. *Artificial Intelligence in Medicine. Elsevier*, 26(3):211–235, 2002.

[23] K. Higuchi, K. Sato, H. Makuuchi, A. Furuse, S. Takamoto, and H. Takeda. Automated diagnosis of heart disease in patients with heart murmurs: application of a neural network technique. *Journal of Medical Engineering & Technology*, 30(2):61–68, 2006.

[24] A. Hyvärinen. Survey on independent component analysis. *Neural Computing Surveys*, 2:94–128, 1999.

[25] S. W. Jacob, C. A. Francone, and W. J. Lossow. *Anatomía y Fisiología Humana*. Nueva Editorial Interamericana, S.A., México, 4 edition, 1982.

[26] A. K. Jain, R. P. W. Duin, and J. Mao. Statistical pattern recognition: A review. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 22(1):4–37, January 2000.

[27] R. A. Johnson and D. W. Wichern. *Applied Multivariate Statistical Analysis*. Prentice Hall, 5 edition, 2002.

[28] D. L. Kasper, E. Braunwald, A. S. Fauci, S. L. Hauser, D. L. Longo, and J. L. Jameson. *Principios de Medicina Interna*, volume 1. McGraw Hill, 16 edition, 2006.

[29] K. King. A model based toolkit for building medical diagnostic support systems in developing countries. In *IEE Colloquium on Intelligent Decision Support Systems and Medicine*, pages 911–916, 1992.

[30] A. König. Dimensionality reduction techniques for multivariate data classification, interactive visualization, and analysis - systematic feature selection vs. extraction. In *Fourth International Conference on Knowledge-Based Intelligent Engineering Systems & Allied Techndogies*, pages 44–55, 2000.

[31] M. Kudo and J. Sklansky. Comparison of algorithms that select features for pattern classifiers. *Pattern Recognition*, 33:25–41, 2000.

[32] H. Matiz and O. Gutiérrez. *Electrocardiografía Básica: Del trazado al paciente.*, volume I. Siglo del hombre, 2004.

[33] N. I. Moiseeva and V. V. Usov. Some medical and mathematical aspects of computer diagnosis. *IEEE Special Issue on Technology and Healt Services*, 57(11):1919– 1925, 1969.

[34] W. Pedrycz, A. Breuer, and N. J. Pizzi. Genetic design of feature spaces for pattern classifiers. *Artificial Intelligence in Medicine*, 32:115—125, 2004.

[35] R. Rangayyan. *Biomedical signal analysis: A case-study approach*. IEEE press in Biomedical Engineering, 2002.

[36] Z. Reynolds, H. Bunke, M. Last, and A. Kandel. Comparing representative selection strategies for dissimilarity representations. *International journal of intelligent systems*, 21:1093–1109, 2006.

[37] R. H. Rozo, editor. *Cardiología*, volume II. Sociedad Colombiana de Cardiología, 1999.

[38] R. H. Rozo, editor. *Cardiología*, volume I. Sociedad Colombiana de Cardiología, 1999.

[39] N. V. Takhor and D. Sherman. *The Biomedical Engineering Handbook*, chapter Wavelet (time-scale) analysis in biomedical signal processing, pages 886–906. CRC Press, 1995.

[40] A. P. Nemirko, A. N. Kalinichenko, P. V. Murashov, N. N. Pogodin, D. V. Goncharov, S. E. Ivanova, C. M. Matous, and A. N. Sankin. Software complex for the recognition of diagnostically significant ECG changes. *Pattern Recognition and Image Analysis*, 16(1):9–11, 2006.

[41] A. van Woerkom and K. Brodman. Statistics for a diagnostic mode. *Biometrics*, 17(2):299–318, 1961.

[42] I. H. Witten and E. Frank. *Data Mining: Practical Machine Learning Tools and Techniques*. Elsevier, 2 edition, 2005.

[43] L. Yu and H. Liu. Efficient feature selection via analysis of relevance and redundancy. *Journal of Machine Learning Research*, 5:1205–1224, 2004.

CAPÍTULO 2

Modelos de representación para las señales cardíacas

> Nihil est in intellectu quod prius non fuerit in sensu.
>
> *(Estratón de Lampsaco)*

La estimación de medidas sobre segmentos de la señal, con el fin de transformar la información clínica en un conjunto de coeficientes, se puede lograr a través de modelos de representación. El objetivo final es representar las señales en un espacio donde alguna métrica minimice la distancia entre patrones de una misma clase y maximice la distancia entre patrones de distinta clase.

2.1. Medidas de diagnóstico: análisis temporal y análisis acústico

De acuerdo a lo expuesto en el anterior capítulo, las señales cardíacas contienen abundante información acerca del estado funcional del corazón. Existe una cantidad de criterios médicos que definen patrones de funcionamiento, a partir del análisis temporal, donde las características principalmente se basan en la morfología de la señal, ondas componentes, eje eléctrico del corazón y estimación de intervalos. En general, las características temporales están determinadas de forma más precisa en las señales ECG que en las FCG, ya que la información fisiológica del ECG se concentra en el eje temporal, mientras que la información contenida en el FCG está regida por parámetros de naturaleza acústica (intensidad, tono y timbre). Así, el espacio de representación queda conformado por medidas inspiradas en el análisis clínico que son tomadas directamente de la señal, quedando el siguiente modelo:

$$\mathbf{X}_1 = \mathcal{MD}\{\mathbf{S}\} \tag{2.1}$$

donde $\mathbf{X}_1$ es la estructura multivariada resultante, $\mathbf{S}$ es el conjunto de señales para el entrenamiento y $\mathcal{MD}$ es un conjunto de operadores que se encargan de extraer las medidas del conjunto de señales.

Para el caso de las señales ECG, usualmente se consideran diferentes mediciones en puntos específicos de la señal, las cuales son usadas tradicionalmente para el diagnóstico clínico (ver Tabla 2.1) [78, 28]. El cálculo de cada uno de estos parámetros requiere inicialmente

Parámetro	**Descripción**
$qrsw$	Ancho del QRS
$pamp$	Amplitud positiva del QRS
$namp$	Amplitud negativa del QRS
$pqrs$	Área positiva del QRS
$nqrs$	Área negativa del QRS
$areat$	Área de la onda T
ivr	Repolarización ventricular invertida
lst	Nivel del segmento ST
$stsl$	Pendiente del segmento ST
pr	Intervalo PR
ta	Amplitud de la onda T
tp	Posición de la onda T respecto al pico R
qtd	Intervalo QT

Tabla 2.1. Medidas de diagnóstico sobre señales ECG

la estimación de los siguientes valores: *picor* (posición del pico R), *iqrs* (inicio del QRS), *fqrs* (final del QRS), *it* (inicio de la onda T), *ft* (fin de la onda T) e *ip* (inicio de la onda P). De acuerdo a esto, las medidas de diagnóstico se obtienen de la siguiente manera:

Ancho del QRS: Se toma como el intervalo desde el inicio hasta el final del complejo QRS:

$$qrsw = fqrs - iqrs$$

Amplitud positiva del QRS: Primero se halla el parámetro *bl* (línea de base) como un punto de la señal que se encuentra 80 ms antes de *picor*. La amplitud positiva del QRS se define como la amplitud del pico R menos el valor encontrado para *bl*.

Amplitud negativa del QRS: Se define como el valor mínimo que hay entre *iqrs* y *fqrs*, restando el valor de *bl*.

Área positiva del QRS: Se define como la sumatoria de todos los puntos positivos de la señal que están entre *iqrs* y *fqrs*, restando el valor de *bl*.

Área negativa del QRS: Se define como la sumatoria de todos los puntos negativos de la señal que están entre *iqrs* y *fqrs*, restando el valor de *bl*.

Área de la onda T: Se define como la sumatoria de todos los puntos de la señal que se encuentran desde *it* hasta *ft*, restando el valor de *bl*.

Repolarización ventricular invertida (IVR):

$$IVR = \frac{\text{Área positiva del } QRS + \text{Área negativa del } QRS}{\text{Área de la onda } T}$$

Nivel del segmento ST: Este valor típicamente se toma 60 ms después de *fqrs*.

Pendiente del segmento ST: Se define como la relación entre la variación en amplitud y la variación en tiempo de un intervalo contenido en el segmento ST. Generalmente, los datos se toma entre los puntos $picor + 180\ ms$ y $picor + 220\ ms$.

Intervalo PR: Este intervalo va desde el inicio de la onda P hasta el comienzo del QRS:

$$pr = iqrs - ip$$

Amplitud de la onda T: Se define como el punto máximo del trozo de señal comprendido entre *it* y *ft* menos el valor de *bl*.

Posición de la onda T respecto al pico R: Se define como el número de puntos que existen entre el pico R y el punto donde se toma la amplitud de la onda T.

Intervalo QT: Este intervalo va desde el comienzo del QRS hasta el final de la onda T:

$$qtd = ft - iqrs$$

El análisis temporal de las señales FCG se basa principalmente en el estudio de la envolvente de la señal. En la Figura 2.1 se muestran diferentes métodos para calcular la envolvente de una señal normalizada y debido a la simetría inherente a la aplicación de las definiciones de energía, solamente se toma en consideración la parte positiva.

Sea x una señal normalizada, la cual tiene valores reales entre -1 y 1. La envolvente puede analizarse mediante los siguientes métodos:
Energía de Shannon: $E = -x^2 \cdot \log x^2$
Entropía de Shannon: $E = -|x| \cdot \log |x|$
Valor absoluto: $E = |x|$
Energía (cuadrática): $E = x^2$
La Figura 2.1 indica que la energía enmascara los sonidos de baja con los de alta intensidad por ampliar la proporción entre las intensidades de valor alto/bajo. La entropía de Shannon resalta los efectos del ruido de baja intensidad, lo cual hace que la envolvente sea muy ruidosa para interpretar, mientras que el valor absoluto da el mismo peso para toda la señal. La energía de Shannon resalta las intensidades medias contenidas en la señal y atenúa el efecto de las intensidades bajas mucho más que el de las altas. Por lo tanto, la energía de Shannon es mejor que el valor absoluto al reducir la diferencia entre las magnitudes de la envolvente obtenida en sonidos de baja y alta intensidad, de manera que facilita la detección de sonidos de baja intensidad [51]. La energía promedio de Shannon está definida como los promedios

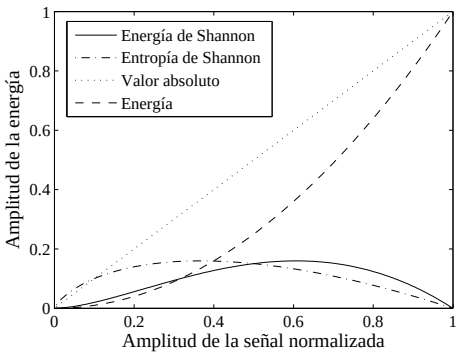


Figura 2.1. Diferentes métodos de envolvente

de energía sobre ventanas sucesivas traslapadas de la señal, esto es:

$$E_s(t) = -\frac{1}{N}\sum_{i=1}^{N} x^2(i)\log x^2(i)$$

donde, x es el valor de la señal y N el número de datos por ventana. Así, la *energía promedio normalizada de Shannon* es calculada como:

$$P_a(t) = \frac{E_s(t) - \bar{E}_s(t)}{S(E_s(t))}$$

donde $\bar{E}_s(t)$ es el valor medio de $E_s(t)$ y $S(E_s(t))$ es la desviación estándar de $E_s(t)$.

Por otro lado, los procedimientos médicos basados en la auscultación, para detectar anormalidades mecánicas del corazón, son altamente dependientes del *entrenamiento auditivo* que tenga el especialista. Debido a esto, los *sistemas automáticos* buscan que el procedimiento de medición involucrado en las apreciaciones médicas a través del ejercicio de la auscultación tengan un carácter objetivo logrado mediante el *análisis acústico* de la señal. Estudios fisiológicos que analiza la capacidad humana para descifrar estructuras en frecuencia, a partir de sonidos, muestran que el *sistema auditivo* no sigue una escala lineal sino que funciona como un banco de filtros donde el módulo de amplitud decae exponencialmente y las frecuencias centrales de cada filtro están espaciadas de forma creciente y logarítmica para capturar características importantes de los sonidos audibles [58]. Los *coeficientes cepstrales en la escala de frecuencia Mel* (MFCC) intentan emular el comportamiento auditivo mediante un banco de filtros triangulares que usa la escala Mel para la determinación de las frecuencias centrales de cada filtro (ver Figura 2.2). De manera que la energía que pasa por cada filtro corresponde a cada MFCC de la señal. Particularmente, para estimar los MFCC de una señal $x[k]$, estos coeficientes pueden ser estimados usando la siguiente expresión [3]:

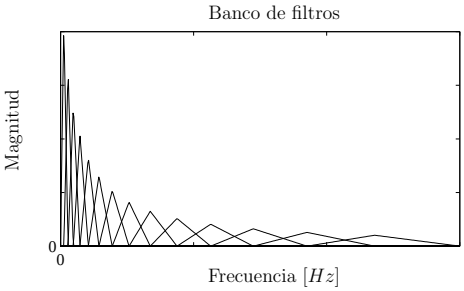


Figura 2.2. Banco de filtros para obtener los MFCC

$$c\,[n] = \sum_{m=0}^{M-1} X_F\,[m]\cos\left(\frac{\pi n - 1/2}{M}\right), \quad 0 \leq n < M \tag{2.2}$$

tomando un número M de filtros y la respuesta en frecuencia de la señal filtrada, notada como $X_F[m]$, puede expresarse como:

$$X_F[m] = \ln\left(\sum_{k=0}^{N-1} |X\,[k]|^2\, H_m\,[k]\right) \tag{2.3}$$

donde, $X[k]$ es la transformada de Fourier de la señal $x[k]$ y $H_m[k]$ es el banco de filtros

triangulares, dado por:

$$H_m[k] = \begin{cases} 0, & k < f[m-1] \\ \frac{2(k-f[m-1])}{(f[m+1]-f[m-1])(f[m]-f[m-1])}, & f[m-1] \leq k \leq f[m] \\ \left(\frac{2(f[m+1]-k)}{(f[m+1]-f[m])}\right), & f[m] \leq k \leq f[m+1] \\ 0, & k > f[m+1] \end{cases}$$

donde, $f[m]$ es la frecuencia central del filtro m.

El *Jitter* es otra medida acústica, propuesta en [13], que también puede usarse para caracterizar señales FCG, la cual toma tramos de la señal que contienen al menos un ciclo de la frecuencia fundamental (f_0) y determina la variación de la f_0 de cada tramo con respecto a la f_0 de la señal [3]. Inicialmente, el procedimiento requiere determinar la longitud de la ventana que contenga al menos un ciclo de la componente de menor frecuencia con magnitud significativa en el espectro promedio de las señales FCG. Luego, se hace un ventaneo de los segmentos $S1$, sístole, $S2$ y diástole, con el fin de estimar la frecuencia fundamental f_0 de cada tramo, tomando desplazamientos de la ventana con algún nivel de solapamiento. Así, se obtiene el vector de frecuencias fundamentales de dimensión N para cada segmento, y se estima el Jitter, como:

$$\text{Jitter} = \frac{\left(\sum_{i=1}^{N} \left| \frac{f_0(i-1)+f_0(i)+f_0(i+1)}{3} - f_0(i) \right| \right)}{\sum_{i=1}^{N} f_0(i)} \tag{2.4}$$

En general, las medidas de diagnóstico son características con alto contenido fisiológico, y es por esto, que desde el punto de vista médico son muy aceptadas, ya que ofrecen alta facilidad para la interpretación de los resultados. Sin embargo, desde el punto de vista automático, estas medidas son altamente dependientes del desempeño de la etapa de preproceso (filtración y segmentación), por lo que algún desacierto en esta etapa preliminar hace que la estimación sea no consistente, y por lo tanto, no representativa de los verdaderos estados funcionales presentes en la señal de análisis.

2.2. Análisis espectro-temporal

Los *sistemas biológicos* exhiben información relacionada a ritmos y periodicidades que son más fáciles de interpretar en el dominio de la frecuencia que en el tiempo. Cuando la señal de análisis está compuesta por eventos discretos, como ocurre en la señal ECG, el ritmo básico o la rata de actividad presente en la señal puede ser valorada directamente en el dominio del tiempo. De otro modo, señales como las de FCG contienen estructuras complejas de interpretar en el tiempo pero proveen múltiples características en el dominio de la frecuencia que permiten realizar análisis interesantes. Por ejemplo, adicionalmente a la presencia de periodicidades o ritmos latido a latido, los sonidos del corazón en un ciclo cardíaco exhibe *resonancia*. Debido al múltiple compartimiento natural del *sistema cardíaco*, se podría esperar que los sonidos cardíacos posean múltiples frecuencias de resonancia. Este razonamiento permite ver la necesidad de describir la señal FCG, no solo en términos

de un ritmo o una frecuencia de resonancia única, sino como un espectro compuesto de múltiples frecuencias dominantes o de resonancia. Además, el flujo de sangre que pasa a través de una válvula defectuosa produce una turbulencia que resulta en un ruido de banda ancha. En el caso de los soplos cardíacos, no es importante identificar ritmos o frecuencias de resonancia, la verdadera necesidad consiste en considerar la distribución de la energía o potencia de la señal sobre un amplio rango de frecuencias. De esta manera, se dirige el estudio hacia la noción conocida como *densidad espectral de potencia* [68].

La representación tiempo-frecuencia ha comprobado mejorar los resultados obtenidos con métodos espectrales y temporales clásicos al ser capaz de reflejar cambios en frecuencia con respecto al tiempo (transitorios espectrales). Como primera aproximación a las transformadas tiempo-frecuencia consiste en analizar la señal con transformadas de Fourier consecutivas y de corta duración (STFT). Así, tomando pequeñas ventanas de la señal se apreciará que la distribución espectral de cada una de ellas varía con el tiempo, permitiendo conocer con más precisión la evolución de las componentes espectrales. La equivalencia directa dentro de las representaciones cuadráticas es el espectrograma, aunque la baja resolución que se obtiene es un problema abierto para resolver [91].

Las representaciones tiempo-frecuencia lineales son ampliamente utilizadas debido a su sencillez y similitud con los métodos espectrales clásicos en cuanto a concepto y estrategia de cálculo. Reciben este nombre ya que en su obtención, la señal a analizar no se multiplica por sí misma como ocurre en las representaciones cuadráticas o bilineales , y además cumplen con el principio de superposición. También se les llama descomposiciones atómicas, ya que proyectan la señal en el plano de forma que se obtienen celdas que conforman la representación global, cuyo tamaño depende del tipo de representación, frecuencia de muestreo, número de puntos elegidos, etc. Las representaciones tiempo-frecuencia lineales usadas con más frecuencia en el reconocimiento de patrones están: la transformada de Fourier de tiempo corto (STFT), la transformada de Gabor, análisis multiresolución y la transformada wavelet [93]. En general, el espacio de representación mediante el modelo de representación espectro-temporal queda conformado usando el siguiente modelo:

$$\mathbf{X}_2 = \mathcal{L}_1\{\mathcal{L}_2\{\mathbf{S}\}\} \text{ tal que, } \mathcal{L}_1 = \{\mathcal{E}_T, \mathcal{E}_S, \ldots\} \text{ y } \mathcal{L}_2 = \{\mathcal{F}, \mathcal{W}, \ldots\} \tag{2.5}$$

donde $\mathbf{X}_2$ es la estructura multivariada resultante, $\mathbf{S}$ es el conjunto de señales para el entrenamiento, $\mathcal{L}_1$ es un operador de envolvente basado por ejemplo en la energía de Shannon [51], energía de Teager [44], *eigen*–frecuencia [4], entre otras y $\mathcal{L}_2$ es un operador de mapeo a un espacio de representación espectro-temporal basado en las transformadas mencionadas anteriormente (STFT, WT, etc).

2.2.1. El espectro de Fourier

La *transformada de Fourier* es la técnica de uso más común para el análisis de las señales en el dominio de la frecuencia. Esto es principalmente porque la transformada de Fourier usa funciones sinusoidales como funciones base. Así, la señal $x(t)$ es proyectada sobre la función base exponencial compleja de frecuencia $\omega = 2\pi f$, obteniéndose la representación de Fourier $X(e^{j\omega})$, como:

$$X(e^{j\omega}) = \int_{-\infty}^{\infty} x(t)e^{-j\omega t}dt$$

El límite inferior del integral será cero si la señal es causal; el límite superior será igual a la duración de la señal para el caso en que la duración sea finita. El valor de $X(e^{j\omega})$ en cada frecuencia de interés ω representa la cantidad de funciones de seno y coseno presentes en la señal $x(t)$. La transformada inversa representa la señal $x(t)$ como una combinación pesada de la función exponencial compleja, y está dada de la forma

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(e^{j\omega}) e^{j\omega t} d\omega$$

En algunas situaciones, la transformada de Fourier está definida con $e^{+j\omega t}$ en el integral y la transformada inversa con $e^{-j\omega t}$, pero esta distinción no es significativa [68]. De otra forma, la *transformada de Fourier en tiempo discreto* de una secuencia $x[n]$, está definida por

$$X\left(e^{j\omega}\right) = \sum_{n=-\infty}^{\infty} x\left[n\right] e^{-j\omega n}$$

En general, $X\left(e^{j\omega}\right)$ es una función compleja de la variable real ω y puede ser escrita como

$$X\left(e^{j\omega}\right) = X_{\mathrm{re}}\left(e^{j\omega}\right) + jX_{\mathrm{im}}\left(e^{j\omega}\right)$$

donde $X_{\mathrm{re}}\left(e^{j\omega}\right)$ y $X_{\mathrm{im}}\left(e^{j\omega}\right)$ son, respectivamente, las partes real e imaginaria de $X\left(e^{j\omega}\right)$, y son funciones reales de ω. Alternativamente, $X\left(e^{j\omega}\right)$ puede ser expresada en la forma

$$X\left(e^{j\omega}\right) = \left|X\left(e^{j\omega}\right)\right| e^{j\theta(\omega)} \text{ donde } \theta\left(\omega\right) = \arg\left\{X\left(e^{j\omega}\right)\right\} \tag{2.6}$$

La expresión $\left|X\left(e^{j\omega}\right)\right|$ es conocida como *función magnitud* y $\theta\left(\omega\right)$ como *función fase*. En muchas aplicaciones, la transformada de Fourier es llamada *espectro de Fourier*. De la ecuación (2.6) se observa que si se reemplaza $\theta(\omega)$ por $\theta(\omega) + 2\pi k$, donde k es cualquier entero, $X\left(e^{j\omega}\right)$ no cambia, lo cual significa que la función fase no es única para cualquier transformada de Fourier, a menos que sus valores sean restringidos en el siguiente rango: $-\pi \leq \theta(\omega) < \pi$. Así, los coeficientes $x[n]$ pueden ser calculados a partir de $X\left(e^{j\omega}\right)$ usando la integral de Fourier, dada por

$$x\left[n\right] = \frac{1}{2\pi} \int_{-\pi}^{\pi} X\left(e^{j\omega}\right) e^{j\omega n} d\omega$$

conocida como *transformada inversa de Fourier en tiempo discreto*.

2.2.2. Densidad espectral de potencia

La transformada de Fourier en tiempo discreto de la función de correlación es comúnmente llamada espectro de potencia o densidad espectral de potencia,

$$S_x\left(e^{j\omega}\right) = \sum_{k=-\infty}^{\infty} R_x(k) e^{-j\omega k}$$

Inversamente, la función de correlación puede ser calculada desde el espectro de potencia, por

$$R_x(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_x\left(e^{j\omega}\right) e^{j\omega k} d\omega$$

La transformada z puede ser usada en lugar de la transformada de Fourier en tiempo discreto, de forma que el espectro de potencia queda expresado por

$$S_x(z) = \sum_{k=-\infty}^{\infty} R_x(k) z^{-k}$$

Para un proceso estacionario en sentido amplio, el espectro de potencia es simétrico,

$$S_x\left(e^{j\omega}\right) = S_x\left(e^{-j\omega}\right)$$

y no negativo,

$$S_x\left(e^{j\omega}\right) \geq 0$$

La potencia media centralizada de un proceso estacionario en sentido amplio, puede ser escrito

$$E\left[|x(n)|^2\right] = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_x\left(e^{j\omega}\right) d\omega$$

2.2.3. Transformada de Fourier de tiempo corto (STFT)

La STFT está basada en el procedimiento clásico de la transformada de Fourier. La idea básica es introducir una noción de localización temporal usando una ventana $w(t)$ que se desplaza sobre toda la señal $s(t)$, de forma que el análisis espectral es desarrollado en los tramos de la señal correspondiente a cada ventana. Así, el análisis de tiempo-frecuencia es obtenido ya que que la localización de la ventana desplazada adiciona la dimensión del tiempo. La expresión para este análisis de tiempo-frecuencia está dada por:

$$S(e^{j\omega}) = \int_{-\infty}^{+\infty} s(t) w(\tau - t) e^{-j\omega\tau} d\tau$$

El espectrograma obtenido mediante la STFT consiste en subdividir la señal en un número de pequeños segmentos traslapados, asumiendo estacionariedad para cada uno de ellos. Estos segmentos son multiplicados por una ventana de longitud adecuada que permita reducir los efectos no deseados inducidos por la transformada de Fourier. La selección de la longitud de la ventana es el parámetro de mayor importancia debido a que es el que asegura la condición de estacionariedad. Teniendo en cuenta que las señales cardíacas son altamente no estacionarias, un problema difícil de resolver es el de asegurar la estacionariedad con una resolución en frecuencia aceptable, ya que la resolución en frecuencia es igual al inverso de la longitud de la ventana de análisis [22].

2.2.4. Transformada de Gabor

Gabor introdujo átomos de Fourier enventanados para medir las variaciones de frecuencia en los sonidos. Una ventana real y simétrica $g(t) = g(-t)$ es trasladada por un factor u y modulada por la frecuencia ϑ:

$$g_{u,\vartheta} = e^{i\vartheta t} g(t-u)$$

$g(t)$ es normalizada, tal que $||g_{u,\vartheta}|| = 1$ para cualquier pareja $(u, \vartheta) \in R^2$. El resultado de la transformada de Fourier enventanada de $f(t) \in \mathfrak{L}^2$, es:

$$Sf(u,\vartheta) = \langle f, g_{u,\vartheta} \rangle = \int_{-\infty}^{+\infty} f(t) g(t-u) e^{-i\vartheta t} dt$$

La multiplicación $f(t-u)$ de la transformada, localiza la integral de Fourier en la vecindad de $t = u$. Así, la densidad de energía (espectrograma), denotado por P_s, puede definirse:

$$P_S f(u,\vartheta) = |Sf(u,\vartheta)|^2 = \left| \int_{-\infty}^{+\infty} f(t) g(t-u) e^{i\vartheta t} dt \right|^2$$

El espectrograma mide la energía de $f(t)$ en la vecindad tiempo-frecuencia correspondiente a (u, ϑ), especificada por la caja *Heisenberg* de $g_{u,\vartheta}$.

Debido a que $g_{u,\vartheta}(t) = e^{i\vartheta t} g(t-u)$ está centrada en u, el tiempo dispuesto alrededor de u es independiente de u y ϑ:

$$\sigma_t^2 = \int_{-\infty}^{+\infty} (t-u)^2 |g_{u,\vartheta}(t)|^2 dt = \int_{-\infty}^{+\infty} t^2 |g(t)|^2 dt$$

Como g es real y simétrica, entonces la transformada de Fourier de g, notada como $\hat{g}_{u,\vartheta}(\omega)$, también es real y simétrica, de forma que:

$$\hat{g}_{u,\vartheta}(\omega) = \hat{g}(\omega - \vartheta) e^{-iu(\omega-\vartheta)}$$

lo cual es una traslación de la frecuencia de la ventana, quedando centrada en ϑ. La frecuencia dispuesta alrededor de ϑ es independiente de u y ϑ:

$$\sigma_\omega^2 = \frac{1}{2\pi} \int_{-\infty}^{+\infty} (\omega - \vartheta)^2 |\, \hat{g}_{u,\vartheta}(\omega)| d\omega = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \omega^2 |\hat{g}(\omega)| d\omega$$

Así, $g_{u,\vartheta}$ corresponde a una caja *Heisenberg* de área $\sigma_t \sigma_\omega$ y centrada en (u, ϑ). El tamaño de esta caja es independiente de (u, ϑ), es decir, la transformada de Gabor tiene la misma resolución a través del plano tiempo-frecuencia [54].

2.2.5. Análisis multiresolución y transformada wavelet

La descomposición multiresolución consiste en aproximaciones que son combinaciones lineales de dilataciones y traslaciones de una función de escalamiento, obteniendo aproximaciones en diferentes niveles de resolución. Este análisis ha cobrado importancia en el procesamiento de señales debido a que permite resaltar cualidades determinantes mediante

la representación que ofrece de la señal. Un análisis multiresolución en $\mathfrak{L}^2(\mathbb{R})$ se compone de una secuencia de subespacios cerrados embebidos $\{V_j : j \in Z\}$, satisfaciendo las siguientes condiciones:

1. $V_j \subset V_{j-1}; \quad j \in \mathbb{Z}$ (embebimiento)
2. $\overline{\bigcup_{j\in\mathbb{Z}} V_j} = \mathfrak{L}^2(\mathbb{R})$ (completitud hacia arriba)
3. $\bigcap_{j\in\mathbb{Z}} V_j = \{0\}$ (completitud hacia abajo)
4. $f(x) \in V_j \leftrightarrow f(2^j x) \in V_0$ (invariante en la escala)
5. $f(x) \in V_0 \leftrightarrow f(x-n) \in V_0 \ \ \forall n \in \mathbb{Z}$ (invariante en el desplazamiento)
6. $\exists\, \varphi \in V_0$ tal que $\{\varphi(x-n)|n \in \mathbb{Z}\}$ es una base ortonormal para V_0 (existencia de una base)

La función $\varphi(t)$ es llamada función de escalamiento e integra la unidad. También, $\varphi_{j,n}(x) = \{2^{j/2}\varphi(2^j x - n)|n \in \mathbb{Z}\}$ es una base para V_{-j}. La ortogonalidad de $\varphi(t)$ no es una condición necesaria, ya que una base no ortonormal puede ser ortogonalizada con la propiedad de desplazamiento. Tomando en cuenta las propiedades de escalamiento y embebimiento puede verificarse que la función $\varphi(t)$ satisface una ecuación bi-escalar. En efecto, como $\varphi(t) \in V_0 \subset V_{-1}$ entonces existe en $\mathfrak{L}^2$ una secuencia única $\{h_n : n \in \mathbb{Z}\}$, llamada secuencia bi-escalar, que satisface la ecuación bi-escalar:

$$\varphi(x) = \sqrt{2}\sum_{n\in\mathbb{Z}} h_n \varphi(2x - n)$$

Se denota el complemento ortogonal de V_{j+1} en V_j como W_{j+1} para cada j, lo cual puede expresarse como:

$$V_j = V_{j+1} \oplus W_{j+1}$$

donde $\oplus$ representa la suma directa con $V_{j+1} \perp W_{j+1}$. Cada subíndice j se relaciona con un nivel de resolución 2^j. De esta forma, la máxima resolución podría notarse como 2^J, de lo cual se obtiene:

$$V_0 = W_1 \oplus W_2 \oplus \cdots \oplus W_J \oplus V_J$$

Los espacios $\{W_i\}_{i=1}^J$ son conocidos como niveles de detalle o espacios wavelet (representación de alta frecuencia) y V_J como el nivel de aproximación (representación de baja frecuencia) [16].

2.3. Análisis de complejidad

El análisis lineal atribuye el comportamiento irregular de un sistema a la naturaleza aleatoria de la señal de entrada, sin embargo, las entradas aleatorias no son las únicas fuentes posibles de irregularidad en la salida de un sistema [34, 45]. Una forma de superar este inconveniente consiste en atribuir las irregularidades de la salida del sistema a estructuras

no lineales impuestas en el modelo dinámico, ya que los *sistemas dinámicos* no lineales se caracterizan porque su respuesta ante un estímulo dado, en algún instante de tiempo, no se produce de forma proporcional a su respectiva excitación de entrada [45]. En ciertas circunstancias, los *sistemas no lineales deterministas* entran en un estado llamado *caos*, cuando la respuesta presenta una notable sensibilidad a las condiciones iniciales del modelo; es por esto, que aunque exista un modelo determinista del sistema no lineal, al entrar en el régimen caótico, se genera un comportamiento impredecible a largo plazo si no se tienen de manera precisa las condiciones iniciales [64]. A finales de los 80's, se desarrolló una hipótesis sobre la naturaleza fractal y el posible comportamiento caótico de las señales biológicas del cuerpo humano [29]. De acuerdo a esto, se propuso el uso de técnicas de dinámica no lineal para el procesamiento y caracterización de las señales fisiológicas [77]. Generalmente, el análisis de señales derivado de técnicas fractales o de dinámica no lineal es denominado *análisis de complejidad* y siguen el siguiente modelo de representación:

$$\mathbf{X}_3 = \mathcal{MC}\{\mathbf{S}\} \tag{2.7}$$

donde $\mathbf{X}_3$ es la estructura multivariada resultante, $\mathbf{S}$ es el conjunto de señales para el entrenamiento, $\mathcal{MC}$ es un operador no lineal basado en diferentes técnicas de complejidad. Es importante anotar que el operador no lineal $\mathcal{MC}$, generalmente mapea la señal a un espacio de embebimiento de dimensión más alta donde se realizan las mediciones respectivas.

2.3.1. Determinación de la no linealidad determinista

El paso previo a la caracterización de señales mediante análisis de complejidad, es determinar si la serie de tiempo contiene estructuras no lineales deterministas, con el fin de asegurar la consistencia en la estimación de las características y que la representación obtenida en realidad corresponde a la dinámica intrínseca de la señal. En [87] se presenta el desarrollo de un método para detectar determinismo de naturaleza no lineal en series de tiempo, conocido como el método de los *datos sustitutos*, el cual es utilizado y aceptado como uno de los mejores métodos para detectar dinámicas no lineales deterministas [66].

El método de los datos sustitutos consiste en tomar una serie de tiempo $x[n]$ y a partir de algún principio de sustitución se forma una nueva secuencia de datos $\hat{x}[n]$, luego se estiman medidas dinámicas sobre $x[n]$ y $\hat{x}[n]$ con el fin de comparar los resultados de estimación. Si los resultados al aplicar medidas dinámicas son significativamente diferentes, antes y después de sustituir los datos, se concluye que existe una estructura no lineal determinista en la serie original. Básicamente, existen dos formas de procedimiento para obtener datos sustitutos. El primero es muy sencillo y consiste en asignar un orden aleatorio a los datos, con lo cual se conserva la distribución de probabilidad, pero se generan funciones de correlación y espectros de potencia diferentes. Este ordenamiento se puede lograr permutando los datos de forma aleatoria sin reemplazo [80]. El segundo método (ver Algoritmo 2.1), consiste en obtener la transformada de Fourier de la serie de tiempo, luego aleatorizar la fase entre 0 y 2π, y posteriormente realizar la transformada inversa de Fourier. Este procedimiento genera una serie de datos sustitutos que mantiene el mismo espectro de Fourier y función de autocorrelación que la serie original, aunque la distribución de probabilidad es diferente [87].

Algoritmo 2.1 Cálculo de los valores sustitutos usando la transformada de Fourier

Require: $x[n]$, $y[n] = 0$, $n = 1, \ldots, N$

$z[n] = x[n] + jy[n]$ ⋆\Conformar arreglo complejo a partir de la serie de entrada\⋆

$Z(w) = X(w) + jY(w) = \frac{1}{N} \sum_{n=1}^{N} z[n] e^{-2\pi j(w-1)(n-1)/N}$ ⋆\Cálculo de la transformada de Fourier\⋆

$\theta_w \in [0, \pi]$, $w = 2, 3, \ldots, \frac{N}{2}$ ⋆\Conjunto de fases aleatorias\⋆

for $w = 1, \ldots, N$ **do**

$$Z(w)' = \begin{cases} Z(w), & w = 1, w = \frac{N}{2} + 1 \\ |Z(w)|\, e^{j\theta_w}, & w = 2, 3, \ldots, \frac{N}{2} \\ |Z(N - w + 2)|\, e^{-j\theta_{N-w+2}}, & w = \frac{N}{2} + 2, \frac{N}{2} + 3, \ldots, N \end{cases}$$ ⋆\Se aleatoriza la fase\⋆

end for

$z[n]' = x[n]' + jy[n]' = \frac{1}{N} \sum_{w=1}^{N} Z(w)' e^{2\pi j(w-1)(n-1)/N}$ ⋆\Cálculo de la transformada inversa de Fourier\⋆

Output: $\hat{x}[n] = x[n]'$ ⋆\Esta es la serie de datos sustitutos\⋆

2.3.2. Reconstrucción del espacio de estados

Un *sistema no lineal determinista* puede definirse mediante un conjunto de ecuaciones diferenciales no lineales de la forma:

$$\dot{\mathbf{x}}(t) = \boldsymbol{f}\left\{\mathbf{x}(t)\right\} \tag{2.8}$$

donde $\boldsymbol{f}$ es una función vectorial no lineal que representa las reglas dinámicas que gobiernan el comportamiento de las variables, y $\mathbf{x}(t) = [x_1(t), \ldots, x_m(t)]$ es el vector de estados con m grados de libertad, los cuales están dados por la cantidad de ecuaciones diferenciales ordinarias autónomas de primer orden que se requieren para describir la dinámica del sistema [1, 34]. Cada una de las soluciones de la ecuación (2.8) corresponde a una *trayectoria* u *órbita* en un espacio con dimensión múltiple conocido como *espacio de estados* o *espacio de fase*, el cual es un espacio vectorial abstracto generado por las variables dinámicas del sistema que representan la evolución en el tiempo [45].

La reconstrucción del espacio de estados, a partir de una serie de tiempo, se basa en el teorema de embebimiento presentado en [85] y puede ser enunciado como sigue [75]:

Teorema 2.1 (Teorema de embebimiento)

Dado un sistema dinámico con un espacio de solución m-dimensional y una solución $h(t)$, sea x alguna observación $x\left(h(t)\right)$. Tómese, $\mathbf{x}(t) = (x_t, x_{t-\tau}, x_{t-2\tau} \ldots, x_{t-(m-1)\tau})$ como el t-ésimo vector de retardo, con dimensión m y tiempo común de retardo τ. Luego, bajo condiciones generales, el espacio de vectores $\mathbf{x}(t)$ generado por la dinámica contiene toda la información del espacio de vectores solución $\mathbf{h}(t)$. El mapeo entre ellos es suave e invertible. Esta propiedad es referida como difeomorfismo y esta clase de mapeo determina un embebimiento. Así, el estudio de la serie de tiempo $x(t)$ es también el estudio de las soluciones del sistema dinámico subyacente $\mathbf{h}(t)$ a través de un sistema particular de coordenadas dado por la x observable.

El Teorema 2.1 establece que, cuando se tienen mediciones muestreadas de la señal proveniente de un sistema, es posible reconstruir un espacio de estados *difeomórfico* al espacio de estados original, aunque desconocido, compuesto por todas las variables dinámicas del sistema [64].

Difeomorfismo

Para enunciar formalmente el concepto de *difeomorfismo*, primero es necesario considerar las Definiciones 2.1 y 2.2 correspondientes a *homeomorfismo* y *variedad diferenciable*, respectivamente.

Definición 2.1 (Homeomorfismo)

> Una aplicación biyectiva $f : X \to Y$ entre dos espacios topológicos es un *homeomorfismo* si f y su inversa son continuas. Dos espacios topológicos son *homeomorfos* si existe un homeomorfismo entre ellos.

Un homeomorfismo induce una biyección entre los abiertos de los espacios, por lo que ambos son topológicamente indistinguibles. Es claro que cualquier propiedad definida exclusivamente a partir de la topología se conserva por homeomorfismos, luego dos espacios homeomorfos tienen las mismas propiedades topológicas.

Definición 2.2 (Variedad diferenciable)

> Un conjunto $S \subset \mathbb{R}^m$ es una *variedad diferenciable* de dimensión $n \leq m$ y de clase C^q, si para cada punto $p \in S$ existe un entorno V de p, un abierto U en $\mathbb{R}^n$ y una función $X : U \to \mathbb{R}^m$ de clase C^q de modo que el rango de la matriz jacobiana de X sea igual a n en todos los puntos y $X : U \to S \cap V$ sea un homeomorfismo. Una aplicación X en estas condiciones se llama *carta* de S alrededor de p.

Etimológicamente, se entiende por *variedad* a un conjunto cuyos elementos vienen determinados por varias coordenadas. En este contexto, una variedad de clase C^q significa que es continua y diferenciable q veces, con q lo suficientemente grande como para que supere cualquier consideración. La palabra *carta* hay que entenderla en el sentido de *mapa*. En efecto, podemos pensar en U como un *mapa plano* de una región de S, y la aplicación X es la que hace corresponder cada punto del mapa con el punto real que representa (ver Figura 2.3). Alternativamente, se puede considerar X^{-1} como una aplicación que asigna a

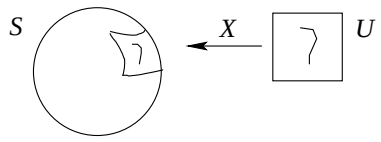


Figura 2.3. Mapa plano de una región

cada punto $p \in S \cap V$ unas coordenadas $x = (x_1, \ldots, x_n) \in U \subset \mathbb{R}^n$, de forma análoga a los *sistemas de coordenadas* en un espacio afín.

Definición 2.3 (Difeomorfismo)

> Un homeomorfismo entre 2 variedades diferenciables se llama *difeomorfismo*. Dos variedades son difeomorfas si existe un difeomorfismo entre ellas.

De esta manera, las cartas de una variedad son difeomorfismos en su imagen.

Parámetros de embebimiento

Un espacio topológico difeomorfo al espacio de estados de un sistema, en el sentido de la Definición 2.3, puede ser reconstruido con el método de vectores de retardo de acuerdo al

Teorema 2.1 [45]. La dimensión de embebimiento m y el tiempo de retardo τ, son comúnmente llamados *parámetros de embebimiento* y el ajuste adecuado de estos parámetros es información obligatoria para obtener la reconstrucción del espacio de estados, o de embebimiento, a partir de una serie de tiempo [85, 75]. Los vectores de retardo, también son conocidos como *regresores* y contienen toda la información disponible de los estados del sistema en un tiempo dado, de manera que si se toma cada regresor como un punto en el espacio de embebimiento, entonces el conjunto de regresores en ese espacio generan una figura geométrica de estructura irregular denominada *atractor* [79]. Sin embargo, la *reconstrucción de atractores* mediante vectores de retardo presenta serias dificultades relacionadas a los parámetros de embebimiento. Por ejemplo, la *dimensión del atractor* d_A es desconocida y no hay idea del valor mínimo que puede tomar m para que el atractor reconstruido sea *difeomorfo* al original. En [85], se presenta una relación que garantiza una dimensión de embebimiento adecuada: $m > 2d_A$; aunque esta relación asume el conocimiento de d_A, es una relación que bajo procedimiento iterativos puede ofrecer buenos resultados. Por otro lado, el retardo τ no está sujeto al Teorema 2.1; en principio, con cualquier valor arbitrario de τ se obtiene un embebimiento, de acuerdo a esto, la principal dificultad está relacionada con la observación de estructuras interesantes en el atractor reconstruido [33]. Si τ toma valores pequeños, elementos sucesivos de los vectores de retardo quedan fuertemente correlacionados y todos los puntos caerían cerca de la diagonal del espacio. Sin embargo, si τ toma valores grandes, vectores de retardo sucesivos aparecerán como causalmente no relacionados y el atractor reconstruido no representará la verdadera dinámica. Estos problemas reciben el nombre de *redundancia* e *irrelevancia*, respectivamente. Un criterio para estimar el valor de τ, propuesto en [72], es haciéndolo igual al tiempo en que la función de autocorrelación de la señal cae por debajo de la cota $1 - 1/e$ de su valor inicial. Aunque también es muy frecuente el método basado en el cálculo de la información mutua promedio (AMI) sobre la señal de análisis $x[n]$:

$$I(\tau) = \sum_{x[n],x[n+\tau]} P(x[n], x[n+\tau]) \log_2 \frac{P(x[n], x[n+\tau])}{P(x[n])P(x[n+\tau])}, \tag{2.9}$$

donde $P(x[n], x[n+\tau])$ es la densidad de probabilidad conjunta para $x[n]$ y $x[n+\tau]$, así como $P(s[n])$ y $P(s[n+\tau])$ son las densidades de probabilidad marginal para $x[n]$ y $x[n+\tau]$, respectivamente. De la expresión (2.9), se puede observar que a mayor dependencia estadística entre los puntos (valores pequeños de τ), el valor de $I(\tau)$ es más alto, mientras que para valores suficientemente grandes de τ, las series $x[n]$ y $x[n+\tau]$ pueden considerarse independientes, por lo que $I(\tau)$ tenderá a cero. El valor de τ se toma entonces donde se presenta el primer mínimo de la serie AMI, $I(\tau)$ [25]. El procedimiento para escoger el primer mínimo de la información mutua promedio es similar al que se utiliza para escoger el primer cero de la función de autocorrelación lineal

$$C(\tau) = \sum_n (x[n] - \bar{x})(x[n+\tau] - \bar{x})$$

donde, $\bar{x} = \frac{1}{N}\sum_{n=1}^{N} x[n]$ y N es el número de puntos de la serie de tiempo $x[n]$. Este método representa la manera lineal óptima para seleccionar τ, ya que se toma desde el

punto de predicción $x[n+\tau]$, a partir del conocimiento de $x[n]$, en el sentido de los mínimos cuadrados [1]. En general, existen formas diferentes para determinar valores adecuados de m y τ, pero ninguno de ellos muestra ser efectivo para obtener resultados óptimos en cada situación [33].

2.3.3. Medidas de complejidad

Análisis de dimensiones en el espacio de estados

El análisis del comportamiento caótico de *sistemas no lineales deterministas* implica el estudio de la *dimensión del atractor* reconstruido en el espacio de estados, o de embebimiento. Este análisis de dimensiones permite estimar el número de variables de estado independientes que serían necesarias para describir toda la dinámica del sistema.

Una estructura geométrica puede ser representada por un conjunto de puntos en un espacio euclídeo $\mathbb{R}^m$, con un valor de m lo suficientemente grande para localizar de forma única la posición de cada punto en la estructura. Cada conjunto en $\mathbb{R}^m$ tiene asignado una dimensión topológica $d_T \in [0, m]$. La *dimensión fractal* o *fraccional* (d_F) es una medida de dimensión que asume valores no enteros, de esta manera, una estructura fractal es un conjunto con dimensión no entera y el uso de la geometría euclídea no es conveniente para estimar su dimensión real, ya que la dimensión euclídea $d_E = D,\ D \in \mathbb{N}$. La importancia de los fractales en la dinámica no lineal, radica en que los atractores son estructuras geométricas de este tipo, y su dimensión d_F se relaciona con el número mínimo de variables necesarias para modelar la dinámica.

Un atractor reconstruido a partir de una serie de tiempo $x[n]$ con $n = 1, \ldots, N$, se puede representar por una secuencia de puntos m-dimensionales, de acuerdo al Teorema 2.1, así:

$$\boldsymbol{S}[k] = [x[k], x[k+\tau], x[k+2\tau], \ldots, x[k+(m-1)\tau]] \tag{2.10}$$

donde $k = 1, \ldots, l$ y $l = N-(m-1)\tau$. Sea v la cantidad de puntos m-dimensionales dentro de una hiperesfera de radio r y centro en $\mathbf{s}$. Se establece algún punto $\boldsymbol{s} \in \boldsymbol{S}$ sobre el atractor reconstruido, y se evalúa el cambio de v a medida que r se hace más pequeño. Es importante anotar que r no puede ser tan pequeño que excluya todos los puntos, ni tan grande que los contenga a todos, esto último es, $r \approx R_A$, siendo R_A el tamaño del atractor. Así, v depende de r y de la ubicación del punto $\boldsymbol{s}$, mediante la relación $v(\boldsymbol{s}, r) \approx r^{d(\boldsymbol{s})}$, donde r puede variar de acuerdo al rango de magnitud $0 < r/R_A < 1$ y $d(\boldsymbol{s})$ es la dimensión del atractor en ese punto (dimensión local). Cuando un atractor posee una estructura con geometría irregular, no tiene una dimensión topológica constante en todos sus puntos y su densidad es heterogénea. Este tipo de estructuras son conocidas como *atractores extraños*. Para el caso en que el atractor fuera una figura geométrica regular de dimensión entera D, entonces se encontraría una cantidad de puntos r^D en cada hiperesfera de radio r alrededor de cada punto $\boldsymbol{s}$, por lo que $d(\boldsymbol{s}) = D$, $\forall \boldsymbol{s}$ [1].

Con el propósito de estimar la *dimensión fractal* de un atractor extraño reconstruido a partir de la expresión (2.10), se calcula la cantidad de puntos contenidos en la hiperesfera de radio r alrededor del punto $\boldsymbol{s}$ en el espacio de estados, mediante la expresión:

$$v(\boldsymbol{s}, r) = \frac{1}{l} \sum_{k=1}^{l} \Theta(r - |\boldsymbol{S}[k] - \boldsymbol{s}|) \tag{2.11}$$

donde $\Theta(\cdot)$ es la función escalón unitario (función de *Heaviside*) y l es el número de puntos del atractor $\boldsymbol{S}[k]$. La expresión (2.11) realiza el conteo de todos los puntos sobre la órbita $\boldsymbol{S}[k]$, dentro de una hiperesfera de radio r y centro en $\boldsymbol{s}$, y esa cantidad es normalizada por el número total de puntos. A partir de esta expresión, se obtiene la *suma de correlación generalizada* $C(q,r)$, entendida como la estimación de la cantidad de puntos cercanos y situados a una distancia r con respecto al total de puntos del atractor, la cual puede expresarse como:

$$C(q,r) = \frac{1}{l}\sum_{i=1}^{l}\left(\frac{1}{l}\sum_{j=1}^{l}\Theta\left(r - |\boldsymbol{S}[i] - \boldsymbol{S}[j]|\right)\right)^{(q-1)} \tag{2.12}$$

Si r es menor que el tamaño del atractor, además mayor que el espacio más pequeño entre los puntos, entonces $C(q,r) \approx r^{(q-1)D_q}$, siendo la dimensión fractal expresada como $d_F = D_q$ [1, 7]. Sin embargo, no es conveniente tomar en cuenta toda la información de $C(q,r)$, sino concentrar el análisis en la pendiente del segmento lineal del gráfico que se obtiene al relacionar $\log C(q,r)$ contra $\log r$. De esta forma, la dimensión fractal D_q está definida por el límite:

$$D_q = \lim_{r\to 0}\frac{\log C(q,r)}{(q-1)\log r} \tag{2.13}$$

A partir de la expresión (2.13), conocida como *dimensión fractal generalizada*, son estimadas frecuentemente para el análisis de complejidad, la dimensión de capacidad D_0 y la dimensión de correlación D_2. La *dimensión de capacidad* ó *dimensión box-counting*, es una forma simple de obtener la dimensión fractal de un atractor, ya que la evaluación consiste en determinar cuántas esferas de radio r son necesarias para cubrir todos los puntos del conjunto de datos, con este fin, se establece una función $N(r)$ para valores de r pequeños. De acuerdo a la expresión (2.13), la dimensión D_0 se expresa como:

$$D_0 = \lim_{r\to 0}\frac{\log N(r)}{\log \frac{1}{r}} \tag{2.14}$$

Así, de (2.14), se puede observar que para valores muy pequeños de r, $N(r) \approx r^{-D_0}$.

Dimensión de correlación

La dimensión fractal del atractor puede estimarse usando la expresión (2.12), cuando $q = 2$, también conocida como *suma de correlación* [30], y que sigue la siguiente aproximación: la probabilidad que dos puntos del conjunto estén en la misma hiperesfera de radio r, es aproximadamente igual a la probabilidad de que dos puntos del conjunto estén separados por una distancia $\rho \leq r$, esto es:

$$\rho\left(|\boldsymbol{S}[i] - \boldsymbol{S}[j]|\right) = \sqrt{\sum_{\gamma=1}^{m}\left(\boldsymbol{S}_\gamma[i] - \boldsymbol{S}_\gamma[j]\right)^2},$$

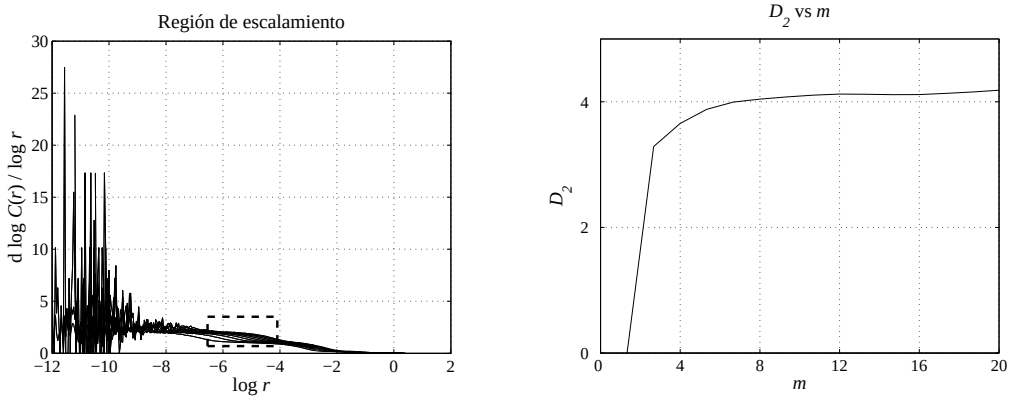
donde γ indica las coordenadas del espacio de estados, por lo cual, la suma de correlación $C(r)$ puede aproximarse a la forma:

$$C(r) \approx \frac{1}{l^2} \sum_{i=1}^{l} \sum_{j=1}^{l} \Theta \left(r - \rho \left| \boldsymbol{S}[i] - \boldsymbol{S}[j] \right| \right) \tag{2.15}$$

La *dimensión de correlación* (D_2) se obtiene a partir de la suma de correlación $C(r)$, determinando la región en la cual la función $\log C\left(r\right)$ es lineal gráficamente con respecto a $\log r$ para cada valor de m. A este rango de valores de $\log r$ se le denomina *región de escalamiento*. La dimensión D_2 se aproxima a la pendiente encontrada en la región de escalamiento. De (2.13), el cálculo de D_2 puede expresarse como sigue:

$$D_2 = \lim_{r \to 0} \frac{\log C(r)}{\log r}$$

Uno de los más grandes inconvenientes presentes en el cálculo de la dimensión de correlación es la determinación adecuada de la región de escalamiento. Primero se debe derivar la función $\log C\left(r\right)$, y se determinan los valores de $\log r$ donde se encuentra la *meseta* de la derivada (Figura 2.4(a)). Posteriormente, de acuerdo al rango de valores de $\log r$ determinados según la observación de la meseta en la gráfica de $d \log C\left(r\right) / d\left(\log r\right)$, se establecen los valores de $\log r$ donde se debe estimar las pendientes de las rectas correspondientes a la función $\log C\left(r\right)$, para cada valor de m. Normalmente, D_2 aumenta con respecto a m, hasta que converge a un valor determinado (ver Figura 2.4(b)), lo cual quiere decir que el atractor reconstruido es acotado en el espacio de estados. Así, la precisión en la estimación de D_2 depende fuertemente de la ubicación adecuada de la meseta mostrada en la Figura 2.4(a), que en este caso se presenta para valores de $\log r$ entre -6 y -4 aproximadamente.



(a) Derivada de la función $\log C\left(r\right)$ (b) Dimensión de correlación

Figura 2.4. Estimación de D_2 para una señal FCG

Máximo exponente de Lyapunov

Una de las principales características de los *sistemas no lineales deterministas* es su *impredecibilidad* durante prolongados periodos de tiempo, debido a la inherente inestabilidad de

las soluciones y la fuerte dependencia a las condiciones iniciales. A medida que las trayectorias de un *sistema caótico* evolucionan temporalmente en el espacio de estados, se presenta una separación entre las mismas que crece a una tasa exponencial. En las series de tiempo, predominantemente periódicas, también se presenta una separación entre las trayectorias durante su evolución temporal, pero es de forma lenta y no exponencial. Por lo tanto, se puede cuantificar esta separación en un instante determinado mediante el cálculo de un exponente que relaciona el incremento de la distancia entre las trayectorias, conocido como *máximo exponente de Lyapunov* [45].

Sean dos trayectorias $\boldsymbol{S}_\alpha[k]$ y $\boldsymbol{S}_\beta[k]$ en el espacio de estados del mismo sistema, pero evolucionadas a partir de condiciones iniciales diferentes (el instante inicial se notará con k_0), $\boldsymbol{S}_\alpha[k_0] = \boldsymbol{\alpha}$ y $\boldsymbol{S}_\beta[k_0] = \boldsymbol{\beta}$, con una separación inicial de $\boldsymbol{u}[k_0] = \|\boldsymbol{\alpha} - \boldsymbol{\beta}\| \ll 1$ (Ver Figura 2.5). En la determinación de los valores de predicción $\boldsymbol{S}_\alpha[k+1] = \boldsymbol{f}\left(\boldsymbol{S}_\alpha[k]\right)$ y $\boldsymbol{S}_\beta[k+1] = \boldsymbol{f}\left(\boldsymbol{S}_\beta[k]\right) = \boldsymbol{f}\left(\boldsymbol{S}_\alpha[k] + \boldsymbol{u}[k]\right)$, se puede emplear una aproximación lineal que describa la desviación de las trayectorias $\boldsymbol{S}_\alpha[k]$ y $\boldsymbol{S}_\beta[k]$, dada por:

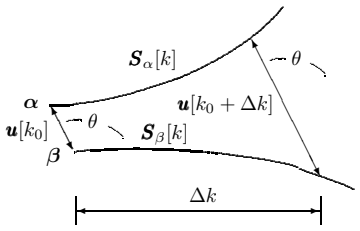


Figura 2.5. Divergencia de trayectorias

$$\boldsymbol{u}[k+1] = \boldsymbol{J}\left(\boldsymbol{S}_\alpha\left[k\right]\right)\boldsymbol{u}[k] \tag{2.16}$$

donde $\boldsymbol{J}(\boldsymbol{S}_\alpha[k])$ es la matriz del jacobiano evaluada desde un punto de $\boldsymbol{S}_\alpha[k]$. Después que las trayectorias $\boldsymbol{S}_\alpha[k]$ y $\boldsymbol{S}_\beta[k]$ han evolucionado temporalmente Δk iteraciones, la distancia de separación entre las trayectorias, está dada por $\boldsymbol{u}[k_0+\Delta k] = \|\boldsymbol{S}_\alpha[k_0+\Delta k] - \boldsymbol{S}_\beta[k_0+\Delta k]\|$. Usando la expresión (2.16), esta distancia de separación también puede expresarse como:

$$\boldsymbol{u}[k_0 + \Delta k] = \boldsymbol{J}[k_0 + \Delta k]\boldsymbol{u}[k_0]$$

donde $\boldsymbol{J}[k_0+\Delta k]$ es la matriz formada por el producto de las matrices jacobianas evaluadas en los estados que componen la trayectoria desde el inicio hasta $\boldsymbol{S}_\alpha[k_0 + \Delta k - 1]$, como sigue:

$$\boldsymbol{J}\left[k_0 + \Delta k\right] = \boldsymbol{J}\left(\boldsymbol{S}_\alpha\left[k_0 + \Delta k - 1\right]\right)\boldsymbol{J}\left(\boldsymbol{S}_\alpha\left[k_0 + \Delta k - 2\right]\right)\cdots\boldsymbol{J}\left(\boldsymbol{S}_\alpha\left[k_0\right]\right)$$

Con el propósito de medir la *separación exponencial* de las trayectorias, se asume que en el futuro distante ($\Delta k \gg 0$), la norma del vector $\boldsymbol{u}[k_0 + \Delta k]$ se comporta como

$$\|\boldsymbol{u}[k_0 + \Delta k]\| = \|\boldsymbol{u}[k_0]\|\, e^{(\lambda_1 \Delta k)} \tag{2.17}$$

donde $\lambda \in \mathbb{R}$, siendo $\boldsymbol{u}[k_0] \ll 1$ y $\Delta k \gg 1$. Se define entonces el máximo exponente de Lyapunov (λ_1), de la forma:

$$\begin{aligned}\lambda_1(\boldsymbol{S}_\alpha[k_0], \boldsymbol{u}[k_0]) &= \lim_{\Delta k \to \infty} \frac{1}{\Delta k} \ln \|\boldsymbol{J}[k_0 + \Delta k]\boldsymbol{u}[k_0]\| \\ &= \lim_{\Delta k \to \infty} \frac{1}{\Delta k} \ln \|\boldsymbol{J}(\boldsymbol{S}_\alpha[k_0 + \Delta k - 1])\boldsymbol{J}(\boldsymbol{S}_\alpha[k_0 + \Delta k - 2]) \cdots \boldsymbol{J}(\boldsymbol{S}_\alpha[k_0])\boldsymbol{u}[k_0]\|\end{aligned}$$

El parámetro $\lambda_1(\boldsymbol{S}_\alpha[k_0], \boldsymbol{u}[k_0])$ caracteriza en el futuro distante ($\Delta k \gg 0$), la separación

exponencial de las trayectorias $\boldsymbol{S}_\alpha[k]$ y $\boldsymbol{S}_\beta[k]$. El máximo exponente de Lyapunov depende entonces de la condición inicial $\boldsymbol{S}_\alpha[k_0]$ correspondiente a la trayectoria de referencia, así como de la separación inicial $\boldsymbol{u}[k_0]$ entre las trayectorias. Si λ_1 es positivo, significa que existe una divergencia exponencial entre las trayectorias, por lo que si las trayectorias son acotadas en el espacio de estados, se indica la presencia de caos. Dos trayectorias no se pueden separar a una distancia mayor que el tamaño del atractor, por lo que (2.17) es válida solo para tiempos Δk en los que $\boldsymbol{u}[k_0 + \Delta k]$ permanece pequeño. La cantidad de exponentes de Lyapunov es la misma que el número de variables de estado, los cuales se ordenan como $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_m$, siendo λ_1 el máximo exponente de Lyapunov. Si $\lambda_1 > 0$ hay expansión, si $\lambda_1 = 0$ existen puntos de bifurcación, donde el comportamiento cambia de regular a caótico y viceversa, y si $\lambda_i < 0$ hay contracción en alguna dirección específica o hay tendencia en la trayectoria a ser periódica.

Exponente de Hurst

El exponente de Hurst (H) cuantifica la dinámica de la serie de tiempo y permite ponderar hasta qué punto los datos pueden ser representados por un movimiento de tipo browniano. Si H existe, sus valores se encontrarán en un rango entre 0 y 1, evidenciando el comportamiento no lineal de la serie de tiempo analizada. La Tabla 2.2 establece el comportamiento de una serie de tiempo de acuerdo al valor del exponente de Hurst [15].

Valor de H	*Comportamiento a largo plazo de la serie de tiempo*
$0 \leq H < 0.5$	Las tendencias del pasado serán inversas en el futuro
$H = 0.5$	Comportamiento similar a ruido blanco o movimiento browniano
$0.5 < H \leq 1$	Las tendencias del pasado permanecerán en el el futuro

Tabla 2.2. Comportamiento de una señal según el exponente de Hurst (H)

En la estimación de H a partir de una serie de tiempo $x[n]$ con $n = 1, \ldots, N$, primero se debe definir un rango de variación comparable con el de las fluctuaciones de la señal, para esto, se obtiene una secuencia $\varrho[i]$ de desviaciones acumuladas sobre segmentos de la señal:

$$\varrho[i] = \sum_{n \in M_i} x[n] - \mu_{M_i}$$

donde M_i es el segmento i de la señal $x[n]$ compuesto de M puntos y μ_{M_i} es la media estimada sobre el conjunto de puntos correspondiente al segmento M_i. De esta forma, el rango de variación $R(M)$ se obtiene como sigue,

$$R(M) = \max_i \{\varrho[i]\} - \min_i \{\varrho[i]\}$$

Así, la estimación de H se obtiene como la pendiente de la relación logarítmica de la siguiente expresión:

$$(cM)^H = \frac{R(M)}{\sigma}$$

donde c es una constante y σ es la desviación estándar estimada sobre todos los puntos

de la señal $x[n]$. El cálculo de $R(M)/\sigma$ puede ser inexacto cuando el tamaño de M es demasiado grande o demasiado pequeño, por lo cual, para la estimación de H se procura ajustar el valor de M por medio de procedimientos experimentales.

Uno de los obstáculos más fuertes en el cálculo de parámetros de complejidad es el gasto de cómputo. Sin embargo, existen algunos algoritmos rápidos basados en el análisis de reescalamiento [10], o análisis de dispersiones [11], para la estimación de H. El primero da buenos resultados para dinámicas con $H > 0.5$, mientras que el segundo da buenos resultados para dinámicas con $H < 0.5$. La virtud de estos procedimientos es que tienen un buen desempeño para señales de corta duración, además de no requerir la reconstrucción del atractor en el espacio de estados, ya que se basan en cálculos estadísticos y espectrales de la señal (se emplea el concepto de autosemejanza).

2.4. Estado del arte en el análisis de señales cardíacas

2.4.1. Filtración

Esta primera etapa de procesamiento consiste en grupo de técnicas que son claramente visibles en el estado del arte para las señales electrocardiográficas, pero en las señales fonocardiográficas, la exclusión de sonidos que no ofrecen información de la actividad valvular cardíaca, usualmente se realiza en el proceso de segmentación. A continuación se expone el conjunto de técnicas más importantes dentro del proceso de filtración de señales ECG.

- Filtración
  - Técnicas clásicas de filtración
    - IIR
    - FIR
  - Técnica de filtración por media móvil
  - Técnica basada en el promedio de latidos
  - Técnica basada en la aproximación mediante funciones
    - Por segmentos
    - Funciones ortogonales
    - Polinomios
    - LMS (mínimos cuadrados)
    - Interpolación
  - Técnica basada en filtros globales
  - Técnicas basadas en la transformada *wavelet*
    - Descomposición y recomposición de niveles
    - Técnicas vectoriales de proyección
    - Adaptativo tiempo-escala
    - *Waveshrink*
    - De acuerdo a la umbralización
    - Por supresión de polinomios
  - Técnicas de filtración adaptativa
    - Estimación incremental
    - Notch adaptativo
    - Filtros adaptativos digitales
    - LMS y CLMS Adaptativo
    - Filtro de Ahlstrom y Tompkins
    - Filtro adaptativo de entrada correlacionada (AICF)
    - Filtros polinomiales deformados en el tiempo (TWPF)
    - Filtración adaptativa de banda completa
    - Filtro *wavelet* adaptativo
    - Filtro ranura adaptativo a frecuencia cero
    - Cancelador adaptativo de interferencia sinusoidal (ASIC)
  - Filtros variantes en el tiempo

– *Técnicas clásicas de filtración:* Uno de los métodos más usuales de reducción del ruido presente en una señal lo constituye el uso de filtros paso bajo. Éste es un método genérico de reducción del ruido de una señal [62], teniendo en cuenta que sus

componentes en frecuencia están por encima de las de la señal, aunque en algunos casos también pueden traslaparse. Estos filtros se diseñan utilizando las técnicas clásicas de diseño de filtros FIR (respuesta al impulso finita) e IIR (respuesta al impulso infinita). Por su escasa selectividad, sobre todo utilizando pocos coeficientes, no suelen usarse en estos casos [57]. En [5] se describen dos filtros paso bajo para eliminar el ruido de alta frecuencia, uno de ellos de primer orden, y el otro de segundo orden.

$$H(z) = \frac{1 - z^{-m}}{1 - z^{-1}} \tag{2.18}$$

que corresponde a la ecuación de diferencias

$$y[n] = y[n-1] + x[n] - x[n-m] \tag{2.19}$$

El problema relacionado con estos filtros es la presencia de lóbulos secundarios y poca selectividad. Para mejorar estos aspectos se puede utilizar el filtro de segundo orden también propuesto en [5]

$$H(z) = \frac{(1 - z^{-4})^2}{(1 - z^{-1})^2} \tag{2.20}$$

cuya ecuación en diferencias es:

$$y[n] = 2y[n-1] - y[n-2] + x[n] - 2x[n-4] + x[n-8] \tag{2.21}$$

En [52] se propone un filtro pasa-bajas el cual elimina las componentes espectrales de interferencia AC (60 Hz), incluyendo los armónicos altos pero también componentes en frecuencia importantes de la señal de ECG. Este es un método genérico de reducción del ruido [62], teniendo en cuenta que sus componentes espectrales están por encima (aunque en algunos casos también pueden solaparse) de la señal. Los filtros rechaza banda remueven la interferencia AC del contenido de una señal por atenuación dentro de un rango estrecho de frecuencia. Algunos filtros FIR también remueven armónicos altos de la interferencia de la red [8]. Los filtros IIR [35] y los filtros análogos rechaza banda no muestran un retardo constante, causando una distorsión indeseable. Los filtros rechaza banda con un pequeño ancho de banda pierden su efectividad en la remoción del ruido de red cuando la frecuencia de interferencia cambia de 60 Hz. Los filtros rechaza banda con anchos de banda simples pueden eliminar cambios en la interferencia pero también eliminarán componentes espectrales de la señal de ECG. Para el caso de la reducción de la perturbación causada por la variación de la linea de base, una primera solución ha sido la utilización de filtros paso banda de 0.5 a 100 Hz, o de paso alto con una frecuencia de corte de 0.5 Hz, ya que estas interferencias normalmente tienen un contenido de frecuencias por debajo de este valor. Los filtros lineales tienen un elevado costo computacional, y para conseguir una banda estrecha es necesario un elevado número de coeficientes. Si se utilizan filtros IIR serán necesarios menos coeficientes, pero al no tener fase lineal distorsionan la señal. Por ejemplo, en [5], se propone un conjunto de filtros digitales

en todas las etapas de preprocesamiento de la señal ECG. Entre ellos se destaca un filtro para reducir las variaciones de la línea base. Este filtro es un paso-alto IIR, con una frecuencia de corte a 0.5 Hz.

– *Técnica de filtración por media móvil:* Otra forma de minimizar el ruido es utilizando filtros de media móvil [62]. Constituyen un método muy sencillo, aunque su selectividad es muy baja, además de enmascarar detalles de alta frecuencia de la señal. Estos filtros tienen una expresión genérica:

$$y[n] = \frac{1}{a+b+1} \sum_{k=n-a}^{k=n+b} x[k] \tag{2.22}$$

de manera que la muestra situada en el centro de un intervalo de anchura k, resulta de la ponderación de las muestras vecinas a la muestra original. Dependiendo de autores, se utiliza este filtro con unos valores de ventana de a y b distintos, además de aplicar distintos pesos a las muestras.

– *Técnica basada en el promedio de latidos:* Aprovechando la semejanza entre los latidos de la señal ECG, su semi-periodicidad, y si la señal es de suficiente duración para poder contar con una cantidad elevada de periodos, se puede hacer un promedio de varios latidos semejantes para reducir el ruido. Este método exige que en primer lugar se haya aplicado un algoritmo para dividir el registro en sus latidos componentes [60], y además utilizar alguna función de alineamiento temporal para poder sumar los latidos con distinta duración. El promediado temporal de latidos es una técnica que se utiliza para mejorar la relación señal a ruido de las señales ECG de alta resolución. Para ello se hace uso de su característica repetitiva y de la no correlación entre el ruido y la señal. El uso del promediado requiere un punto de sincronismo como referencia, ya que el desalineamiento de las señales a promediar produce un efecto indeseable de filtrado paso bajo.

– *Técnica basada en la aproximación mediante funciones:* Otras técnicas, basadas en el reconocimiento de formas utilizan aproximaciones de la señal como un medio de reducir el ruido. Entre estas técnicas estarían el ajuste de una curva mediante segmentos [47] y el ajuste de curvas mediante funciones ortogonales y polinomios [82, 61], las cuales se utilizan también para la compresión de datos. En [61] se describe la representación de esta señal mediante expansiones ortogonales de un reducido número de coeficientes utilizando el producto escalar y estimación adaptativa mediante el algoritmo LMS (mínimos cuadrados), donde el método elegido depende de la relación señal a ruido inicial. En cuanto a la reducción del ruido por la desviación de la línea de base, hay que decir que éstos son filtros son más exactos, permiten variaciones en la línea base de mayor frecuencia, aunque su coste computacional es mayor. En [56] se explica un método para la reducción de la línea base mediante aproximación de la señal con splines cúbicos. Este método se basa en la interpolación entre los niveles isoeléctricos estimados en los intervalos $P - R$. De esta forma, su precisión dependerá en gran medida de la precisión en la selección de los puntos adecuados, la cual muchas veces no es la más conveniente por la dificultad inherente a la localización

de éstos. Otro inconveniente es que el método pierde prestaciones cuando los puntos entre los cuales se interpola están relativamente distanciados. En cuanto a la aproximación mediante polinomios, en [18] se describe un método utilizando polinomios de Chebyshev, similar al utilizado con la aproximación mediante wavelets [63]. Para ello se basa en el hecho de que cualquier función continua puede ser aproximada por un conjunto de funciones polinómicas $\{\phi_k(t)\,|k = 0, 1, ..., n\}$, donde el subíndice k representa el grado del polinomio. El objetivo de este método consiste en obtener los coeficientes y las funciones que ofrezcan la mejor aproximación de la señal. Una vez logrado esto, se calcula la curva resultante y se resta de la señal electrocardiográfica inicial para eliminar las variaciones de la línea base.

– *Técnica basada en filtros globales:* El procedimiento general se realiza con una estimación simple de la frecuencia, amplitud y fase contenida en la señal de ruido de la línea de alimentación en un registro, y se remueve este contenido estimado de la señal original. Un método usado es Least Mean Square (LMS) [55]; otro método propuesto [9] para el filtrado de la señal de ECG por suavizado de la interferencia en el dominio de la frecuencia, se realiza usando la Transformada Rápida de Fourier (FFT). Éstos producen un efecto no deseado cuando la amplitud de la interferencia cambia abruptamente produciendo un retardo a la salida no siendo apropiado para aplicaciones en tiempo real.

– *Técnicas basadas en la transformada wavelet:* Se encuentran diversos métodos basados en la transformada wavelet para reducir el ruido que frecuentemente se presenta en la señal ECG (considerando las variaciones de línea base y línea de alimentación como ruido blanco gaussiano). Dichos métodos son aplicados después de la adquisición de la señal ECG tratándose en todos los casos como métodos de procesamiento de señales fuera de línea [18]. Existen algunos métodos propuestos en [19, 20, 41, 39, 12], para la reducción del ruido mediate procedimientos similares de descomposición y recomposición, variando únicamente la cantidad de niveles de acuerdo a la necesidad y procedimientos que difieren al llevar a cabo la filtración mediante parámetros estadísticos, y la manera de aplicarlos a los coeficientes de detalle. La mayoría se basan en modificaciones del método propuesto en [19]. En este trabajo se establecen los principios básicos para la reducción del ruido, donde se determina que la señal resultante del proceso debe ser tan suave como la señal original minimizando el error cuadrático medio. A partir de este trabajo se derivan otros, donde se estudian variaciones del método anterior para conseguir mejorar la relación SNR (relación señal-ruido). Por ejemplo, en [39], se describe un método basado en la transformada *wavelet*, añadiendo técnicas vectoriales de proyección sobre conjuntos y así evitar el solapamiento que existe entre los filtros de la transformación, provocando que la umbralización no sea la ideal. En [12] se describen algunas modificaciones a la umbralización para aplicar el método en algunos casos donde el ruido no corresponde a ruido blanco gaussiano, tal como se supone en la mayoría de ocasiones. En [41] también aplica una variación a la selección del umbral. En este caso se utiliza la validación cruzada de *splines* para calcular este umbral. Como principal desventaja de esta variante se destaca el costo computacional de los cálculos. En [95], se presenta una nueva for-

ma de filtración adaptativa tiempo-escala, basada en técnicas de compresión wavelet. Una clase de funciones de compresión suave y el SURE (Estimador de riesgo imparcial de *Stein*) se usan para lograr un filtrado adaptativo tiempo-escala. Un método llamado *waveshrink* [74], para la extracción de la señal de ruido en señales ha sido probado como una poderosa herramienta. Este método es especialmente eficiente para la estimación espacial en señales homogéneas. Un paso clave en este procedimiento es la selección del parámetro de umbralización. Donoho and Johnstone proponen un método de selección del umbral basado en el principio *minimax* [20]. En [27] se sugiere una variante del método anterior, con la diferencia que es aplicado otro método de umbralización, $semi - flexible$, el cual generaliza los umbrales flexibles y rígidos propuestos por Donoho y Johnstone. Para la selección del umbral, en [88], el umbral se estima mediante cuatro procedimientos: `SURE`, `HEURISTIC SURE`, `FIXTHRES` y `MINIMAX`. `SURE` es una regla adaptativa de selección del umbral definida como umbral $\delta = \sqrt{2 \cdot log_e(n \cdot log_2(n))}$, donde n es el número de muestras de la señal. Con este aprovechamiento se obtienen riesgos y son minimizados con respecto a los valores δ dados por la selección del umbral. El método es adaptativo buscando a través de un nivel umbral por cada nivel de descomposición *wavelet*. Un acercamiento del umbral fijo `FIXTHRES` calcula el umbral con respecto a la longitud de la señal y el umbral estimado está dado por $\delta = \sqrt{2 \cdot log_e(n)}$. El `HEURISTIC SURE`, aproximándose a ser una variante del primero, reemplaza en condiciones de mucho ruido el estimador `SURE` con `FIXTHRES`. El procedimiento `MINIMAX` aplicado a un umbral fijado $\delta = 0.3936 + 0.1829 \cdot log(n)$ para producir el también llamado desarrollo `MINIMAX` para el error cuadrático medio contra un caso ideal. En [19] se implementó un método mediante umbralización en el cual se utiliza la transformada wavelet con el umbral $\delta = \sqrt{2log(N)}\hat{\sigma}$ donde la umbralización es realizada con la expresión:

$$C_\delta(i,j) = \begin{cases} 0, si\,|C(i,j) < \delta| \\ \text{sgn}(C(i,j))(|C(i,j)| - \delta), si\,|C(i,j) \geq \delta| \end{cases} \tag{2.23}$$

donde $C(i,j)$ representa los coeficientes de los detalles obtenidos mediante la transformada wavelet. El valor de $\hat{\sigma}$ utilizado para este umbral viene dado por la expresión:

$$\hat{\sigma} = median(|C(i,j)|)/0.6745 \tag{2.24}$$

Para aplicar este método hay que tener en cuenta que las señales utilizadas son de larga duración. En este caso, los métodos de reducción de ruido aplicados pueden fallar debido a que a lo largo de la señal el nivel de ruido muy probablemente cambiará. Una posible solución es dividir la señal en una serie de intervalos donde el nivel de ruido se considere constante y procesar cada uno de ellos por separado [18]. En [88] fueron usadas 50 simulaciones independientes para evaluar el desempeño de los métodos de filtrado de ECG con tipos de ruido específicos: gaussiano y ruido blanco uniforme, generados con un modelo autorregresivo (AR) de orden 4. Las amplitudes del ruido fueron escaladas de tal manera que la SNR fuera 5 dB para todas las señales. El desempeño de los métodos se estudió obteniendo el error con 2600 muestras de ECG y más específicamente obteniendo el error cuando 6 complejos QRS se

extraen manualmente del registro. El último método permite investigar el desempeño de los métodos de filtración que manejan una parte importante del espectro del ECG. Cuando se consideran diferentes familias wavelet ortogonales con soporte compacto tales como *Daubechies*, *Symlets* y *Coiflets* con transformada distreta, se obtienen mejores resultados. El desempeño del filtrado fue muy aproximado entre las familias sin embargo las bases *Coiflet* muestran ligeramente un mejor desempeño. La wavelet *coiflet* de orden 5 fue la que menor error tuvo entre otras funciones. El desempeño del error fue medido con la norma $L2$. Por otro lado, tomando la capacidad de las wavelet para suprimir un polinomio, dependiendo de los momentos nulos que posea, se pueden utilizar para reducir las variaciones de la línea base. Así, si $x^k\psi(x)$ es cero para $k = 0, \ldots, n$, entonces la wavelet posee $n+1$ momentos nulos y los polinomios de grado n pueden suprimirse por esta wavelet. Entonces, dentro de las características de la WT se encuentra la capacidad en la reducción de las variaciones de la línea base. Sea la señal adquirida como

$$y[n] = x[n] + \mu e[n] + s[n] \tag{2.25}$$

donde $x[n]$ es la señal sin ruido, $e[n]$ representa ruido blanco Gaussiano con media nula y varianza unitaria, μ su nivel y $s[n]$ representa la línea base. El objetivo del método en [18] es obtener la estimación de la línea base, $\hat{s}[n]$, de forma que su sustracción de $y[n]$, sea una señal sin variaciones en la línea base. En este caso se utiliza filtrado lineal, de forma que todos los detalles de la transformada wavelet serán eliminados, quedándose únicamente con los coeficientes de aproximación. La señal $\hat{s}[n]$ se obtiene directamente del efecto de filtro paso bajo de la aproximación de la señal $y[n]$ de cierto nivel, el cual debe ser establecido *a priori*. La desventaja de esta técnica es que es usada solo para este tipo de perturbaciones y la señal ECG contiene un número considerable de perturbaciones, a la hora de diseñar un filtro adecuado.

– *Técnicas de filtración adaptativa:* Los filtros adaptativos fueron los primeros propuestos para el filtrado de ECG por Widrow et al [92]. Mortara [59] propuso una clase especial de filtrado adaptativo AC (60 Hz) llamado estimación incremental, el cual calcula la salida del ruido como referencia de las 3 últimas muestras de la señal ECG contaminada. Este filtro tolera desviaciones de frecuencia de la línea de red desde un valor indicado hasta cerca de 100 mHz dependiendo de algunos parámetros de adaptación. La interferencia de línea de potencia también puede ser tratada ya sea con un filtro notch adaptativo o simplemente mediante técnicas de substracción de la señal de ruido, estimada de forma adaptativa de la señal ruidosa original. En [23] se proponen tres diferentes filtros notch adaptativos. Para el primer caso un filtro digital FIR rechaza-banda de segundo orden adaptativo es diseñado para seguir los cambios de la frecuencia central de la interferencia. Para el segundo caso los ceros de un filtro digital IIR rechaza-banda de segundo orden adaptativo se colocan sobre el circulo unitario y los polos se adaptan para obtener un ancho de banda óptimo para eliminar el ruido sobre un nivel de atenuación predefinido. El tercer filtro presentado en [23], tanto los polos como los ceros del filtro digital IIR rechaza-banda de segundo orden adaptativo son adaptados para seguir la variación de la frecuencia central

dentro de un ancho de banda óptimo. El proceso adaptativo se ve considerablemente simplificado por el diseño de los filtros notch mediante la reubicación de polos y ceros en el circulo unitario usando algunas reglas sugeridas en [43, 17]. Además de esto se usó el algoritmo CLMS para el proceso adaptativo. En [31] se ha comparado la complejidad y desempeño de dos filtros notch de 60 Hz, uno adaptativo [5] con una generación interna de referencia y uno no adaptativo de segundo orden. Ambos filtros son implementados más eficientemente a una tasa de muestreo de 360 mps (muestras por segundo). En este caso el filtro adaptativo no requiere multiplicaciones y puede ser implementado efectivamente con enteros de 16-bits. una implementación razonable de un filtro no adaptativo requiere dos multiplicaciones de 16-bits, y su desempeño variará con la precision de la implementación. A otra tasa de muestreo, el filtro adaptativo y el no adaptativo pueden ser implementados con una y tres multiplicaciones de 16-bits respectivamente. En cualquier caso, el filtro Ahlstrom y tompkins [5] representa una mayor eficiencia computacional que un filtro de notch de 60 Hz. Con una tasa de muestreo de 360 mps, y con ausencia de ruido, el filtro adaptativo introduce menos distorsión en una señal típica de ECG. La distorsión más significativa producida por el filtro no adaptativo puede ser una vibración después de un complejo QRS. Debido al aumento en la adaptación usada en el filtro Ahlstrom y Tompkins la salida del filtro adaptativo no es afectada significativamente por el complejo QRS, el cual tiene una gran amplitud y muy corta duración, y por lo tanto no se observa vibración. La distorsión observada en la señal de ECG en la ausencia de ruido tiende a decrecer a medida que la respuesta transitoria del filtro se incrementa. Este decrecimiento en la distorsión de la señal puede ser so-pesado con la habilidad del filtro de rastrear los cambios en la amplitud del ruido. Se evaluó la efectividad de los dos filtros en señales reales determinando sus efectos en la entropía residual de una señal de ECG con ruido de 60 Hz de bajo nivel. Se encontró que este ruido de bajo nivel tiene un efecto significativo en los bits por muestra requeridos para almacenar una señal de ECG comprimida en aplicaciones como diagnóstico de ECG. Para las variaciones de las líneas de base, la literatura ha propuesto un filtrado adaptativo en cascada de dos pasos. El primero consiste en un filtrado adaptativo constante, y el segundo, en un filtrado tomando como referencia los instantes de aparición en cada latido. Para ambos se utiliza el algoritmo LMS. El primer filtrado pretende eliminar la componente DC y las frecuencias inferiores que las correspondientes al segmento ST. Después de esto y teniendo en cuenta el comportamiento repetitivo del ECG, se aplica un filtrado adaptativo con impulsos correlacionados, adecuado para la estimación de señales periódicas o de señales que se repiten por bloques en el tiempo. Este filtrado elimina el remanente de las variaciones en la línea base que no estén correlacionadas con el QRS para así preservar las componentes del QRS que sí lo estén. Los filtros adaptativos digitales pasa-bajas eficientes computacionalmente son introducidos en el procesamiento de la señal como detectores de cruce por cero [89]. El propósito de estos filtros es extraer la señal sinusoidal fundamental del ruido y perturbaciones impulsivas, dejando la señal de salida en fase con la entrada. Estos filtros constan de coeficientes arreglados, los cuales son multiplicados por parámetros adaptativos con el fin de ajustar instantáneamente la frecuencia. Sólo dos parámetros adaptativos son usados en este método, y además son considerados por los filtros

básicos dos ventajas de fase lineal: diseños con una buena atenuación rechaza banda y simplicidad computacional. Los filtros polimoniales deformados en el tiempo (TWPF) [65], son filtros adaptativos para la remoción de ruido estacionario de señales biomédicas no estacionarias. El filtro ajusta polinomios deformados para segmentos largos de señales. Este puede ser interpretado como un filtro pasa-bajas con variación en el tiempo de la frecuencia de corte. En condiciones óptimas, la frecuencia de corte del filtro es igual al ancho de banda de la señal. Este posee importantes ventajas sobre las otras técnicas de eliminación: reacciona inmediatamente a los cambios en las propiedades de la señal, independientemente de la reducción deseada del ruido, no requiere de una función de referencia y puede ser aplicado a señales no periódicas. En el caso de funciones cuasi-periódicas, aplicando TWPF para períodos individuales de la señal permite una óptima reducción del ruido. Muchas de las técnicas adaptativas para la remoción del ruido son basadas en el principio *recursive least squares* (RLS). Estos filtros gradualmente reducen el error cuadrático medio entre la señal de entrada principal y alguna señal de referencia dada. Las técnicas de ruido y remoción de interferencia pueden ser clasificadas ampliamente en dos categorías, dependiendo en cuales es registrada o no, por una señal de referencia:

a) Estimación del ruido en la señal de entrada mediante una entrada de referencia para su sustracción. En la mayoría de las técnicas, la señal de referencia consiste en ruido correlacionado con el ruido en la primera entrada [24, 94, 86]. En ciertas técnicas, la señal de referencia consiste en un registro de posición diferente a la señal biomédica [86]. En este caso, el ruido en la entrada principal y la entrada de referencia no deben tener correlación.

b) Técnicas usando una señal de referencia artificial: En este caso de entrada sencilla, la señal de referencia es tomada de la señal registrada. Por ejemplo, los filtros predictores usan una versión retardada de la señal como una referencia.

El método de filtración adaptativo de banda completa, se basa en la técnica clásica de filtración adaptativa para la cancelación de ruido, en el que se hace una estimación de una función lineal variante en el tiempo, entre dos derivaciones del ECG. Las características de las derivaciones de entrada al filtro a partir de la cuales se hacen las estimaciones lineales son [92]: $D1 = s1 + r1$, $D2 = s2 + r2$, donde $s1$ y $s2$ representan la actividad auricular (ondas P), $r1$ y $r2$ representan la AV (actividad ventricular) para las derivaciones DI y DII respectivamente, $s2$ tiene escaso contenido de energía debido a la ubicación de los electrodos en la derivación DII. Como la señal de error a la salida del esquema se halla a partir de una derivación con escasa energía de actividad auricular, se garantiza que la cancelación se llevará a cabo principalmente sobre AV, lo que permite que las ondas P solapadas con los complejos QRS o la onda T sean reflejadas a la salida. Existen otras técnicas como la wavelet adaptativa, de la cual ya se mencionó anteriormente, el filtro ranura adaptativo a frecuencia cero y el cancelador adaptativo de interferencia sinusoidal (ASIC), los cuales tienen un buen rendimiento en la tarea de filtrar señales ECG y el costo computacional permite hacer aplicaciones en tiempo real.

– *Filtros variantes en el tiempo:* Han sido aplicados para la reducción de las perturba-

ciones debidas a la desviación de la línea base. En este caso, la frecuencia de corte varía según el análisis de las componentes espectrales en cada latido respecto al promedio de los mismos. Se parte de la expresión de un filtro patrón, que puede cambiar su frecuencia de corte de forma simple variando alguno de sus parámetros. En [81] se presentan dos métodos: El primer método adapta la frecuencia de corte al nivel de línea base presente en la señal, que ha sido estimada a partir del error entre la salida del filtro actual y la salida producida por el filtro de frecuencia de corte máxima. En el segundo método, la frecuencia de corte se estima a partir de la longitud del intervalo RR de entre 5 y 10 latidos anteriores al latido bajo estudio.

2.4.2. Segmentación

A continuación se expone el conjunto de técnicas más relevantes dentro del proceso de segmentación de señales electrocardiográficas y fonocardiográficas.

- Segmentación
  - Técnicas basadas en la transformada *wavelet*
    - Algoritmo de Mallat para multiresolución
    - DWT con modelo de QRS
    - *Best basis*
    - Descomposición y recomposición de niveles
  - Técnicas basadas en modelos ocultos de Markov y Semi-Markov
  - Técnicas basadas en la representación de señales
    - Mejor base trigonométrica
    - Estimador adaptativo múltiple (MMAE)
  - Transformaciones no lineales
    - Multiplicación de diferencias hacia atrás
    - Algoritmo de Joeng

– *Técnicas basadas en la transformada wavelet:* En señales electrocardiográficas, [6] expone la detección y ubicación de los puntos característicos del complejo QRS de la señal ECG, mediante la transformada wavelet diádica en registros de ECG Holter, utilizando el algoritmo de Mallat para multirresolución con aproximación y detalles a un nivel $j = 2$. Se caracterizan los mínimos y máximos de funciones que resultan de la descomposición. Este algoritmo se prueba con la base de datos MIT ofreciendo medidas de sensibilidad y nivel de predicción positiva. En [84] se presenta un algoritmo de detección de eventos realizado para implementación de dispositivos cardíacos como marcapasos y desfibriladores, en especial para la detección del complejo QRS. Para evaluar el desempeño del algoritmo en ambientes de ruido se combinaron la base de datos EMG y una base de datos de interferencias. Para la detección de eventos se emplea un modelado básico de la señal $\mathbf{x} = \mathbf{h} + \mathbf{w}$, que tiene como restricción el modelo de un sólo QRS, lo cual es una limitación severa debido a la alta naturaleza de variación de morfologías del complejo QRS. En orden a desarrollar un modelo más general de la señal, se identifican dos principales características del complejo QRS:

a. Diferentes composiciones de morfologías simétricas y asimétricas.

b. diferentes duraciones del complejo QRS.

El modelo supone que el QRS está compuesto por una combinación lineal de señales representativas.

$$\mathbf{H} = [\mathbf{h}_1 \dots \mathbf{h}_p] \tag{2.26}$$

Para representar la matriz **H** del modelo representado en la expresión (2.26), se debe escoger un tipo que se asemeje a las morfologías básicas del QRS en la EMG. Inicialmente se opta por la transformada KL, que caracteriza de manera óptima la energía de un ensamble de señales, pero tiene desventajas en la complejidad de implementación, en la no recursión de cálculo de las bases y el formato de punto flotante de los datos, implicando longitudes de palabra demasiado grandes como para efectos de implementación en DSP. Por lo anterior se utiliza la información morfológica *a priori* que tiene el complejo QRS, compuesta por una forma de onda monofásica o bifásica. Una transformada que puede ser implementada eficientemente es la transformada wavelet diádica, aunque por ser una transformada local, brinda información del comportamiento local de una señal. Con lo anterior se plantea el uso de filtros propios, en donde el banco de filtros a obtener debe tener una respuesta al impulso monofásica y bifásica. Al banco de filtros se aplica el algoritmo de Mallat basado en la transformada wavelet DWT. Se plantea entonces una estructura eficiente del detector con baja complejidad usando transformada wavelet diádica con coeficientes de filtro enteros adecuados, seguido por una prueba de relación de verosimilitud generalizada (GLRT), en orden a tomar alguna decisión. Los resultados muestran que una detección óptima puede ser obtenida moderando los altos niveles de ruido para algunas fuentes de ruido común. En términos de probabilidad de una detección falsa y probabilidad de una falsa alarma, en promedio el desempeño fue menor que 0,7 % y 0,1 % respectivamente, conseguido para niveles de ruido moderados de cinco tipos de ruido. En [14] y [48], de determina mediante criterios basados en las características intrínsecas de la señal, la técnica de la mejor base wavelet para segmentar las ondas y complejos relevantes de la señal ECG (P, QRS, T y ST). Se analizan las desventajas de emplear únicamente el criterio de la entropía encontrado como el método adecuado en la teoría de la información y se propone dos nuevos criterios que mejoran la segmentación de las diferentes morfologías de la señal ECG. Se ha demostrado experimentalmente que criterios de representación parsimonia ciertamente pueden alcanzar una segmentación aceptable para un número de clases de señales. Esto sin embargo, no es un criterio universal, en el sentido de que una segmentación deseada puede estimarse debido a características propias de la morfología de una señal más, que por su representación parsimonia. Se presenta primero el problema de la segmentación en la búsqueda de una base ortonormal adaptativa y, en desarrollar un criterio determinista, cuya matemática incorpore de la mejor forma posible la información que se estime, tenga un valor visual. El algoritmo de *Best Basis* (BB) ha sido aplicado exitosamente para bases wavelet packet (mejor segmentación espectral) como también en bases trigonométricas locales (LCB) (mejor segmentación temporal). La búsqueda de la wavelet se realiza de acuerdo a las características morfológicas de los componentes de la señal ECG con formas de onda como el coseno de diferente espectro y en diferente ubicación temporal. Los resultados preliminares obtenidos con este método de análisis ECG son prometedores, siendo este método hasta ahora una variación de las técnicas existentes. Aunque existen puntos que quedan por resolver, en particular, la presente limitación en el método diádico de bases ortonormales.

– *Técnicas basadas en modelos ocultos de Markov y ocultos de Semi-Markov:* En [38]

estudia el uso de Modelos Ocultos de Markov (HMM) y Modelos Ocultos de Semi-Markov (HSMM), para la segmentación del electrocardiograma en sus ondas características constitutivas. Se usa la trasformada wavelet undecimada (UWT) para proveer una representación completa de la señal que es más apropiada para su posterior modelado. La gran mayoría de algoritmos para el análisis automatizado de ECG se basan en métodos de umbralización para predecir el fin de la onda T, como el punto donde la onda cruza un umbral previamente calculado. Basados en el hecho natural de que la señal ECG es el resultado de un proceso generativo en el que cada onda característica es generada por un correspondiente estado cardiológico del corazón, además cada estado depende únicamente del estado anterior, se ha optado por el uso de HMM. Las ondas que se detectaron fueron: la onda P, la línea de base 1, el complejo QRS, la onda T, la línea de base 2 y la onda U. Inicialmente se entrenó el modelo de una forma supervisada con una base de datos etiquetada manualmente. Aunque se obtuvieron resultados razonables en la clasificación del complejo QRS y la onda T, en clasificación de la onda P los resultados no fueron muy consistentes. Con el objeto de mejorar el comportamiento del modelo se requirió codificar la señal para capturar las principales características de la onda. Para tal propósito se utilizó la UWT ya que ésta provee una descripción de la señal tiempo-frecuencia en una base muestra a muestra. Con el fin de encontrar la wavelet más efectiva se examinaron wavelets tipo *Daubechies*, *Symlet*, *Coiflet* y *Bi-ortogonales*. Se encontró que la wavelet *Coiflet* de segundo orden ofrece mejores resultados en la clasificación. La limitación es el peso computacional y que las patologías cardíacas son muy diversas y el modelo nunca sería lo suficientemente robusto para una segmentación precisa en presencia de patologías.

– *Técnicas basadas en la representación de señales ECG:* En [14] se utiliza un algoritmo de la mejor base dejando de lado la heurística. Debido a la gran variabilidad del SNR local se propone el criterio de entropía estándar que produce una segmentación demasiado fina, por lo que se introduce un nuevo criterio de optimalidad, el cual se basa en una combinación lineal de medidas de entropía y una función de suavizado de forma general. Se probó el algoritmo en la base de datos de arritmias de la MIT-BIH y en mapas de potenciales de superficie corporales. En [48] se propone una técnica de segmentación basada en la mejor base trigonométrica localizada. Se muestra por medio de ejemplos de datos reales que el criterio de entropía el cual alcanza la mejor representación parsimonia de una señal da como resultado una segmentación fina de la señal ECG, que establece la necesidad de un criterio más comprensivo. Se introduce una nueva búsqueda del criterio de la mejor base el cual se basa en una combinación lineal de medidas de entropía y medidas locales de suavidad y curvatura. Se probó el algoritmo en la base de datos de arritmias de la MIT-BIH. En [36] se presenta un nuevo algoritmo de proceso para ECG basado en un modelo múltiple de estimador adaptativo (MMAE) para sistema de monitoreo fisiológico. De la base de datos MIT fueron empleadas 20 señales de ECG para desarrollar el modelo del sistema para el MMAE. Los segmentos característicos del ECG como la onda P, el complejo QRS y la onda T fueron usados para desarrollar bancos de filtros hipotéticos. Incluyendo un algoritmo de umbral filtro-conmutado a la implementación convencional MMAE,

el dispositivo imita la forma como un especialista analiza la búsqueda de complejos de la señal ECG con un acotamiento temporal y extiende la búsqueda en otras ondas componentes claves y su tiempo. Las 20 señales y un experimento con una señal animal fueron empleadas para la prueba del algoritmo. Usando un algoritmo condicional de prueba de hipótesis, el MMAE identificó correctamente a los segmentos de la señal ECG correspondientes a los modelos de hipótesis con un 96.8 % de efectividad para 11539 posibles segmentos probados. La robustez del algoritmo MMAE también detecta cualquier desalineamiento en el filtro hipotético y automáticamente reinicia los filtros con el MMAE para sincronizar la hipótesis con la señal entrante. Finalmente, el MMAE selecciona el banco de filtros óptimos basados en las mediadas del ECG entrante. El algoritmo también provee información crítica relacionada con el corazón como rata del corazón, intervalos QT, PR de la señal ECG. Este analizador puede fácilmente añadido como actualización de software de monitores estándar fisiológicos empleados universalmente en vehículos de emergencia.

– *Técnicas basadas en transformaciones no lineales:* En [49] se propone un algoritmo para la detección de parámetros que puedan clasificar el ritmo cardíaco en línea. Los parámetros a ser extraídos son:

  a. Onda P: Tiempo de inicio, máximo, fin; amplitud.
  b. Complejo QRS: Tiempo de inicio, pico R, fin; amplitud.
  c. Onda T: Tiempo del máximo/mínimo (dependiendo de la polarización), segundo extremo (si está disponible), fin; amplitud.
  d. Segmento ST: desviación y pendiente.
  e. Parámetros derivados: intervalo $R-R$, intervalo PQ, longitud QT.

  La detección del complejo QRS está basada en el algoritmo de *Multiplicación de Diferencias hacia Atrás (MOBD)*. La desviación ST se deriva 80 ms después del punto final del complejo QRS. La detección de las ondas P y T se realiza con el algoritmo de Joeng [42], el cual busca la onda T después de que se ha detectado un complejo QRS. Los algoritmos de detección de todos los parámetros son evaluados basándose en la base de datos QT. A pesar que el algoritmo es rápido, su eficiencia baja en presencia de patologías severas.

Con respecto a la segmentación de señales FCG, en [51] separa el ciclo cardíaco en cuatro partes: el primer sonido($S1$), el periodo sistólico, el segundo sonido cardíaco($S2$) y el periodo diastólico. Primero se ubica las localizaciones e intervalos de los picos dominantes del primer sonido cardíaco para ser computados, a continuación los intervalos del periodo sistólico y diastólico son obtenidos consecuentemente. El algoritmo de segmentación es basado en la envolvente de la energía, calculada por medio de la energía normalizada de Shannon, la cual atenúa los efectos del ruido de bajo valor, y toma los sonidos de intensidad baja fácilmente. También acorta la diferencia de envolvente de la intensidad, entre sonidos de intensidad alta y sonidos de intensidad baja. A partir del envelograma calculado se toma un umbral con el fin de eliminar señales de baja intensidad que no superan a este valor y así determinar los picos que indican cuales son los ruidos $S1$ y $S2$ de la señal fonocardiográfica; de esta manera se eliminan además perturbaciones de artefacto. La adquisición de las bioseñales

se realizó por medio de un PC equipado con un estetoscopio digital de 16 bits de precisión y 11025 Hz de frecuencia de muestreo. Se utilizaron 37 grabaciones, 14 soplos patológicos y 27 soplos fisiológicos, de niños entre 1 y 13 años de edad con duraciones de 7 a 12 s. Los resultados del experimento son del 93 % de éxito para los 515 ciclos cardíacos evaluados (ver Tabla 2.3). Otro algoritmo de segmentación de sonidos cardíacos utilizado es por medio

	Número de ciclos	Porcentaje (%)
correctos	479	93.01
perdidos	30	5.83
incorrectos	6	1.17
Total	515	100

Tabla 2.3. Resultados de la segmentación basada en el envelograma. Tomado de [51]

de descomposición y reconstrucción de Wavelets [53]. El algoritmo separa la bioseñal en cuatro partes: el primer sonido $S1$, el período sistólico, el segundo sonido $S2$ y el período diastólico. Los puntos claves del algoritmo de la segmentación está en que los $S1$ y los $S2$ son identificados primero, los intervalos de los $S1$'s y los $S2$'s son computados, y basado en esta información, los intervalos del período sistólico y diastólico son por lo tanto obtenidos. Los detalles y las aproximaciones, que correspondieron a diversas bandas de frecuencia, fueron obtenidos usando la descomposición y la reconstrucción Wavelet discreta. Para cada señal de la banda de frecuencia, un método similar de segmentación basado en la envolvente de la señal correspondiente fue utilizado para conseguir los resultados de la segmentación. Entonces el mejor resultado de las diversas bandas fue elegido para ser el resultado final de la segmentación. El funcionamiento del algoritmo se ha evaluado usando 1165 períodos cardíacos a partir de 77 grabaciones fonocardiográficas digitales incluyendo sonidos de corazón normales y anormales. El algoritmo ha demostrado un 93 % de éxito (ver Tabla 2.4). La detección del primer y segundo ruido cardíaco S1 y S2 también se puede realizar

	No. de ciclos	Porcentaje (%)
correctos	498	96.7
perdidos	10	1.9
incorrectos	7	1.4
total	515	100

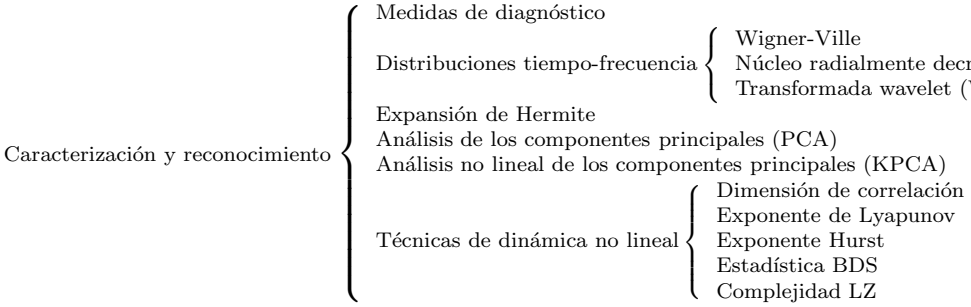
Tabla 2.4. Resultados de segmentación usando descomposición y reconstrucción wavelet. Tomado de [53]

a través de análisis de tiempo-frecuencia. Este algoritmo utiliza la transformada wavelet debido a que las señales del corazón son no estacionarias [73]. Se selecciona una wavelet madre basada en la energía de los coeficientes de detalle, dentro de un grupo de wavelets. Para escoger la mejor wavelet se analiza cual de ellas contiene la mayor concentración de energía en un número reducido de coeficientes. El método final de detección es comparado con la energía hallada a partir de la transformada de Hilbert, comparando los detalles importantes. Debido a los resultados de niveles energéticos,la wavelet $Db5$, fue escogida para el análisis de señales FCG. Por lo tanto este es un apoyo significativo en la extracción

de características de enfermedades cardíacas como es le caso de la estenosis mitral. A través de una señal FCG normal, se puede observar que la debuachies5, contiene mejor concentración de coeficientes en los coeficientes $b3$, $b4$, $b5$, $b6$, cerca de un nivel energético del 86 % de toda la energía analizada. Este método detecta correctamente los componentes $S1$ y $S2$ de la señal fonocardiográfica. Los productos de los coeficientes wavelets $e4$, $e5$, $e6$, poseen una alto componente energético en la temporización de la diástole. En el momento del análisis de un soplo sistólico, estenosis aórtica, el algoritmo encontró correctamente los componentes del ciclo cardíaco $S1$ y $S2$.

2.4.3. Caracterización y reconocimiento

La *caracterización* es el proceso mediante el cual se buscan atributos que brinden alguna medida de separabilidad entre las clases de análisis. Son muchas las técnicas que son usadas para la caracterización de señales, ya que cada vez son más las herramientas matemáticas, que se involucran en las áreas de aplicación, particularmente, como medidas de caracterización de señales. A continuación se expone un grupo de técnicas que han ofrecido un buen desempeño en la caracterización orientada al reconocimiento de estados funcionales a partir de señales electrocardiográficas y fonocardiográficas.



– *Medidas de diagnóstico:* En [37] se presenta una investigación en la extracción de características eficientes de la señal de ECG para mejorar el comportamiento de la detección automática y clasificación de arritmias cardíacas. El vector de características que se seleccionó se compone del intervalo RS, área del QRS, intervalo RR, amplitud de la onda R, área del segmento ST, amplitud de la onda T, energía de la señal, energía del complejo QRS, coeficiente de autocorrelación, amplitud máxima del histograma de la señal, además de 13 muestras comprimidas, para un total de 23 características. Empleando este método se logró una tasa promedio de reconocimiento de 0.95 sobre 3 diferentes formas de onda ECG. En [50], una vez se tienen detectados los puntos característicos, se selecciona un conjunto de parámetros significativos que permitan una clasificación exitosa. Basados en consultas con médicos y de acuerdo a la bibliografía se escogieron como parámetros la amplitud y duración de la onda P, duración del intervalo PR, duración del complejo QRS, duración de la onda S, amplitud de la onda R, duración del intervalo QT, duración de la onda T, intervalo ST y área del intervalo QT. En general, los trabajos presentados con este método presentan buenos resultados, sin embargo, la debilidad del método consiste en la dependencia que tiene de la etapa de segmentación y la variabilidad que tiene las señales

biológicas en sus estados tanto normales como patológicos, aunque con una base de datos completamente etiquetada, se puede entrenar bien el sistema.

– *Distribuciones tiempo-frecuencia:* El análisis conjunto en tiempo-frecuencia es preferible en la caracterización de procesos no estacionarios. Por esta razón han sido desarrolladas algunas otras técnicas de procesamiento para resolver la estructura tiempo-frecuencia de señales complejas o para realizar tareas particulares, entre estas se cuentan distribuciones como: Wigner Distribution(WD), Continuous Wavelet Transform (CWT), Choi Williams Distribution (CWD), Binomial Distribution (BD). Las dos últimas son llamadas Reduced Interference Distributions (RID) y pertenecen a una clase de distribuciones llamadas clase de Cohen, la cual es una extensión de la Wigner-Ville Distribution [70]. La Continuous Wavelet Transform (CWT) ha recibido especial atención por su habilidad para localizar singularidades de una señal, lo cual la ha hecho una atractiva herramienta en el análisis de señales no estacionarias [67]. Usando representaciones gráficas se pueden comparar fácilmente varias señales para encontrar similitudes o diferencias entre ellas y clasificarlas según algún criterio. En [71] se propone una combinación de estrategias de técnicas clásicas y modernas para la selección de parámetros extraídos de distribuciones tiempo-frecuencia, enfocados a mejorar el comportamiento de los algoritmos de detección de fibrilación ventricular (VF). Inicialmente se procesa la señal para obtener 25 parámetros tiempo-frecuencia de la *Distribución de Wigner-Ville*. Las 4 clases de ECG que se quisieron clasificar fueron: ritmo sinusal normal, fibrilación ventricular y flutter, taquicardia ventricular y otros ritmos. Se concluyó que los métodos de selección propuestos simplifican la solución y mejoran los índices de clasificación. El trabajo presentado en [90], tiene como objetivo la implementación de un sistema capaz de realizar un diagnóstico de isquemia sin necesidad de esperar a la etapa de angina, recurriendo a la valoración de las manifestaciones fisiopatológicas propias de la isquemia y apoyándose en las técnicas que ofrece el procesado digital de señales. Para lograr este objetivo, han diseñado una técnica de procesado integral que permite obtener, a partir de la señal ECG, y más concretamente de la serie de latidos detectados (Frecuencia Cardíaca), una estimación espectral fiable (Variabilidad de la Frecuencia Cardíaca, VFC). Mediante un análisis teórico-práctico, dilucidan la más efectiva de las técnicas existentes en algunas etapas de procesado y, hacen aportaciones originales en aquellas en las cuales las técnicas existentes eran deficientes. Se propone una técnica para la representación tiempo-frecuencia de la señal: *Distribución de Núcleo Gaussiano Radialmente Decreciente Controlada Instantáneamente*. En la etapa de validación emplean registros seleccionados de la base de datos $ST-T$ Europea. La conclusión más importante de este estudio es la confirmación de la relación entre VFC, control neuro-vegetativo del corazón e isquemia. En [69] se propone el uso de varias *funciones wavelet* para cuantificar el número de singularidades que se presentan dentro del complejo QRS. Los potenciales tardíos ventriculares son indicadores de conductividad ventricular anormal, asociados a pacientes que han sobrevivido al infarto agudo de miocardio (IAM). Estos potenciales son formas de onda muy cambiantes por lo que se sugiere el uso de 4 funciones wavelet con el objeto de detectarlos. Los resultados de este estudio fueron comparados con el análisis uni-wavelet, en el que se utilizó la función Morlet,

las primeras 6 derivadas de la función Gaussiana, B-spline y Meyer. Para el análisis multi-wavelet se utilizó Morlet, Meyer, Daubechies de segundo y cuarto orden. Con el objeto de detectar las singularidades después de aplicar la transformada wavelet (WT), se sugieren los siguientes pasos:

a) Detección de los máximos locales.

b) Localización de la conectividad de los máximos locales.

c) Asignación de los máximos locales como singularidad.

Para la detección de las singularidades se fijaron las siguientes constantes:

a) Umbral de dispersión.

b) Número mínimo de crestas interconectadas.

c) Umbral de amplitud.

Los resultados llegaron hasta el 92 % de acierto en todos los casos.

– *Expansión de Hermite:* En [32], se presenta el uso de Redes Neuronales Artificiales (ANN) para detectar signos de Infarto Agudo de Miocardio (IAM) en la señal ECG. El ECG de 12 derivaciones se expresa como una serie de las funciones bases de Hermite y los coeficientes resultantes son usados como entradas a las ANN. Mediante un *análisis de sensitividad* se indica qué regiones temporales del ECG son críticas para la salida del clasificador basado en ANN. El desempeño del clasificador es medido en términos del área bajo la curva de la *Característica Receptora Operante* (ROC). Bajo este parámetro se estudia la detección de IAM basado en la representación del ECG en términos de los coeficientes de la expansión Hermite, junto a clasificadores con ANN, encontrándose una área ROC de 83.4 %, lo cual es casi tan bueno como el método basado en medidas de diagnóstico, cuya área ROC es de 84.3 %; sin embargo, la primera representación tiene la ventaja de ser invertible.

– *Análisis de componentes principales:* Fujimura analiza el ECG de 12 derivaciones [26]. El algoritmo desarrollado se denomina *extracción de características de propósito orientado*: los datos primero son ortogonalizados y reducidos mediante *Análisis de Componentes Principales* (PCA), luego se extrae un conjunto apropiado de características a través de combinaciones lineales de los componentes reducidos. En [40] se muestra una propuesta de extracción de características y modelo de representación del vector de patrones del complejo *QRS* y del segmento *ST* en ECG ambulatorio, usando la *Transformada de Karhunen-Loève* (KLT). Se describe la derivación de matrices de covarianza robustas usando un método de aproximación Kernel, un método para estimar la dimensión de los vectores de características de la KLT, además de una técnica para derivar y representar series de tiempo de vectores de características. Como conclusión se tiene que la representación de la forma de onda en términos de los coeficientes de la KLT es superior a las mediciones temporales tradicionales.

– *Análisis no lineal de componentes principales:* En [83] se desarrolla un algoritmo para la extracción de características del segmento *ST* basado en *Análisis No Lineal de Componentes Principales* (NLPCA). Mediante una red neuronal de dos capas se implementan las funciones de mapeo de la señal a un espacio de características. Se

hicieron experimentos de clasificación usando 34 archivos de la base de datos $ST-T$ Europea. Las redes neuronales fueron entrenadas únicamente con latidos normales. El método de clasificación se basa en redes de función base radial (RBFN). Se consideran dos clases: normales y anormales, la última incluyendo latidos isquémicos y artefactos. El artículo muestra los resultados de clasificación para diferentes archivos de la base de datos $ST-T$ Europea. El índice de clasificación fue de 79.32 % para latidos normales y 75.19 % para anormales.

– *Técnicas de dinámica no lineal:* Se presenta como una herramienta alternativa de extracción de características en señales ECG y FCG [15, 4]. Una ventaja que tiene con respecto a las otras técnicas es su capacidad de capturar la dinámica intrínseca de la señal, por lo cual es posible encontrar patrones característicos en los atractores multidimensionales, obtenidos mediante medidas de complejidad, impredecibilidad y nivel de caos, a través de indicadores como: la dimensión de correlación, máximo exponente de Lyapunov, exponente Hurst, entre otros. La debilidad de esta técnica frente a las otras, es que realizar medidas en espacios multidimensionales, exige mucho gasto computacional.

En cuanto a las señales FCG, se prefiere la caracterización mediante el análisis conjunto en tiempo-frecuencia debido a que la acústica cardíaca es altamente no estacionaria. Sin embargo, estas señales en la etapa de adquisición son altamente propensas a ser contaminadas con ruido acústico ambiental [2]. El análisis acústico pierde efectividad proporcionalmente a la cantidad de ruido acústico de adquisición que se traslapa espectralmente en las zonas donde se encuentra la información discriminante [76]. El análisis de señales FCG a través de Transformada Wavelet Continua (CWT) es una buena técnica para arreglar los problemas de resolución de la STFT es usar la transformada de wavelet. El objetivo del estudio es realizar un comparación de seis wavelet (Harr, Meyer, Morlet, Mexican Hat, Gauss, Daubechies), en su desempeño comparando características espectrales de energía de señales fonocardiográficas, con el modelo de auto-regresión (AR), el cual ha demostrado que presenta un muy buen desempeño en la representación de características espectrales de señales (FCG) [21]. Las funciones de la transformada wavelet son cambiadas de acuerdo aun prototipo llamado wavelet madre. Cada wavelet madre diferente presenta características diferentes desde el punto de vista de tiempo y frecuencia y valores de amplitud para la señal FCG. En otras palabras la wavelet madre es el punto mas importante en el análisis de (TFR). Mientras que una wavelet de escala ancha obtiene características de alta frecuencia, una wavelet de escala angosta obtiene características de baja frecuencia de la señal. Con el fin de encontrar las características espectrales de energía se usa *wavelet power spectrum estimation* (WPSE), combinando así en una sola gráfica, información de distribución de energía (ED), amplitud instantánea, y distribución de frecuencia (FD), poder espectral. Los resultados mas importantes obtenidos son los siguientes: La wavelet Harr y la mexican hat, presenta un espectro de energía incorrecto para altas frecuencias. Si solamente las distribuciones de frecuencia son tomadas en consideración, las wavelets Meyer, Morlet, Gaus , Daubechies, presentan buenos resultados para determinar anomalías en las señales FCG. La mayoría de wavelets presentan errores en cuanto al análisis de la ED, estos errores en comparación del modelo AR, son medidos por medio de la raíz cuadrada del error cuadrático medio normalizado (NRMSE). De acuerdo a los resultados del NRMSE, el menor error

lo presenta Morlet (ver Tabla 2.5). Por lo tanto, se tiene que esta base wavelet es la más adecuada para el análisis de señales FCG, ya que con sus coeficientes se puede construir un espacio de representación discriminante para distinguir, a partir de los sonidos del corazón ($S1$ y $S2$), anormalidades asociadas a los soplos cardíacos. Adicionalmente, el *análisis*

	Haar	Meyer	Morlet	M. Hat	Gauss(5)	Daub.(5)
NRMSE	81.76	77.43	45.32	64.25	79.96	74.87

Tabla 2.5. Raíz cuadrada del error cuadrático medio normalizado. Tomado de [21]

espectral de los sonidos cardíacos proveen información acerca del estado de las válvulas involucradas en la dinámica cardíaca, tal como la calcificación o el posible deterioro. Particularmente, el estudio del pico de frecuencia dominante a altas frecuencias es de gran importancia para este tipo de análisis. El algoritmo denominado MUSIC (clasificación de señal múltiple) y el FOS (búsqueda ortogonal rápida) pueden demostrar eficientemente si existe un mal funcionamiento de una válvula implantada por medio del análisis de los principales picos de frecuencia [46]. Como resultados se muestra por los dos métodos que los pacientes con picos de frecuencias dominantes mayores a 300 Hz pueden presentar daños en las válvulas implantadas (ver la Tablas 2.6 y 2.7). No obstante el método FOS tiene una desventaja cuando solo uno o dos de los picos de frecuencia dominante son importantes.

Paciente No.	MUSIC	FOS
paciente 1	281.2	250
paciente 2	218.8	200
paciente 3	234.4	200
paciente 4	234.4	200
paciente 5	140.6	150
paciente 6	355.5	350

Tabla 2.6. Primer pico de frecuencia dominante (Hz). Tomado de [46]

Paciente No.	MUSIC	FOS
paciente 1	234.4	–
paciente 2	281.2	100
paciente 3	156.2	400
paciente 4	140.6	250
paciente 5	218.8	250
paciente 6	902.3	900

Tabla 2.7. Segundo pico de frecuencia dominante (Hz). Tomado de [46]

BIBLIOGRAFÍA

[1] H. D. I. Abarbanel. *Analysis of observed chaotic data*. Springer-Verlag, 1996.

[2] T. R. R. abd N. E. Reed and P. Fritzson. Heart sound analysis for symptom detection and computer-aided diagnosis. *Simulation Modelling Practice and Theory*, (12):129–146, 2004.

[3] A. Acero and H. W. Hon. *Spoken Language Processing*. Prentice Hall, 2001.

[4] C. Ahlstrom, P. Hult, P. Rask, J. E. Karlsson, E. Nylander, U. Dahlström, and P. Ask. Feature extraction for systolic heart murmur classification. *Annals of Biomedical Engineering*, 34(11):1666–1677, 2006.

[5] M. L. Ahlstrom and J. Tompkins. Digital filters for real-time ECG signal processing using microprocessors. *IEEE Transactions on Biomedical Engineering*, 32(9):708–713, 1985.

[6] R. Almeida, S. Gouveia, P. Rocha, P. Lago, and F. De-Freitas. Wavelet transform based matlab system for the detection and delineation of QRS complexes in ambulatory ECG recordings. In *6th Portuguese Conference on Biomedical Engineering*, 2001.

[7] Y. Almog, O. Oz, and S. Akselrod. Correlation dimension estimation: Can this nonlinear description contribute to the characterization of blood pressure control in rats? *IEEE Transactions on Biomedical Engineering*, 46(5):535–547, 1999.

[8] J. A. Alste and T. S. Schilder. Removal of base-line wander and power-line interference from the ECG by an efficient FIR filter with a reduced number of taps. *IEEE Transaccions on Biomedical Engineering*, BME-32(12):1052–1060, 1985.

[9] R. Barr and M. Ferdjallah. Frequency-domain digital filtering techniques for selective noise elimination in biomedical signal processing. In *Proceedings of the Twelfth Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 813–814, 1990.

[10] J. B. Bassingthwaighte and G. M. Raymond. Evaluating rescaled range analysis for time series. *Annals of Biomedical Engineering*, 22(4):432–444, 1994.

[11] J. B. Bassingthwaighte and G. M. Raymond. Evaluation of the dispersional analysis method for fractal time series. *Annals of Biomedical Engineering*, 23(4):491–505, 1995.

[12] K. Berkner and R. O. Wells. Wavelet transforms and denoising algorithms. In *Thirty-Second Asilomar Conference on Signals, Systems & Computers*, volume 2, pages 1639–1643, 1998.

[13] R. Beyar, S. Levkovitz, S. Braun, and Y. Paltiu. Heart-sound processing by average and variance calculation – physiologic basic and clinical implications. *IEEE Transaction on Biomedical Engineering*, 31(9):591–596, 1984.

[14] D. Brooks, H. Krim, J. Pesquet, and R. McLeod. Best basis segmentation of ECG signals using novel optimality criteria. In *Proceedings of IEEE Acoustics, Speech, and Signal Processing*, pages 2750–2753, 1996.

[15] R. Carvajal, M. Vallverdú, and P. Caminal. Análisis no lineal de la variabilidad de la frecuencia cardíaca en casos normales y cardiopatías. *Revista Mexicana de Ingeniería Biomédica*, 21(2):29–34, 2000.

[16] R. Chandramouli and K. M. Ramachandran. *Wavelets and Signal Processing*, chapter Wavelets for Statistical Estimation and Detection, pages 103–135. Birkhäuser, 2003.

[17] C. T. Chen. *One-Dimensional Digital Signal Processing*. Marcel-Dekker, Inc., 1985.

[18] D. Cuesta-Frau, D. Novak, V. Eck, J. C. Pérez-Cortés, G. Andreu-García, C. Sastre, and M. Llorca. Reducción del ruido en señales electrocardiográficas mediante la transformada wavelet. In *Congreso Anual de la Sociedad Española de Ingeniería Biomédica*, pages 103–106, 2000.

[19] D. L. Donoho. De-noising by soft-thresholding. *IEEE Transactions on Information Theory*, 41(3):612–627, 1995.

[20] D. L. Donoho and I. M. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association*, 90(432):1200–1224, 1995.

[21] B. Ergen and Y. Tatar. Optimal continuous wavelet analisys of periodograms signals. In *IJCI Proceedings of International Conference on Signal Processing*, volume 1, 2003.

[22] B. Ergen and Y. Tatar. Time-frequency analysis of phonocardiogram. In *4th International Conference MEASUREMENT*, pages 222–225, 2003.

[23] M. Ferdjallah and R. E. Barr. Adaptive digital notch filter design on the unit circle for the removal of powerline noise from biomedical signals. *IEEE Transactions on Biomedical Engineering*, 41(6):529–536, 1994.

[24] E. Ferrara and B. Widrow. Fetal electrocardiogram enhancement by time sequenced adaptive filtering. *IEEE Transaccions on Biomedical Engineering*, 29(6):458–460, 1982.

[25] A. M. Fraser and H. L. Swinney. Independent coordinates for strange attractors from mutual information. *Physical Review A*, 33:1134–1140, 1986.

[26] S. Fujimura and S. Kiyasu. Application of feature extraction scheme to the discrimination of electrocardiogram (ECG). *T.IEE Japan*, 121-A(8):725–730, 2001.

[27] H. Y. Gao and A. Bruce. Waveshrink with semisoft shrinkage. Technical Report 39, StatSci Division of MathSoft, 1995.

[28] J. García, G. Wagner, L. Sörnmo, P. Lander, and P. Laguna. Identification of the occluded artery in patients with myocardial ischemia induced by prolonged percutaneous transluminal coronary angioplasty using traditional vs transformed ECG-based indexes. *Computers and Biomedical Research*, 32(5):470–482, 1999.

[29] A. L. Goldberger, D. R. Rigney, and B. J. West. Chaos and fractal in human physiology. *Scientific American*, 262:42–49, 1990.

[30] P. Grassberger and I. Procaccia. Measuring the strangeness of strange attractors. *Physica D*, 9:189–208, 1983.

[31] P. S. Hamilton. A comparison of adaptive and nonadaptive filters for reduction of a power line interference in the ECG. *IEEE Transactions on Biomedical Engineering*, 43(1):105–109, 1996.

[32] H. Haraldsson, L. Edenbrandt, and M. Ohlsson. Detecting acute myocardial infarction in the 12-lead ECG using hermite expansions and neural networks. *Artificial Intelligence in Medicine*, 32(2):127–136, 2002.

[33] R. Hegger, H. Kantz, and E. Olbrich. *Nonlinear analysis of physiological data*, chapter Problems in the reconstruction of high-dimensional deterministic dynamics from time series, pages 23–47. Springer, 1998.

[34] B. Henry, N. Lovell, and F. Camacho. *Nonlinear Biomedical Signal Processing*, volume II, chapter Nonlinear dynamics time series analysis. IEEE Press Series on Biomedical Engineering, 2001.

[35] L. G. Herrera. Real-time digital filters for ECG signals: Evaluation and new designs. In *Computers in Cardiology*, pages 133–136, 1992.

[36] G. S. Hoffman, M. M. Miller, M. Kabrisky, P. S. Maybeck, and J. F. Raquet. Novel electrocardiogram segmentation algorithm using a multiple model adaptive estimator. In *Proceedings of the 41st IEEE Conference on Decision and Control*, volume 3, pages 2524–2529, 2002.

[37] H. G. Hosseini and H. Nazeran. Efficient features for ann-based ECG classifiers. In *International Conference IEEE Engineering in Medicine and Biology Society*, 1999.

[38] N. P. Hughes, L. Tarassenko, and S. J. Roberts. Markov models for automated ECG interval analysis. In S. Thrun, L. Saul, and B. Schoelkopf, editors, *Advances in Neural Information Processing Systems 16*. MIT Press, 2004.

[39] H. Inoue and A. Miyazaki. A noise reduction method for ECG signals using the dyadic wavelet transform. *IEICE Trans. Fundamentals*, E81-A(6):1001–1007, 1998.

[40] F. Jager. Feature extraction and shape representation of ambulatory electrocardiogram using the karhunen-loève transform. *Electrotechnical Review*, 69(2):83–89, 2002.

[41] M. Jansen. Generalized cross validation for wavelet thresholding. *Signal Processing*, 56(1):33–44, 1997.

[42] H. K. Joeng, K. K. Kim, S. C. Hwang, and M. H. Lee. A new algorithm for p-wave detection in the ECG signal. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 42–43, 1989.

[43] C. R. Johnson. Adaptive IIR filtering: Current results and open issues. *IEEE Transactions on Information Theory*, IT-30:237–249, 1984.

[44] J. F. Kaiser. On a simple algorithm to calculate the "energy" of a signal. In *International Conference on Acoustics, Speech, and Signal Processing*, pages 381–384, 1990.

[45] H. Kantz and T. Schreiber. *Nonlinear Time Series Analysis*. Cambridge University Press, 2002.

[46] S. H. Kim, H. J. Lee, J. M. Huh, and B. C. Chang. Spectral analysis of heart valve sound for detection of prosthetic heart valve diseases. *Yoisei Medical Journal*, 39(4):302–308, 1998.

[47] A. Koski and M. Juhola. Segmentation of digital signals based on estimated compression ratio. *IEEE Transactions on Biomedical Engineering*, 43(9):928–938, 1996.

[48] H. Krim and D. Brooks. Feature-based segmentation of ECG signals. In *Proceedings of the IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis*, pages 97–100, 1996.

[49] U. Kunzmann, G. Von-Wagner, J. Schöchlin, and A. Bolz. Parameter extraction of ECG signals in real-time. *Biomedizinische Technik*, 47(1):875–878, 2002.

[50] L. Lhotská, M. Fejtová, J. Macek, and D. Novák. Biological data preprocessing: A case study. In *Intelligent and Adaptive Systems in Medicine*, pages 77–99, 2003.

[51] H. Liang, S. Lukkarinen, and I. Hartimo. Heart sound segmentation algorithm based on heart sound envelogram. In *Computers in Cardiology*, volume 24, pages 105–108, 1997.

[52] R. Limacher. Removal of power-line interference from the ECG signal by an adaptive digital filter. In *Proceedings of European Telemetry Conference*, pages 300+, 1996.

[53] S. Lukkarinen, H. Liang, and I. Hartimo. A heart sound segmentation algorithm using wavelet decomposition and reconstruction. In *19th Annual International Conference IEEE Engineering in Medicine and Biology Society*, pages 1630–1633, 1997.

[54] S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, 1998.

[55] D. McMarrus, K. D. Neubert, and E. Cramer. Characterization and elimination of AC noise in electrocardiograms: A comparision of digital filtering methods. *Computers and Biomedical Research*, 26(1):48–67, 1993.

[56] C. R. Meyer and H.Ñ. Keiser. Electrocardiogram baseline noise estimations and removal using cubic splines and state-space computation techniques. *Computers and Biomedical Research*, 10(5):459–470, 1977.

[57] S. K. Mitra. *Digital Signal Processing: A computer-Based Approach*. McGraw-Hill, 2 edition, 2001.

[58] S. Molau, M. Pitz, R. Schluter, and H. Ney. Computing mel-frequency cepstral coefficients on the power spectrum. In *IEEE Acoustics, Speech, and Signal Processing*, volume 1, pages 73–76, 2001.

[59] D. W. Mortara. Digital filter for ECG signals. In *Computers in Cardiology*, pages 511–514, 1977.

[60] M. Okada. A digital filter for the QRS complex detection. *IEEE Transactions on Biomedical Engineering*, 26(12):700–703, 1979.

[61] S. Olmos, J. García, R. Jané, and P. Laguna. Análisis de la señal ECG con expansiones ortogonales de reducido número de coeficientes. In *XV Congreso Anual de la Sociedad Española de Ingeniería Biomédica*, pages 89–92, 1997.

[62] A. V. Oppenheim and R. W. Schafer. *Discrete-Time Signal Processing*. Prentice-Hall International, 1989.

[63] N. J. Outram, E. C. Ifeachor, P. W. Van-Eetvelt, and J. S. H. Curnow. Techniques for optimal enhancement and feature extraction of fetal electrocardiogram. *IEE Proceedings*, 142(6):482–489, 1995.

[64] T. S. Parker and L. O. Chua. *Practical Numerical Algorithms for Chaotic Systems*. Springer-Verlag, 1989.

[65] I. W. Philips. Adaptive noise removal from biomedical signals using warped polynomials. *IEEE Transaccions on Biomedical Engineering*, 43(5):480–492, 1996.

[66] D. Popivanov and A. Mineva. Testing procedures for non-stationarity and non-linearity in physiological signals. *Mathematical Biosciences*, 157:303–320, 1999.

[67] I. Provaznik. *Wavelet analysis for signal detection applications to experimental cardiology research*. PhD thesis, Brno University of Technology, 1996.

[68] R. Rangayyan. *Biomedical signal analysis: A case-study approach*. IEEE press in Biomedical Engineering, 2002.

[69] M. Reyna and R. Jané. Análisis multi-wavelet para la detección de conductividad ventricular anormal en señales ECG de alta resolución. *Biomédica*, 12(2):98–110, 2001.

[70] J. Ritola and S. Lukkarinen. Comparison of time-frecuency distributions in the heart sounds analysis. *Medical and Biological Engineering and Computing*, 34(Suppl 1. Part 1):89–90, 1996.

[71] A. Rosado, G. Camps, J. Guerrero, J. V. Francés, J. Muñoz, and A. J. Serrano. Enhancing feature extraction for VF detection using data mining techniques. In *Computers in Cardiology*, pages 209–212, 2002.

[72] M. Rosenstein, J. Collins, and C. D. Luca. A practical method for calculating largest lyapunov exponents from small data sets. *Physica D*, 65(1-2):117–134, 1993.

[73] M. Santos. Detecção da primeira e segunda bulhas ($s1$ e $s2$) dos sons cardíacos através de análise tempo-freqüência. In *NAIS DO CBEB´2000 Processamento de Sinais Biológicos*, 2000.

[74] S. Sardy. Minimax threshold for denoising complex signals with waveshrink. *IEEE Transactions on Signal Processing*, 48(4):1023–1028, 2000.

[75] T. Sauer, J. Yorke, and M. Casdagli. Embeddology. *Journal of Statistical Physics*, 65:579–616, 1991.

[76] O. Say, Z. Dokur, and T. Olmez. Classification of heart sounds by using wavelet transform. In *IEEE Engineering in Medicine and Biology Society (EMBS/BMES)*, pages 128–129, 2002.

[77] T. Schreiber. *Nonlinear analysis of physiological data*, chapter Processing of Physiological Data, pages 7–21. Springer, 1998.

[78] R. Silipo. Investigating electrocardiographic features in fuzzy models for cardiac arrhythmia classification. Technical report, International Computer Science Institute, Berkeley, USA, 1999.

[79] G. Simon and M. Verleysen. High-dimensional delay selection for regression models with mutual information and distance-to-diagonal criteria. *Neurocomputing*, 70:1265–1275, 2007.

[80] J. Sprott and G. Rowlands. *Chaos Data Analyzer: The Professional Version 2.1*. American Institute of Physics, Physics Academic Software, New York, 1998.

[81] L. Sörnmo. Time-varying digital filtering of ECG baseline wander. *Medical and Biological Engineering and Computers*, 1993.

[82] L. Sörnmo, P. O. Börjesson, M. Nygards, and O. Pahlm. A method for evaluation of QRS shape features using a mathematical model for the ECG. *IEEE Transactions on Biomedical Engineering*, 28(10):713–717, 1981.

[83] T. Stamkopoulos, K. Diamantaras, N. Maglaveras, and M. Strintzis. ECG analysis using nonlinear PCA neural networks for ischemia detection. *IEEE Transactions on Signal Processing*, 46(11):3058–3067, 1998.

[84] M. Åström. *Detection and Classifcation in Electrocardiac Signals*. PhD thesis, Lund University, Sweden, 2003.

[85] F. Takens. *Dynamical Systems and Turbulence, Lecture Notes in Mathematics*, volume 898, chapter Detecting strange attractors in turbulence, page 366–381. Springer-Verlag, 1981.

[86] N. Thakor and Y. S. Zhu. Applications of adaptive filtering to ECG analysis: Noise cancellation and arrhytmia detection. *IEEE Transaccions on Biomedial Engineering*, 38(8):785–794, 1991.

[87] J. Thelier, S. Eubank, A. Longtin, B. Galdrikian, and D. Farmer. Testing for nonlinearity in time series: The method of surrogate data. *Physica D*, 58:77–94, 1992.

[88] P. E. Tikkanen. Nonlinear wavelet and wavelet packet denoising of ECG signal. *Biological Cybernetics*, 80(4):259–267, 1999.

[89] O. Vainio and S. Ovaska. Adaptive lowpass filters for zero-crossing detectors. In *28th Annual Conference of Industrial Electronics Society*, volume 2, pages 1483–1486, 2002.

[90] J. Vila. *Análisis de la variabilidad de señales fisiológicas, integración en un sistema de monitorización inteligente*. PhD thesis, Universidad de Santiago de Compostela, Departamento de Electrónica y Computación, 1996.

[91] G. G. Walter and X. Shen. *avelets and other orthogonal systems*. Chapman & Hall/CRC, 2 edition, 2001.

[92] B. Widrow, J. R. Glover, J. M. McCool, J. Kaunitz, C. S. Williams, R. H. Hearn, J. R. Zeidler, E. Dong, and R. C. Goodlin. Adaptive noise cancelling: Principles and applications. *Proceedings of the IEEE*, 63(12):1692–1716, 1975.

[93] W. J. Williams. *Time Frequency and Wavelets in Biomedical Signal Processing*, chapter Recent Advances in Time-Frequency Representations: Some Theoretical Foundation. IEEE press, 1998.

[94] M. Yelderman, B. Widrow, J. Cioffi, E. Hesler, and J. Leddy. ECG enhancement by adaptive cancellation of electrosurgical interference. *IEEE Transaccions on Biomedical Engineering*, BME-30(7):392–398, 1983.

[95] X. P. Zhang and Z. Q. Luo. A new time-scale adaptive denoising method based on wavelet shrinkage. In *Proceedings of the Acoustics, Speech, and Signal Processing*, pages 1629–1632, 1999.

CAPÍTULO 3

Análisis de relevancia en estructuras multivariadas

Non sunt entia multiplicanda
praeter necessitatem.

(Guillermo de Ockam)

Este capítulo comienza con una discusión acerca del modelo general de *sistemas gaussianos* para luego tratar acerca de la representación de datos multidimensionales en un espacio reducido de variables como una etapa precedente a la clasificación en el reconocimiento de patrones. La razón por la cual se realiza el procedimiento de reducir dimensiones, tiene varios aspectos: disminuir la complejidad de la representación, mejorar los resultados de clasificación, remover la información redundante e irrelevante, y en algunos casos, descubrir la estructura esencial de los datos para obtener una representación gráfica. Todo esto se obtiene a través de un *análisis de relevancia*, el cual se encarga de dirigir las proyecciones en el espacio de representación.

3.1. Modelo general de sistemas gaussianos

El modelo básico considera un *sistema dinámico* lineal de tiempo discreto con ruido gaussiano y se asume que los estados del sistema pueden en cualquier momento ser resumidos mediante un vector $\mathbf{\Phi}$ de *variables de estado*, ó *causas*, de dimensión q, el cual no puede ser observado directamente. Así, el sistema también produce en cada intervalo de tiempo una *salida observable*, notada por un vector $\mathbf{\Psi}$ de dimensión p, del cual sí se tiene acceso. Se asume que el estado $\mathbf{\Phi}$ evoluciona de acuerdo a una dinámica markoviana simple de primer orden. Cada vector de salida ($\mathbf{\Psi}$) es generado a partir de un estado actual mediante un proceso de observaciones lineal simple. Tanto la evolución de los estados como el proceso de observaciones son perturbados por ruido gaussiano aditivo el cual también está oculto. Si las variables de estado $\mathbf{\Phi}$ usan valores continuos, el modelo generativo básico puede

escribirse como:

$$\mathbf{\Phi}_{t+1} = \mathbf{G}\mathbf{\Phi}_t + \boldsymbol{\chi}_t = \mathbf{G}\mathbf{\Phi}_t + \boldsymbol{\chi}_\bullet \qquad \boldsymbol{\chi}_\bullet \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}) \tag{3.1}$$

$$\mathbf{\Psi}_t = \mathbf{C}\mathbf{\Phi}_t + \boldsymbol{\eta}_t = \mathbf{C}\mathbf{\Phi}_t + \boldsymbol{\eta}_\bullet \qquad \boldsymbol{\eta}_\bullet \sim \mathcal{N}(\mathbf{0}, \mathbf{R}) \tag{3.2}$$

donde $\mathbf{G}_{q\times q}$ es la *matriz de transición de estados* y $\mathbf{C}_{p\times q}$ es la *matriz generadora de observaciones* o *medidas*.

Los vectores $\boldsymbol{\chi}$ de dimensión q y $\boldsymbol{\eta}$ de dimensión p son variables aleatorias que representan los ruidos en la evolución de los estados y en las observaciones respectivamente, siendo independientes entre sí como también de los valores de $\mathbf{\Phi}$ y $\mathbf{\Psi}$. Ambas fuentes de ruido son temporalmente blancas (no correlacionadas de un intervalo de tiempo a otro) y distribuidas espacialmente de una forma gaussiana con media cero y matrices de covarianza denotadas como $\mathbf{Q}$ y $\mathbf{R}$ respectivamente. Se toman las notaciones $\boldsymbol{\chi}_\bullet$ y $\boldsymbol{\eta}_\bullet$ en lugar de $\boldsymbol{\chi}_t$ y $\boldsymbol{\eta}_t$ para enfatizar que los procesos de ruido no tiene ningún conocimiento del índice de tiempo. Sin pérdida de generalidad se restringen las fuentes de ruido a media cero. Dado que el ruido de evolución de los estados es gaussiano y sus dinámicas son lineales, entonces $\mathbf{\Phi}_t$ es un *proceso aleatorio gauss-markoviano de primer orden*. Nótese que los procesos de ruido son elementos esenciales del modelo: sin el proceso de ruido $\boldsymbol{\chi}_\bullet$, el estado $\mathbf{\Phi}_t$ siempre podría contraerse o expanderse exponencialmente en la dirección del eigenvector principal de $\mathbf{G}$; de igual forma en la ausencia del ruido de observación $\boldsymbol{\eta}_\bullet$, el estado no podría esconderse por mayor tiempo. La Figura 3.1 ilustra el modelo básico.

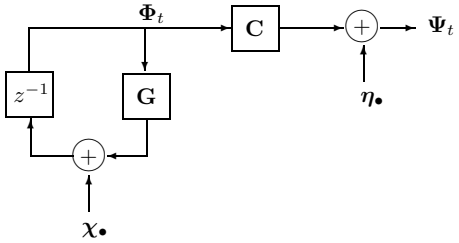


Figura 3.1. Modelo general de un sistema dinámico lineal [37]

Nótese que el modelo puede degenerarse: toda la estructura en la matriz $\mathbf{Q}$ puede moverse al interior de las matrices $\mathbf{G}$ y $\mathbf{C}$. Esto significa que se puede trabajar con modelos en los cuales $\mathbf{Q}$ es la matriz identidad sin que haya pérdida de generalidad. Desde luego, $\mathbf{R}$ no se puede restringir de la misma manera dado que los valores $\mathbf{\Psi}_t$ son observados y por tanto no hay posibilidad para blanquearlos o, por otra parte, reescalarlos. Finalmente, los componentes del vector de estado pueden reordenarse arbitrariamente; esto corresponde a intercambiar las columnas de $\mathbf{C}$ y también de $\mathbf{G}$. Típicamente se escoge un orden basado en las normas de las columnas de $\mathbf{C}$ lo cual resuelve este degeneramiento.

La idea principal de los modelos en las ecuaciones (3.1) y (3.2) es que la secuencia escondida de estado $\mathbf{\Phi}_t$ podría ser una *proyección informativa de menor dimensión*, o una forma de *explicación*, de la compleja secuencia de observaciones $\mathbf{\Psi}_t$. Con la ayuda del modelo dinámico y de ruido, los estados deben capturar las causas inherentes de los datos más fácilmente que considerando las observaciones en sí mismas. Por esta razón, frecuentemente se trabaja con dimensiones de estado mucho menores que el número de observaciones, es decir, $q \ll p$. Se asume que tanto $\mathbf{G}$ como $\mathbf{C}$ son de rango q y que $\mathbf{Q}$ y $\mathbf{R}$

son siempre de rango completo.

La popularidad de los *modelos gaussianos* lineales es debida a dos propiedades importantes:

1. La suma de dos cantidades con *distribución gaussiana* independientes es también gaussiana.

2. La salida de un *sistema lineal* cuando la entrada tiene distribución gaussiana es también gaussiana.

De acuerdo a esto, si se asume el estado inicial $\mathbf{\Phi}_1$ del sistema distribuido de forma gaussiana:

$$\mathbf{\Phi}_1 \sim \mathscr{N}(\boldsymbol{\mu}_1, \mathbf{Q}_1)$$

entonces todos los estados futuros $\mathbf{\Phi}_t$ y observaciones $\mathbf{\Psi}_t$ serán también distribuciones gaussianas. En efecto, se pueden escribir de manera explícita las expresiones para las esperanzas condicionales de los estados y las observaciones, así:

$$P\left(\mathbf{\Phi}_{t+1} \mid \mathbf{\Phi}_t\right) = \mathscr{N}(\mathbf{G}\mathbf{\Phi}_t, \mathbf{Q}) \mid_{\mathbf{\Phi}_{t+1}}$$

$$P\left(\mathbf{\Psi}_t \mid \mathbf{\Phi}_t\right) = \mathscr{N}(\mathbf{C}\Phi_t, \mathbf{R}) \mid_{\mathbf{\Psi}_t}$$

Por otro lado, los modelos de variable latente tienen un amplio rango de aplicación en el análisis de datos. En algunos casos se conocen exactamente cuales son los estados ocultos que deben tomarse en cuenta para luego estimarlos. En esos casos, usualmente, se hace el registro a priori de las matrices que contienen la información relacionada a la evolución de las observaciones y/o estados, basados en el conocimiento que se tiene de la estructura del problema, o en la física. El énfasis consiste en lograr una inferencia ajustada de la información oculta a partir de alguna información que sí se tiene. Existen otros casos en los que se requieren descubrir causas o explicaciones para los datos y no hay un modelo explícito que facilite esta información, teniendo en cuenta que los procesos de evolución de los estados y las observaciones son completamente desconocidas. Alternativamente, el énfasis consiste en realizar un *entrenamiento* robusto con un conjunto pequeño de parámetros que modelen bien los datos observados. Esos dos objetivos: la estimación de los estados ocultos dadas las observaciones y un modelo, y el entrenamiento de los parámetros del modelo; típicamente se relacionan con la resolución de dos problemas distintos: *inferencia* e *identificación de sistemas*.

La *inferencia* o *filtración/suavizado* siguen el siguiente planteamiento: dado un modelo de parámetros fijo $\{\mathbf{G}, \mathbf{C}, \mathbf{Q}, \mathbf{R}, \boldsymbol{\mu}_1, \mathbf{Q}_1\}$ ¿Qué puede decirse acerca de la secuencia desconocida de estados ocultos a partir de algunas observaciones? Un procedimiento básico consiste en calcular la probabilidad total de una secuencia de observaciones:

$$P\left(\{\mathbf{\Psi}_1 \ldots \mathbf{\Psi}_\wp\}\right) = \int_{\text{Todos los posibles } \{\mathbf{\Phi}_1 \ldots \mathbf{\Phi}_\wp\}} P\left(\{\mathbf{\Phi}_1 \ldots \mathbf{\Phi}_\wp\}, \{\mathbf{\Psi}_1 \ldots \mathbf{\Psi}_\wp\}\right) d\{\mathbf{\Phi}_1 \ldots \mathbf{\Phi}_\wp\} \tag{3.3}$$

se asume que $\wp$ es tanto el número de observaciones como de variables de estado. Este procedimiento requiere una vía eficiente de integración de las probabilidades conjuntas sobre todas las posibilidades en el espacio de estados. Una vez esta integración es obtenida no es difícil calcular la distribución condicional para cada una de las secuencias de los estados ocultos propuestos, dadas las observaciones, mediante la división de la probabilidad conjunta entre la probabilidad total de las observaciones:

$$P\left(\{\mathbf{\Phi}_1 \dots \mathbf{\Phi}_\wp\} \mid \{\mathbf{\Psi}_1 \dots \mathbf{\Psi}_\wp\}\right) = \frac{P\left(\{\mathbf{\Phi}_1 \dots \mathbf{\Phi}_\wp\}, \{\mathbf{\Psi}_1 \dots \mathbf{\Psi}_\wp\}\right)}{P\left(\{\mathbf{\Psi}_1 \dots \mathbf{\Psi}_\wp\}\right)}$$

A menudo, el interés radica en la distribución de los estados ocultos en el momento t. En procedimientos de filtración se calcula la probabilidad condicional a posteriori,

$$P\left(\mathbf{\Phi}_t \mid \{\mathbf{\Psi}_1 \dots \mathbf{\Psi}_t\}\right)$$

dadas todas las observaciones incluyendo el tiempo t. En procedimientos de suavizado se calcula la distribución sobre $\mathbf{x}_t$,

$$P\left(\mathbf{\Phi}_t \mid \{\mathbf{\Psi}_1 \dots \mathbf{\Psi}_\wp\}\right)$$

dadas todas las secuencias enteras de las observaciones.

El otro problema interesante con modelos gaussianos lineales es el *entrenamiento* e *identificación de sistemas*: dada solamente una secuencia de las observaciones de salida $\{\mathbf{\Psi}_1 \dots \mathbf{\Psi}_\wp\}$, encontrar los parámetros $\{\mathbf{G}, \mathbf{C}, \mathbf{Q}, \mathbf{R}, \boldsymbol{\mu}_1, \mathbf{Q}_1\}$, los cuales maximizan la probabilidad de las observaciones de acuerdo a la expresión (3.3).

Tomando en cuenta el modelo básico y el proceso de entrenamiento se encuentra que la consideración de casos lineales específicos del modelo, donde la variable de estado oculta $\mathbf{\Phi}$ es continua y los procesos de ruido son gaussianos, permite explicar la relación que hay entre diferentes métodos de representación de datos como son: el análisis de factores, análisis de componentes principales y filtro Kalman. El estudio requiere dividir los modelos en aquellos que generan *datos estáticos* y *datos dinámicos*. Los datos estáticos son independientes del tiempo (ninguna información se perdería si se permuta el orden de los datos $\mathbf{\Psi}_t$), mientras que para los datos dinámicos el orden temporal es relevante.

En este capítulo solo se tendrá en cuenta el análisis de datos estáticos orientado a la representación reducida de observaciones multidimensionales, ya que en muchas situaciones se puede asumir que el conjunto de datos ha sido generado idéntica e independientemente, es decir, no existe un ordenamiento temporal de los datos sino que son tomados como una simple colección de observaciones. Para esto, se asume que el vector $\mathbf{\Phi}$ no es dinámico, por lo tanto la matriz $\mathbf{G}$ es una matriz nula y $\mathbf{\Phi}$ es una constante perturbada por ruido. El nuevo modelo generativo es expresado como:

$$\mathbf{G} = \mathbf{0} \quad \Rightarrow \quad \mathbf{\Phi}_\bullet = \boldsymbol{\chi}_\bullet \qquad \boldsymbol{\chi}_\bullet \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}) \tag{3.4}$$

$$\mathbf{\Psi}_\bullet = \mathbf{C}\mathbf{\Phi}_\bullet + \boldsymbol{\eta}_\bullet \qquad \boldsymbol{\eta}_\bullet \sim \mathcal{N}(\mathbf{0}, \mathbf{R}) \tag{3.5}$$

Se observa que $\mathbf{\Phi}_t$ es afectado solamente por el ruido $\boldsymbol{\chi}_\bullet$ y $\mathbf{\Psi}_t$ depende solamente de $\mathbf{\Phi}_t$, de tal manera que toda la dependencia temporal ha desaparecido. De esta manera,

orientado al reconocimiento de patrones, el modelo estático de representación puede tratarse como una estructura multivariante, con el fin de hallar un nuevo espacio de caracterización compuesto por *variables relevantes*, de forma que las variables redundantes, correlacionadas e irrelevantes pueden descartarse usando funciones de evaluación generadas mediante un *análisis de relevancia*.

3.2. Relevancia como medida de representación

La representación efectiva de patrones relacionados al comportamiento fisiológico, a partir del análisis de múltiples características extraídas de bioseñales, generalmente es obtenida mediante procedimientos basados en técnicas de *reducción de dimensiones*. De la Figura 3.2 se puede notar que el análisis de relevancia tiene como objetivo distinguir las variables que de manera efectiva están representando la dinámica del fenómeno fisiológico subyacente de acuerdo a una medida de evaluación, llamadas *características relevantes*; además busca descartar las variables que tienen información repetida, *características redundantes* y las que no aportan al objetivo de representación, *características irrelevantes*.

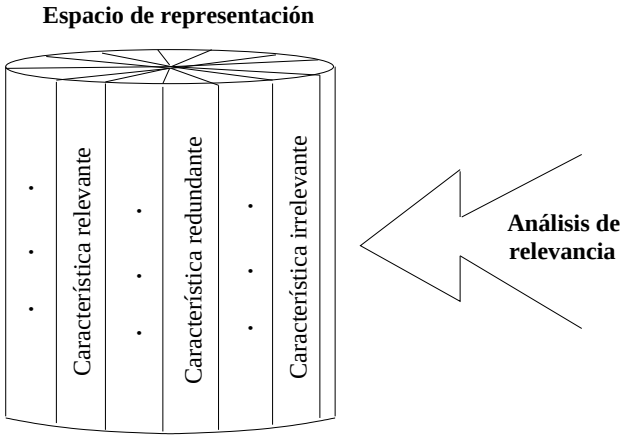


Figura 3.2. Análisis de relevancia sobre el espacio de representación

3.2.1. Definición generalizada de la relevancia

Sea $\mathcal{X}$ un espacio de objetos donde cada uno de los elementos está descrito por el conjunto de variables aleatorias $\boldsymbol{\xi} = \{\xi_1, \xi_2, \ldots, \xi_p\}$ y está asociado a una y sólo una etiqueta del conjunto de etiquetas $\boldsymbol{c} = \{c_k | k \in K\}$. De forma que, $\xi_i : \mathcal{X} \to \mathbb{R}$ y $c : \mathcal{X} \to \{c_k | k \in K\}$. Es decir, si $\boldsymbol{x} \in \mathcal{X}$ entonces $\mathbf{x} = [\xi_1(\boldsymbol{x}), \xi_2(\boldsymbol{x}), \ldots, \xi_p(\boldsymbol{x})]$ y $c(\boldsymbol{x}) \in \boldsymbol{c}$.

Definición 3.1 (Relevancia generalizada)

Sea $\boldsymbol{\xi}$ un conjunto de variables, $\varphi(\boldsymbol{\xi})$ partes de $\boldsymbol{\xi}$ y $\mu : \varphi(\boldsymbol{\xi}) \to \mathbb{R}$ una función tal que $\forall \check{\boldsymbol{\zeta}} \in \varphi(\boldsymbol{\xi})$, $\mu(\check{\boldsymbol{\zeta}}) \leq \mu(\xi)$. Sea $[\boldsymbol{\xi}] = \left\{\check{\boldsymbol{\zeta}} \in \varphi(\boldsymbol{\xi}) : \mu(\check{\boldsymbol{\zeta}}) = \mu(\xi)\right\}$, como $\boldsymbol{\xi}$ es un conjunto finito entonces $[\boldsymbol{\xi}]$ también es finito, de manera que $[\boldsymbol{\xi}] = \left\{\hat{\boldsymbol{\zeta}}_i\right\}_{i=1}^{\mathcal{D}}$. Luego, todo $\hat{\boldsymbol{\zeta}}_i \in [\boldsymbol{\xi}]$ es llamado **subconjunto relevante de variables** con respecto a la función μ.

Observación 3.1 (Estratificación de la relevancia)

Si $\mathcal{D} = 1$, todas las variables del conjunto $\boldsymbol{\xi}$ serían relevantes. Si $\mathcal{D} > 1$, tomando en cuenta que $[\boldsymbol{\xi}]$ es finito, los subconjuntos $\hat{\boldsymbol{\zeta}}_i$ se pueden ordenar por contenencia, tomando $\{\boldsymbol{\zeta}_i\}_{i=1}^{d}$ como

el conjunto de subconjuntos minimales. Se denomina **relevancia fuerte** al procedimiento de distinguir a partir de los subconjuntos minimales $\boldsymbol{\zeta}_i$ las variables que tienen mayor cantidad de información (de acuerdo a μ) no repetida y equivalente al de todo el conjunto $\boldsymbol{\xi}$. En efecto, las *características fuertemente relevantes* quedan en la intersección de todos los subconjuntos minimales $\boldsymbol{\zeta}_i$. Se denomina **relevancia débil** al procedimiento que distingue las variables que aunque hacen parte de los subconjuntos relevantes $\hat{\boldsymbol{\zeta}}_i$, su información puede ser inferida a partir de otra u otras. En efecto, las variables que quedan en la unión de todos los subconjuntos minimales $\boldsymbol{\zeta}_i$ menos la intersección, son llamadas *características débilmente relevantes* de $\boldsymbol{\xi}$ con respecto a μ.

Como se puede observar en la Figura 3.3, la intersección de los subconjuntos minimales puede ser nula, lo cual indicaría que la información relevante no se concentra en un subconjunto único de variables, sino que existen varias combinaciones minimales que ofrecen la misma medida de relevancia asociada a μ. En este caso cuando todas las características de los subconjuntos minimales son redundantes, se toma cualquiera de los subconjuntos $\boldsymbol{\zeta}_i$ para representar $\boldsymbol{\xi}$. Es importante anotar que en orden de construir un espacio de representación relevante de un *sistema fisiológico*, se toma el subconjunto minimal de características que signifique menos costo de estimación.

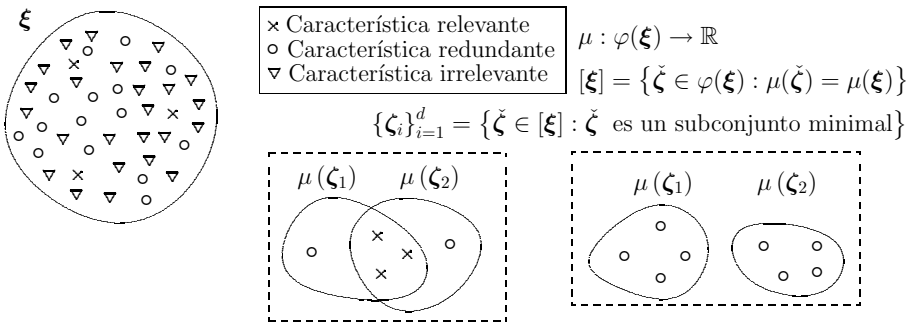


Figura 3.3. Interpretación gráfica de la relevancia generalizada

Es posible obtener casos particulares de la Definición 3.1, cuando la *medida de relevancia* μ se asocia a una medida estadística, informativa, espacial, de consistencia o de clasificación determinando la relevancia estadística, informativa, algebraica, por consistencia o por clasificación respectivamente.

Relevancia estadística

Se asocia μ a una medida que compara las propiedades estadísticas del conjunto de entrenamiento, por ejemplo, los valores de aleatoriedad o estadísticas de alto orden. Para este tipo de relevancia, μ puede definirse como:

$$\mu(\check{\boldsymbol{\zeta}}) = -\|P\left(\boldsymbol{c}|\boldsymbol{\xi}\right) - P\left(\boldsymbol{c}|\boldsymbol{\xi} \setminus \check{\boldsymbol{\zeta}}\right)\|$$

donde $\check{\boldsymbol{\zeta}}$ es cualquier subconjunto de características que pertenece a $\varphi(\boldsymbol{\xi})$ y P es la función de distribución de probabilidad. Se dice que un subconjunto $\check{\boldsymbol{\zeta}}_1$ es más *relevante* que el subconjunto $\check{\boldsymbol{\zeta}}_2$ si $\mu(\check{\boldsymbol{\zeta}}_1) > \mu(\check{\boldsymbol{\zeta}}_2)$.

Relevancia algebraica

El contexto operativo de la *relevancia*, puede ser definido usando aproximaciones algebraicas para determinar su influencia en la conformación de agrupaciones en los datos. De acuerdo a esto, se puede construir una función de energía que presente su valor óptimo al considerarse el subconjunto deseado de variables [55]. Se asocia μ a una medida que compara las propiedades espaciales del *conjunto de entrenamiento* $\mathbf{X}$. Sea $\mathbf{A}_{\boldsymbol{\xi}}$ la matriz de afinidad para el conjunto inicial de variables, determinada como:

$$\mathbf{A}_{\boldsymbol{\xi}} = \sum_{i=1}^{p} \xi_i \xi_i^{\top}$$

donde ξ_i es tomado como un vector de dimensiones $n \times 1$ y $\xi_i \xi_i^{\top}$ es la matriz de rango unitario definida por el producto externo entre ξ_i y sí mismo. La coherencia de las agrupaciones que presentan los datos del conjunto de entrenamiento $\mathbf{X}$, puede ser medida mediante el análisis de las propiedades espectrales de $\mathbf{A}_{\boldsymbol{\xi}}$. Para el caso de r agrupaciones, los r autovalores más grandes de $\mathbf{A}_{\boldsymbol{\xi}}$ representan las correspondientes coherencias de los grupos y los componentes de un autovector representa la participación de las observaciones en cada agrupación. Los autovalores λ_j de $\mathbf{A}_{\boldsymbol{\xi}}$ decrecen como la interconexión de los puntos dentro de las agrupaciones se esparcen [41, 31]. De esta manera, se obtiene la siguiente medida:

$$\mu(\boldsymbol{\xi}) = \sum_{j=1}^{r} \lambda_j^2$$

De esta manera, un subconjunto $\check{\boldsymbol{\zeta}}_1$ de características es más *relevante* que el subconjunto $\check{\boldsymbol{\zeta}}_2$, si $\mu(\check{\boldsymbol{\zeta}}_1) > \mu(\check{\boldsymbol{\zeta}}_2)$.

Relevancia informativa

Un subconjunto de características A puede considerarse relevante a nivel informativo, si existe una función de comparación μ, con una medida de ganancia de información asociada, de forma que μ presente algún nivel predeterminado de significación para A, indicando alta capacidad de representación del fenómeno aleatorio analizado. El eje central del modelo de información es el concepto de *información relevante*, también denominado *concepto de pertinencia*, el cual se obtiene a través de procedimientos estadísticos basados en *probabilidades*. En general, a menor conocimiento implícito de los datos de una variable correspondiente a un proceso para el cual se quiere obtener un modelo, existe una mayor relevancia informativa en el caso de revelarse los datos de dicha variable, ya que el *sistema de entrenamiento* valora como especialmente informativos los datos que desconoce previamente, y la relación existente entre el estado de conocimiento a priori y a posteriori del sistema, determinará la relevancia informativa que poseen los datos de la variable. Por lo tanto, la cantidad de información respecto a la ocurrencia de un evento, es inversamente proporcional a su probabilidad. Una probabilidad de ocurrencia grande, ofrece una cantidad de información menor, y por ello implica menor relevancia informativa [26]. En general, la relevancia informativa es una expresión no lineal de la relevancia estadística, donde los términos de evaluación se asocian a un concepto denominado *redundancia*, R.

Para este tipo de relevancia se usa una función de entropía condicional que cuantifica la incertidumbre de la variable de interés $\boldsymbol{c}$ dados los datos de entrenamiento:

$$H(\boldsymbol{c}|\boldsymbol{\xi}) = -\sum P\left(\boldsymbol{\xi}, \boldsymbol{c}\right) \log P\left(\boldsymbol{c}|\boldsymbol{\xi}\right)$$

donde P es la función de distribución de probabilidad. Así, es posible obtener una medida del nivel de redundancia existente en el conjunto de variables $\boldsymbol{\xi}$, mediante la siguiente expresión:

$$R(\boldsymbol{\xi}) = \frac{H_{\text{máx}} - H(\boldsymbol{c}|\boldsymbol{\xi})}{H_{\text{máx}}} = \frac{\log n - H(\boldsymbol{c}|\boldsymbol{\xi})}{\log n}, \quad 0 \leq R \leq 1$$

De acuerdo a esto, para un subconjunto $\check{\boldsymbol{\zeta}}$ de características puede considerarse la medida $\mu(\check{\boldsymbol{\zeta}}) = -R(\check{\boldsymbol{\zeta}})$. Por lo tanto, se dice que un subconjunto $\check{\boldsymbol{\zeta}}_1$ es más *relevante* que un subconjunto $\check{\boldsymbol{\zeta}}_2$, si $\mu(\check{\boldsymbol{\zeta}}_1) > \mu(\check{\boldsymbol{\zeta}}_2)$.

Relevancia por consistencia

Se asocia a una medida generada a partir del concepto denominado *relación de inconsistencia*, y puede estimarse tomando en cuenta las siguientes consideraciones [7]:

a) Una observación $\mathbf{x}_\ell \in \mathbf{X}$, tomada como un punto multidimensional ($\mathbf{x}_\ell \in \mathbb{R}^p$), se considera *inconsistente* si existe al menos otra observación contenida en $\mathbf{X}$ en la que coincide el valor de las ordenadas del punto pero no coincide la etiqueta de clase $c_k \subset \boldsymbol{c}$.

b) Dado el conjunto $\mathcal{Z} = \{\boldsymbol{z}_\ell \in \mathbb{R}^q : \ell = 1, \ldots, n\}$ que resulta después de representar las observaciones con cada uno de los subconjuntos de características que pertenecen a $\varphi(\boldsymbol{\xi})$, el valor de inconsistencia correspondiente a $\boldsymbol{z}_\ell$ se estima tomando el número de veces que se repite en todo el conjunto, menos el número más grande para el que tome una sola etiqueta de clase.

c) La *relación de inconsistencia* para $\check{\boldsymbol{\zeta}} \in \varphi(\boldsymbol{\xi})$ se denota por $I_R\left(\check{\boldsymbol{\zeta}}\right)$ y corresponde a la suma de los valores de inconsistencia que presentan todas las observaciones del conjunto, dividido por el número total de observaciones.

En este caso, la medida para un subconjunto $\check{\boldsymbol{\zeta}}$ de características puede ser tomada como $\mu(\check{\boldsymbol{\zeta}}) = -I_R(\check{\boldsymbol{\zeta}})$. Por lo tanto, un subconjunto $\check{\boldsymbol{\zeta}}_1$ es más *relevante* que un subconjunto $\check{\boldsymbol{\zeta}}_2$, si $\mu(\check{\boldsymbol{\zeta}}_1) > \mu(\check{\boldsymbol{\zeta}}_2)$.

Relevancia por clasificación

Se asocia μ a una medida de *precisión de clasificación*. Los procedimientos que usan en su función de evaluación este tipo de relevancia son conocidos con el nombre de *wrapper* (es decir, dado un subconjunto de características $\check{\boldsymbol{\zeta}}$ y un conjunto de etiquetas de clase $\boldsymbol{c}$, la función μ es la misma función de clasificación). El uso de este tipo de medida es suficientemente preciso debido a que implica el análisis del propio clasificador para el cual se diseña la reducción de dimensiones, pero la exigencia computacional comparada con otras medidas puede llegar a ser bastante alta [6]. Por lo tanto, un subconjunto $\check{\boldsymbol{\zeta}}_1$ es más *relevante* que un subconjunto $\check{\boldsymbol{\zeta}}_2$, si $\mu(\check{\boldsymbol{\zeta}}_1) > \mu(\check{\boldsymbol{\zeta}}_2)$.

Comparación

Una descripción cualitativa del desempeño de los diferentes conceptos de relevancia para tareas concernientes al *entrenamiento de sistemas*, se muestra en la Tabla 3.1, donde se tienen en cuenta los siguientes parámetros:

1. *Capacidad de generalización*. Indica el nivel de aplicabilidad a diversos conjuntos de características (sin importar su origen) y las facilidades que ofrece como criterio de separabilidad para diferentes clasificadores, esto es, que no se restrinja el buen rendimiento a un sólo tipo de clasificadores.

2. *Complejidad computacional*. El tiempo empleado en el procedimiento que ejecuta la selección del subconjunto óptimo o sub-óptimo de características.

3. *Precisión*. Indica el mayor rendimiento posible del clasificador usando el subconjunto seleccionado.

Función μ	*Capacidad de generalización*	*Complejidad computacional*	*Precisión*
Relevancia algebraica	Sí	Bajo	–
Relevancia estadística	Sí	Bajo	–
Relevancia informativa	Sí	Bajo	–
Relevancia por consistencia	Sí	Moderado	–
Relevancia por clasificación	No	Alto	Muy alto

Tabla 3.1. Comparación de las diferentes medidas de relevancia

La notación (–) significa que no se puede concluir nada acerca de la precisión del clasificador con respecto a la respectiva función μ.

3.2.2. Función de evaluación en el análisis de relevancia

La razón por la que existe una amplia diversidad de conceptos que encierra la relevancia, es la fuerte dependencia con relación a la siguiente pregunta: "¿relevante a qué?", es por esto que algunas connotaciones pueden ser más apropiadas que otras dependiendo del objetivo del análisis [2]. De ahí la importancia de tomar en cuenta que el análisis de relevancia se encarga de direccionar las proyecciones usadas en la reducción de dimensiones cuando se busca una representación efectiva orientada al reconocimiento de patrones, de forma que el contexto de esta reducción depende de la medida de representación asociada a una *función de evaluación* del espacio inicial de entrenamiento. Esta función de evaluación se relaciona directamente con la función de relevancia μ, para lo cual, se usará la siguiente notación: Dado un espacio de objetos $\mathbf{X}$ donde cada objeto está asociado a una y solo una etiqueta de clase del conjunto de etiquetas $\boldsymbol{c} = \{c_k : k \in K\}$; se obtiene una medida de relevancia μ sobre un subconjunto de variables $\check{\boldsymbol{\zeta}} \in \varphi(\boldsymbol{\xi})$, de forma que si esta medida de relevancia toma en cuenta las etiquetas de clase a la que pertenecen los objetos representados por el subconjunto de variables, esta medida será llama *función de evaluación* del subconjunto de variables y será notada como $f_{\check{\boldsymbol{\zeta}}} \equiv f(\boldsymbol{c}, \check{\boldsymbol{\zeta}})$. Así, cada $f_{\check{\boldsymbol{\zeta}}}$ tiene asociada implícitamente la medida de relevancia $\mu(\check{\boldsymbol{\zeta}})$ en el sentido de la Definición 3.1.

3.3. Reducción de dimensiones y niveles de operación

En general, la reducción de dimensiones busca encontrar la representación sobre alguna variedad en un *sistema coordenado*, que permita obtener una representación alterna y compacta con la menor dimensión posible, optimizando una *función de evaluación* específica en el espacio resultante y regida por una *medida de relevancia* μ encargada de dirigir el contexto de representación.

Definición 3.2 (Reducción de dimensiones)

Sea el espacio inicial de análisis dado en forma de un conjunto $\mathbf{X} = \{\mathbf{x}_\ell \in \mathbb{R}^p : \ell = 1, \ldots, n\}$, tal que los valores que componen cada observación $\mathbf{x}_\ell$ corresponden a la estimación del *conjunto inicial de características* $\boldsymbol{\xi}$ de dimensión p. Se llama *conjunto reducido de características* al conjunto $\boldsymbol{\zeta}$, con dimensión q, $q < p$, tan pequeña como sea posible, asociado a un espacio reducido de representación $\mathbf{Z}$ que satisface las siguientes condiciones:

1. Existe una función o mapeo de reducción de dimensión $\mathcal{G}$, descrita de la forma:

$$\begin{aligned} \mathcal{G} : \mathbf{X} &\rightarrow \mathbf{Z} \\ \boldsymbol{\xi} &\mapsto \boldsymbol{\zeta} = \mathcal{G}\{\boldsymbol{\xi}\} \end{aligned}$$

2. Existe una función o mapeo de reconstrucción suave y no singular $\mathfrak{H}$, descrita de la forma:

$$\begin{aligned} \mathfrak{H} : \mathbf{Z} &\rightarrow \mathbf{Y} \subset \mathbf{X} \\ \boldsymbol{\zeta} &\mapsto \boldsymbol{\xi} = \mathfrak{H}\{\boldsymbol{\zeta}\} \end{aligned}$$

de tal manera que se cumplan las restricciones:

(a). La variedad $\mathbf{Y} \triangleq \mathfrak{H}\{\mathbf{Z}\}$ debe contener aproximadamente toda la información sobre el espacio inicial de análisis, $\mathbf{X} = \{\mathbf{x}_\ell : \ell = 1, \ldots, n\} \overset{\subseteq}{\sim} \mathbf{Y}$.

(b). Dada una métrica $m : \mathbb{R}^p \times \mathbb{R}^p \rightarrow \mathbb{R}^+ \cup \{0\}$ y considerando que $\hat{\mathbf{x}}_\ell = \mathfrak{H}\{\mathcal{G}\{\mathbf{x}_\ell\}\}$, se obtiene que $\boldsymbol{d} = \{d_\ell : d_\ell = m(\mathbf{x}_\ell, \hat{\mathbf{x}}_\ell); \ell = 1, \ldots, n\}$, para lo cual existe una función $\boldsymbol{\varepsilon} : \mathbb{R} \rightarrow \mathbb{R}$, de tal forma que $\boldsymbol{\varepsilon}(\boldsymbol{d})$ corresponde al error de reconstrucción obtenido entre $\mathbf{X}$ y $\mathbf{Y}$, y se debe cumplir que $\boldsymbol{\varepsilon}(\boldsymbol{d}) \rightarrow 0$.

De acuerdo a las Definiciones 3.1 y 3.2 se puede resumir la tarea de reducción de dimensiones a la siguiente expresión:

$$\mathbf{Z} = f(\boldsymbol{c}, \mathbf{X}) \rightarrow \underset{\mu(\varphi(\boldsymbol{\xi})),[\boldsymbol{\xi}]}{\text{Optimizar}} \quad f_{\check{\boldsymbol{\zeta}}}$$

Debido a las ventajas que ofrece la reducción de dimensiones orientada al reconocimiento de patrones para el análisis de estructuras multivariadas, puede tomarse en 3 capas o niveles de operación definidos de acuerdo a su aplicación (ver Figura 3.4): proyección, interpretación y visualización.

3.3.1. Nivel 1: Proyección

De acuerdo a la Definición 3.2, las observaciones de alta dimensión pueden ser representadas en un espacio de menor dimensión mediante una función de transformación o mapeo de los datos, optimizando algún criterio específico en el espacio resultante y usando todas las variables disponibles [4]. La literatura suele llamar al hallazgo de este espacio imagen: *extracción de características* [10, 53, 47]. El interés de esta reducción no sobrepasa el hecho

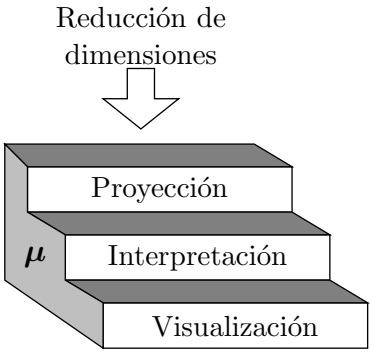


Figura 3.4. Niveles de operación en la reducción de dimensiones.

de encontrar un espacio de menor dimensión que permita disminuir el costo computacional o incrementar la tasa de acierto en la clasificación.

La transformación $\mathcal{G}$ puede ser una combinación lineal o no lineal de las variables originales de forma supervisada o no supervisada. Si el mapeo es lineal, la función de mapeo está bien definida y la tarea consiste en encontrar los coeficientes de tal manera que se optimice una función de evaluación, $f_{\tilde{\zeta}}$. Por lo tanto, si se tiene un criterio apropiado para evaluar la efectividad de las características, se pueden usar técnicas derivadas del álgebra lineal para criterios simples, o para el caso de criterios complejos, se pueden aplicar técnicas iterativas con el fin de determinar los coeficientes del mapeo. Desafortunadamente, en muchas aplicaciones de reconocimiento de patrones, hay características importantes que no son obtenidas a través de funciones lineales del espacio inicial de mediciones, por el contrario, son obtenidas mediante funciones altamente no lineales. Así, el problema se resume a encontrar una función de mapeo no lineal apropiada para los datos de entrenamiento.

Existen varias razones para realizar la extracción de características a partir de una estructura multivariada,

1. Reducir el ancho de banda de los datos de entrada.

2. Proveer un conjunto de variables relevante para un clasificador, permitiendo aumentar el desempeño, especialmente de clasificadores simples.

3. Disminuir la redundancia.

4. Obtener nuevas variables subyacentes o latentes que sean significativas para describir los procesos que generan las observaciones.

5. Representar los datos en una dimensión menor, con mínima pérdida de información.

Técnicas frecuentemente usadas en este nivel de reducción de dimensiones son: el análisis de componentes principales (PCA), en su forma estándar y probabilística (PPCA), el escalamiento multidimensional (MDS), el análisis factorial (FA), el análisis de componentes independientes (ICA) y la búsqueda de la proyección (PP).

Análisis de componentes principales A partir de las expresiones (3.4) y (3.5), se considera el caso para cuando $\mathbf{R} = \lim_{\rho \to 0} \rho \mathbf{I}$ y $\mathbf{Q} = \mathbf{I}$, donde se obtiene un modelo de representación multivariada conocido como *análisis de componentes principales* (PCA). La dirección de las proyecciones están dadas por $\mathbf{A} = \left(\mathbf{C}^{\mathsf{T}}\mathbf{C}\right)^{-1}\mathbf{C}^{\mathsf{T}}$, de forma que el nuevo espacio queda formado por componentes no correlacionados y obtenidos de acuerdo a la variabilidad de los datos originales. Si se tiene un conjunto de observaciones $\mathbf{X}$, donde cada punto está formado por un vector de características $\boldsymbol{\xi}$ de dimensión p, con vector de medias dado por $\boldsymbol{\mu}_{\boldsymbol{\xi}} = E\{\boldsymbol{\xi}\}$ y matriz de covarianza $\boldsymbol{\Sigma}_{\boldsymbol{\xi}} = E\left\{(\boldsymbol{\xi} - \boldsymbol{\mu}_{\boldsymbol{\xi}})^{\mathsf{T}}(\boldsymbol{\xi} - \boldsymbol{\mu}_{\boldsymbol{\xi}})\right\} > 0$, de forma que algunos de sus elementos o componentes están correlacionados. El vector $\boldsymbol{\xi}$ puede ser representado en un vector $\boldsymbol{\zeta}$ de menor dimensión y variables linealmente independientes, el cual representa la información estadística considerada importante y contenida en $\boldsymbol{\xi}$, mediante la siguiente expresión,

$$\boldsymbol{\zeta} = \mathbf{A}^{\top}\boldsymbol{\xi} \tag{3.6}$$

donde, $\mathbf{A}_{p\times q}$ es una matriz compuesta por q autovectores ortonormales obtenidos a partir de los q mayores autovalores λ_l de la matriz de covarianza $\boldsymbol{\Sigma}_{\boldsymbol{\xi}}$. Para el caso $q < p$, se obtiene la reducción dimensional del espacio inicial de representación. La función criterio de este principio de transformación consiste en hallar el espacio ortogonal que mejor represente la variabilidad de los datos. Los elementos de representación de la variable $\boldsymbol{\zeta}$ se denominan los *componentes principales* de $\boldsymbol{\xi}$, y son obtenidos mediante, $\boldsymbol{\zeta}_l = \boldsymbol{\lambda}_l^{\mathsf{T}}\boldsymbol{\xi}$, siendo $\boldsymbol{\lambda}_l$ el autovector asociado al autovalor λ_l, para los cuales de (3.6) se cumple que:

$$\begin{cases} \boldsymbol{\Sigma}^2_{\boldsymbol{\xi}_l} = \lambda_l. \\ \operatorname{cov}\{\zeta_m, \zeta_n\} = 0, \forall m \neq n. \end{cases}$$

PCA probabilístico De acuerdo a las expresiones (3.4) y (3.5), si se considera $\mathbf{Q} = \mathbf{I}$ y, en lugar de tomar $\mathbf{R} = \lim_{\rho \to 0} \rho \mathbf{I}$, se considera simplemente $\mathbf{R} = \alpha \mathbf{I}$, entonces se obtiene un modelo de representación para estructuras multivariadas llamado *análisis de componentes principales probabilístico* (PPCA). Las columnas de $\mathbf{C}$ componen el *subespacio principal* (el mismo subespacio formado por PCA). En general, PCA no contempla un modelo probabilístico para los datos observados, en este sentido, PPCA permite construir un esquema de trabajo que tiene en cuenta la estimación de la densidad de probabilidad de los datos, y tiene como principales ventajas [48, 49]:

a) La definición de una medida de verosimilitud que permite la comparación con otras técnicas probabilísticas, además facilita la aplicación de contrastes estadísticos.

b) La posibilidad de aplicar métodos de inferencia Bayesiana al combinar la verosimilitud con una probabilidad a priori.

c) Permite obtener las proyecciones de los datos para los componentes principales cuando hay valores de datos perdidos.

d) Se puede utilizar como un modelo general de densidad Gaussiano, en donde, los estimados de máxima verosimilitud para los parámetros asociados con la matriz de

covarianza, se pueden calcular eficientemente a partir de los componentes principales de los datos. Lo cual es útil en la reducción de dimensión, *sistemas de clasificación* y detección de *outliers*.

La determinación de las direcciones principales de un conjunto de datos observados es posible a través de la estimación de los parámetros de máxima verosimilitud en un modelo de variables latentes, el cual incluye la representación de un grupo de variables observadas por medio de un grupo menor de variables no observadas, de la forma,

$$\boldsymbol{\Psi} = \mathbf{g}\left(\boldsymbol{\Phi}, \mathbf{C}\right) + \boldsymbol{\eta} \tag{3.7}$$

donde $\mathbf{g}\left(\cdot\right)$ es una función que relaciona las variables latentes $\boldsymbol{\Phi}$ con los parámetros $\mathbf{C}$ y $\boldsymbol{\eta}$ es un proceso de ruido independiente de las variables $\boldsymbol{\Phi}$. Al definir una distribución a priori sobre $\boldsymbol{\Phi}$ junto con la distribución de $\boldsymbol{\eta}$, la ecuación (3.7) induce una distribución correspondiente en el espacio de los datos, por tanto, los parámetros del modelo pueden ser determinados por medio de técnicas de máxima verosimilitud.

Escalamiento multidimensional La técnica de escalamiento multidimensional (MDS - *Multi-Dimensional Scaling*) es una generalización del método de componentes principales cuando en lugar de contar con una matriz de observaciones $\mathbf{X}_{n\times p}$, se tiene se tiene una matriz cuadrada $\mathbf{M}$ de orden n, cuyos elementos m_{ij} representan una medida de relación (cercanía, diferencia, distancia o similaridad) entre las observaciones i y j de un conjunto de datos $\mathbf{X}$. Se puede construir el ordenamiento de $\mathbf{M}$ en forma de un conjunto de n puntos sobre un espacio con dimensión q, tal que las distancias entre los puntos i y j aproxime m_{ij}, o alguna función de m_{ij}. El análisis de componentes principales se puede emplear para obtener la aproximación de los puntos en el espacio con dimensión q, sobre el cual se realiza la proyección que minimice la suma de los cuadrados de la divergencia de las distancias, entre el espacio inicial de representación y el reducido de transformación. Las propiedades de distancia infieren la conformación de la matriz $\mathbf{M}$ simétrica, por lo cual puede ser descompuesta en la forma $\mathbf{Z}\mathbf{Z}^{\top}$, entonces las columnas de $\mathbf{Z}$ pueden ser tomadas en calidad de coordenadas de representación ó *coordenadas principales*. En este caso, $\mathbf{Z}^{\top}\mathbf{Z} = \boldsymbol{\Lambda}$ es una matriz diagonal. Por lo tanto, los ejes obtenidos de las coordenadas son los ejes principales de la configuración de los puntos. De aquí mismo, se obtiene que los valores propios $\{\lambda_i : i = 1, \ldots, n\}$, $\lambda_n = 0$, ordenados en orden descendente, corresponden a las sumas de los cuadrados, calculadas para los ejes secuenciales, que pueden ser empleadas para la determinación de la dimensión adecuada del espacio de representación, como en el caso de PCA [32]. Este análisis también es conocido como *análisis de similaridad*, la cual usa una medida de representación geométrica asociada a una función de evaluación $f_{\boldsymbol{\zeta}}$ que direcciona la función ó mapeo de reducción dado por $\mathcal{G}$.

Análisis factorial Si se restringe la matriz $\mathbf{R}$, de la expresión (3.5), a ser diagonal (control del ruido en las observaciones) y la matriz $\mathbf{Q}$, de la expresión (3.4), a ser la matriz identidad, entonces se obtiene un modelo estadístico conocido como *análisis factorial* (FA). El objetivo principal de este modelo es explicar un conjunto de variables observadas a partir de combinaciones lineales de un conjunto menor de *variables latentes* o *factores* que son de naturaleza aleatoria. Los factores son construcciones subyacentes no medibles que generan las variables observadas $\boldsymbol{\Psi}$, las cuales si presentan algún grado de correlación, es posible

inferir que existe un conjunto menor de variables que explique el fenómeno con menos redundancia [32, 35, 9]. De (3.5), los estados desconocidos $\mathbf{\Phi}$ son llamados *factores* y se representan en un vector de dimensión q, la matriz $\mathbf{C}$ de dimensiones $p \times q$ contiene los coeficientes que describen la forma en que los factores $\mathbf{\Phi}$ afectan las variables observadas $\mathbf{\Psi}$, también es denominada *matriz de cargas* y $\boldsymbol{\eta}$ es un vector de perturbaciones no observadas, en el cual se recoge el efecto de todas las variables distintas de los factores que influyen sobre $\mathbf{\Phi}$ y además se supone que tiene distribución $\mathcal{N}(\mathbf{0}, \mathbf{R})$. Generalmente, se incluye el término $\boldsymbol{\mu}$ que determina la media de las observaciones $\mathbf{\Psi}$, ya que tanto los factores como las perturbaciones tienen media cero. Así, el modelo de análisis factorial establece que el vector de datos observados se genera mediante la relación:

$$\mathbf{\Psi} = \boldsymbol{\mu} + \mathbf{C}\mathbf{\Phi} + \boldsymbol{\eta} \tag{3.8}$$

La expresión (3.8) implica que dada una muestra aleatoria generada por el modelo factorial y conformada por un conjunto de observaciones $\mathbf{X} = \{\mathbf{x}_i \in \mathbb{R}^p : i = 1, \ldots, n\}$, donde cada punto $\mathbf{x}_i$ es formado por la estimación de un vector de características $\boldsymbol{\xi}$ de dimensión p, los datos x_{ij} pueden escribirse como:

$$x_{ij} = \mu_j + c_{j1}\phi_{1i} + c_{j2}\phi_{2i} + \ldots + c_{jq}\phi_{qi} + \eta_{ij} \quad i = 1, \ldots, n \quad j = 1, \ldots, p$$

Los efectos de los factores sobre x_{ij} son el producto de los coeficientes $c_{j1}, c_{j2}, \ldots, c_{jq}$, que dependen de la relación entre cada factor y la variable j (iguales para todos los elementos de la muestra), por los valores de los q factores en el elemento muestral i, $\phi_{1i}, \phi_{2i}, \ldots, \phi_{qi}$. De esta manera, la matriz $\mathbf{X}_{n\times p}$, puede escribirse como:

$$\mathbf{X} = \mathbf{1}\boldsymbol{\mu}^\top + \tilde{\mathbf{\Phi}}\tilde{\mathbf{C}}^\top + \tilde{\boldsymbol{\eta}}$$

donde $\mathbf{1}$ es un vector $n \times 1$ de unos, $\tilde{\mathbf{\Phi}}$ es una matriz $(n \times q)$ que contiene los q factores para los n elementos de la muestra, $\tilde{\mathbf{C}}^\top$ es la transpuesta de la matriz de carga $(q \times p)$ y $\tilde{\boldsymbol{\eta}}$ es una matriz $(n \times p)$ de perturbaciones.

Análisis de componentes independientes A partir del modelo estadístico obtenido para variables latentes, expuesto en la expresión (3.8), se puede asumir sin pérdida de generalidad que la media de los vectores aleatorios se ha removido (como se consideró inicialmente en la expresión (3.5)), quedando

$$\mathbf{\Psi} = \mathbf{C}\mathbf{\Phi} + \boldsymbol{\eta} \tag{3.9}$$

Si a la expresión (3.9) se le impone la restricción de que las variables latentes $\mathbf{\Phi} = [\phi_1, \ldots, \phi_q]$ sean independientes mutua y estadísticamente, entonces se considera un modelo estadístico de representación conocido como *análisis de componentes independientes* (ICA) [17]. Lo que distingue a ICA de otros métodos es la búsqueda de componentes estadísticamente independientes y de distribución no gaussiana, lo cual hace que este modelo sea ampliamente usado para hallar factores ocultos independientes a partir de estructuras multivariadas y con posibles mezclas entre las variables [18]. Sin embargo, el problema definido en la expresión (3.9) tomando en consideración las restricciones para ICA no es fácil de resolver. Por tal motivo, la gran mayoría de los procedimientos se basan en una

definición más simple, en la cual el modelo ICA se plantea libre de ruido (*noise-free ICA model*), el cual parece ser suficiente para muchas aplicaciones [18], y está dado por,

$$\mathbf{\Psi} = \mathbf{C}\mathbf{\Phi} \tag{3.10}$$

El modelo ICA puede ser tomado como un modelo generador, ya que describe la forma en que los datos son generados por un proceso de mezcla de componentes ϕ_i. Los coeficientes de la matriz $\mathbf{C}_{(p\times q)}$, conocidos como *coeficientes de mezcla* se asumen desconocidos. Como se mencionó anteriormente, los componentes independientes deben ser no gaussianos, sin embargo, según el teorema del límite central es conocido que si se tiene un grupo numeroso de variables independientes con distribuciones arbitrarias, la suma de ellas tiende a distribuirse de forma normal [18]. Entonces, buscar las variables generadoras menos gaussianas implica intuitivamente que éstas no provienen de la combinación de otras variables, lo que a su vez se puede entender como independencia. Con el fin de asegurar que el modelo de ICA dado por (3.10) puede ser estimado, deben imponerse ciertas restricciones y suposiciones:

a) Las componentes (variables latentes) se asumen estadísticamente independientes. Básicamente, se dice que las variables aleatorias $\phi_1, \phi_2, \ldots, \phi_q$ son independientes si la información sobre el valor de ϕ_i no da ninguna información sobre el valor de ϕ_j para cualquier $i \neq j$. La independencia puede definirse por las densidades de probabilidad. Sea $P(\phi_1, \phi_2, \ldots, \phi_q)$ la función de densidad de probabilidad conjunta y $P_i(\phi_i)$ es la función de densidad de probabilidad marginal de ϕ_i. Entonces se puede decir que las variables ϕ_i son (mutuamente) independientes, si y sólo si la función de densidad de probabilidad conjunta se puede factorizar de la forma,

$$P(\phi_1, \phi_2, \ldots, \phi_q) = P(\phi_1)\, P(\phi_2) \cdots P(\phi_q)$$

b) Todas las componentes independientes deben tener distribuciones no gaussianas, con la posible excepción de una sola componente. Esto debido a que los cumulantes de orden superior son iguales a cero para las distribuciones gaussianas, y la estimación del modelo ICA emplea estadística de alto orden.

c) El número de variables observadas p debe ser por lo menos igual al número de variables latentes (componentes independientes), $p \geq q$.

d) La matriz $\mathbf{C}$ debe ser de rango completo en las columnas.

Por otra parte, de la ecuación (3.10) es posible ver que las siguientes indeterminaciones necesariamente existen:

1. No se pueden determinar las varianzas (potencia de las componentes independientes). Porque como se puede ver en la expresión (3.11), debido a que $\mathbf{C}$ y $\mathbf{\Phi}$ son desconocidos, cualquier escalar multiplicado por una de las variables latentes ϕ_i siempre puede cancelarse al dividir la correspondiente columna $\mathbf{c}_i$ de $\mathbf{C}$ entre el mismo escalar.

$$\mathbf{\Phi} = \sum_{i=1}^{q} \left(\frac{1}{\alpha_i} \mathbf{c}_i \right) (\phi_i \alpha_i) \tag{3.11}$$

2. No se puede determinar el orden de los componentes independientes. La razón es que se puede cambiar libremente el orden de los términos de la sumatoria, definida implícitamente en la operación matricial (3.10), y darle el nombre de primer componente independiente a cualquiera de ellos.

La estimación del modelo ICA usualmente se desarrolla a través de una función de evaluación $f_{\mathbf{\Phi}}$ y el respectivo algoritmo para optimizarla. La función de evaluación es la encargada de definir las propiedades estadísticas del método ICA por medio de medidas como: consistencia, varianza asintótica, robustez, etc. Por otra parte, el algoritmo encargado de optimizar (maximizando o minimizando) la función de evaluación, puede usar propiedades como la velocidad de convergencia, los requerimientos de memoria o la estabilidad numérica en la implementación del procedimiento.

Búsqueda de la proyección Una manera particular de resolver el problema definido por la expresión (3.10), es invertirlo hacia la búsqueda de direcciones mediante el ajuste de los coeficientes de una matriz $\mathbf{A}$ que permita proyectar los datos observados $\mathbf{\Psi}$ en un espacio de menor dimensión, de forma que las variables $\mathbf{\Phi}$ capturan la estructura original de los datos:

$$\mathbf{\Phi} = \mathbf{A}\mathbf{\Psi}$$

Este modelo de representación es conocido con el nombre de *búsqueda de la proyección* (PP), el cual para una nube de puntos de alta dimensión $\mathbf{X}$, selecciona de manera lineal proyecciones ortogonales *interesantes* sobre un espacio de dimensión más baja $\mathbf{Z}$, a través de la optimización de una función de evaluación llamada *índice de proyección*. En este contexto, el término *proyección interesante* se refiere a aquella que contiene una estructura alejada de la distribución normal. Para el caso en que se busque maximizar una función de evaluación consistente en la variabilidad de los datos y sujeta a la ortogonalidad de las proyecciones, se obtiene PCA, donde la expresión (3.6) para el conjunto de datos de alta dimensión $\mathbf{X}$, queda de la forma

$$\mathbf{Z} = \mathbf{X}\mathbf{A} \tag{3.12}$$

Un índice de proyección, ϱ, es un funcional real en el espacio de distribuciones $\mathbb{R}^q$:

$$\varrho : \ f \in \mathfrak{L}_2(\mathbb{R}^q) \rightarrow \rho = \varrho(f) \in \mathbb{R}$$

donde, $f = F_{\mathbf{A}}$ es la distribución que resulta después de proyectar los puntos de una variable aleatoria $\boldsymbol{\xi}$ de dimensión p con distribución F mediante una matriz de proyección $\mathbf{A} = [\mathbf{a}_1\ \mathbf{a}_2\ \cdots\ \mathbf{a}_q]$, quedando una variable aleatoria $\boldsymbol{\zeta}$ de dimensión q. La búsqueda de la proyección intenta encontrar direcciones de proyección $\mathbf{a}_i$ para una distribución F dada, las cuales producen la local óptima de ϱ. Para hacer que el problema de optimización sea independiente de la magnitud de los vectores de proyección, además de obtener direcciones no correlacionadas, se restringen los vectores $\mathbf{a}_i$ a ser de magnitud unitaria y mutuamente ortogonales (es decir, los vectores columna de $\mathbf{A}$ deben ser ortonormales). De esta manera, el problema de optimización se resume a satisfacer la siguiente condición:

Condición 3.1 (Direccionamiento ortonormal)

Optimizar $\varrho(\mathbf{A})$ sujeto a:

$$\mathbf{a}_i^\top \mathbf{a}_j = \delta_{ij} = \begin{cases} 1, i = j, \\ 0, i \neq j. \end{cases}$$

En general, PP es un método lineal que, a diferencia de PCA y FA, puede incorporar información más alta que la de segundo orden, y por lo tanto, es útil para conjuntos de datos con distribución no-normal. La complejidad computacional resulta más costosa que la que involucran los métodos de segundo orden. De otra parte, los índices de proyección usualmente miden algún aspecto de la no-normalidad de la distribución de los datos observados (representados por la matriz $\mathbf{X}$); partiendo de que la distribución normal es la menos interesante, se buscan las proyecciones que revelen las estructuras más interesantes contenidas en $\mathbf{X}$ sobre el espacio reducido $\mathbf{Z}$, optimizando una función de evaluación $f_{\check{\boldsymbol{\zeta}}}$, que para este caso es el índice de proyección, y sin dejar de lado la Condición 3.1. Este modelo, sin embargo, no es efectivo al analizar estructuras de naturaleza altamente no lineal [12, 28].

3.3.2. Nivel 2: Interpretación

Este nivel constituye una capa más profunda en la reducción de dimensiones, ya que no solo se busca un espacio de representación de menor dimensión, sino ejercer algún tipo de distinción entre las variables de entrada o establecer la conexión del espacio de salida con el de entrada para así determinar cuales son las variables del espacio inicial de características que contribuyen de manera efectiva en la representación. A este conjunto de variables se le conoce con el nombre de *características relevantes* y este nivel de reducción de dimensiones es conocido con el nombre de *selección de características*. Tomando en cuenta la Definición 3.1 y el comentario de la Sección 3.2.2, la reducción de dimensiones puede obtenerse mediante una *selección de características* relevantes usando procedimientos restringidos a una función de evaluación $f_{\check{\boldsymbol{\zeta}}}$, la cual está asociada a una medida de relevancia μ.

Observación 3.2 (Selección de características)

Es una forma particular de reducir dimensiones, y se presenta cuando la medida de relevancia μ asociada a la función de evaluación $f_{\check{\boldsymbol{\zeta}}}$ es involucrada en la función de mapeo de reducción $\mathcal{G}$, obteniéndose así respecto a μ, un subconjunto efectivo de variables $\boldsymbol{\zeta} = \{\xi_i : i = 1, \ldots, q\}$, las cuales corresponden al espacio inicial de características $\boldsymbol{\xi} = \{\xi_i : i = 1, \ldots, p\}$, orientado a maximizar la capacidad de representación y minimizar el coste operativo.

En general, el problema de selección de características consiste en encontrar un subconjunto $\check{\boldsymbol{\zeta}} \subset \boldsymbol{\xi}$, tal que, si el cardinal de $\check{\boldsymbol{\zeta}}$ es q y en $\check{\boldsymbol{\xi}} = \left\{\check{\boldsymbol{\zeta}}_j \subset \boldsymbol{\xi} : j = 1, \ldots, M\right\}$ están todos los subconjuntos de cardinal q, se procura hallar el subconjunto $\boldsymbol{\zeta} \in \check{\boldsymbol{\xi}}$, para el cual se optimice la función de evaluación $f_{\check{\boldsymbol{\zeta}}_j}$, dado un conjunto de etiquetas de clase $\boldsymbol{c}$ [19],

$$f\left(\boldsymbol{c}, \boldsymbol{\zeta}\right) = \max_{\check{\boldsymbol{\xi}} \subset \boldsymbol{\xi}} f\left(\boldsymbol{c}, \check{\boldsymbol{\zeta}}_j\right)$$

Diagrama general

A partir del espacio inicial de entrenamiento, dado en forma de una estructura multivariada $\mathbf{X} = \{\mathbf{x}_\ell \in \mathbb{R}^p : \ell = 1, \ldots, n\}$, tal que los valores que componen cada observación $\mathbf{x}_\ell$ corresponden a la estimación del *conjunto inicial de características* $\boldsymbol{\xi}$ de dimensión p y tiene asignada una etiqueta de clase del conjunto de etiquetas $\boldsymbol{c} = \{c_k : k \in K\}$, la selección de características contempla las siguientes tres etapas básicas de proceso: generación de subconjuntos, evaluación de subconjuntos y validación (ver Figura 3.5).

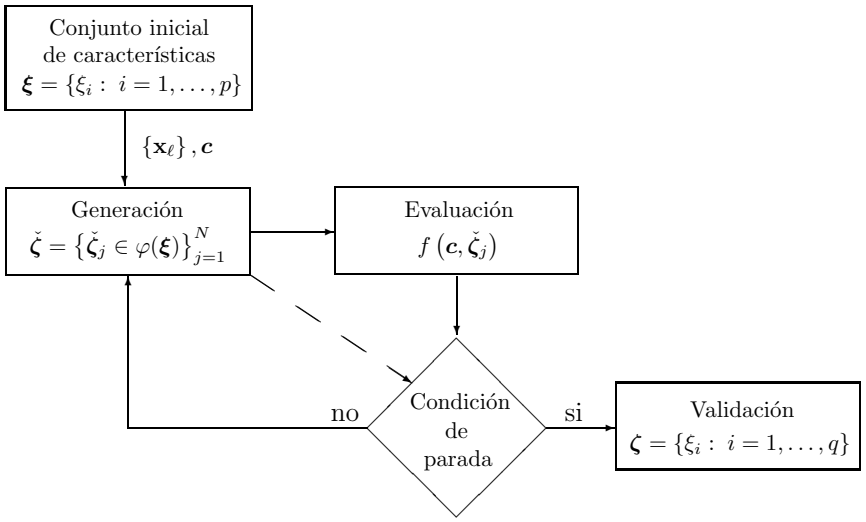


Figura 3.5. Diagrama general de la selección de características [6]

Generación de subconjuntos En esta etapa se origina cada nuevo subgrupo de variables $\check{\boldsymbol{\zeta}}_j \in \varphi(\boldsymbol{\xi})$ con $j = 1, \ldots, N$ que va a ser analizado, y las variables son tomadas directamente del espacio inicial de características $\boldsymbol{\xi} = \{\xi_i : i = 1, \ldots, p\}$. En algunos casos particulares se generan subespacios conseguidos a través de algún principio de transformación sobre el espacio inicial de entrenamiento: $\check{\boldsymbol{\zeta}}_j = \mathcal{G}_j\{\boldsymbol{\xi}\}$. Generalmente, la generación de subconjuntos se basa en alguno de los siguientes procedimientos [6]:

Completa. El generador de subconjuntos realiza la búsqueda sobre todas las posibles combinaciones que se pueden formar con el conjunto completo de características $\boldsymbol{\xi}$, para encontrar el subconjunto óptimo de acuerdo a la función dada de evaluación $f_{\check{\boldsymbol{\zeta}}}$. La búsqueda completa es *exhaustiva*, debido a que para un conjunto de características con dimensión p, el número total de subconjuntos candidatos inducirían una complejidad equivalente[1] a $\Theta(2^p)$, lo cual puede ser muy costoso y, en términos prácticos, prohibitivo para valores de $p > 50$ [25]. Por lo anterior, los procedimientos de generación completa no presentan mayor interés para su implementación práctica y, por tanto, no se analizan con mayor detalle.

[1] $\Theta(\cdot)$ denota una estimación de la complejidad computacional de manera exacta excepto por la multiplicación de alguna constante.

Heurística. Del conjunto inicial de características $\boldsymbol{\xi}$ con p variables, se seleccionan algunos de los posibles grupos sub-óptimos $\check{\boldsymbol{\zeta}}_j$ con dimensión q, tal que $q < p$, y supere un umbral δ de aceptación para una función dada de evaluación, $f_{\check{\boldsymbol{\zeta}}_j} > \delta$ [45]. Estos métodos están basados en técnicas de búsqueda que operan en concordancia con reglas empíricas y están destinados a reducir la complejidad computational, sin disminuir el rendimiento del sistema. Además, requieren de una condición de parada para prevenir que la búsqueda de subconjuntos se vuelva exhaustiva. La generación de subgrupos mediante criterios heurísticos puede tener variaciones en cuanto al espacio de búsqueda, por cuanto ésta puede ser realizada directamente del conjunto inicial de características o desde el nuevo espacio obtenido a través de algún principio de transformación $\check{\boldsymbol{\zeta}}_j = \mathcal{G}_j\{\boldsymbol{\xi}\}$. Inicialmente, se determina el punto (o puntos) de partida en el espacio inicial de características, los cuales darán la dirección de la búsqueda y se usan procedimientos basados en técnicas estadísticas o evolutivas para generar los estados siguientes [2].

Aleatoria. Este procedimiento halla aleatoriamente el espacio inicial de búsqueda y luego usando algoritmos basados en elecciones probabilísticas, o medidas de consistencia de los datos, es guiado a una solución óptima respecto a una función dada de evaluación $f_{\check{\boldsymbol{\zeta}}}$ [3]. Si bien el espacio de búsqueda[2] es $\mathcal{O}(2^p)$, estos métodos típicamente buscan en un número más reducido de conjuntos que 2^p; para ello establecen un número máximo de iteraciones posible. En esta categoría el hallazgo del subconjunto óptimo de características depende de los recursos disponibles, pues aunque la generación de los subconjuntos en un principio se hace de manera aleatoria, después se basa en procedimientos matemáticos para la continua generación de los subconjuntos siguientes, es por esto que este método requiere el valor de algunos parámetros o de una función objetivo para la continua construcción de subconjuntos.

Estos tres procedimientos son comúnmente utilizados para la generación de subconjuntos de características y se diferencian básicamente en la conformación del subconjunto inicial al momento de comenzar la búsqueda. Es decir, el procedimiento denominado *completo* hace un barrido por todas las posibles maneras de conformación de subconjuntos usando algún orden pre-establecido comenzando desde el primero hasta el último; la estrategia denominada *heurística* usa métodos como pueden ser los estadísticos o evolutivos para inicializar la búsqueda con un subconjunto sub-óptimo y luego continuar la búsqueda mediante reglas heurísticas; finalmente la estrategia *aleatoria* elige el primer subconjunto de búsqueda aleatoriamente, es por esto que, así como puede encontrar el subconjunto óptimo rápidamente, también puede llegar a ser exhaustivo, o no efectivo, en casos cuando el número de características es muy grande [7].

Evaluación de subconjuntos En esta etapa se involucra la medida obtenida mediante una función de evaluación $f_{\check{\boldsymbol{\zeta}}}$, la cual determina el *contexto* de reducción de dimensiones mediante una medida de relevancia μ. De acuerdo a la función de evaluación, existen dos tipos de selección de características frecuentemente reportados en la literatura:

1. *Wrapper.* Cuando $f_{\check{\boldsymbol{\zeta}}}$ usa información del procedimiento de clasificación buscando minimizar el error del clasificador.

[2] $\mathcal{O}(\cdot)$ denota una estimación de la complejidad computacional para el cual solo se conoce el límite superior.

2. *Filtro.* Consiste en un preprocesamiento de los datos mediante la optimización de $f_{\check{\boldsymbol{\zeta}}}$ respecto a alguna medida de representación donde se descartan las variables dependientes, redundantes e irrelevantes. Este proceso es independiente del clasificador y tiene la ventaja que su implementación es simple, aunque tiene el problema de que requiere la suposición de ciertos parámetros, como es el tipo de distribución.

De otra parte, la selección de características orientada a la clasificación puede llevarse a cabo mediante alguno(s) de los siguientes criterios de efectividad tenidos en cuenta en la evaluación de los subconjuntos $\check{\boldsymbol{\zeta}}_j$:

a) Similitud en la estructura de aleatoriedad de los conjuntos analizados. La selección de características entrega un subconjunto con dimensión reducida, tal que la distribución estadística sea lo más cercana posible a la distribución original del espacio inicial. Bajo esta restricción puede tomarse la independencia estadística de las características o la carga informativa de las mismas [24].

b) Error de reconstrucción de las observaciones. Dado el subespacio con menor dimensión de características obtenido a través de algún principio de transformación $\mathcal{G}$, se analiza la capacidad de reconstrucción del espacio inicial de características $\boldsymbol{\xi}$, mediante un criterio de error de reconstrucción $\boldsymbol{\varepsilon}$ establecido previamente, de acuerdo a lo enunciado en la restricción *(b)* de la Definición 3.2 [29].

c) Rendimiento de proceso. Cuando el subconjunto de características mejora la tasa de rendimiento del clasificador o disminuye el tamaño de la estructura sin reducir significativamente la precisión del clasificador construido, usando solamente las características seleccionadas [24].

d) Complejidad computacional. Cuando se obtiene un subconjunto sub-óptimo en un determinado número de iteraciones, o con un número de características previamente fijado [25].

También es importante no dejar de lado la *condición de parada*, ya que existen diferentes formas para establecerla. Por ejemplo, se puede dejar de adicionar o remover características cuando ninguno de los subconjuntos mejore la medida obtenida con la función de evaluación $f_{\check{\boldsymbol{\zeta}}}$; aunque se puede continuar evaluando los subconjuntos de características hasta que la la medida de evaluación no se degrade; o se puede seguir generando subconjuntos mientras se alcanza el final del espacio de búsqueda y entonces se selecciona el mejor. Una condición simple de parada para una selección tipo *wrapper* es detenerse cuando cada combinación de valores para las características seleccionadas converjan a valores simples de clase, pero esto asume datos de entrenamiento libres de ruido. Una alternativa más robusta ordena las características de acuerdo con algún nivel de significación de relevancia, luego se fija un umbral para determinar el punto final de parada [7].

Validación Los procedimientos de validación usados por los métodos de selección de características, puede emplear dos tipos de datos de entrenamiento: *a*) datos artificiales y *b*) datos obtenidos de aplicaciones reales. Los conjuntos de datos artificiales son construidos teniendo en cuenta las clases como concepto objetivo, así que todas las características relevantes para este concepto son conocidas al igual que las características irrelevantes/redundantes o ruidosas. El método de selección de características se realiza sobre

este conjunto de datos y el resultado se compara con el esperado. Para el segundo procedimiento, se escogen conjuntos de datos adquiridos en aplicaciones reales, los cuales requieren ser etiquetados por algún experto. Como en este caso el subconjunto óptimo de características es desconocido, se prueba la precisión de los subconjuntos seleccionados con la ayuda de algún clasificador calificado para la tarea. La precisión y eficacia obtenida debe ser comparada con algún método suficientemente conocido. Es por esto, que si los clasificadores están orientados a estructuras particulares de datos se generan problemas en su generalización [6]. A continuación se describen técnicas de clasificación frecuentemente empleadas en la etapa de validación para discriminar señales a partir de subconjuntos de características.

Análisis discriminante lineal. El objetivo principal del análisis discriminante lineal (LDA) es separar grupos usando funciones lineales (funciones discriminantes) de las variables para describir o revelar las diferencias entre dos o más clases. Así, el análisis discriminante consiste en identificar la contribución relativa de las variables de entrada a la separabilidad de las clases y hallar el plano óptimo en el cual los puntos pueden ser proyectados para concebir mejor la configuración de los grupos [35]. Matemáticamente, se definen dos medidas para las muestras en todas las clases: (i) *Matriz de dispersión intra-clase*

$$S_w = \sum_{j=1}^{r} \sum_{i=1}^{N_j} \left(x_i^j - \mu_j \right) \left(x_i^j - \mu_j \right)^T \tag{3.13}$$

donde x_i^j es la i-ésima muestra de la clase j, μ_j es la media de la clase j, r es el número de clases, y N_j es el número de muestras en la clase j, y (ii) *Matriz de dispersión entre-clases*

$$S_b = \sum_{j=1}^{r} \left(\mu_j - \mu \right) \left(\mu_j - \mu \right)^T \tag{3.14}$$

en donde μ representa la media de todas las clases. El objetivo es maximizar el valor entre clases, mientras se minimiza el valor intra-clase. Una forma de hacerlo es maximizando la relación $\det |S_b| \,/\, \det |S_w|$. La ventaja de utilizar esta relación es que se ha comprobado que, si S_w es una matriz no singular, dicha relación se maximiza cuando los vectores columna de la matriz de proyección $\mathbf{W}$ son los autovectores de $S_w^{-1} S_b$.

Clasificador bayesiano. El criterio de trabajo de este clasificador consiste en minimizar la probabilidad de error en un problema de clasificación [10]. El algoritmo de decisión bayesiana evalúa el punto a clasificar en cada una de las funciones discriminantes construidas para cada clase. Sea $\mathbf{X}_k$ la matriz que contiene los hiperpuntos de cada clase, de tamaño n_c (muestras por clase) $\times$ p (número de características) $\times$ r (número de clases), para la cual se procede del siguiente modo:

- Se calcula el vector de medias $\boldsymbol{\mu}_k$ de $\mathbf{X}_k$.
- Se calcula la matriz de covarianza $\boldsymbol{\Sigma}_k$ de $\mathbf{X}_k$.

– Se calculan los coeficientes de la función discriminante para cada clase:

$$\begin{aligned} \mathbf{W}_k &= -\frac{1}{2}\boldsymbol{\Sigma}_k^{-1} \\ \mathbf{w}_k &= \boldsymbol{\Sigma}_k^{-1}\boldsymbol{\mu}_k \\ w_{k0} &= -\frac{1}{2}\boldsymbol{\mu}_k\boldsymbol{\Sigma}_k^{-1}\boldsymbol{\mu}_k^{\top} - \frac{1}{2}\ln \mid \boldsymbol{\Sigma}_k \mid + \ln P(\omega_k) \end{aligned} \tag{3.15}$$

donde $P(\omega_k)$ es la probabilidad de ocurrencia de la clase k.

– Se construye la función discriminante para cada clase con los coeficientes calculados en (3.15):

$$g_k(\mathbf{x}) = \mathbf{x}^{\top}\mathbf{W}_k\mathbf{x} + \mathbf{w}_k^{\top}\mathbf{x} + w_{k0}$$

El punto pertenece a aquella clase que da un mayor valor al evaluarlo en su función discriminante.

Máquinas de soporte vectorial. Las máquinas de soporte vectorial (SVM) están sustentadas en el principio de minimización de riesgo estructural (SRM), propuesto en [50]. Un subconjunto de funciones encontradas en el proceso de optimización minimizan el riesgo actual del problema, de manera que entrenando una serie de máquinas para el objetivo dado, se minimizan el riesgo y la confiabilidad de la dimensión Vapnik-Chervonenkis, la cual implica los requerimientos de almacenamiento de la técnica de aprendizaje y la calidad de sus respuestas para responder a un problema de clasificación. De acuerdo a esto, este método busca un hiperplano de máxima separación entre las clases, para lo que se requiere hacer una transformación no lineal de los datos de entrada (datos de entrenamiento) hacia un espacio de características de dimensión mayor donde las clases tengan máxima separación. En el hiper-espacio resultante se procede a realizar la búsqueda del hiperplano que divide las clases. Sea el conjunto de datos de entrenamiento:

$$(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_n, y_n) \in \mathbf{X} \times \{\pm 1\}$$

donde cada $\mathbf{x}_i \in \mathbb{R}^p$ es un elemento del conjunto $\mathbf{X}$ y y_i es la etiqueta de $\mathbf{x}_i$. A este conjunto $\mathbf{X}$ se aplica una transformación no lineal de los datos, a través de una función kernel:

$$\begin{aligned} &\kappa : \mathbf{X} \times \mathbf{X} \to \mathbb{R} \\ &\quad (\mathbf{x}, \mathbf{x}') \mapsto \kappa(\mathbf{x}, \mathbf{x}') := (\mathbf{x}\mathbf{x}') \end{aligned}$$

En el espacio resultante se conservan las mismas etiquetas y_i para cada $\mathbf{x}_i$ y se hallan los elementos promedio de cada clase,

$$m_{c_1} = \frac{1}{r_{c_1}} \sum_{\{i:y_i=1\}} \mathbf{x}_i \tag{3.16}$$

$$m_{c_2} = \frac{1}{r_{c_2}} \sum_{\{i:y_i=-1\}} \mathbf{x}_i \tag{3.17}$$

donde r_{c_1} y r_{c_2} es el número de elementos que corresponden a la etiqueta $y_i = 1$ y $y_i = -1$ respectivamente, de manera análoga para los promedios de clase m_{c_1} y m_{c_2}. De acuerdo a esto, se obtiene el vector $\mathbf{w}$ cuyos extremos se ubican en los dos valores medios m_{c_1} y m_{c_2}, donde la magnitud de $\mathbf{w}$ informa acerca de la distancia que existe entre las medias de las clases. Al tomar una muestra de prueba $\mathbf{d}$ se procede a hallar el vector $\mathbf{q}$ que relaciona la distancia entre $\mathbf{d}$ y el punto medio ubicado entre las medias de las clases, $\mathbf{q} = \mathbf{d} - \frac{m_{c_1}+m_{c_2}}{2}$. Se analiza el ángulo θ formado por $\mathbf{q}$ y $\mathbf{w}$, de tal manera que si $\frac{\pi}{2} > \theta$ y $\theta < -\frac{\pi}{2}$ la muestra $\mathbf{d}$ pertenece a una clase, y si $\frac{\pi}{2} < \theta$ y $\theta > -\frac{\pi}{2}$ pertenece a la otra clase. Por lo tanto, la predicción de la etiqueta de la muestra de prueba es:

$$y = \text{sign}\left\{\left(\mathbf{d} - \frac{m_{c_1} + m_{c_2}}{2}\right) \cdot (m_{c_1} - m_{c_2})\right\} \tag{3.18}$$

$$= \text{sign}\left\{(\mathbf{d} \cdot m_{c_1}) - (\mathbf{d} \cdot m_{c_2}) + b\right\} \tag{3.19}$$

donde,

$$b = \frac{1}{2}(||m_{c_1}||^2 - ||m_{c_2}||^2)$$

Por último se re-escribe la ecuación (3.19) basados en (3.16) y (3.17)

$$y = \text{sign}\left\{\frac{1}{r_{c_1}} \sum_{\{i:y_i=1\}} (\mathbf{d} \cdot \mathbf{x}_i) - \frac{1}{r_{c_2}} \sum_{\{i:y_i=-1\}} (\mathbf{d} \cdot \mathbf{x}_i) + b\right\}$$

de forma general se obtiene que:

$$y = \text{sign}\left\{\frac{1}{r_{c_1}} \sum_{\{i:y_i=1\}} \kappa(\mathbf{x}, \mathbf{x}_i) - \frac{1}{r_{c_2}} \sum_{\{i:y_i=-1\}} \kappa(\mathbf{x}, \mathbf{x}_i) + b\right\}$$

y b puede considerarse como:

$$b := \frac{1}{2}\left\{\frac{1}{r_{c_1}^2} \sum_{\alpha} \kappa(\mathbf{x}_i, \mathbf{x}_j) - \frac{1}{r_{c_2}^2} \sum_{\beta} \kappa(\mathbf{x}_i, \mathbf{x}_j)\right\}$$

donde α es el conjunto de índices $\{i, j : y_i = y_j = 1\}$ y β el conjunto de índices $\{i, j : y_i = y_j = -1\}$. Se puede observar entonces que las máquinas de soporte vectorial son útiles en la clasificación de patrones, solo basta con hacer una elección adecuada del kernel de transformación.

k-vecinos más cercanos. Unos de los procedimientos de clasificación más simples que ofrece buenos resultados para poblaciones no normales está basado en la regla de los *vecinos más cercanos*. En el Algoritmo 3.1 se presenta el procedimiento básico para clasificar un punto de prueba $\mathbf{x}_0$ usando la regla de los vecinos más cercanos [32]. En general, esta técnica de clasificación basa su procedimiento en que una observación debe estar

situada cerca de otras que pertenecen a la misma clase en el espacio de ejemplos. Por lo tanto, se buscan los k vecinos más cercanos y se clasifica el ejemplo de acuerdo a la clase más frecuente entre los vecinos. En el caso particular de $k = 1$, el método asigna la etiqueta de clase de un ejemplo a toda la población que tenga más próxima.

Algoritmo 3.1 Clasificador basado en los k vecinos más cercanos

Require: $\mathbf{X} = \{\mathbf{x}_i : i = 1, \ldots, n\}$, $\boldsymbol{c}$, $\mathbf{x}_0$, k

1. Definir una medida de distancia entre los puntos de entrenamiento $\mathbf{x}_i$.
2. Calcular las distancias del punto a clasificar, $\mathbf{x}_0$, a todos los demás puntos.
3. Seleccionar los k puntos muestrales más próximos al que se pretende clasificar.
4. Calcular la proporción en que los k puntos pertenecen a cada una de las poblaciones.
5. Clasificar el punto $\mathbf{x}_0$ en la población con mayor frecuencia entre los k puntos.

Output: $\mathbf{x}_0 \in \boldsymbol{c}$

Procedimientos frecuentes en el reconocimiento de patrones

La reducción de dimensiones, en este nivel de operación, permite descubrir cuales parámetros obtenidos con los diferentes modelos de representación, son los que determinan el patrón de comportamiento inmerso en el conjunto de bioseñales para el reconocimiento de estados funcionales. Es importante anotar que no siempre existe una conexión directa entre los parámetros de representación y el comportamiento fisiológico, por lo que la interpretación fisiológica en muchas ocasiones se dificulta. Sin embargo, a partir de los modelos de representación se puede construir una interpretación matemática que de cierta manera intente explicar los principios fisiológicos influyentes directa o indirectamente en el comportamiento biológico involucrado en los estados de funcionamiento. De otra parte, este nivel de operación permite reducir los costos de análisis, ya que define cuales variables son las que se requieren estimar para crear el espacio efectivo de representación. Es importante anotar que existen procedimientos del nivel de proyección que pueden extenderse a este nivel de interpretación, como es el caso de PCA, que después de proyectar los datos en un espacio ortogonal (optimizando una función de evaluación $f_{\boldsymbol{\zeta}}$ basada en una medida de la variabilidad de los datos), ofrece la posibilidad de descubrir las características que más aportan en la formación de las componentes principales, para destacarlas como características relevantes.

Orientado al reconocimiento de patrones, existen procedimientos frecuentes para llevar a cabo la reducción de dimensiones en este nivel de operación y a continuación se describen algunos de ellos que son usados independiente o conjuntamente.

Búsqueda heurística de subconjuntos discriminantes Se realiza directamente sobre el conjunto de características mediante la generación de subconjuntos usando criterios heurísticos. Inicialmente, se determina el punto (o puntos) de partida en el espacio de características. Puede existir un ordenamiento natural parcial en este espacio, donde cada etapa actual tiene exactamente una característica más que la etapa anterior. Esto sugiere que se puede comenzar con una característica e ir agregando sucesivamente, o también se puede

comenzar con todo el conjunto de características e ir quitando sucesivamente. La primera aproximación se llama selección hacia adelante ó progresiva (SFS – *sequential forward selection*), mientras que la otra es conocida como eliminación hacia atrás ó selección regresiva (SBS – *sequential backward selection*). Se pueden tener variaciones a estas aproximaciones, por ejemplo en [8], se expone un operador que adiciona d características y descarta una; y mediante operadores genéticos se obtienen mutaciones que producen diferentes tipos de conectividad. De acuerdo a una función de evaluación f_{ζ} se pueden desarrollar técnicas de generación de subconjuntos conocidas como *flotantes*, las cuales consisten en la combinación de las técnicas de generación progresiva y regresiva buscando minimizar f_{ζ}. Es importante decir que los procedimientos de selección tipo *wrapper* presentan mejor desempeño hallando el subconjunto óptimo de características, que los de tipo *filtro*, debido a que f_{ζ} depende de los resultados de clasificación, sin embargo, la complejidad computacional es más alta. La literatura reseña rutinas *wrapper* que realizan la búsqueda en un número considerablemente menor de subconjuntos, entre las cuales se encuentra la selección flotante hacia adelante (SFFS – *sequential forward floating selection*) [19, 53]. En esta técnica, cada paso incluye una nueva variable por medio de un procedimiento secuencial hacia adelante, pero luego realiza la exclusión de las variables menos significativas, una por una, hasta que la tasa de error de clasificación correcta disminuya. Una vez que ya no se puede seguir excluyendo variables se hace otro paso hacia delante para incluir otra variable y nuevamente se realiza la exclusión de variables, si es posible. El proceso es recurrente hasta que ya no se puedan efectuar más pasos hacia adelante debido a que la precisión de clasificación ya no se incremente. Esta rutina se expone con más detalle en el Algoritmo 3.2.

Generación de subconjuntos por árboles de decisión Un árbol de decisión consiste en una estructura jerárquica cuyos nodos representan las características, las ramas son los valores posibles que puede tomar cada característica y las hojas determinan la clase correspondiente. Los árbol de decisión son frecuentemente usados para clasificar ejemplos, operando en forma de filtro de manera descendente hasta asignar a cada ejemplo una hoja. En general, un árbol de decisión representa una disyunción de conjunciones que determinan las restricciones relacionadas a los posibles valores que pueden tomar las características de entrenamiento. Cada rama que va desde la raíz del árbol a una hoja, representa una conjunción de tales restricciones, mientras, el árbol mismo representa la disyunción de esas conjunciones. Un procedimiento que desarrolla la búsqueda descendente y egoísta en el espacio de posibles árboles de decisión, corresponde al algoritmo ID3, propuesto en [34], el cual evalúa cada una de las características con base a una función dada de evaluación f_{ζ}, midiendo el rendimiento del clasificador durante el entrenamiento. Dicha característica es ubicada como nodo del árbol, donde se fragmentan los ejemplos para cada uno de los valores que puede tomar, siendo éstos las ramas del árbol. El procedimiento se hace en forma recursiva hasta que los ejemplos de entrenamiento compartan la misma clase, obteniéndose una hoja del árbol. Esta rutina se ilustra en la Figura 3.6. La función de evaluación utilizada en el algoritmo ID3, propuesta para evaluar cada característica, es la ganancia de información o medida de la cantidad de información presente. En particular, se propone la entropía como una manera de cuantificar la bondad de cada característica. Así, sean los L posibles valores que puede tomar una característica $\xi_i = \{v_j : j = 1, \ldots, L\}$, con probabilidad de ocurrencia $p(v_i)$, entonces la entropía H de la respuesta actual se define

Algoritmo 3.2 Algoritmo SFFS

Require: $\boldsymbol{\xi}$, $\boldsymbol{c}$, $\jmath = 1$, $\hat{\boldsymbol{\xi}}_0 = \emptyset$, $q \in \mathbb{N}$.
  while $\jmath \leq q$ **do**
    1. Adicionar a $\hat{\boldsymbol{\xi}}_\jmath$ la característica ξ más significante de acuerdo a la función de evaluación $f_{\hat{\boldsymbol{\xi}}_\jmath}$:
       $\xi^+ = \underset{\{\boldsymbol{\xi} \backslash \hat{\boldsymbol{\xi}}_\jmath\}}{\text{máx}} f\left(\boldsymbol{c}, \left\{\hat{\boldsymbol{\xi}}_\jmath, \xi\right\}\right)$
       $\hat{\boldsymbol{\xi}}_{\jmath+1} = \hat{\boldsymbol{\xi}}_\jmath + \xi^+$; $\jmath = \jmath + 1$.

    2. Hallar la característica ξ que minimice el valor de $f_{\hat{\boldsymbol{\xi}}_\jmath}$

    3. **if** ξ es la misma ξ^+ del paso 1 **then**
         No se modifica el conjunto $\hat{\boldsymbol{\xi}}_\jmath$.
         Regresar al paso 1.

    4. **else**
         $\xi^- = \underset{\{\boldsymbol{\xi} \backslash \hat{\boldsymbol{\xi}}_\jmath\}}{\text{máx}} f\left(\boldsymbol{c}, \left\{\hat{\boldsymbol{\xi}}_\jmath \backslash \xi\right\}\right)$
         if $f_{\left\{\hat{\boldsymbol{\xi}}_\jmath \backslash \xi\right\}} > f_{\hat{\boldsymbol{\xi}}_\jmath}$ **then**
           $\hat{\boldsymbol{\xi}}_{\jmath-1} = \hat{\boldsymbol{\xi}}_\jmath - \xi^-$; $\jmath = \jmath - 1$.
           Regresar al paso 2
         else
           Regresar al paso 1
         end if

    5. **end if**
  end while
  $\boldsymbol{\zeta} \leftarrow \hat{\boldsymbol{\xi}}_\jmath$
Output: $\boldsymbol{\zeta}$ = {subconjunto efectivo de características} $\star \backslash$Efectivo respecto a $f_{\hat{\boldsymbol{\xi}}_\jmath}\backslash \star$

como $H = -\sum_{i=1}^{L} p(v_i) \log_2 p(v_i)$. Por lo tanto, la ganancia de información que corresponde a la reducción de la entropía causada por fragmentar el conjunto de entrenamiento $\mathbf{X}$ respecto a una característica ξ_i, se expresa como:

$$I_G(\mathbf{X}, \xi_i) = H(\mathbf{X}) - \sum_{v_j \in \xi_i} \frac{N(\mathbf{X}_{v_j})}{N(\mathbf{X})} H\left(\mathbf{X}_{v_j}\right)$$

donde $N(\mathbf{X})$ es el conjunto de ejemplos de entrenamiento para los cuales la característica ξ_i toma el valor v_j, mientras que $N\left(\mathbf{X}_{v_j}\right)$ es el tamaño del conjunto $\mathbf{X}_{v_j}$.

Algoritmos genéticos La selección de características, frecuentemente, es tratada como un problema de optimización, donde los algoritmos genéticos desempeñan un papel importante en la tarea de reducir dimensiones restringido a una función criterio o de evaluación [30]. El *algoritmo genético simple* (*simple genetic algorithm - SGA*) es un caso especial de búsqueda heurística aleatoria [52], para el cual el espacio de búsqueda Ω corresponde a:

$$\Omega = \underbrace{\mathcal{Z}_2 \times \cdots \times \mathcal{Z}_2}_{\iota \text{ veces}} \tag{3.20}$$

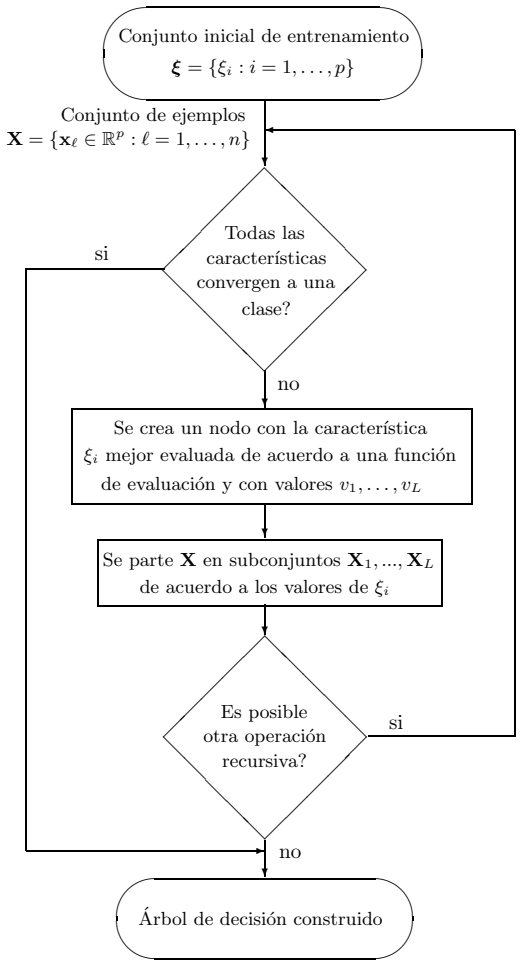


Figura 3.6. Diagrama de flujo del algoritmo ID3

donde $\mathcal{Z}_2$ es el conjunto de los enteros módulo 2 [21]. La expresión (3.20) corresponde exactamente a la representación binaria de los enteros en el intervalo $[0, 2^\iota)$; por tanto, Ω está compuesto por $n = 2^\iota$ cadenas binarias de longitud ι. La longitud ι se conoce como *longitud del cromosoma*. La exploración en Ω se realiza mediante la aplicación de mecanismos denominados *operadores*, que actúan sobre los elementos de un subconjunto P_i dado perteneciente al espacio de búsqueda. Cada subconjunto P_i recibe el nombre de i-ésima *población* o i-ésima *generación*, e.g.

$$P_i = \underbrace{\{\overbrace{10\cdots01}^{\iota \text{ bits}}, 11\cdots00, \ldots, 11\cdots10\}}_{r \text{ individuos}} \tag{3.21}$$

El parámetro r en (3.21) representa el *tamaño de la población*, i.e. la cardinalidad de cada generación P_i. A medida que r se incrementa se posee una mayor diversidad genética para la exploración sobre Ω.

Operadores genéticos. Son los mecanismos evolutivos que generan la secuencia de poblaciones

$$P_0 \xrightarrow{\tau} P_1 \xrightarrow{\tau} P_2 \xrightarrow{\tau} \cdots$$

donde τ es una *regla de transición* que se implementa mediante la combinación de tres operadores básicos: selección, mutación y recombinación ó cruzamiento.

- *Selección*. Es el proceso por el cual se eligen miembros de una población de acuerdo con una *función de adaptabilidad* $f : \Omega \to \mathbb{R}$ que se desea maximizar. Los individuos elegidos serán los *padres* de la próxima generación. El uso de la función f determina el esquema de selección. Los principales esquemas son: selección proporcional, selección jerárquica y selección por torneo [15, 52]. En las aplicaciones de clasificación y reconocimiento de patrones, la función de adaptabilidad se asigna usualmente a la precisión de la clasificación [14] o a una medida que incluye la precisión de clasificación y una o varias medidas adicionales del desempeño del clasificador [33].
- *Mutación*. Consiste en alterar aleatoriamente cada gen (bit) con una probabilidad típicamente pequeña [1]. Este mecanismo asegura que la búsqueda heurística pueda realizarse en puntos de Ω que no pueden ser explorados mediante la reproducción de la población inicial.
- *Recombinación*. Los cromosomas de los individuos padres intercambian bits en aquellas posiciones donde una máscara de recombinación es 1, produciendo dos nuevos cromosomas denominados *hijos*. La recombinación no se aplica a todas las parejas seleccionadas, su ejecución está sujeta a una distribución de probabilidad. Los tipos clásicos de recombinación son: recombinación en un punto y recombinación uniforme [1].

Parámetros de control. Los parámetros básicos del SGA que deben ser ajustados son: el tamaño de la población (r); la probabilidad o tasa de mutación, denotada por $\mathcal{P}_m \in [0, 1]$; y la probabilidad o tasa de recombinación, denotada por $\chi \in [0, 1]$. Estos parámetros no son independientes entre si [11] y tienen gran influencia sobre el desempeño del algoritmo genético. Para la sintonización de estos parámetros se han propuesto diferentes métodos basados en el concepto de *meta-algoritmo* [16, 56].

Los pasos básicos para implementar el SGA están descritos en el Algoritmo 3.3. En este trabajo se ha utilizado una traducción y extensión en lenguaje C del SGA original en Pascal presentado en [15]. La distribución se denomina `SGA-C 2.0` y se encuentra disponible en [46]. La ampliación de este tema se expone en el Apéndice A.

Análisis de dependencia lineal Sea el vector de características $\boldsymbol{\xi}$ con cardinal p, el cual se asocia a un conjunto de observaciones $\mathbf{X} = \{\mathbf{x}_i \in \mathbb{R}^p : i = 1, \ldots, n\}$. El análisis de dependencia tiene como objetivo descubrir la relación que puede existir entre las p variables aleatorias basado en medidas de similaridad entre ellas. Para este respecto se analizan las consecuencias de que la matriz $\boldsymbol{\Sigma}_{\mathbf{X}}$ sea singular. Si existe algún vector $\mathbf{w}$, tal que, $\mathbf{w}^\top \boldsymbol{\Sigma}_{\mathbf{X}} \mathbf{w} = 0$, entonces la variable escalar $v_i = \mathbf{w}^\top (\mathbf{x}_i - \overline{\mathbf{x}})$ tiene media y varianza

Algoritmo 3.3 Algoritmo genético simple

Require: Inicializar N_{gen}, χ, $\mathcal{P}_m$, r cromosomas de ι bits ⋆\N_{gen}: máximo número de generaciones\⋆
  repeat
    Determinar la función de adpatabilidad de cada cromosoma, f_i, $i = 1, \ldots, r$
    Jerarquizar los cromosomas
    repeat
      Seleccionar dos cromosomas con el mayor puntaje
      if $\text{rand}[0, 1) < \chi$ **then**
        cruzar la pareja en un bit aleatoriamente escogido
      else
        Cambiar cada bit con probabilidad $\mathcal{P}_m$
        Remover los cromosomas padres
      end if
    until que hayan sido creados r descendientes
  until se haya alcanzado N_{gen} o la población converja
Output: el cromosoma con mayor f en todas las generaciones ⋆\Elitismo\⋆

nula, así, esta variable siempre tomará el valor de cero. Por tanto, para cualquier i:

$$\sum_{l=1}^{p} w_l \left(x_{il} - \overline{x}_l\right) = 0 \ , \quad \forall i.$$

Esta expresión implica que las p variables no son independientes, ya que se puede despejar alguna en función de las demás:

$$x_{i1} = \overline{x}_1 - \frac{w_2}{w_1}\left(x_{i2} - \overline{x}_2\right) - \ldots - \frac{w_p}{w_1}\left(x_{ip} - \overline{x}_p\right).$$

Por tanto, si existe algún vector $\mathbf{w}$ que haga $\mathbf{w}\boldsymbol{\Sigma}_{\xi}^{\top}\mathbf{w} = 0$, existe una relación lineal entre las variables. Lo cual puede extenderse para cualquier número de valores propios nulos si $\boldsymbol{\Sigma}_{\xi}$ tiene rango $q < p$, por lo cual existen $r = p - q$ variables redundantes que pueden eliminarse, provocando que las observaciones puedan representarse con solo q variables [42, 32]. El procedimiento puede ejecutarse haciendo uso de medidas de dependencia lineal entre las variables, entre las cuales una de las más conocidas está el coeficiente de correlación lineal o simple [9] y usa el criterio de relevancia fuerte asociado a esta medida estadística.

Análisis de varianza En la selección de características, es necesario cuantificar el nivel de interacción entre las variables desde el escenario de múltiple influencia. Con este fin, se pueden emplear métodos para determinar el nivel de interacción entre las diferentes características o factores usados para el entrenamiento de sistemas de clasificación. El modelo para este caso, se puede representar de la siguiente manera:

$$\begin{aligned}\left(\text{Observación}\right) = &\sum \left[\text{Características que describen el fenómeno}\right] + \\ &+ \sum \left[\text{Variables aleatorias que describen efectos no determinados}\left(\text{residuales}\right)\right]\end{aligned}$$

En principio, entre mayor cantidad de características se emplee, menor debe ser la variabilidad residual del modelo [20]. En este caso, se debe determinar el aporte de los factores que impliquen mayor información del modelo y no mayor aporte a la variabilidad residual.

La determinación de las características que más aportan a la separabilidad entre las clases, a partir del conjunto inicial de características $\boldsymbol{\xi} = \{\xi_i : i = 1, \ldots, p\}$, puede obtenerse mediante el *análisis univariado de la varianza* (ANOVA). Sea y el valor estimado de una variable aleatoria (alguna de las características ξ_i) con distribución normal para r clases de n observaciones y varianzas iguales según se presenta en la Tabla 3.2. Se asume que las

	Clase 1 $\mathcal{N}(\hat{\mu}_1, \sigma^2)$	*Clase* 2 $\mathcal{N}(\hat{\mu}_2, \sigma^2)$	$\cdots$ $\cdots$	*Clase* r $\mathcal{N}(\hat{\mu}_r, \sigma^2)$
	y_{11}	y_{21}	$\cdots$	y_{r1}
	y_{12}	y_{22}	$\cdots$	y_{r2}
	$\vdots$	$\vdots$		$\vdots$
	y_{1n}	y_{2n}	$\cdots$	y_{rn}
Total	$y_{1\Sigma}$	$y_{2\Sigma}$	$\cdots$	$y_{r\Sigma}$
Media	$\bar{y}_{1\Sigma}$	$\bar{y}_{2\Sigma}$	$\cdots$	$\bar{y}_{r\Sigma}$
Varianza	${s_1}^2$	${s_2}^2$	$\cdots$	${s_r}^2$

Tabla 3.2. Conjunto de observaciones con distribución normal

r clases son independientes. Se emplea el subíndice Σ en la notación para los totales y las medias de cada grupo:

$$y_{i\Sigma} = \sum_{j=1}^{n} y_{ij}, \quad \bar{y}_{i\Sigma} = \frac{1}{n} \sum_{j=1}^{n} y_{ij}$$

Así, el modelo para cada observación puede escribirse como

$$y_{ij} = \hat{\mu}_i + \epsilon_{ij} \qquad i = 1, \ldots, r;\ j = 1, \ldots, n$$

donde $\hat{\mu}_i$ es la media de la i-ésima población y ϵ es la desviación con respecto a la media poblacional del valor estimado de la variable para cada observación. El procedimiento requiere la comparación entre las medias de la variable para todas las clases $\bar{y}_{i\Sigma}$, con el objetivo de comprobar si son suficientemente diferentes, dando con esto a entender que las medias de la población difieren y por lo tanto son separables o discriminantes. La función de evaluación f_ξ se relaciona con una prueba de hipótesis donde la hipótesis nula se expresa como $H_0 : \hat{\mu}_1 = \hat{\mu}_2 = \cdots = \hat{\mu}_r$ [35]. Debido a que la evaluación se basa en la separabilidad entre clases, la relevancia asocia una medida de distancia en el espacio de representación. También se considera el caso conocido como *análisis de varianza multivariada* (MANOVA), donde frecuentemente se tienen mediciones que involucran varias características dependientes de cada unidad experimental. Por tal motivo, se asume que las r muestras aleatorias independientes de tamaño n son obtenidas de poblaciones con distribución normal p-variadas con matrices de covarianza iguales. Así, la hipótesis nula no es tomada sobre valores puntuales, sino sobre los vectores de medias,

$$H_0 : \hat{\boldsymbol{\mu}}_1 = \hat{\boldsymbol{\mu}}_2 = \cdots = \hat{\boldsymbol{\mu}}_r \tag{3.22}$$

De esta manera, la función de evaluación $f_{\check{\boldsymbol{\zeta}}}$ que actúa sobre los subconjuntos de características que se quieren analizar, hace uso de una medida de distancia multivariada para determinar los subconjuntos de características que ofrecen separabilidad entre las clases.

Agrupación no supervisada de datos A este nivel, también pertenecen los procedimientos que realizan *clustering* de datos, ya que permiten la representación de un conjunto determinado de datos, en un conjunto reducido de agrupaciones basados en alguna medida de relevancia. Entre las más conocidas, se considera la técnica iterativa de análisis de datos auto-organizados (Isodata – *Iterative Self-Organizing Data Analysis Technique*), donde la función de evaluación $f_{\check{\boldsymbol{\zeta}}}$, incluye una medida de distancia espacial para agrupar elementos suficientemente cercanos. Esta técnica de agrupación no supervisada usa un algoritmo iterativo basado en la regla de los vecinos más cercanos e incluye una consideración heurística que le permite funcionar adecuadamente cuando el conocimiento a priori sobre el número de *clusters* no es bueno. Adicionalmente, permite que el número de agrupaciones sea ajustado automáticamente durante las iteraciones mediante la fusión de agrupaciones similares y con desviaciones estándar grandes. Un procedimiento comúnmente utilizado en tareas de *clustering* de datos es expuesto detalladamente en el Algoritmo 3.4.

Algoritmo 3.4 Algoritmo Isodata

Require: $\mathbf{V} = \{\mathbf{v}_i : i = 1, \ldots, p\}, K, I, P, \delta_N, \delta_S, \delta_C$. $\star\backslash\mathbf{V}$ es un conjunto de p puntos multidimensionales, K es el número deseado de agrupaciones, I es el número máximo de iteraciones permitida, P es el máximo número de mezclas por iteración, δ_N es el umbral para el mínimo número de elementos que cada agrupación puede tener y δ_C es el umbral de distancia máxima para la unión de agrupaciones$\backslash\star$

1. Escoger de manera arbitraria tanto el valor de k como los centroides iniciales $\aleph_1, \ldots, \aleph_k$ $\star\backslash$No necesariamente k es igual a $K\backslash\star$

2. Asignar a cada $\mathbf{v}_i$ el centroide de la agrupación $\boldsymbol{\upsilon}_j$ más próxima: $\mathbf{v}_i \sim \boldsymbol{\upsilon}_j$ para $j = 1, \ldots, k$.

3. Descartar las agrupaciones cuyo número de miembros es menor que δ_N, asignando estos miembros a los centroides más cercanos y por cada descarte: $k \leftarrow k - 1$.

4. Actualizar los centroides: $\aleph_j = \frac{1}{N_j} \sum_{\mathbf{v}_i \sim \boldsymbol{\upsilon}_j} \mathbf{v}_i$ para valores de $j = 1, \ldots, k$. $\star\backslash N_j$ es el número de miembros de la agrupación $\boldsymbol{\upsilon}_j\backslash\star$

5. Estimar la distancia promedio D_j entre los miembros de cada agrupación $\boldsymbol{\upsilon}_j$ y su centroide $\aleph_j$ correspondiente: $D_j = \frac{1}{N_j} \sum_{\mathbf{v}_i \sim \boldsymbol{\upsilon}_j} \operatorname{dist}\{\mathbf{v}_i, \aleph_j\}$ para $j = 1, \ldots, k$.

6. Estimar la distancia promedio global D entre los miembros y el centroide al que pertenecen: $D = \frac{1}{p} \sum_{j=1}^{k} N_j D_j$

7. **if** $k \leq K/2$ **then**

   Encontrar el vector de desviación estándar para cada agrupación: $\boldsymbol{\Sigma}_j = \left[\sigma_1^{(j)}, \ldots, \sigma_{N_j}^{(j)}\right]^\top$

   for $j = 1 : k$ **do**

   Encontrar el máximo componente de $\boldsymbol{\Sigma}_j$ y notarlo como $\sigma_{max}^{(j)}$

   end for

   if Para cualquier $\sigma_{max}^{(j)}$, es cierto que $\sigma_{max}^{(j)} > \delta_S$, $D_j > D$ y $N_j > 2\delta_N$ **then**

   Dividir $\aleph_j$ en dos nuevos centroides $\aleph_j^+$ y $\aleph_j^-$ mediante la adición de $\pm\eta$ al componente de $\aleph_j$ correspondiente a $\sigma_{max}^{(j)}$, donde η puede ser $\alpha\sigma_{max}^{(j)}$, para algún $\alpha > 0$. Borrar $\aleph_j$ y asignar $k \leftarrow k + 1$. Regresar al paso 2.

   else

   Avanzar al paso 10.

   end if

8. **else**

   Calcular las distancias entre los centros de todas las agrupaciones por parejas: $D_{ij} = \operatorname{dist}\{\aleph_i, \aleph_j\}$ para todo $i \neq j$ y organizar las $k(k-1)/2$ distancias en orden ascendente.

   Hallar no más de P distancias D_{ij} que sean las más pequeñas y que también sean menores que δ_C, guardándolas en orden ascendente: $D_{i1j1} \leq D_{i2j2} \leq \cdots \leq D_{iPjP}$.

   for $l = 1 : P$ **do**

   if No han sido usadas en esta iteración $\aleph_{il}$ ni $\aleph_{jl}$ **then**

   Fusionar para formar un nuevo centroide: $\aleph = \frac{1}{N_{il}+N_{jl}} \left[N_{il}\aleph_{il} + N_{jl}\aleph_{jl}\right]$

   Borrar $\aleph_{il}$ y $\aleph_{jl}$ y asignar $k \leftarrow k - 1$.

   Regresar al paso 2.

   end if

   end for

9. **end if**

10. Asignar a la variable m el valor de k. Terminar si el número máximo de iteraciones I ha sido alcanzado. En caso contrario regresar al paso 2.

Output: $\boldsymbol{\Upsilon} = \{\boldsymbol{\upsilon}_j : j = 1, \ldots, m\}$ $\star\backslash m$ es el número de agrupaciones $\boldsymbol{\upsilon}_j\backslash\star$

3.3.3. Nivel 3: Visualización

La estimación y la extracción/selección de características puede disponerse en cascada para realizar la *clasificación* de datos, donde se produce una partición del espacio final de características. Este espacio de características debería ser óptimo con respecto a las restricciones de compacidad y separabilidad, de forma que la dimensionalidad resultante es dependiente de la aplicación y no debería fijarse con antelación para la clasificación. En contraste la *visualización de características* requiere una reducción de dimensiones muy restrictiva a una representación bidimensional o tridimensional [27, 43]. En la Figura 3.7 se muestra la taxonomía extraída de la literatura compuesta por los métodos de proyección frecuentemente usados para reducir dimensiones buscando la visualización de datos. Es importante anotar que estos métodos son usados usualmente en cascada con técnicas de selección para obtener un espacio resultante suficientemente pequeño con alto nivel de representación.

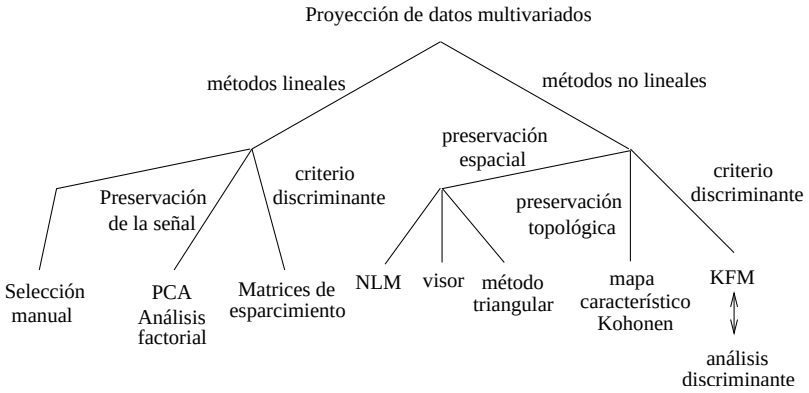


Figura 3.7. Taxonomía de los métodos de proyección

Los mapeos de reducción de dimensiones usados para la extracción de características, proyección, visualización y análisis interactivo de datos multivariados han sido temas de interés por más de tres décadas [39, 40]. Especialmente se reseña el trabajo pionero en este tema, publicado por Sammon en 1969, el cual sirvió de inspiración para la creación de los sistemas actuales de procesamiento de información. En sus trabajos, la reducción de dimensiones se combinó con aplicaciones de análisis visual interactivo de datos, clasificación y el diseño de clasificadores visuales interactivos [40]. Recientemente, hay un fuerte y renovado interés en este tópico incentivado por la minería de datos, registro y procesamiento de variables, entre otros muchos ejemplos. Para lidiar con flujos enormes de información provenientes de bases de datos que crecen rápidamente y de recursos computacionales relacionados, que requieren de métodos avanzados para el entrenamiento de máquinas, reconocimiento de patrones, análisis y visualización de datos, a fin de descubrir estructuras y correlaciones interesantes en los datos. Adicionalmente, los *sistemas inteligentes* gozan de una creciente aceptación industrial con aplicaciones en diferentes campos *e.g.*, reconocimiento de manuscritos, inspección e identificación visual en la industria, vigilancia e identificación, control de acceso usando sistemas biométricos y tecnología de tarjetas integradas, contadores por visión, sistemas de conteo de pasajeros, sistemas aeroportuarios y

tareas de asistencia a conductores en el ambiente automotor. Especialmente la integración de sistemas dedicados con sensores apropiados, *e.g.*, sensores de imagen para sistemas de visión, poseen un significativo potencial económico. Además de las restricciones de aplicaciones tales como, *e.g.*, tamaño, velocidad y desempeño, la minimización del consumo de energía se está volviendo un asunto crucial. Así, se puede observar una fuerte demanda por acelerar y aliviar, aún automatizar el diseño de dichos sistemas a partir del concepto y la arquitectura bajo la implementación microelectrónica.

Con frecuencia, la ayuda visual de gráficos permite tener conocimiento acerca de la estructura con que son regidos los datos en espacios de dimensión múltiple. En este sentido, la reducción de dimensiones es más exigente, ya que se debe capturar la estructura correspondiente de la manera más aproximada posible. El análisis visual se usa generalmente para aprovechar la habilidad humana en descubrir patrones sobre proyecciones de baja dimensión (1-D a 3-D), con lo cual se obtienen representaciones gráficas de la densidad de los datos proyectados (estimaciones de densidad suavizada, gráficas de dispersión, gráficas de contorno, etc) y pueden ser inspeccionados para determinar la presencia de estructuras. Aunque la opción de una disposición 3D puede ser muy útil, y en casos de una alta *dimensión intrínseca* puede brindar beneficios prácticos, generalmente las consideraciones se limitan a representaciones orientadas a visualizaciones 2D; la razón principal es que el grado de libertad que adicionan las representaciones tridimensionales a la visualización 2D, usados en la actualidad, suman complejidad al usuario [22]. Por lo tanto, además de los aspectos computacionales, las representaciones 2D son más fáciles de analizar, manipular e interpretar. El análisis visual facilita el reconocimiento de patrones ya que se explota la percepción y asociación.

Después de obtener un mapeo de calidad aceptable, la visualización de los datos parece ser un asunto trivial. Típicamente, un gráfico de esparcimiento 2D se halla entre las herramientas más comúnmente usadas en la investigación. La información de categoría o clase típicamente es representada utilizando marcadores individuales, *e.g.*, puntos, cuadrados, cruz o círculo. Sin embargo, algo como un visualizador estático pierde todos los beneficios que podrían ser incluidos al emplear interactividad y CAD con el usuario, como es la funcionalidad. Más aún, las características correspondientes al análisis de los datos pueden ser adicionadas, al mismo tiempo como se puede establecer una conexión a la base de datos para obtener una interfaz hombre-máquina (MMI) transparente e intuitiva para el trabajo relacionado al reconocimiento de patrones [40] o computación en general. Con este fin, se desarrolla el concepto conocido como observador de pesos (*Weight Watcher* – WW), el cual fue implementado primero en la estación de trabajo SUN y luego en PC como parte del *sistema QuickCog* [23]. Inicialmente, el WW fue usado en el análisis visual de redes neuronales. Pero el enfoque pronto se amplió para ser utilizado en la visualización y análisis de datos multivariados, con el fin de hacer una óptima explotación de las capacidades humanas de percepción y asociación.

En la Figura 3.8 se muestra un diagrama de bloques del sistema de análisis y visualización de datos multivariados desarrollado, incorporando el WW. Un atributo adicional de disposición debe complementar los puntos de proyección por etiquetas de clase. Aunque la información obtenida mediante los campos de cada punto de proyección es importante, el tamaño requerido puede ir en detrimento, así como los campos contiguos van obstruyendo la vista de sus vecinos. Especialmente el modo de visualización empleando mosaicos es-

pecíficos de Voronoi, en los que se colorean las clases, es de mucha ayuda para descubrir acercamientos en diferentes regiones de clase, *e.g. outliers*. Como no solo el punto individual de proyección sino la correspondiente celda de Voronoi completa es coloreada en este modo, mientras los puntos de proyección son mostrados como pequeños puntos, se da una muy buena visualización de la estructura y distribución de los datos.

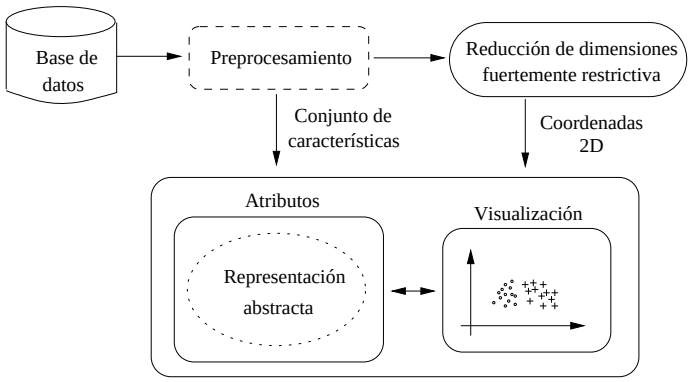


Figura 3.8. Diagrama de bloques del sistema de análisis y visualización de datos multivariados

Estos mecanismos anteriormente descritos facilitan la primera vista de la estructura de los datos. Sin embargo, especialmente para conjuntos de datos muy grandes con considerables variaciones de densidad local, no es factible el análisis en una escala sencilla. Por lo tanto, la búsqueda de la proyección es soportada mediante funciones de ampliación y ajuste, usando una ventana adicional de revisión o búsqueda. La ventana de búsqueda muestra la representación del conjunto completo de datos. Adicionalmente, el nivel actual de ampliación es visible en la ventana de búsqueda por el tamaño de un rectángulo que corresponde a la parte de la proyección visible del WW principal. La posición actual en el WW principal también puede identificarse fácilmente en la ventana de búsqueda. Esta característica permite la exploración interactiva de los datos de aspectos globales a locales. De acuerdo con la densidad local de los datos, el factor de ampliación se puede escoger, de forma que sean visibles los detalles apropiados y sus atributos correspondientes. El foco de interés puede variarse convenientemente a través de las escalas de ampliación, y en la misma escala cambiando la ventana de búsqueda.

Los mapeos proveen proyecciones que revelan la estructura subyacente global de los datos. Sin embargo, si se tiene una *dimensión intrínseca* creciente, ocurren fallas en el mapeo que llevan a distorsiones y giros en la representación 2D. Para una interpretación acertada, el WW provee un *visualizador de vecindario local actual* (ALND). Para esta técnica, se computan los k vecinos mas cercanos del patrón seleccionado. Normalmente, los k vecinos más cercanos obtenidos son representados por una poli-línea roja en el orden de su posición actual en el vecindario del espacio de características o son conectados mediante un rayo de líneas de grosor variante con el punto actual de mapeo seleccionado. El grosor de la línea implica la posición actual en el vecindario en este caso. Así, si la proyección es errónea y tiene giros locales, el ALND aleja el error del mapeo y provee el vecindario correcto para el análisis. Por ejemplo, la lista de vecinos más cercanos puede ser cruzada en orden ascendiente o descendiente. Para cada vecino actual en el foco de interés, se pueden

disponer atributos adicionales y comparar con los otros vectores representados.

La literatura correspondiente a la visualización de características reseña otros métodos importantes como Scatter Plot [5], S-Isomap [13] y Weight Watcher [22], los cuales permiten obtener interpretación visual de datos que residen en espacios de alta dimensión. También es importante destacar que muchos procedimientos de los niveles anteriores de operación, pueden extenderse a este nivel de visualización, imponiendo una restricción fuerte de reducción a 2 ó 3 dimensiones, como sucede con PCA cuando se toman diferentes rotaciones de los 2 ó 3 componentes principales para abstraer estructuras que informan sobre el comportamiento de los datos del espacio inicial (agrupaciones, tendencias, etc).

En este trabajo, el concepto de *visualización* se toma desde un punto de vista aún más exigente, ya que no sólo se busca encontrar las estructuras de la información embebida en el espacio de características, sino que se generan relaciones entre las estructuras y el comportamiento fisiológico subyacente, así, la reducción de dimensiones orientada a este sentido de la visualización, de acuerdo a lo expuesto en el punto 2. y la restricción (b). de la Definición 3.2, está sujeta a las consideraciones expuestas en la Condición 3.2.

Condición 3.2 (Visualización interpretativa de características)

> La visualización de la información embebida en espacios de alta dimensión conectada al fenómeno fisiológico subyacente puede lograrse mediante procedimientos de reducción de dimensiones siempre y cuando exista un mapeo de reconstrucción suave y no singular $\mathfrak{H}$, de forma que el error de reconstrucción $\varepsilon(\boldsymbol{d}) \to 0$.

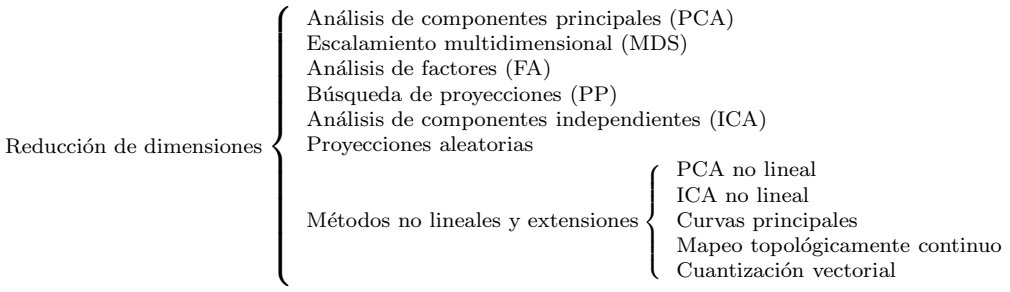
Tomando en cuenta la Condición 3.2, se pueden plantear procedimientos de visualización que sigan el esquema matemático dado por la siguiente expresión:

$$\mathbf{Z} = f_r \circ f_{r-1} \circ \cdots \circ f_2 \circ f_1 \left(\boldsymbol{c}, \mathbf{X}\right) \quad \text{tal que} \quad \exists\, f_1^{-1}, f_2^{-1}, \ldots, f_{r-1}^{-1}, f_r^{-1} \tag{3.23}$$

donde cada f es una función de evaluación con una medida de relevancia μ asociada.

3.4. Estado del arte en la reducción de dimensiones

Algunos de los problemas de mayor importancia en los *sistemas de reconocimiento* automático de patrones son, la reducción del costo computacional y el aumento de la *precisión del sistema* (tasa de aciertos en la etapa de clasificación). Para conseguir estos dos objetivos es necesario realizar eliminación de características irrelevantes, redundantes y correlacionadas; estas últimas son las que en la etapa de clasificación inducen error significativo. Las actividades anteriores se relacionan en un sistema de reconocimiento como la etapa de *reducción de dimensiones* mediante extracción y/o selección de características. La extracción se consigue a través de principios de transformación o mapeo a un espacio reducido de representación y la selección se podría definir básicamente como el proceso a través del cual se busca un subconjunto reducido de q características a partir de los p atributos iniciales que componen el vector de características extraídas de los registros ECG y FCG. A continuación se exponen las técnicas de reducción de dimensiones frecuentemente encontradas en la literatura.



En [12] se relaciona el estado del arte tradicional y los métodos actuales de reducción de dimensiones.

- *Análisis de componentes principales (PCA):* Es la mejor técnica de reducción de dimensiones, en el sentido del error cuadrático medio. En esencia, PCA busca reducir la dimensión del conjunto de datos, hallando unas pocas combinaciones lineales de las variables originales con mayor varianza.

- *Escalamiento multidimensional (MDS):* Las técnicas de escalado multidimensional son una generalización de la idea de componentes principales cuando en lugar de de disponer de una matriz de observaciones por variables, se dispone de una matriz cuadrada de distancias o disimilaridades entre los n elementos del conjunto.

- *Análisis de factores (FA):* Es igualmente otro método lineal, basado en el sumario de datos de segundo orden. La inspección común sugiere que, el FA asume que las variables medidas dependen de algunos frecuentes y desconocidos factores comunes. Ejemplos típicos incluyen como variables definidas, varios resultados de pruebas individuales. Tales resultados son reservados para ser relacionados a un factor de *inteligencia* común. La meta del FA es encontrar tales relaciones, y así usarlos para reducir la dimensión de los conjuntos de datos siguiendo el modelo de factores.

- *Búsqueda de la proyección (PP):* Es un método lineal que, a diferencia de PCA y FA, puede incorporar información a niveles superiores al de segundo orden, y así es útil para conjuntos de datos no-Gaussianos. Ésta una complejidad computacional más alta que los métodos de segundo orden. Dado un índice de la proyección que define lo *interesante* de una dirección, PP busca las direcciones que optimizan el índice. Como la distribución normal es la distribución menos interesante. Los índices de proyección miden varios aspectos de no-normalidad.

- *Análisis de componentes independientes (ICA):* Es un método que busca proyecciones lineales de más alto orden, no necesariamente ortogonales entre sí, pueden ser *casi*

independientes a nivel estadístico, tanto como sea posible. La independencia estadística es una condición mucho más fuerte que la no-correlación.

- *Análisis de componentes principales no lineal:* El análisis no lineal de componentes principales introduce la no linealidad en la función criterio, pero las componentes resultantes son aún combinaciones lineales de las variables originales. Este método puede ser visto como un caso especial de ICA con función objetivo *multi-unidad*.

- *Proyecciones aleatorias:* El método de proyecciones aleatorias es una simple pero potente técnica de reducción de dimensiones, mediante proyecciones aleatorias, se mapean los datos en un espacio dimensional más bajo.

- *Métodos no lineales y extensiones:*
  - Análisis de componentes independientes no lineal.
  - Curvas principales.
  - Mapeo topológicamente continuo.
  - Redes neuronales.
  - Cuantización vectorial.
  - Algoritmos genéticos y evolutivos.
  - Regresión.

En la literatura reciente, se encuentra que la reducción de dimensiones tiene sus tres campos más grandes de aplicación en la clasificación de datos multivariantes, el análisis de datos y la visualización, considerando que los procedimientos para la visualización tienen como objetivo, la mayoría de las veces, facilitar la interpretación de espacios multidimensionales [22]. Muchos procedimientos han sido propuestos para reducir la dimensión, tales como Análisis de Componentes Principales (PCA), Análisis de Componentes Independientes (ICA) y Escalamiento Multidimensional (MDS) [35]. En PCA, la idea principal es encontrar la proyección que preserva al máximo la varianza de los datos originales. ICA es similar a PCA excepto que los componentes tienen la restricción de que deben ser estadísticamente independientes y MDS dirige el esfuerzo a encontrar el espacio inmerso de menor dimensión que mejor preserve las distancias entre todas las parejas de los datos originales. Estos métodos son de fácil implementación, y al mismo tiempo, están bien entendidas las maneras de optimizar las rutinas conservando la eficiencia para reducir dimensiones. Debido a estas ventajas, estas técnicas han sido ampliamente usadas en visualización y clasificación de datos multidimensionales; desafortunadamente, tienen en común una limitación inherente relacionada a la estructura de los datos sobre los cuales son aplicadas: son técnicas lineales mientras que la distribución de la mayoría de los datos reales son no lineales [13]. El principal cometido al reducir dimensiones es preservar al máximo la información representativa de los datos originales de acuerdo a algún criterio de optimización (relevancia). La selección de características se puede tomar como una forma particular de reducir dimensiones donde el subconjunto óptimo o sub-óptimo de variables contribuye

en maximizar la capacidad de representación y minimizar el coste operativo [14]. En [25] se presenta un trabajo orientado a solucionar el problema de selección de características cuando el espacio inicial de entrenamiento está compuesto por un número de características mayor a cincuenta; se realizó un estudio comparativo entre las diferentes técnicas de selección de características con respecto a la complejidad computacional, tipo de objetivo y tipo de búsqueda (ver Tabla 3.3). Donde, $\Theta(\cdot)$ denota una estimación de la complejidad (de manera exacta excepto por la multiplicación de alguna constante), $\mathcal{O}(\cdot)$ denota una estimación de la complejidad para el cual solo se conoce el límite superior. En cuanto a los tipos de objetivo, A se refiere a buscar el mejor subconjunto de un tamaño dado, B se refiere a buscar el subconjunto más pequeño satisfaciendo una condición, C se refiere a encontrar un subconjunto cuya combinación de tamaño y rata de error es óptima. El tipo de búsqueda puede ser serial (S) o paralelo (P). En el estudio, la bondad de cada subconjunto de características es medido mediante la precisión de clasificación usando la conocida estrategia de vecinos más cercanos. Se muestra que los métodos de búsqueda

Algoritmo	**Complejidad computacional**	**Tipo de objetivo**	**Tipo de búsqueda**
SFS, SBS	$\Theta\left(n^2\right)$	A	S
GSFS(g), GSBS(g)	$\Theta\left(n^{g+1}\right)$	A	S
PTA(l, r)	$\Theta\left(n^2\right)$	A	S
GPTA(l, r)	$\Theta\left(n^{\text{máx}\{l+1, r+1\}}\right)$	A	S
SFFS, SBFS	$\mathcal{O}\left(2^n\right)$	A	S
BAB, BAB$^+$, BAB^{++}	$\mathcal{O}\left(2^n\right)$	A	S
RBAB, RBABM	$\mathcal{O}\left(2^n\right)$	B	S
GA	$\Theta(1)\left(\Theta(n)\right)$	C	P
PARA	$\Theta(n)\left(\Theta\left(n^2\right)\right)$	C	P

Tabla 3.3. Comparación de algoritmos de selección de características. Tomado de [25]

flotante secuencial progresiva o regresiva (SFFS, SBFS) son adecuados para conjuntos de características de pequeña y mediana escala, mientras que los algoritmos genéticos (GA) son adecuados para conjuntos de características de gran escala; sin embargo, en términos de interpretación para el análisis de datos multivariantes no se reflejan facilidades destacables. En [54] se expone un trabajo que reúne los conceptos de selección de variables y algoritmos de reducción de dimensiones mediante principios de transformación. Los algoritmos de selección de variables presentan deficiencias en el desempeño cuando los datos están altamente correlacionados y los algoritmos de reducción de dimensiones mediante la transformación del *espacio inicial* de entrenamiento (e.g. PCA, ICA, etc.) tienden a combinar todas las variables, dejando de lado la selección del subconjunto de variables significantes. La combinación del algoritmo de selección de variables con el de reducción de dimensiones logra seleccionar las características con estructura estadística compleja y, al mismo tiempo, la proyección en un subespacio con las bondades inherentes al criterio de transformación. Este algoritmo híbrido no reseña facilidades de visualización, ni la conexión del espacio final reducido con el espacio inicial para interpretar el resultado desde el punto de vista fisiológico.

Por otro lado, tomando en cuenta las dificultades que presentan las técnicas lineales para reducir dimensiones frente a conjuntos de datos reales, dos métodos han ganado aceptabi-

lidad en los últimos años para tratar el problema no lineal de reducción de dimensiones; se conocen con el nombre de Isomap [44] y *Local Linear Embedding* (LLE) [38]. Ambos métodos buscan preservar tanto como sea posible la vecindad local de cada objeto mientras trata de obtener inmersiones altamente no lineales. Debido a esto, han sido categorizados como una clase aparte de técnicas de reducción de dimensiones denominada inmersiones locales (*local embeddings*) [51]. La idea central de estas técnicas es usar el ajuste lineal local para resolver los problemas no lineales globales, la cual está basada en la suposición que los datos caen en una variedad no lineal que puede ser vista como lineal en áreas locales. Tanto Isomap como LLE han sido usados en visualización y clasificación, encontrándose resultados prometedores cuando el conjunto de entrenamiento contiene un nivel muy reducido de ruido, pero ante conjuntos de datos reales con niveles naturales de ruido de adquisición o estimación, estás técnicas pierden su efectividad. En [13], se presenta un trabajo que propone una variación a la técnica Isomap, llamada Isomap supervisado (S-Isomap), la cual utiliza la información sobre la clase para guiar el procedimiento de reducción no lineal de dimensiones. En S-Isomap, la gráfica de la vecindad de los datos de entrada es construida de acuerdo a cierto tipo de disimilaridad entre los puntos, facilitando integrar la información de la clase. La disimilaridad de los datos de entrada contribuye a descubrir verdaderas vecindades, ya que la información geométrica del *espacio inicial* de entrenamiento aunque induce más complejidad computacional es menos vulnerable al ruido; en este sentido, S-Isomap se presenta como una técnica robusta tanto para la visualización como para la clasificación, especialmente en aplicaciones con datos reales. En cuanto a la visualización, S-Isomap es comparado con Isomap y LLE, encontrándose que S-Isomap presenta mejor desempeño. Sin embargo, cuando los datos de entrenamiento están muy dispersos en agrupaciones distantes, las vecindades pueden llegar a desconectarse, en este caso, S-Isomap no es adecuado para tratar este tipo de datos, los cuales son frecuentes en casos reales. En [36] se expone una técnica de alto rendimiento para la reducción de dimensiones en aplicaciones de minería de datos. La técnica es llamada *representative dissimilarity*, mediante la cual se crea un espacio de inmersión donde espacios de alta dimensión con patrones complejos pueden ser simplificados a un conjunto de puntos en un espacio euclidiano de dimensión menor. La técnica obtiene con tan solo unas variables representativas incrementar la velocidad de convergencia en la clasificación sin sacrificar la precisión. Cuando la selección de variables se hace de manera aleatoria, la creación del espacio de inmersión es significativamente reducido sin penalizar la precisión de clasificación. Sin embargo, en términos de interpretación de espacios multidimensionales no se reseñan ventajas, considerando que la interpretación se orienta a revelar sobre cuales variables o características se carga la mayor información correspondiente a la representación de patrones, y de esta manera, seleccionar el grupo de características efectivo para el sistema de reconocimiento.

BIBLIOGRAFÍA

[1] D. Beasley, D. R. Bull, and R. R. Martin. An overview of genetic algorithms: Part 1, fundamentals. *University Computing*, 15(2):58–69, 1993.

[2] A. L. Blum and P. Langley. Selection of relevant features and examples in machine learning. *Artificial Intelligence*, 97(1-2):245–271, 1997.

[3] G. Brassard and P. Bratley. *Fundamentals of Algorithms*. Prentice Hall, New Jersey, 1996.

[4] I. Cohen, Q. Tian, X. S. Zhou, and T. S. Huang. Feature selection and dimensionality reduction using principal feature analysis. In *IEEE International Conference on Image Processing*, 2002.

[5] P. Craig and J. Kennedy. Coordinated graph and scatter-plot views for the visual exploration of microarray time-series data. In *IEEE Symposium on Information Visualization*, pages 173–180, 2003.

[6] M. Dash and H. Liu. Feature selection for classification. *Intelligent Data Analysis. Elsevier*, 1:131–156, 1997.

[7] M. Dash and H. Liu. Consistency-based search in feature selection. *Artificial Intelligence*, 151:155–176, 2003.

[8] P. A. Devijver and J. Kittler. *Pattern Recognition: A Statistical Approach*. Prentice Hall, 1982.

[9] I. Doltsinis, F. Rau, and M. Werner. *Stochastic Analysis of Multivariate Systems in Computational Mechanics and Engineering*, chapter Analysis of Random Systems, pages 9–159. International Center for Numerical Methods in Engineering, Barcelona, Spain, 1999.

[10] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. John Wiley Sons, Inc., 2 edition, 2001.

[11] A. E. Eiben, R. Hinterding, and Z. Michalewicz. Parameter control in evolutionary algorithms. *IEEE Transactions on Evolutive Computation*, 3(2):124–141, 1999.

[12] I. K. Fodor. A survey of dimension reduction techniques. Technical Report UCRL-ID-148494, Lawrence Livermore National Laboratory, 2002.

[13] X. Geng, D.-C. Zhan, and Z.-H. Zhou. Supervised nonlinear dimensionality reduction for visualization and classification. *IEEE Transaction on Systems, Man, and Cybernetics - Part B: Cybernetics*, 35(6):1098–1107, 2005.

[14] L. F. Giraldo, E. Delgado, J. C. Riaño, and G. Castellanos. Feature selection using hybrid evaluation approaches based on genetic algorithms. In *IEEE Computer Society Press Proceedings - Electronics, Robotics and Automotive Mechanics Conference (CERMA'06)*, volume 2, pages 242–247, 2006.

[15] D. E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley, 1989.

[16] J. J. Grefenstette. Optimization of control parameters for genetic algorithms. *IEEE transactions on Systems, Man, and Cybernetics*, 16(1):122–128, 1986.

[17] A. Hyvärinen. Survey on independent component analysis. *Neural Computing Surveys*, 2:94–128, 1999.

[18] A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. Wiley series on adaptive and learning systems for signal processing, comunications and control. John Wiley & Sons, Inc, New York, 2001.

[19] A. Jain and D. Zongker. Feature selection: Evaluation, application, and small sample performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(2):153–158, 1997.

[20] N. L. Johnson and F. C. Leone. *Statistic and experimental design*, volume II. John Wiley & Sons, 2 edition, 1977.

[21] T. Kaczynski, K. Mischaikow, and M. Mrozek. Algebraic topology: A computational approach. Available in http://books.pdox.net/, 2000. Chapter 1.

[22] A. König. Dimensionality reduction techniques for multivariate data classification, interactive visualization, and analysis - systematic feature selection vs. extraction. In *Fourth International Conference on Knowledge-Based Intelligent Engineering Systems & Allied Techndogies*, pages 44–55, 2000.

[23] A. Köning, M. Eberhard, and R. Wensel. A transparent and flexible development environment for rapid design of cognitive systems. In *Proceedings of the EUROMICRO conference, Workshop Computational Intelligence*, pages 655–662, 1998.

[24] D. Koller and M. Sahami. Toward optimal feature selection. In *Proceedings of the 13th International Conference on Machine Learning*, pages 284–292, 1996.

[25] M. Kudo and J. Sklansky. Comparison of algorithms that select features for pattern classifiers. *Pattern Recognition*, 33:25–41, 2000.

[26] M. Last, A. Kandel, and O. Maimon. Information-theoretic algorithm for feature selection. *Pattern Recognition Letters*, 22:799–811, 2001.

[27] R. C. T. Lee, J. R. Staggle, and H. Blum. A triangulation method for the sequential mapping of points form n-space to two-space. *IEEE Transactions on Computers*, C-26:288–292, 1977.

[28] M. Mizuta. *Handbook of Computational Statistics*, chapter Dimension Reduction Methods. Springer, 2004.

[29] M. Ángel Carreira-Perpiñán. *Continuous latent variable models for dimensionality reduction and sequential data reconstruction*. PhD thesis, Department of Computer Science, University of Sheffield, UK, 2001.

[30] M. Orozco-Alzate. Selección efectiva de características wavelet en la identificación de bioseñales 1-D y 2-D usando algoritmos genéticos. Master's thesis, Universidad Nacional de Colombia, Manizales, Colombia, 2005.

[31] M. Pavan and M. Pelillo. A new graph-theoretic approach to clustering and segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 145–152, 2003.

[32] D. Peña. *Análisis de datos multivariantes*. Mc Graw Hill, 2002.

[33] W. F. Punch, E. D. Goodman, M. Pei, L. Chia-Shun, P. Hovland, and R. Enbody. Further research on feature selection and classification using genetic algorithms. In *Proceedings of the International Conference on genetic Algorithms*, pages 557–564, 1993.

[34] J. R. Quinlan. Induction of decision trees. *Machine Learning*, 1(1):81–106, 1986.

[35] A. C. Rencher. *Methods of Multivariate Analysis*. Wiley Series in Probability and Statistics, New York, 2 edition, 2002.

[36] Z. Reynolds, H. Bunke, M. Last, and A. Kandel. Comparing representative selection strategies for dissimilarity representations. *International journal of intelligent systems*, 21:1093–1109, 2006.

[37] S. Roweis and Z. Ghahramani. A unifying review of linear gaussians models. *Neural Computation*, 11(2):305–345, 1999.

[38] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by local linear embedding. *Science*, 290(5500):2323–2326, 2000.

[39] J. W. Sammon. A nonlinear mapping for data structure analysis. *IEEE Transactions on Computers*, C-18(5):401–409, 1969.

[40] J. W. Sammon. Interactive pattern analysis and classification. *IEEE Transactions on Computers*, C-19(7):594–616, 1970.

[41] S. Sarkar and K. L. Boyer. Quantitative measures of change based on feature organization: eigenvalues and eigenvectors. *Computer Vision and Image Understanding*, 71(1):110–136, 1998.

[42] S. Sharma. *Applied multivariate techniques*. Wiley, 1996.

[43] W. Siedlecki, K. Siedlecka, and J. Sklansky. An overview of mapping techniques for exploratory pattern analysis. *Pattern Recognition*, 21(5):411–429, 1994.

[44] J. B. Tenenbaum, V. de Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.

[45] P. M. Narendra and K. Fukunaga. A branch and bound algorithm for feature selection. *IEEE Transactions on Computers*, C-26(9):917–922, 1977.

[46] J. W. P. Ng and L. Hanzo. SGA-C20 Software: A Simple Genetic Algorithm in C version 2.0. Department of Electronics and Computer Science, The University of Southampton, 2001. Available in http://www-mobile.ecs.soton.ac.uk/comms/_private/jwpn.htm.

[47] S. Theodoridis and K. Koutroumbas. *Pattern Recognition*. Elsevier, 3 edition, 2006.

[48] M. Tipping and C. Bishop. Probabilistic principal component analysis. Technical Report NCRG/97/010, Neural Computing Research Group, Aston University, 1997.

[49] M. E. Tipping and C. M. Bishop. Mixtures of probabilistic principal component analysers. *Neural Computation*, 11(2):443–482, 1999.

[50] V. Vapnik. *The Nature of Statistical Learning Theory*. Springer, 1995.

[51] M. Vlachos, C. Domeniconi, D. Gunopulos, G. Kollios, and N. Koudas. Non-linear dimensionality reduction techniques for classification and visualization. In *Proceedings 8th ACM SIGKDD International Conference Knowledge Discovery and Data Mining*, page 645–651, 2002.

[52] M. D. Vose. *The Simple Genetic Algorithm: Foundations and Theory*. The MIT Press, Cambridge, MA, 1999.

[53] A. R. Webb. *Statistical Pattern Recognition*. John Wiley & Sons Ltd, 2 edition, 2002.

[54] L. Wolf and S. Bileschi. Combining variable selection with dimensionality reduction. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 801–806, 2005.

[55] L. Wolf and A. Shashua. Feature selection for unsupervised and supervised inference: the emergence of sparsity in a weighted-based approach. In *Proceedings of the Ninth IEEE International Conference on Computer Vision*, pages 378–384, 2003.

[56] S. J. Wu and P. T. Chow. Genetic algorithms for nonlinear mixed discrete-integer optimization problems via meta-genetic parameter optimization. *Engeenering Optimization*, 24(2):137–159, 1995.

CAPÍTULO 4

Marco experimental

Ignoranti, quem portum petat,
nullus suus ventus est.

(Lucio Anneo Séneca)

En este capítulo se determina la realización de un conjunto de experimentos que involucra el uso de procedimientos a diferentes niveles del esquema general propuesto para reducir dimensiones. Además, se describen los modelos requeridos para el preprocesamiento de las señales ECG y FCG y la generación de características, que tiene como fin la detección de patrones funcionales relacionados a eventos isquémicos y deficiencias valvulares.

4.1. Esquema metodológico

En la Figura 4.1 se muestra el esquema general de procedimiento usado en este trabajo, donde a partir de registros de la actividad cardíaca (ECG y FCG) se obtiene la estructura multivariada de representación y mediante un análisis de relevancia se encuentra el espacio de representación reducida que determina los patrones de funcionamiento relacionados a isquemia y soplos cardíacos.

4.2. Bases de datos

4.2.1. Señales electrocardiográficas BD-UNAL-GCPDS-ECG

Esta base de datos correspondiente a señales electrocardiográficas (ECG) fue creada en la Universidad Nacional de Colombia Sede Manizales, durante el período de Julio de 2003 y Junio de 2004, la cual está diseñada para ser utilizada en la evaluación de algoritmos entrenados en la detección de cambios de la señal debidos a eventos isquémicos. Los registros fueron tomados en los centros hospitalarios ubicados en la ciudad de Manizales (Hospital de Caldas, Hospital Santa Sofía y grupo Telesalud de la Universidad de Caldas). La base de datos consta de 89 registros ECG de 12 derivaciones durante 10 minutos. Fueron tomados a 60 hombres con edades entre 35 y 55 años, y 29 mujeres entre 45 y 65 años de edad. El registro de cada paciente está almacenado en una carpeta que contiene 4 archivos de texto, donde está contenida la información de las 12 derivaciones. Los criterios de selección

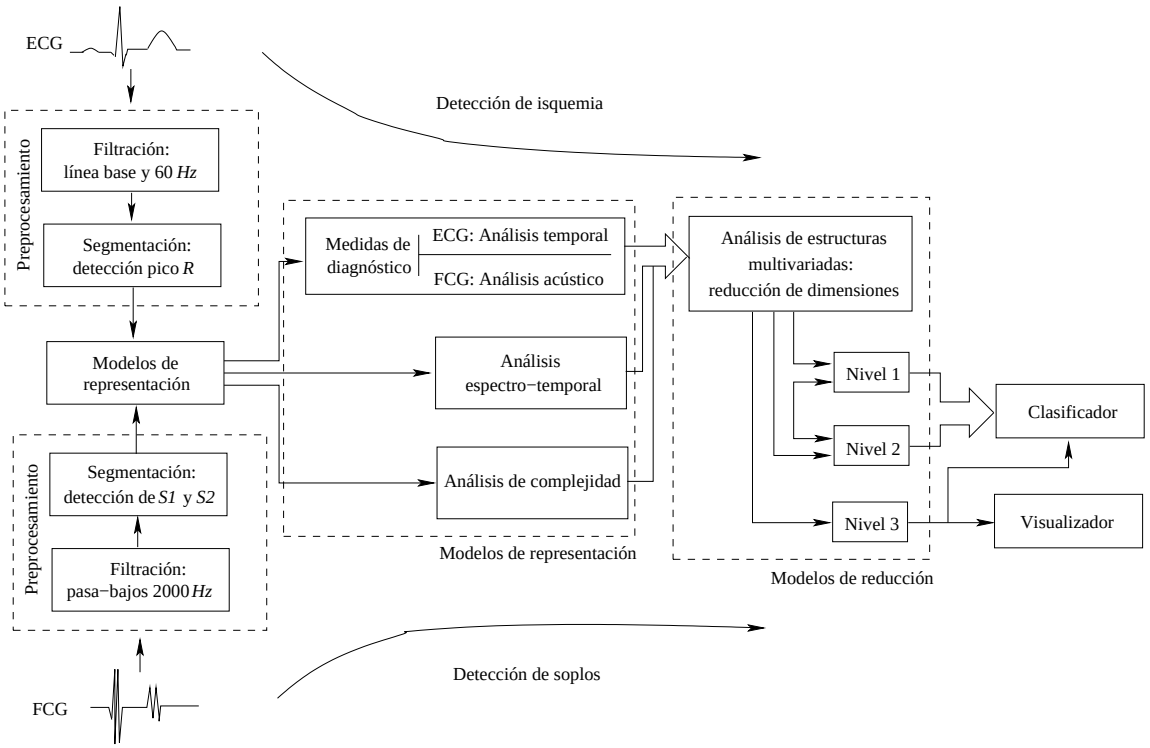


Figura 4.1. Esquema general de procedimiento

adicional fueron establecidos con el propósito de obtener una selección representativa de anormalidades en la señal, tales como desplazamientos de la línea base del segmento *ST* debidas a hipertensión, disquinesia ventricular y efectos de los medicamentos. Adicionalmente, la base de datos cuenta con una breve descripción de la historia clínica, en formato manuscrito, donde se relacionan variables como factores de riesgo, antecedentes familiares, diagnóstico de ingreso, número de días de hospitalización y medicamentos prescritos. El dispositivo utilizado en la adquisición de los registros en formato digital fue el *Cardio Card PC Based ECG / RESTING PC ECG / PC EKG System* fabricado por Nasiff Associates®. Esta tarjeta de adquisición fue utilizada bajo los siguientes parámetros de operación: frecuencia de muestreo $f_s = 500Hz$, resolución de 13 bits, rango de 10 mV, 3 canales y 12 derivaciones. Dentro de los registros con cardiopatías se encuentran patologías cardíacas tales como infarto antiguo, bloqueo auriculoventricular, flutter auricular, bloqueo de rama derecha, arritmias, entre otros.

4.2.2. Señales electrocardiográficas BD-MIT E-STT

La base de datos ST-T Europea está diseñada para ser usada en la evaluación de algoritmos para el análisis de cambios en el *ST* y en la onda *T*. Esta base de datos consta de 90 registros de dos canales de ECG cada uno y de dos horas de duración, tomados a partir de cintas analógicas obtenidas de grabaciones realizadas con 79 sujetos. Los sujetos fueron 70 hombres con edades entre los 30 a los 84 años, y 8 mujeres con edades entre los 55 a los 71 [32]. La grabación de las cintas analógicas se obtuvo mediante una variedad de

grabadores Holter de 2 canales. Estas cintas fueron digitalizadas a 250 Hz por canal, usando hardware específico para ello. Se usó para este propósito un conversor A/D de 12 bits con un rango de $\pm 10mV$. Dos cardiólogos trabajaron independientemente para anotar cada registro latido a latido y los cambios en la morfología del segmento ST y de la onda T, tomando en cuenta el ritmo y la calidad de la señal. Los cambios en el segmento ST y la onda T fueron identificados en ambas derivaciones, y sus valores de inicio, amplitud máxima y fin fueron anotados. Posteriormente se compararon las anotaciones hechas por los dos cardiólogos y las divergencias las resolvió otro cardiólogo perteneciente al grupo de coordinación [31]. La mayoría de los episodios ST de la base de datos y muchos de los episodios T están relacionados con el diagnóstico o la sospecha de la existencia de isquemia de miocardio, pero algunos pueden ser producidos por cambios de posición del paciente. Se establecieron criterios de selección adicionales para obtener una selección representativa de las anormalidades del ECG en la base de datos, incluyendo desplazamiento del segmento ST resultante de condiciones tales como hipertensión, disquinesia ventricular y efectos de los medicamentos. La base de datos incluye 367 episodios de cambios del segmento ST y 401 episodios de cambios de la onda T, con duraciones que van desde los 30 segundos a varios minutos y amplitudes pico que van desde los 100 μV hasta más de 1 mV. Además, 11 episodios de desplazamiento del eje que resultaban en un aparente cambio del ST y 10 episodios de desplazamiento del eje que resultaban en un aparente cambio de la onda T, fueron marcados. Cada registro aparece documentado con un informe clínico. Estos informes, que se encuentran en los archivos de cabecera (`.hea`) asociados con cada registro, incluyen la patología, los medicamentos, desequilibrios electrolíticos e información técnica de cada registro. Los archivos de cabecera incluyen información acerca de las derivaciones usadas, la edad del paciente, el sexo, los medicamentos, los hallazgos clínicos y el equipo de registro [32]. De cada registro se obtuvieron varios latidos promediados (promedios de 20 latidos aprox.) obteniéndose así un conjunto de 900 latidos normales y 900 latidos con evidencias isquémicas.

4.2.3. Señales fonocardiográficas BD-UNAL-GCPDS-FCG

Esta base de datos pertenece a la Universidad Nacional de Colombia y está conformada por registros FCG tomados a sujetos normales y a pacientes que evidencian la presencia de soplos cardíacos causados por valvulopatías cardíacas (estenosis aórtica, regurgitación mitral, etc). A cada paciente se le tomaron 8 registros correspondientes a los cuatro focos de auscultación (mitral, tricuspídeo, aórtico y pulmonar) en fase de apnea post-espiratoria y post-inspiratoria, cada uno con una duración de 12 segundos, en posición decúbito dorsal y a una frecuencia de muestreo de 44100 Hz. El tiempo no se pudo extender más debido a que los pacientes con problemas cardíacos son incapaces de sostener la apnea post-inspiratoria y post-espiratoria por más tiempo. Las muestras fueron adquiridas a 50 sujetos normales y 98 pacientes con evidencia de soplo cardíaco con un estetoscopio electrónico marca WelchAllyn® modelo Meditrón. La toma de los registros se realizó sobre personas con edades comprendidas entre los 18 y 35 años y buscando que la población estuviera balanceada con respecto al género. Todos los registros normales se adquirieron dentro de la cabina acústica del Hospital Infantil Universitario "Rafael Henao Toro", después de haber valorado a los sujetos mediante un chequeo médico para comprobar su estado auscultatorio e indicarles el propósito del estudio. De esta base de datos etiquetada

se usaron 164 registros distribuidos así: 81 registros con etiqueta normal y 83 registros con etiqueta de soplo. Finalmente, con ayuda de los médicos especialistas se eligieron 360 latidos representativos, 180 normales y 180 con evidencia de soplo cardíaco. Es importante anotar que el análisis de estacionariedad de los registros FCG se realiza con el objetivo de encontrar si cada ciclo cardíaco tiene las mismas características estadísticas con respecto a los otros, y por lo tanto, la ventana cuasi-estacionaria sobre la cual se debe realizar la estimación corresponde al tiempo de duración de un ciclo cardíaco. Esto se realiza, calculando los valores de media y desviación estándar para cada uno de los ciclos cardíacos.

4.2.4. Conjuntos de datos artificiales

El análisis de estructuras multivariadas requiere inicialmente la realización de pruebas con estructuras artificiales donde hay conocimiento a priori sobre cuales variables son relevantes o irrelevantes al concepto de clasificación.

Monk-1. Es un conjunto de datos multidimensionales compuesto de dos clases no linealmente separables, 6 características $\{\xi_i\}_{i=1}^{6}$ y 432 observaciones. Solamente ξ_1, ξ_2 y ξ_5 son relevantes al concepto de clasificación.

Monk-3. Es un conjunto de datos multidimensionales compuesto de dos clases con mayor facilidad de separación que el conjunto anterior, 6 características $\{\xi_i\}_{i=1}^{6}$ y 432 observaciones. Solamente ξ_2, ξ_4 y ξ_5 son relevantes al concepto de clasificación.

SynthData. Es un conjunto de datos obtenido a partir de un generador de estructuras multivariadas disponible en Internet [21], con las siguientes especificaciones: 2 clases, 14 características $\{\xi_i\}_{i=1}^{14}$ y 500 observaciones. Solamente las 10 primeras características (ξ_1 a ξ_{10}) son relevantes al concepto de clasificación.

4.3. Preproceso y adecuación de señales

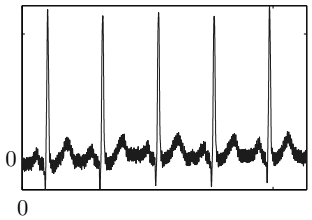
Como puede notarse en la Figura 4.1, esta fase consta de dos etapas importantes: la reducción de perturbaciones y la detección de eventos dentro del registro.

4.3.1. Reducción de perturbaciones

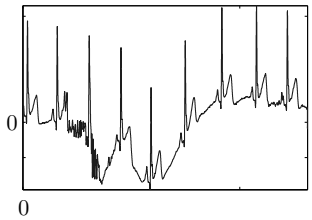
Señales electrocardiográficas

La amplificación y el registro de los bio-potenciales presentan con frecuencia problemas de interferencia, originados por la red de distribución eléctrica como se puede apreciar en la Figura 4.2(a). Las fuentes de esta interferencia pueden ser clasificadas como magnéticas y eléctricas, incluyendo luces, cables de AC, tomas de corriente u otros equipos que operen cerca. Esta interferencia también puede presentarse por inducción magnética, conexiones de equipos o debida a las corrientes parásitas que fluyen a través del cuerpo de la persona y de los cables de conexión. Las diferencias de impedancia entre los electrodos, también generan esta interferencia, al convertir los voltajes de interferencia en modo común, en señales diferenciales no deseadas [8]. Los parámetros típicos de la interferencia de la línea de potencia indican que su frecuencia fundamental está ubicada en 60 Hz y sus armónicos, modelados como sinusoides y combinación de sinusoides, tienen una amplitud hasta del 50 % de la amplitud pico a pico de la señal ECG [12] con un SNR del orden de 3 dB [22]. De otra parte, también se consideran los artefactos inducidos por movimiento (ver Figura 4.2(b)),

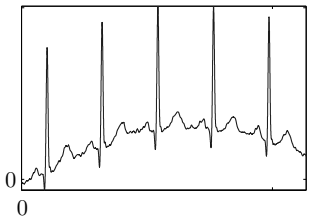
los cuales generan cambios transitorios en la línea base por los cambios en la impedancia electrodo-piel. Como esta impedancia cambia, el amplificador del ECG observa una fuente de impedancia diferente, la cual forma un divisor de voltaje con la impedancia de entrada del amplificador. Por lo tanto, el voltaje de entrada del amplificador depende de la impedancia de la fuente, la cual cambia al modificarse la posición del electrodo. Estos artefactos se caracterizan por tener una amplitud del 30 % de la amplitud pico a pico del ECG y una duración de $100 - 500\ ms$. La desviación de la línea base también puede ser causada por la respiración y se representa como una componente sinusoidal en la frecuencia de respiración, adicionada a la señal electrocardiográfica. La amplitud y la frecuencia de la componente sinusoidal pueden ser variables [12]. La amplitud de la señal electrocardiográfica puede variar cerca del 15 % debido a la respiración. La variación puede producir una modulación de amplitud del ECG por la componente sinusoidal que es adicionada a la línea base (ver Figura 4.2(c)). El espectro de la señal contaminada muestra componentes adicionales de 0.15 a 0.30 Hz. Es de especial cuidado el tratamiento de este tipo de ruido, debido a que se encuentra ubicado en la banda de componentes importantes del ECG, por lo que su filtrado puede eliminar componentes importantes en el registro.



(a) Interferencia de 60 Hz.



(b) Artefactos por movimiento.



(c) Desviación de la línea base.

Figura 4.2. Perturbaciones frecuentes en las señales ECG

Existen criterios de evaluación para los sistemas de proceso de señales ECG, con los cuales se determina la calidad de la señal digitalizada con respecto a la señal real. Un criterio de evaluación importante corresponde a la medida de fidelidad con respecto a la señal original. La *American Heart Association* (AHA) y el *American National Standards Institute* (ANSI) han definido los siguientes estándares relacionados con criterios de fidelidad:

1. *Criterios para la rutina de lectura visual*
   - La desviación del registro de salida con respecto a una representación lineal de la señal de entrada no puede exceder 25 μV ó 5 % del valor máximo de la señal (AHA).

- La desviación del registro de salida con respecto a una representación lineal de la señal de entrada no puede exceder 50 μV ó 10 % del valor máximo de la señal (ANSI).

2. *Criterios para el diagnóstico morfológico*

   - La raíz cuadrada del error cuadrático medio (valor eficaz) de la amplitud de los complejos P, QRS y T, no debe exceder los 10 μV.
   - El error en las deflexiones de amplitud pico a pico no debe exceder los 10 μV ó el 2 %.
   - El valor eficaz dividido por la amplitud cuadrática media de cualquier deflexión no puede exceder el 1 %.
   - Las deflexiones de amplitud del complejo QRS, con una amplitud $\leq$ 20 μV y una duración mayor a 6 ms, deben ser detectadas.
   - El error relativo del pico máximo no puede exceder el 10 % sobre el intervalo de cualquier deflexión de amplitud del $QRS \geq 20$ μV y mayor de 12 ms.

A continuación se presentan los filtros adaptativos que fueron usados en este trabajo para la reducción de la interferencia de la línea de potencia (60 Hz): ASIC e IE.

Filtro ASIC (*adaptive sinusoidal interference canceller*). El problema de la cancelación de interferencias sinusoidales se puede plantear de la siguiente manera: Dada una señal de entrada $r(kT_s)$, tal que:

$$r(kT_s) = s(kT_s) + \alpha \cos(\omega_0 kT_s + \phi)$$

donde $s(kT_s)$ es la señal deseada, T_s es el período de muestreo, ω_0 es la frecuencia conocida de la interferencia, mientras que α y ϕ representan respectivamente, la amplitud y la fase desconocidas de la señal sinusoidal de interferencia. Sin pérdida de generalidad, se asume que $\alpha > 0$ y $\phi \in [0, 2\pi)$. El objetivo es extraer $s(kT_s)$ de la señal contaminada $r(kT_s)$ sin ninguna distorsión [30]. Cuando α y ϕ se obtienen de forma exacta, $s(kT_s)$ puede ser recuperada substrayendo una sinusoide sintética, la cual es caracterizada por los estimados de fase y amplitud de $r(kT_s)$. Usando esta idea de cancelación en el dominio del tiempo, se define una función de error, $e(kT_s)$, que tiene la forma:

$$e(kT_s) = r(kT_s) + \tilde{\alpha} \cos(\omega_0 kT_s + \tilde{\phi})$$

donde $\tilde{\alpha}$ y $\tilde{\phi}$ denotan los estimados de α y ϕ respectivamente. Es de esperar, que cuando sea minimizado el error cuadrático medio de $e(kT_s)$, con respecto a estas dos variables, $\tilde{\alpha} \rightarrow \alpha$ y $\tilde{\phi} \rightarrow \phi$, entonces $e(kT_s)$ se volverá idéntico a $s(kT_s)$. En el algoritmo ASIC propuesto en [30], los parámetros de fase y amplitud se ajustan de forma iterativa para minimizar el valor cuadrático instantáneo de $e(kT_s)$. Similar al algoritmo LMS [36], el ASIC usa estimados estocásticos del gradiente que se obtienen derivando $e^2(kT_s)$ con respecto a $\tilde{\alpha}(kT_s)$ y $\tilde{\phi}(kT_s)$. Los parámetros son adaptados directamente y de forma independiente,

de acuerdo a las siguientes ecuaciones:

$$\tilde{\alpha}((k+1)T_s) = \tilde{\alpha}(kT_s) + \mu_\alpha e(kT_s)\cos(\omega_0 kT_s + \tilde{\phi}(kT_s))$$
$$\tilde{\phi}((k+1)T_s) = \tilde{\phi}(kT_s) - \mu_\phi e(kT_s)\sin(\omega_0 kT_s + \tilde{\phi}(kT_s))$$

donde μ_α y μ_ϕ son escalares positivos que controlan la tasa de convergencia y aseguran la *estabilidad* del sistema. Es importante anotar que a partir de este procedimiento, se conoce también su forma extendida llamada *Adaptive Sinusoidal Interference Multiple Canceller* (ASIMC), la cual tiene básicamente la misma estructura del ASIC, aunque se diferencia en que toma en cuenta M interferencias sinusoidales.

$$r\left(kT_s\right) = s\left(kT_s\right) + \sum_{i=1}^{M} \alpha_i \cos\left[\left(\omega_i + \Delta\omega_i\right) kT_s + \phi_i\right] \tag{4.1}$$

$$e\left(kT_s\right) = r\left(kT_s\right) - \sum_{i=1}^{M} \hat{\alpha}_i\left(kT_s\right) \cos\left[\omega_i kT_s + \hat{\phi}_i\left(kT_s\right)\right]$$

Con el fin de evitar las operaciones de seno y coseno, se construye una tabla de búsqueda de coseno. Esta tabla consiste en un vector $\mathbf{\Gamma}$ de longitud L que tiene los siguientes elementos:

$$\mathbf{\Gamma} = \left[1 \quad \cos\left(\frac{\pi}{L}\right) \quad \cos\left(\frac{2\pi}{L}\right) \cdots \cos\left(\frac{(L-1)\pi}{L}\right)\right]$$

El algoritmo es eficiente porque sólo requiere 5 M multiplicaciones, 3 M sumas y 2 M operaciones de búsqueda, en cada intervalo de muestreo [30].

Filtro adaptativo IE (*incremental estimation*). El filtro de *estimación incremental* se basa en la estimación de la amplitud de la interferencia, ignorando la señal deseada. Se puede demostrar que la amplitud $e(n)$ en una muestra de una interferencia sinusoidal, se puede calcular en términos de la amplitud de las dos muestras anteriores. Sea la interferencia de la forma:

$$e(k) = A\sin\left(\frac{2\pi f_0}{f_s}\right)$$

luego, el valor actual de $e(k)$ se puede calcular de la forma:

$$e(k) = 2e(k-1)\cos\left(\frac{2\pi f_0}{f_s}\right) - e(k-2)$$

donde $e(k)$ es la interferencia estimada en el tiempo $t(k)$, f_0 es la frecuencia de la interferencia y f_s es la frecuencia de muestreo. La salida del filtro es la diferencia entre la muestra de entrada y la estimación de la interferencia:

$$y(k) = x(k) - e(k)$$

Si la predicción es correcta, la muestra de interferencia estimada $e(k)$ será igual a la muestra de la señal $x(k)$, más un posible *offset*. La mejor estimación de este offset es la diferencia previa entre la muestra de la señal y la estimación de la interferencia $(x(k-1) - e(k-1))$. Por lo tanto, si la función:

$$\varepsilon(k) = (x(k) - e(k)) - (x(k-1) - e(k-1))$$

es positiva, es decir $\varepsilon > 0$, la estimación de $e(k)$ ha sido muy pequeña y, por lo tanto, debe ser incrementada antes de estimar el próximo punto y viceversa. El valor de $e(k)$ es corregido al procesar cada muestra de acuerdo a la siguiente regla:

$$e(k) = e(k) + \mu\,\mathrm{sgn}(\varepsilon(n))$$

El parámetro μ determina el tiempo de adaptación. Al aumentar μ se disminuye el tiempo de adaptación, pero aparecen mayores distorsiones en el ECG y algunos problemas de estabilidad. Por el contrario, disminuyéndolo mejora la señal electrocardiográfica pero se vuelve más lenta la convergencia. Este algoritmo se basa en el LMS de signo (sgn-LMS) [10]. El parámetro μ se conoce como el *incremento* y es una cantidad fija, que se tiene que elegir tomando en cuenta la siguiente restricción: Si μ es grande, el filtro se adaptará más rápido al ruido, si es pequeño, el error medio resultante será más pequeño (error en estado estacionario). Aumentando μ se incrementa el rango de frecuencias que se pueden eliminar satisfactoriamente. La máxima desviación de frecuencia se puede calcular aproximadamente como:

$$\Delta f_{o_{max}} = \frac{\mu}{4\pi A} f_s$$

donde A es la amplitud de la interferencia.

Con respecto a la reducción de perturbaciones relacionadas a la desviación de línea base se implementó un filtro basado en la transformada wavelet.

Filtración mediante descomposición wavelet Debido a que las variaciones de línea base consisten en interferencias de baja frecuencia que se añaden a la señal y que pueden ser consideradas como la tendencia a largo plazo de la misma, se utiliza la capacidad de las wavelets para suprimir un polinomio, dependiendo de los momentos nulos que posea, permitiendo de esa manera reducir las variaciones de la línea base. Así, si $x[n]$ es cero para $n = 0, \ldots, L$, entonces la wavelet posee $L + 1$ momentos nulos y los polinomios de grado n pueden suprimirse por la wavelet, por lo tanto, escogiendo una wavelet madre y un nivel de descomposición adecuado se puede reducir el efecto del ruido de linea base en las señales ECG. Tomando el modelo que se propone en [7] para representar la señal de ECG contaminada con ruido electromiográfico y ruido de línea base,

$$y[n] = x[n] + s[n]$$

$x[n]$ es la señal original y $s[n]$ corresponde al ruido por desplazamiento de línea base. El objetivo del método es obtener una estimación de la línea base, $\hat{s}[n]$, de forma que su substracción de $y[n]$ dé como resultado una señal sin variaciones en la línea base. En este caso se utiliza una filtración lineal, de forma que todos los detalles de la transformada wavelet

serán eliminados, quedándose únicamente con la aproximación. La señal $\hat{s}[n]$, se obtiene directamente del efecto de filtro paso bajo de la aproximación de la señal $y[n]$ de cierto nivel, el cual debe ser establecido a priori. Un nivel demasiado bajo puede provocar una sobreaproximación de la señal, por lo cual además de la línea base, se incluirían componentes del ECG. Por el contrario, si el nivel de la aproximación es demasiado elevado, la estimación de la línea base se aleja demasiado y los resultados obtenidos no son satisfactorios. Por todo esto, es evidente que la elección automática del nivel de aproximación adecuado es un paso relevante del método [7]. Para realizar el procedimiento se contamina una señal sin variación de línea base con una señal simulada con amplitud, frecuencia y fase conocidas $s[n]$. Se seleccionan diferentes niveles de aproximación del resultado de la descomposición de la señal, los cuales se reconstruyen haciendo cero los detalles restantes, representando así $\hat{s}[n]$. Se compara la dispersión del espectro de la señal $s[n]$ con el espectro de la señal $\hat{s}[n]$ y se obtiene un parámetro. Se calcula la varianza de $\hat{y}[n] = y[n] - \hat{s}[n]$ y se obtiene un segundo parámetro. El cruce de dichos parámetros en una gráfica formada por varios niveles de descomposición, permite determinar el nivel óptimo para estimar $\hat{s}[n]$. Las pruebas experimentales revelaron que el nivel adecuado de aproximación es 8 y, la wavelet que presenta mejores resultados es la *Daubechies* de orden 4 ($db4$). Todos los resultados fueron validados por un especialista en cardiología.

Criterios para medir la efectividad de los filtros A continuación se define una serie de medidas que permiten medir la efectividad de las diferentes técnicas de filtración implementadas en este trabajo.

PRD (percentage root difference). Representa una medida convencional de distorsión y se utiliza generalmente en algoritmos de compresión de señales. Se define como:

$$\mathrm{PRD} = \sqrt{\frac{\sum_{n=1}^{N} \left(x(n) - \tilde{x}(n)\right)^2}{\sum_{n=1}^{N} x^2(n)}} \times 100 \tag{4.2}$$

donde $x(n)$ es la señal original, $\tilde{x}(n)$ es la señal filtrada, y N es el tamaño de la ventana sobre el cual se calcula el PRD. También se puede usar otra medida de PRD con diferente normalización, y se utiliza para comparar señales con media diferente de cero:

$$\mathrm{PRD}' = \sqrt{\frac{\sum_{n=1}^{N} \left(x(n) - \tilde{x}(n)\right)^2}{\sum_{n=1}^{N} \left(x(n) - \bar{x}\right)^2}} \times 100 \tag{4.3}$$

donde $\bar{x}$ es la media de la señal. La definición (4.3) es independiente del nivel DC de la señal original. (4.2) y (4.3) son por supuesto iguales para señales con media cero [37]. Así, el PRD representa el porcentaje de la raíz cuadrada de la relación entre la energía del error y la energía de la señal original, en un número N de muestras dado. A menor

error, se obtiene menor índice PRD, por lo cual un valor bajo de este índice representa un buen desempeño del algoritmo de filtración.

SNR (*signal to noise ratio*). Esta medida se puede expresar como:

$$\mathrm{SNR} = 10 \log \left(\frac{\sum_{n=1}^{N} (x(n) - \bar{x})^2}{\sum_{n=1}^{N} (x(n) - \tilde{x}(n))^2} \right) \tag{4.4}$$

donde se observa que a menor error entre la señal original y la señal filtrada, se incrementa el SNR. La relación entre el SNR (4.4) y el PRD (4.3) puede expresarse como: $\mathrm{SNR} = -20 \ \log(0.01 \ \mathrm{PRD})$.

Coeficiente de correlación. Con el fin de obtener una medida que tenga en cuenta el grado de similitud entre las señales original x y filtrada y, se emplea el coeficiente de correlación cruzada de Pearson R_{xy}, calculado mediante la siguiente expresión:

$$R_{xy} = \frac{C_{xy}}{\sqrt{C_{xx} C_{yy}}} \tag{4.5}$$

donde C_{xy} es la covarianza cruzada. La correlación mide tanto la relación lineal entre dos variables como su sentido (si es directo o inverso). Cuando la relación es perfectamente lineal la expresión (4.5) equivale a 1 ó -1. Cuando el coeficiente tiene un valor próximo a cero, o bien no existe relación entre las variables analizadas o dicha relación no es lineal [23].

Error cuadrático medio (potencia media del error). Este tipo de error mide el nivel de aproximación que existe entre dos señales mediante una expansión o representación ortogonal. Puede definirse como:

$$\bar{\varepsilon}^2 = \frac{1}{N} \sum_{n=1}^{N} |x(n) - \tilde{x}(n)|^2 \tag{4.6}$$

de forma que $\bar{\varepsilon}^2 \geq 0$. Cuanto más $\bar{\varepsilon}^2$ tienda hacia 0, es porque existe un mayor ajuste entre la señal original x y la señal filtrada $\tilde{x}$.

Señales fonocardiográficas

Con el fin de eliminar componentes de alta frecuencia que pueden causar alteraciones en la caracterización espectral de señales FCG y tomando en consideración que esta señal no contiene información acústica en frecuencias superiores a 1000 Hz, se implementó un filtro pasa-bajos con frecuencia de corte en 2000 Hz. Luego se normalizó la señal con respecto al máximo del valor absoluto, con el fin de obtener valores entre $[-1, 1]$. El procedimiento para lograr la reducción de las perturbaciones acústicas en los fonocardiogramas consistió en diseñar algoritmos del tipo *denoising*, usando cada una de las reglas de selección del umbral

con los métodos de reescalamiento. Estas combinaciones se realizaron variando el nivel de descomposición desde 1 hasta 10 para diferentes familias wavelets (*Haar*, *Daubechies*, *Symlet* y *Coiflet*). Posteriormente, un grupo de tres médicos expertos en auscultación cardíaca escucharon los registros originales y los filtrados mediante *denoising*. A partir de los conceptos emitidos por el grupo de especialistas, se descartaron los algoritmos que alteraron de forma inadecuada las características sonoras de la señal y se eligió el que de forma más clara permitió percibir tato los sonidos cardíacos como los soplos.

4.3.2. Detección de eventos

Señales electrocardiográficas

Como se requiere un algoritmo de segmentación que responda de manera efectiva en cualquiera de las 12 derivaciones, se propone analizar las derivaciones donde las componentes de la señal ECG presentan mayor energía respecto a otras (por ejemplo la onda P, se aprecia con mayor intensidad en la derivación DII). Seguidamente los puntos de segmentación encontrados para la derivación analizada, se proyectan sobre las derivaciones restantes para finalizar el proceso de segmentación. Esto es debido a que como las derivaciones son proyecciones eléctricas de diferentes áreas del corazón, se conserva la correspondencia entre la dinámica cardíaca y el trazo de las doce derivaciones del ECG. Por ejemplo, la onda P informa sobre la despolarización de las aurículas y el intervalo de puntos que corresponde a la onda P en la derivación $V1$ (por citar un ejemplo), tiene su inicio, su final y el mismo número de puntos de igual forma en las demás derivaciones. De acuerdo a esto, con detectar los eventos adecuadamente en una sola derivación, los puntos de segmentación obtenidos en dicha derivación tienen una correspondencia temporal con las otras.

El algoritmo de segmentación inicialmente busca encontrar los puntos donde inicia y termina el complejo *QRS*. La detección de este complejo se obtiene mediante la transformada wavelet usando las escalas 2^1 a 2^4 de una spline cuadrática, debido a que el complejo *QRS* produce dos módulos máximos con signos opuestos y un cruce por cero en el pico R [29]. Por lo tanto, los puntos que determinan el inicio y final del complejo *QRS* son obtenidos aplicando reglas de decisión (umbrales) a la transformada wavelet de la señal ECG. La mayor energía del complejo *QRS* está entre 3 Hz y 40 Hz (ver Tabla 4.1), el rango de frecuencias alrededor de $3dB$ de la transformada de Fourier de la wavelet indican que la mayor parte de energía del complejo *QRS*, se encuentra entre las escalas 2^3 y 2^4, siendo más grande en 2^4, de forma que la energía decrece si la escala es más grande. La energía de artefactos debidos al movimiento y al error de línea base se incrementa para escalas mayores a 2^5. Por lo tanto es conveniente usar las escalas $2^1 - 2^4$ de la transformada wavelet [29]. Debido a que el complejo *QRS* corresponde a dos módulos máximos con signo opuesto de

Escala	Frecuencia bajo 3 dB	Frecuencia sobre 3 dB
2^1	32.1 Hz	92.1 Hz
2^2	18 Hz	65.4 Hz
2^3	8 Hz	33.1 Hz
2^4	4 Hz	16.2 Hz
2^5	2 Hz	7.8 Hz

Tabla 4.1. Respuesta en frecuencia a diferentes escalas

la transformada wavelet, el módulo máximo correspondiente a la onda R fue determinado usando el siguiente procedimiento:

- Los módulos máximos de la escala 2^4, que cruzan un umbral predeterminado Th_4 (donde Th_j es el umbral para la transformada wavelet en la escala 2^j) y las posiciones en la escala de la señal son marcadas como n_k^4 (donde n_k^j es la posición en la escala 2^j y k es el contador de módulos máximos que sobrepasan el umbral).
- El módulo máximo en la cercanía de n_k^4 en la escala 2^3 es determinado y su localización es marcada como n_k^3. Si existen varios módulos máximos, entonces se selecciona el más grande. Si no existen módulos máximos, entonces, n_k^3, n_k^2, n_k^1 son puestos a cero.
- La localización de los módulos máximos de las escalas n_k^4, n_k^3, n_k^2, n_k^1, en sus respectivas escalas es determinada, con el fin de reducir el efecto del ruido de alta frecuencia, que se presenta en las las escalas más bajas.

El pico R se ubica entre el par de módulos máximos consecutivos y con signos opuestos (un máximo y un mínimo) de la transformada wavelet. En ciertos latidos ectópicos o ante la presencia de ruido pueden ocurrir dos o más módulos máximos del mismo signo, de los cuales sólo uno se relaciona con el pico R. Si se presentan dos módulos negativos notados como MIN1 y MIN2, cerca de un módulo positivo a una distancia L1 y L2, y se toma A1 y A2 como los valores absolutos de amplitud para MIN1 y MIN2; entonces se aplica la siguiente regla para saber cual de los mínimos debe ser descartado [29]:

1. si $\frac{A1}{L1} > 1.2\ \frac{A2}{L2}$, MIN2 debe ser descartado.
2. si $\frac{A2}{L2} > 1.2\ \frac{A1}{L1}$, MIN1 debe ser descartado.
3. si MIN1 y MIN2 están en el mismo lado del máximo, entonces el módulo negativo más alejado del máximo debe ser descartado.

Este mismo procedimiento puede también realizarse cuando se presenten dos módulos positivos y uno negativo. Así, de acuerdo a la relación existente entre la señal y su transformada wavelet, el cruce por cero en la escala 2^1 corresponde al pico R, útil para el cálculo de la frecuencia cardíaca del registro. Si ocurre una perturbación que hace el pico R más elevado y puntiagudo, entonces el valor del módulo máximo será muy grande. Para evitar estos errores, se establece el parámetro A_j^{m+1} (una estimación del módulo máximo para calcular el umbral del próximo complejo QRS), el cual conserva su valor anterior ($A_j^{m+1} = A_j^m$) si $\left|Wf(2^j, n_k^j)\right| \geq 2A_j^m$, de lo contrario, $A_j^{m+1} = \left(\frac{7}{8}\right) A_j^m + \left(\frac{1}{8}\right) \left|Wf(2^j, n_k^j)\right|$. Adicionalmente, la exactitud del procedimiento puede mejorarse tomando en cuenta que dos complejos QRS no deben ocurrir en un tiempo inferior a 200 ms [14], entonces se establece un período refractario de 200 ms. Para encontrar el inicio del complejo QRS, se toma una ventana de 100 ms a la izquierda del pico R (módulo máximo negativo) con el fin de hallar un punto máximo. Para el final del QRS, se toma una ventana de 60 ms a la derecha del pico R con el fin de encontrar un punto mínimo, después de hallado ese punto, se establece un umbral de sobrepaso para determinar el final del módulo máximo, ubicado en el 25 % del valor mínimo. El instante correspondiente al sobrepaso de este umbral determina el final

del QRS. Una vez que se tienen los puntos que determinan el complejo QRS, se procede a realizar la detección de las ondas P y T usando el análisis wavelet. Estas ondas tienen su espectro de potencia en el rango de 0.5 Hz a 10 Hz, mientras que el ruido de línea base y el ruido de artefactos tienen su frecuencia en el rango de 0.5 Hz a 7 Hz; por lo tanto se utiliza la escala 2^4 para reducir los efectos de estas perturbaciones. Para la detección de las ondas P y T se usa la escala 2^4, tomando una ventana de 200 ms antes del inicio del complejo QRS de la escala 2^1, la transformada wavelet en esta escala genera de forma similar un par de módulos máximos como los que se generan en el QRS, pero en menor proporción, facilitando así determinar el inicio y el final de las ondas P y T. El pico y el ancho de la onda P se puede encontrar de la siguiente manera:

a) Los módulos máximos son puntos donde la expresión $\left|Wf(2^4, p_k^4)\right|$ se maximiza.

b) El cruce por cero entre el par de módulos máximos corresponde al pico de la onda P.

c) Para obtener el inicio de la onda P, se busca hacia atrás del módulo máximo, que está a la izquierda del cruce por cero, hallando el punto donde $\left|Wf(2^4, p_k^4)\right|$ equivale al 5 % del módulo máximo; este punto determina el inicio de la onda P.

d) Para el final de la onda P se realiza una búsqueda hacia adelante del módulo máximo que está a la derecha del cruce por cero y el procedimiento es similar al anterior.

La detección de la onda T tiene el mismo procedimiento que el de la onda P. De esta manera, se obtienen los puntos que determinan el inicio y final de la onda P, el complejo QRS y la onda T. Para la detección del complejo ST-T se procede de la siguiente manera:

– El inicio del complejo ST-T se establece en el punto final del complejo QRS.

– El final del complejo ST-T se establece con una ventana de 280 ms a la derecha del final del QRS.

Señales fonocardiográficas

Segmentación entre-latidos La segmentación de los latidos a partir de registros FCG presentó dificultades debidas a la dependencia que existe entre la estructura acústica del latido y el foco donde se ausculta. De manera estándar se consideran los 4 focos de auscultación: Mitral, Tricuspídeo, Aórtico y Pulmonar. En la mayoría de las veces los soplos no se escucha en los cuatro focos de auscultación, a menos que sea demasiado intenso. Por lo tanto, cuando se tiene un registro con evidencia de soplo cardíaco puede provenir de cualquiera de los focos de auscultación. De manera general, en el Mitral y Tricuspídeo $S1$ es más intenso que $S2$ y en el Aórtico y Pulmonar $S2$ es más intenso que $S1$, sin embargo no es una regla porque lo contrario también puede ocurrir. En la clínica, el médico usualmente se ayuda de alguna derivación de la señal ECG para ubicar el sonido $S1$, ya que el comienzo del primer sonido cardíaco coincide con el inicio del complejo QRS del ECG. Debido a esto, se tuvo en consideración que el equipo de adquisición con que se obtuvieron los registros FCG usados para este estudio incluyera sincrónicamente la derivación DII del ECG a cada registro FCG, ya que esta derivación presenta una morfología del complejo QRS fácil de detectar. El algoritmo de segmentación diseñado para este estudio usa la información del

ECG para coincidir el inicio del complejo QRS con el inicio del sonido $S1$. Así, el punto inicial del complejo QRS del ECG determina el inicio de cada latido en la señal FCG. Para la detección del pico R en la señal ECG se usa el mismo algoritmo expuesto en la sección (4.3.2 – *Señales electrocardiográficas*), basado en el análisis wavelet, debido a que los módulos máximos y el cruce por cero de la transformada wavelet corresponden a cambios abruptos en la señal.

Segmentación intra-latido Después de tener segmentado cada latido FCG, se procede a segmentar el sonido $S1$ y el sonido $S2$, de forma que las regiones restantes son asignadas a la sístole y a la diástole. Para la segmentación intra-latido, se utilizó un procedimiento basado en la estadística de tiempos de recurrencia, el cual se deriva del *análisis de complejidad*. Después de reconstruir el atractor correspondiente a la señal de análisis, como se explicó en la Sección 2.3.2, la *estadística de tiempos de recurrencia* [1] estima la cantidad de puntos (estados) que quedan dentro de una esfera de radio r y centrada en algún punto del espacio de estados tomado como referencia. La cantidad de puntos dentro de la esfera es dividida entre el número total de puntos del atractor. De esta manera se obtiene una medida probabilística de la densidad del atractor en los puntos de referencia, la cual tiende a cero si el radio de la esfera es muy pequeño, y tiende a 1 cuando este radio tiende a infinito. En el caso particular de las señales FCG, los tiempos de recurrencia $\Delta(r)$ se calculan utilizando una ventana móvil de 2000 puntos sobre la señal muestreada a 44100 Hz, con un solapamiento del 90 %, buscando buena resolución en el dominio del tiempo. Otros dos parámetros importantes para el cálculo de $\Delta(r)$, son los parámetros de embebimiento: la dimensión de embebimiento m y el tiempo de retardo τ. El primer parámetro se calculó utilizando el método de los falsos vecinos, para una determinada cantidad de señales FCG, y se encontró experimentalmente que una dimensión de 5 representa adecuadamente el atractor de la señal. Por otro lado, el tiempo de retardo τ se calcula minimizando la información mutua entre cada una de las variables de estado, mientras se maximiza la distancia entre los puntos del atractor y la diagonal principal del espacio de estados. Se obtuvo un valor de $\tau = 100$. Después de reconstruido el atractor, el procedimiento de segmentación intra-latido puede tomarse como sigue:

Detección del sonido $S1$. Con la segmentación entre-latidos se obtiene el inicio cada $S1$ del registro FCG y el análisis queda restringido a cada ciclo cardíaco (ó latido). Para obtener el final de $S1$, se calcula $\Delta(r)$ con un radio $r = 0.35$. A partir de esto, se obtiene un gráfico que se minimiza cuando la señal presenta mayor actividad con respecto al punto de referencia (en este caso se toma el origen como punto de referencia). Para detectar el final de $S1$, se invierte $\Delta(r)$ quedando $\hat{\Delta}(r) = 1 - \Delta(r)$, con el fin de convertir el mínimo de la función en un máximo como se puede observar en la Figura 4.3. Como el final de $S1$ no se encuentra fisiológicamente más allá de 163 ms después de su inicio, entonces se vuelve cero el intervalo desde 163 ms hasta el final del latido de $\hat{\Delta}(r)$. Luego, se realiza una verificación para comprobar si se ha elegido un radio de esfera adecuado, evaluando si el máximo de $\hat{\Delta}(r)$ es mayor a un umbral determinado experimentalmente en 0.8. Si se cumple esta condición se normaliza $\hat{\Delta}(r)$ para continuar con el siguiente paso del proceso. En caso de cumplirse la condición, significa que el radio escogido es muy grande, y la estadística de tiempos de recurrencia no está brindando información suficiente de la señal, para lo cual,

iterativamente se reduce el radio en una cantidad de 0.03 hasta que se satisfaga esta condición de umbral. De esta manera, se obtendría una gráfica con un único lóbulo y un máximo, correspondientes a $S1$. Como se desea detectar el final de dicho lóbulo, para facilitar el análisis se hacen cero los valores de $\hat{\Delta}(r)$ anteriores a 68 ms después del inicio de $S1$. La detección del final de $S1$, puede presentar algunos inconvenientes

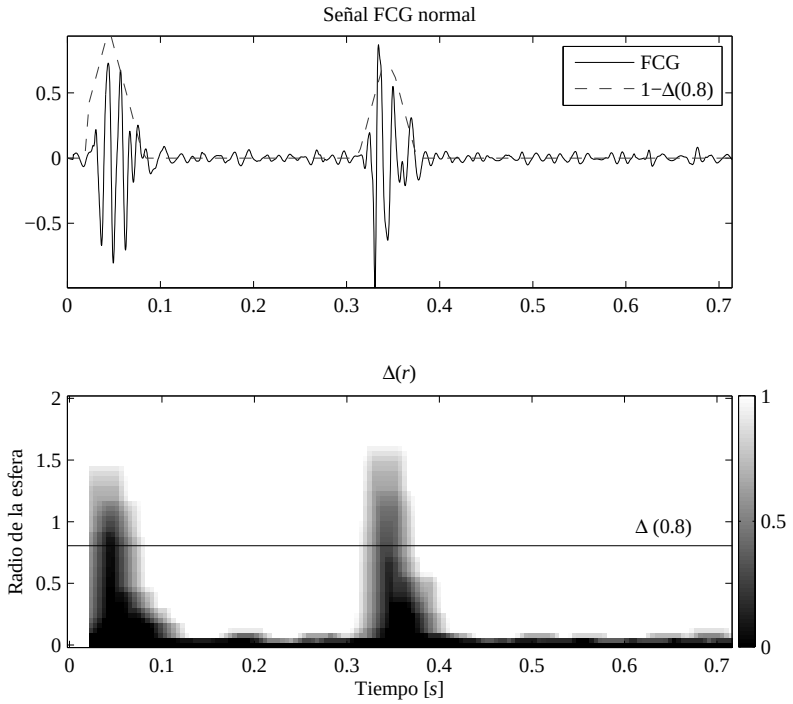


Figura 4.3. Estadística de tiempos de recurrencia sobre una señal FCG normal

cuando la señal FCG no es normal y un soplo sistólico sigue inmediatamente a $S1$, considerando incluso que la amplitud del soplo pueda ser mayor que la de $S1$. De acuerdo a esto, si en el intervalo desde 136 a 163 ms, cualquier punto de $\hat{\Delta}(r)$ supera un umbral predefinido en 0.8, se asume que la señal puede tener un soplo sistólico, y por tal motivo es necesario remover el soplo para considerar exclusivamente las componentes correspondientes a $S1$. La remoción del soplo se realiza mediante una búsqueda hacia atrás, eliminando los valores de la señal $\hat{\Delta}(r)$ desde el punto ubicado en 163 ms hasta encontrar que el valor de $\hat{\Delta}(r)$ sea inferior a 0.9 con pendiente negativa o hasta que la envolvente de la señal tome un valor de 0.5. La búsqueda hacia atrás, también finaliza en 113 ms, para evitar remover información de $S1$ en el caso que ninguna de las condiciones anteriores se llegue a cumplir. Una vez se remueven los soplos se asume que la señal restante debe corresponder únicamente a $S1$, entonces, se detectan los máximos locales de la señal $\hat{\Delta}(r)$, buscando los cruces por cero de la primera derivada. A partir del primer máximo local, se comienza la búsqueda para hallar el final de $S1$. Se considera final de $S1$, si se cumple alguna de

las siguientes condiciones:

1. Se alcanza un umbral mínimo de 0.02.
2. Si ningún punto de la señal entre 91 y 163 ms alcanza a estar por debajo de otro umbral mínimo igual a 0.15, entonces se considerará que el final de $S1$ se encuentra cuando la señal tiene una pendiente positiva mayor a 0.01.

Detección del sonido $S2$. El procedimiento para la detección de $S2$ es un poco más complejo que el usado para la detección de $S1$, ya que es necesario identificar tanto el inicio, como el final de $S2$, el cual es difícil de identificar en señales patológicas porque casi siempre el soplo cardíaco está justo antes o después de $S2$. En este caso la ventana en la cual se puede encontrar $S2$ está entre 250 y 610 ms luego del inicio de $S1$, así, todos los valores que se encuentran por fuera de este intervalo, se hacen iguales a cero para facilitar el análisis. Se hace el cálculo de $\hat{\Delta}(r)$ de forma similar a como se hizo para $S1$, buscando el radio de la esfera de forma recursiva hasta que $\hat{\Delta}(r)$ supere un umbral determinado, el cual para este caso se ajusta a un valor mínimo de 0.1. Este umbral es mucho menor que el usado en la detección de $S1$, ya que es posible que $S2$ tenga una magnitud muy pequeña, y por mas que se disminuya el radio r, puede ser muy difícil que se alcance un umbral más alto. Generalmente los soplos sistólicos son los que tienen mayor intensidad, y por eso es necesario removerlos así como se realizó en la detección de $S1$. Es necesario considerar que puede existir el final de un soplo sistólico justo antes de $S2$, si en el intervalo de 250 a 263 ms hay al menos un punto que supera el umbral fijo de 0.8. Si sucede esto, entonces es necesario eliminar este primer lóbulo correspondiente al soplo sistólico. Para eliminarlo, se hace una búsqueda desde 250 ms, haciendo igual a cero los puntos de la señal $\hat{\Delta}(r)$ hasta que se alcance un umbral fijo δ_x, el cual se establece de acuerdo a determinadas condiciones, con el fin de evitar eliminar toda la información de la señal $\hat{\Delta}(r)$:

a) $\delta_x = 0.5$ si en el intervalo entre 250 y 363 ms, la señal no alcanza a descender por debajo de 0.1.

b) $\delta_x = 0.1$ si en el intervalo entre 250 y 363 ms, la señal tiene valores menores a 0.1.

Es posible que la señal pueda estar tan distorsionada por el soplo cardíaco, que pueda ser imposible removerlo, ya que se correría el riesgo de remover también a $S2$. Si esto sucede, y la señal $\hat{\Delta}(r)$ queda completamente anulada, es conveniente restaurar los valores iniciales. En la Figura 4.4 se muestra la remoción del lóbulo considerado como soplo, con el fin de tener una mejor precisión en la detección de $S2$. A diferencia de la detección de $S1$, en este caso es necesario detectar lóbulos completos en el diagrama de $\hat{\Delta}(r)$, es decir, hay que detectar tanto el inicio como el final de segmentos de la señal que superen determinados umbrales. En este proceso de detección de lóbulos se utilizan dos umbrales, uno de mayor valor al inicio ($\delta_{si} = 0.5$), y otro de valor bajo ($\delta_{sf} = 0.1$) para detectar el final. Esto se hace con el fin de ser precisos en la estimación del inicio y el final de cada lóbulo, sin hacer una detección errónea de lóbulos causados por ruido y otras componentes de la señal FCG diferentes a $S2$. Para cada lóbulo detectado, se halla el punto máximo y el tiempo de ocurrencia. La semiología cardíaca reporta que $S2$ no puede comenzar más allá de 430 ms después

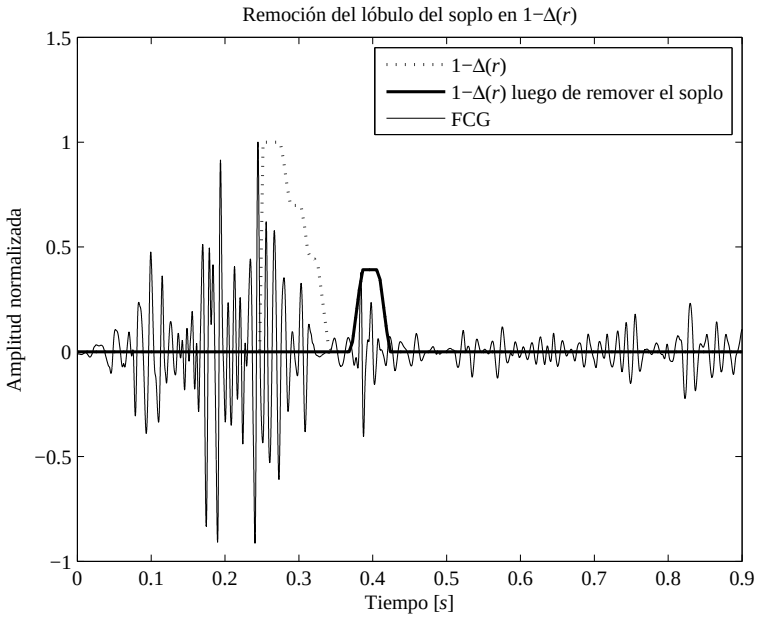


Figura 4.4. Remoción de soplo sistólico y detección de $S2$

del inicio de $S1$, así que se eliminan aquellos lóbulos que comiencen después de ese tiempo. Después de remover el lóbulo correspondiente al soplo cardíaco, en el caso de que exista y sea posible identificarlo y eliminarlo; y después de remover también los lóbulos que comiencen más allá de 430 ms, entonces se asume que el lóbulo de mayor amplitud restante, es el que corresponde a $S2$, así que se procede a determinar con una mayor precisión los límites del comienzo y del final. Para determinar el inicio de $S2$, se hace una búsqueda hacia atrás desde 14 ms antes del punto máximo, hasta que se cumple una de las siguientes condiciones:

1. $\hat{\Delta}(r)$ es igual a cero.
2. $\hat{\Delta}(r)$ es menor que 0.6 y la pendiente es descendente.

Para detectar el final de $S2$ se realiza un procedimiento similar al anteriormente descrito para hallar el inicio, con la diferencia que se hace una búsqueda hacia adelante 14 ms luego del máximo, y en la segunda condición la pendiente es ascendente.

Segmentación alternativa. El procedimiento anterior para detectar $S2$ ofrece un gran desempeño en señales FCG normales, sin embargo, en las señales patológicas puede darse el caso de que la detección no sea adecuada. Así, es necesario identificar si la detección establecida por el procedimiento basado en la estadística de tiempos de recurrencia es correcta o no. Si no lo es, entonces es necesario aplicar otro criterio de segmentación, con el fin de dar una estimación más acertada de la ubicación de $S2$. Los criterios usados para identificar posibles fallas en la segmentación, se expresan a continuación:

– La longitud de $S2$ no debe ser menor que 35 ms.

- La longitud de $S2$ no debe ser mayor que 240 ms.
- El inicio de $S2$ no debe ser antes de 265 ms luego del inicio de $S1$.
- El inicio de $S2$ no debe ser después de 400 ms luego del inicio de $S1$.
- En el proceso de detección de lóbulos, no se detectaron lóbulos definidos.

El procedimiento alternativo se basa en filtros, descomposición wavelet y energía de Shannon [19]. Inicialmente, se realiza un filtrado pasa-altas de la señal, con el fin de resaltar las componentes de alta frecuencia de los sonidos cardíacos. Este filtrado se realiza con un filtro FIR de orden 50, con banda de rechazo por debajo de 20 Hz, y banda de paso a partir de 30 Hz. Luego, sobre la señal filtrada se realiza una descomposición wavelet a nivel 6 con la wavelet madre $db6$. A los detalles de la descomposición se le aplica la energía de Shannon, y se filtra con un filtro pasa-bajas, con el fin de encontrar la envolvente E_s de la señal. El filtro pasa-bajas es de orden 50, con banda de paso por debajo de 5 Hz y banda de rechazo por encima de 10 Hz. Las reglas de decisión para la detección de $S2$ se aplican sobre la envolvente E_s. En la Figura 4.5 se muestra una señal FCG, con el correspondiente envelograma E_s. Después

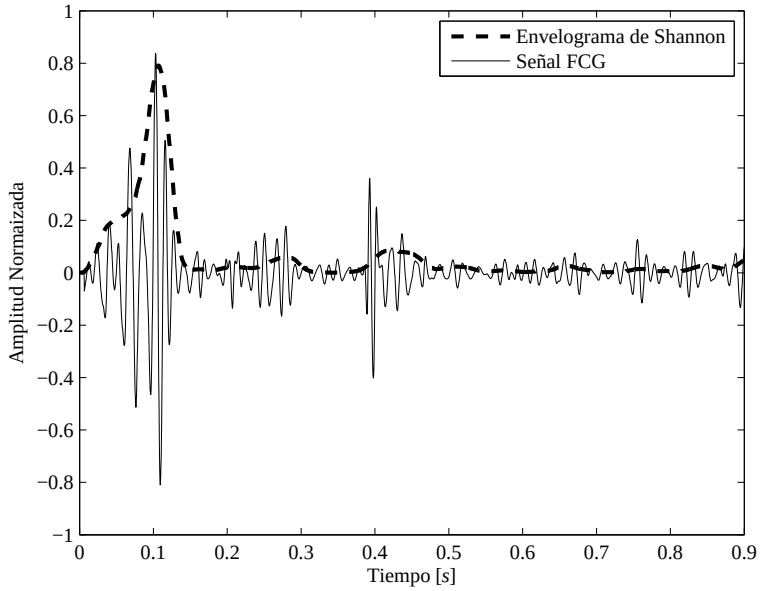


Figura 4.5. Envelograma de Shannon de una señal FCG

de obtener E_s, se realiza un procedimiento similar al utilizado con $\hat{\Delta}(r)$ para eliminar el lóbulo de energía correspondiente a un posible soplo sistólico. Si algún valor entre 250 y 263 ms es mayor que 0.8, entonces se vuelven ceros los puntos de la envolvente hasta que ésta tenga una magnitud menor a 0.3 ó la pendiente sea positiva. Luego de eliminar el lóbulo del soplo sistólico, se procede a detectar los lóbulos de energía restantes en la señal. En este caso se utilizan umbrales simétricos para detectar el inicio y final de $S2$, el cual se determina experimentalmente en 0.3. Para cada lóbulo se halla el máximo y el tiempo en que ocurre. Posteriormente, se eliminan aquellos

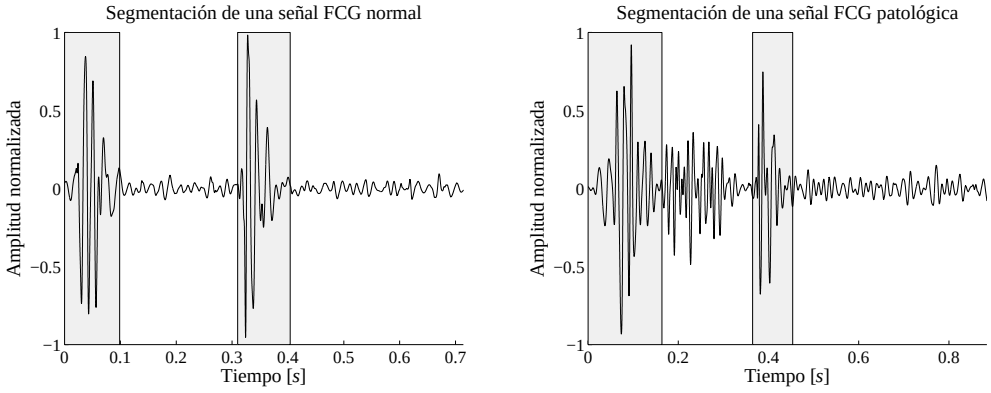
lóbulos que comiencen antes de 260 ms y después de 440 ms. Finalmente, se toma como lóbulo principal, el de mayor valor en el máximo, y se realiza una búsqueda hacia atrás (desde 8 ms antes del máximo) y hacia adelante (desde 8 ms después del máximo), con el objetivo de encontrar el inicio y el final del lóbulo. La búsqueda se realiza hasta que se cumple cualquiera de las siguientes condiciones:

1. E_s es menor que 0.3.
2. E_s es menor que 0.4 y la pendiente es descendente (ascendente para el final).

Si la señal es muy compleja, y al hacer la eliminación de lóbulos de energía no queda alguno para analizar, entonces se establece uno de acuerdo a valores fisiológicos promedio a partir del inicio de $S1$:

- comienzo: 265 ms
- final: 610 ms
- tiempo en el que se presenta el máximo: 440 ms
- valor máximo: 1

En la Figura 4.6 se muestran los resultados de segmentación para una señal normal y para una señal patológica.



(a) (b)

Figura 4.6. Segmentación intra-latido de señales FCG

Desempeño del algoritmo de segmentación de ECG y FCG

El rendimiento de los algoritmos de segmentación es medido con criterios de evaluación recomendados por la AAMI (*Association for the Advancement of Medical Instrumentation*). Es importante decir que usualmente estos criterios también son usados para medir el desempeño de clasificadores [11]. La notación usada es como sigue:

- TP: Número de verdaderos positivos, es decir, segmentos correctamente detectados.
- FP: Número de falsos positivos, es decir, segmentos marcados erróneamente.
- FN: Número de falsos negativos, es decir, segmentos no detectados.

Valor predictivo positivo. Es una medida que informa sobre cuantos segmentos detectados son verdaderos.

$$\mathrm{P} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \times 100\,\% \tag{4.7}$$

Sensibilidad. Es una medida que informa sobre cuantos eventos se detectan entre todos los latidos analizados por el algoritmo de segmentación.

$$\mathrm{S} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \times 100\,\% \tag{4.8}$$

4.4. Extracción de las medidas de representación

Como se puede observar en la Figura 4.1, el espacio de características que representa la información contenida en las señales ECG y FCG está conformado por 3 modelos de representación: Medidas de diagnóstico, análisis espectro-temporal y análisis de complejidad.

4.4.1. Medidas de diagnóstico

De acuerdo a lo que se expuso en la Sección 2.1 y tomando en cuenta la expresión (2.1), de las señales ECG se obtiene el conjunto $\mathbf{X}_{1-ECG} = \mathcal{MD}\{\mathrm{ECG}\}$ compuesto por las medidas de diagnóstico que se citan en la Tabla 2.1, para lo cual se requiere de una etapa adecuada de segmentación. De otra parte, el conjunto $\mathbf{X}_{1-FCG} = \mathcal{MD}\{\mathrm{FCG}\}$ correspondiente a las medidas de diagnóstico sobre las señales FCG está compuesto por medidas que usan la energía promedio normalizada de Shannon y medidas derivadas del análisis acústico, como se exponen en la Tabla 4.2.

Análisis temporal

Las siguientes medidas de energía fueron extraídas de las señales FCG:

- E_1: es el valor máximo entre dos relaciones de energía. La primera relación consiste en dividir la energía de la sístole contra quien tenga mayor energía entre $S1$ y $S2$. La segunda relación consiste en dividir la energía de la diástole contra quien tenga mayor energía entre $S1$ y $S2$. Lo cual puede expresarse como:

$$E_1 = \text{máx}\left(\frac{E_{sistole}}{\text{máx}(E_{S1}, E_{S2})}, \frac{E_{diastole}}{\text{máx}(E_{S1}, E_{S2})}\right)$$

- E_2: es la relación entre la energía de la sístole y la energía de $S1$.

$$E_2 = \frac{E_{sistole}}{E_{S1}}$$

- E_3: es la relación entre la energía de $S2$ y la energía de $S1$.

$$E_3 = \frac{E_{S2}}{E_{S1}}$$

– E_4: es la relación entre la energía de la diástole y la energía de $S1$.

$$E_4 = \frac{E_{diastole}}{E_{S1}}$$

Análisis acústico

La caracterización acústica sobre las señales de FCG se realizó usando un ventaneo de longitud ι y desplazamiento ψ, de manera que el valor estimado de cada característica para todo el registro se hizo igual al promedio de los valores obtenidos en las ventanas. los valores de ι y ψ fueron estimados en la etapa de entrenamiento de los algoritmos.

Estimación de los MFCC. En este trabajo se calculan 14 MFCC para cada segmento intra-latido de la señal FCG ($S1$, sístole, $S2$ y diástole). Se diseñan 14 filtros triangulares cuyas frecuencias centrales van desde 0 hasta $fs/2$ de forma logarítmica, donde fs es la frecuencia de muestreo de la señal ($fs = 3\ kHz$). Las amplitudes decaen exponencialmente hacia $fs/2$. De esta manera, la representación de Fourier de cada segmento del latido, X, se multiplica por los coeficientes de cada filtro H_m, usando la expresión (2.3). Finalmente, se determinan los MFCC $c[n]$ (1 por cada filtro) de acuerdo a la expresión (2.2), considerando $M = 14$ el número de coeficientes MFCC. También se estimó el MFCC relativo, tomado como el valor máximo entre el cuarto MFCC sistólico y el cuarto MFCC diastólico, debido a que la frecuencia central de cuarto MFCC coincide con la mayor componente en frecuencia de los soplos cardíacos.

Estimación del Jitter. Inicialmente, el procedimiento requiere determinar la longitud de ventana que contenga al menos un ciclo de la componente de menor frecuencia con magnitud significativa en el espectro promedio de las señales FCG. Este valor se estimó en 20 *Hz*, por lo tanto la longitud de la ventana se tomó de 5 *ms*. Luego, se realizó un ventaneo de los segmentos intra-latido ($S1$, sístole, $S2$ y diástole), y mediante la transformada rápida de Fourier se estimó la frecuencia fundamental f_0 de cada tramo, desplazando la ventana con un solapamiento del 30 %. Así, usando la expresión (2.4) se obtuvo la estimación del Jitter.

Compendio. El conjunto total $\mathbf{X}_{1-FCG}$ compuesto por las medidas de diagnóstico extraídas de las señales FCG se citan en la Tabla 4.2.

4.4.2. Análisis espectro-temporal

De acuerdo a lo que se expuso en la Sección 2.2 y tomando en cuenta la expresión (2.5), de las señales ECG se obtiene el conjunto $\mathbf{X}_{2-ECG} = \mathcal{L}_2\{\text{ECG}\}$ compuesto por los coeficientes derivados de representaciones tiempo-frecuencia que se listan más adelante. De otra parte, el conjunto $\mathbf{X}_{2-FCG} = \mathcal{L}_1\{\mathcal{L}_2\{\text{FCG}\}\}$ correspondiente al conjunto de características espectro-temporales obtenidos mediante dos operadores en cascada, donde es $\mathcal{L}_1$ es un operador de envolvente y $\mathcal{L}_2$ es un operador de transformación.

Señales electrocardiográficas

Una vez se tienen los puntos de segmentación de la señal ECG, es necesario conformar el espacio de análisis mediante una estructura compuesta por las 12 derivaciones, de forma que cada registro tenga un número de puntos adecuado de acuerdo a la técnica de

Parámetro	**Descripción**
E_1	$E_1 = \text{máx}\left(\frac{E_{sistole}}{\text{máx}(E_{S1},E_{S2})}, \frac{E_{diastole}}{\text{máx}(E_{S1},E_{S2})}\right)$
E_2	$E_2 = \frac{E_{sistole}}{E_{S1}}$
E_3	$E_3 = \frac{E_{S2}}{E_{S1}}$
E_4	$E_4 = \frac{E_{diastole}}{E_{S1}}$
$MS1_{\{1,2,...,14\}}$	Estimación de 14-MFCC sobre $S1$
$MS2_{\{1,2,...,14\}}$	Estimación de 14-MFCC sobre $S2$
$Msist_{\{1,2,...,14\}}$	Estimación de 14-MFCC sobre la sístole
$Mdiast_{\{1,2,...,14\}}$	Estimación de 14-MFCC sobre la diástole
$Mrel$	$Mrel = \text{máx}\{Msist_4, Mdiast_4\}$
$JS1$	Estimación del Jitter sobre $S1$
$JS2$	Estimación del Jitter sobre $S2$
$Jsyst$	Estimación del Jitter sobre la sístole
$Jdiast$	Estimación del Jitter sobre la diástole

Tabla 4.2. Medidas de diagnóstico extraídas de las señales FCG

caracterización tiempo-frecuencia que se vaya a usar. Debido a que los métodos empleados en este trabajo se basan en transformaciones de la señal original, es necesario que la longitud de los latidos sea constante. Como primera aproximación se pueden truncar los latidos a partir de una referencia. En [13] se ubica la marca fiducial del pico R y se toman n muestras a la derecha y m a la izquierda. El inconveniente que se presenta al truncar el latido, es la pérdida de la información que se encuentra en los extremos de la señal y que puede ser relevante, como es el caso de la onda T y la onda P, por consecuencia de una baja frecuencia cardíaca. De otro modo cuando la frecuencia cardíaca es muy alta se pueden tener puntos del latido siguiente y el anterior, considerándose información errónea en el análisis. Una forma de reducir el efecto de los inconvenientes mencionados consiste en ubicar el latido en una ventana de w puntos, ubicando el pico R en la posición r de la ventana, hacia la izquierda los puntos van hasta el inicio de la onda P y hacia la derecha los puntos llegan hasta el final de la onda T. Las posiciones de la ventana que puedan quedar sin valores se rellenan con el último valor del extremo correspondiente. los valores w y r son estimados de forma experimental, usando la base de datos de análisis. De esta manera se abarca un amplio espectro de valores de frecuencia cardíaca, se evita cortar la señal donde exista posible información relevante y se tiene una longitud fija de los latidos para realizar el cálculo de las medidas de representación. Antes de ubicar el latido en la nueva ventana se extrae la media y se normaliza con respecto a la amplitud.

Para este trabajo se propone el uso de la transformada wavelet discreta calculada a partir de bancos de filtros, de forma que se permita resaltar cualidades determinantes mediante la representación que ofrece de la señal (coeficientes wavelet), facilitando el reconocimiento de estados funcionales. En general, una señal se representa como una serie de *aproximaciones* (baja frecuencia) y *detalles* (alta frecuencia) en diferentes resoluciones. Debido a que existe una abundante variedad de funciones wavelet madre, se planteó la conformación del espacio de características mediante la colección de coeficientes wavelets correspondientes a diferentes familias, y cada familia, a distintos niveles de descomposición, con el objetivo de verificar el desempeño del algoritmo propuesto en la identificación de variables redun-

dantes, ya que la variación en cuanto a la representación que ofrecen las diferentes wavelets no debe ser notable. Así, el espacio de representación espectro-temporal $\mathbf{X}_{2-ECG}$ quedó determinado por las configuraciones wavelet expuestas en la Tabla 4.3. De forma que se

WT	Orden–N
Daubechies	2–10
Symlet	2–10
Biorthogonal	13, 15, 24, 26, 28, 55, 68
Reverse Biorthogonal	22, 24, 26, 28, 44, 55, 68

Tabla 4.3. Representación espectro-temporal de las señales ECG

tomaron los coeficientes del tercer nivel de aproximación y los 4 coeficientes con máximo valor absoluto de cada escala de aproximación, realizando la descomposición wavelet hasta el sexto nivel [33].

Señales fonocardiográficas

Los métodos basados en tiempo-frecuencia permiten la construcción del espectrograma de la señal FCG, con el cual se obtiene la distribución de energía tanto en el tiempo como en frecuencia. La forma de detectar soplos usando estos métodos es obtener la energía total en las bandas de tiempo y frecuencia donde la señal presenta actividad. En tiempo se limita el análisis a la sístole y diástole, y en la frecuencia se observa la energía de la señal entre 0 y 500 Hz, que es el ancho de banda donde se encuentran los soplos. De esta forma, el latido cuya energía en estas bandas supere determinado umbral, será clasificado como latido con evidencia de soplo. Los *espectrogramas* de la señal se calcularon con dos métodos: STFT y transformada Gabor. Para ambos casos la señal original con frecuencia de muestreo de 44100 Hz fue sub-muestreada a 4000 Hz. Se eligieron ventanas de 64 puntos en el tiempo y 100 puntos para el análisis en frecuencia.

Energía del espectrograma obtenido con la STFT Una forma comúnmente utilizada para caracterizar una señal en tiempo-frecuencia es mediante la transformada de Fourier en tiempo corto (STFT). En la Figura 4.7, se muestra el promedio de los espectrogramas para las 180 señales FCG normales y las 180 señales con soplo. La gran parte de las señales patológicas que se tienen evidencian soplos sistólicos y por eso es que se observa una concentración de energía mucho mayor en la sístole que en la diástole. La suma de energía en los sectores correspondientes al soplo se toma como una característica usada en la discriminación de las señales. Con el fin de identificar si los soplos son sistólicos o diastólicos se utiliza el máximo entre la energía de la sístole y la diástole del espectrograma de la señal FCG.

Energía del espectrograma obtenido con la transformada de Gabor Mediante la transformada de Gabor se puede modificar el ancho de la ventana en el tiempo con el fin de manejar la resolución tanto en tiempo como en la frecuencia, pues para una ventana gaussiana, entre más grande sea la desviación en el tiempo va a ser menor la desviación en el dominio de la frecuencia y viceversa. Variando el ancho de la ventana de forma inversa a la varianza de la señal, se aumenta la resolución en el tiempo para los segmentos de la señal en donde se presenta el soplo y es resaltada la energía en el domino de la frecuencia.

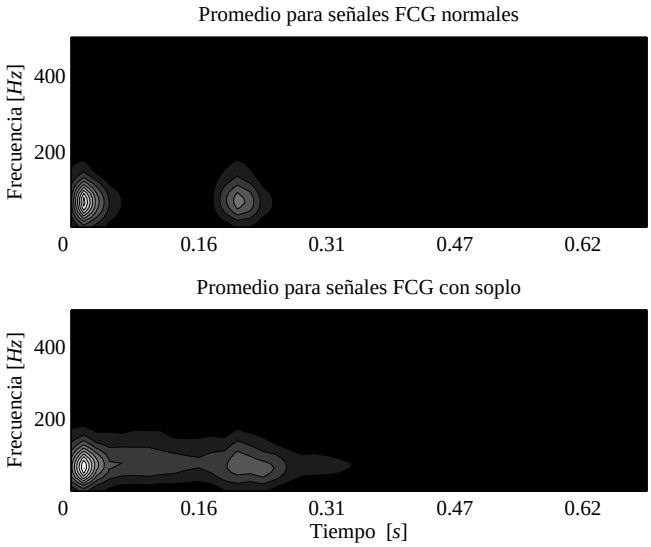


Figura 4.7. Promedio de los espectrogramas obtenidos usando STFT

En la Figura 4.8 se muestra el espectrograma obtenido usando la transformada Gabor con la ventana variable, y se observa una diferencia notoria entre las señales normales y las que evidencian soplo cardíaco. La característica que se propone usando este espectro se obtiene de forma similar a la obtenida con el espectro STFT (sumatoria de la energía en la sístole y en la diástole.

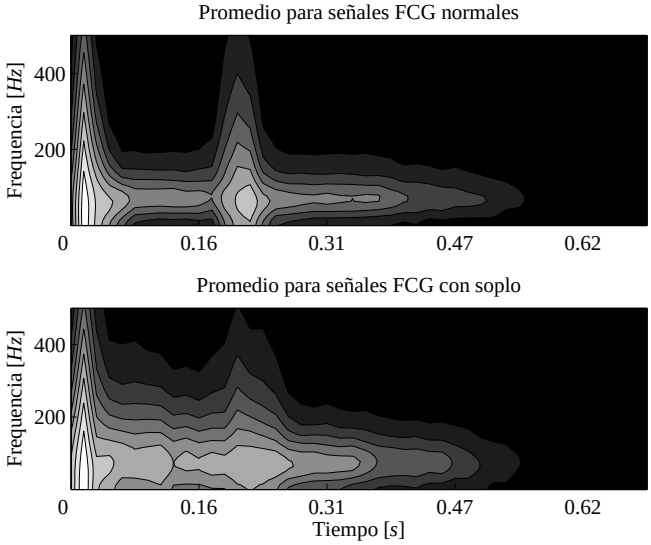


Figura 4.8. Promedio de los espectrogramas obtenidos usando transformada Gabor

Descomposición wavelet La descomposición wavelet (WT) es una herramienta útil para resaltar los soplos que no son muy notorios. En la Figura 4.9 se puede observar que el soplo de la señal patológica es más notorio cuando se toma la suma de los detalles correspondientes a los niveles 1 y 2 de la descomposición wavelet *db*4. De acuerdo a esto, se extraen 3 características: las dos primeras se obtienen estimando el valor máximo entre la energía de la sístole y la diástole usando el operador de Teager [18], sobre la suma de los detalles de nivel 1 y 2 de la descomposición WT correspondiente a *db*1 y *db*4; la otra característica se obtiene estimando la energía sumada del espectrograma obtenido con la STFT aplicada sobre la suma de los detalles 1 y 2 de la WT–*db*4.

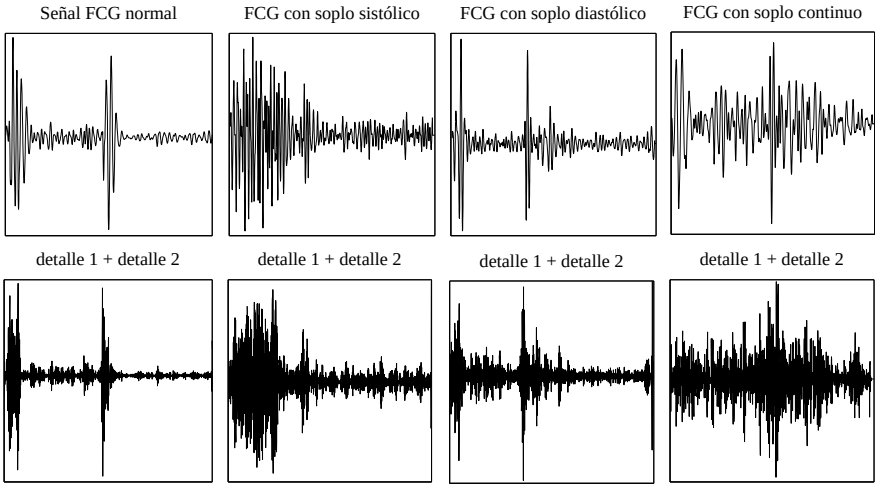


Figura 4.9. Suma de los detalles 1 y 2 de la descomposición WT–*db*4

Caracterización de los espectrogramas Los espectrogramas obtenidos con STFT, transformada de Gabor, transformada wavelet y distribución Wigner-Ville, son arreglos bidimensionales (i.e., dimensiones $m \times n$) que permiten la aplicación de procedimientos comúnmente usados en el procesamiento digital de imágenes. De acuerdo a esto, se propone el uso del análisis de componentes principales (PCA) sobre una matriz $\mathbf{X}_{t-f}$, compuesta en las columnas por los píxeles de cada espectrograma (i.e., cada columna queda de mn filas), de forma que para M señales normales y N señales patológicas, se obtiene una matriz $\mathbf{X}_{t-f}$ con dimensiones $(M+N) \times mn$. Luego de restar la media a cada una de las columnas de $\mathbf{X}_{t-f}$, se obtiene la proyección PCA usando la expresión (3.12). Las 20 componentes principales (i.e., las 20 primeras columnas de $\mathbf{Z}_{t-f}$ siguiendo la notación de (3.12)) se toman como características de los espectrogramas. Este tipo de representación permite descubrir relaciones que caracterizan la dinámica contenida en los espectrogramas, esto es, PCA revela la existencia de correlaciones que han permanecido ocultas en la matriz que representa la variabilidad de los eventos en locaciones específicas de tiempo-frecuencia.

4.4.3. Análisis de complejidad

En este trabajo, también se proponen los modelos de representación basados en técnicas de dinámica no lineal para caracterizar la dinámica cardíaca a partir de las señales ECG y FCG. El problema que inicialmente hay que resolver para este tipo de análisis es el ajuste óptimo de los parámetros de embebimiento (m y τ), los cuales están involucrados en la reconstrucción del espacio de estados que debe ser difeomórfico al que se reconstruiría si se tuvieran todas las variables de estado del sistema original, tal como se planteó en la Sección 2.3.2. Un segundo problema consiste en que cada registro ECG y FCG provienen de personas diferentes, y por lo tanto, de *sistemas cardíacos* diferentes. Aunque la dinámica cardíaca es similar para las diferentes personas, el corazón de cada persona debe tomarse como un sistema diferente, lo cual repercute en la reconstrucción del atractor, ya que el valor de los parámetros de embebimiento pueden cambiar de una persona a otra, sin embargo es importante anotar que la variabilidad de estos valores no debe ser alta. Para resolver el primer problema, se propone un procedimiento basado en algoritmos genéticos sobre un conjunto predeterminado de valores de m y τ (i.e., $m \in [2, 34]$ y $\tau \in [2, 128]$). Una forma para evaluar si el espacio de estados reconstruido realmente corresponde al del sistema de análisis, se propone en [2], la cual consiste en usar la capacidad que tiene el atractor reconstruido de predecir puntos de la señal original. De acuerdo a esto, para cada combinación de parejas (m, τ), la función de evaluación propuesta consiste en medir la capacidad de predicción que tiene el atractor reconstruido mediante una medida estadística conocida como Kullback-Leibler entre los histogramas de los puntos que fueron resultado de la predicción usando el atractor reconstruido y los puntos correspondientes de la señal original (ver Figura 4.10). Este procedimiento es tomado como de *bajo nivel* y se

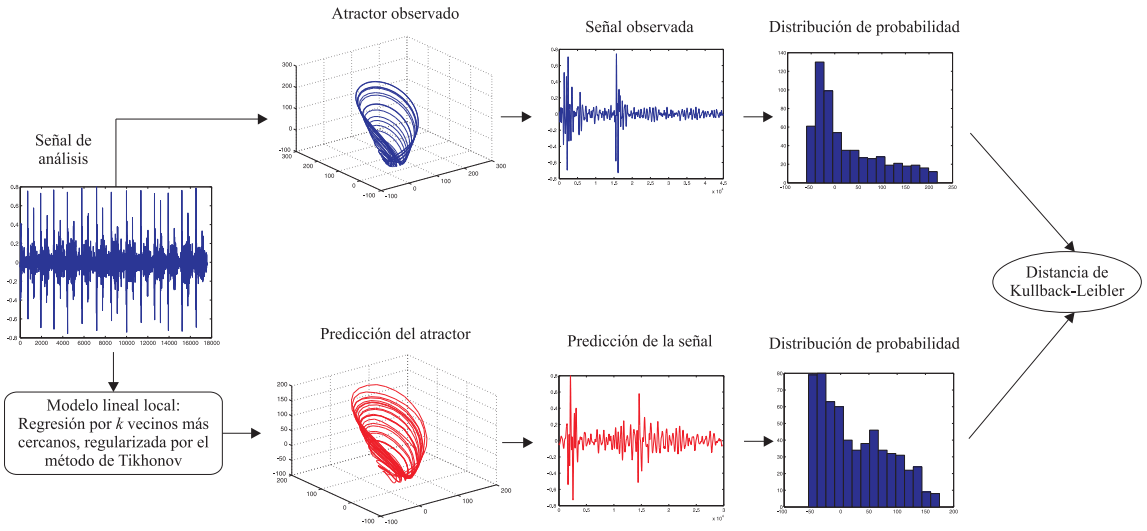


Figura 4.10. Función de evaluación

expone detalladamente en el Algoritmo 4.1. Aunque este procedimiento es una solución al primer problema planteado, incluye una alta complejidad computacional, lo cual es menos favorable al tomar en cuenta el segundo problema, donde se requiere la ejecución de este

Algoritmo 4.1 Algoritmo genético de bajo nivel

Require: $x(t)$, $m \in [2, 34]$ y $\tau \in [2, 128]$. $\star\backslash x(t)$ es la señal de análisis$\backslash\star$

Población inicial de cromosomas (arreglo de bits). $\star\backslash$Codificación de varios valores para m y τ (aleatoriamente)$\backslash\star$

repeat

1. Cálculo de $f_i\{$cromosoma$\}$ $\star\backslash$Siendo f_i, $i = 1 \ldots r$, la función de evaluación basada en la distancia de Kullback-Leibler (ver Figura 4.10)$\backslash\star$
2. Selección del mejor cromosoma evaluado usando muestro estocástico restante
3. Desarrollar la recombinación de un solo punto entre cromosomas apareados aleatoriamente $\star\backslash$Con probabilidad $\chi\backslash\star$
4. Se aplica el operador de mutación sobre cada bit de la población $\star\backslash$Con probabilidad $\mathcal{P}_m\backslash\star$

until El $95\,\%$ de los miembros de la población sean iguales [9]

Output: Valores de m y τ.

algoritmo para obtener los parámetros de embebimiento correspondientes al registro de cada persona. Para obtener una mayor eficiencia del Algoritmo 4.1, se requiere de un procedimiento alterno, conocido como *meta-algoritmo* ó algoritmo genético de *alto nivel*, que ajuste fuera de línea los parámetros de control del algoritmo genético de bajo nivel (i.e., tamaño de la población r, probabilidad de recombinación χ y probabilidad de mutación $\mathcal{P}_m$). El Algoritmo 4.1 es ejecutado varias veces con diferentes conjuntos de parámetros de control. Por cada ejecución se toma registro tanto de la evaluación correspondiente a la solución encontrada como del número de generaciones requerida para hallar dicha solución, con el fin de construir dos superficies de regresión usando la técnica de regresión por vectores de soporte (SVR). Así, sin necesidad de ejecutar el algoritmo de bajo nivel, la evaluación de la solución encontrada y el número de generaciones puede ser estimado usando las superficies de regresión. De esta forma, el algoritmo genético de alto nivel es optimizado con las superficies obtenidas, ya que con ellas se minimiza el número de generaciones y se maximiza el valor de evaluación en un tiempo de procesamiento más bajo (ver Figura 4.11).

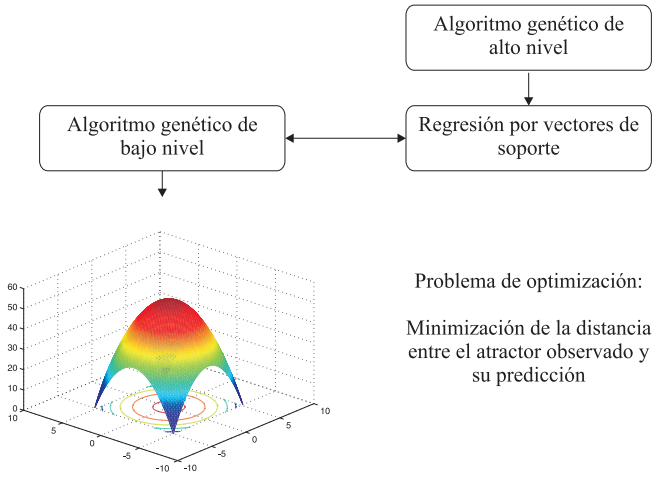


Figura 4.11. Metodología de optimización para los parámetros de control

Después de hallar los valores de m y τ para los registros ECG y FCG de cada sujeto, se estimaron las siguientes medidas de complejidad sobre los atractores reconstruidos para conformar el espacio de representación, de acuerdo al modelo expresado en (2.7):

Dimensión de correlación (D_2): La estimación de D_2 es realizada usando el método propuesto en [28], donde se requiere una estimación previa de la suma de correlación, $C(r)$. La función $lnC(r)$ contra $ln(r)$ es evaluada en cada latido ECG y FCG, estimando su región de escalamiento mediante la derivada, $dlnC(r)/dln(r)$.

Máximo exponente de Lyapunov (λ_1): Esta medida de complejidad se estima como la rata promedio de separación de los vecinos más cercanos, como se propone en [28].

Exponente de Hurst (H): La estimación de H se obtiene usando los procedimientos propuestos en [3] y [4].

4.5. Preproceso experimental de datos

4.5.1. Normalización

En muchas situaciones prácticas un diseño es confrontado con los valores que no expresan lo mismo en ciertos rangos dinámicos. Así las características con un gran número de valores pueden tener una gran influencia en el costo de la función de las características con valores pequeños, aunque esto no necesariamente es reflejado en el diseño del clasificador. Un procedimiento comúnmente usado es la *normalización estadística*, la cual para una variable $\boldsymbol{\xi}$ puede expresarse como:

$$\boldsymbol{\xi}' = \frac{\boldsymbol{\xi} - \mu_{\boldsymbol{\xi}}}{\sigma_{\boldsymbol{\xi}}}$$

donde $\mu_{\boldsymbol{\xi}}$ y $\sigma_{\boldsymbol{\xi}}$ son respectivamente la media y la desviación estándar de $\boldsymbol{\xi}$. Una forma alternativa de procedimiento es la *normalización geométrica*, la cual se obtiene mediante:

$$\boldsymbol{\xi}' = \frac{\boldsymbol{\xi}}{\text{máx}\{\text{abs}\{\boldsymbol{\xi}\}\}}$$

4.5.2. Remoción de valores atípicos

Muchas de las técnicas de análisis multivariado están basadas en modelos paramétricos, por lo tanto, existen claras restricciones en cuanto al tipo de distribución a la que dichas variables deben aproximarse. Por otra parte, es también importante asegurar la homogeneidad de la muestra, mediante el análisis de la posible presencia de valores atípicos debidos a errores de medida u otras causas de heterogeneidad.

Existen diferentes formas de comprobar la homogeneidad de la muestra, la cual refleja si los datos se separan mucho, o por el contrario se concentran alrededor de la media. Una forma de detectar dicha homogeneidad es analizar el coeficiente de kurtosis, para ello, sea $\mathbf{X}$ la matriz original de datos de dimensión $n \times p$, donde las filas corresponden a las observaciones y las columnas a las variables, además, se denota x_{ij} al elemento genérico de

esta matriz, que representa el valor de la variable escalar j sobre la observación i, es decir:

$$\mathbf{X} = \{x_{ij}\} \text{ donde } \begin{array}{ll} i = 1, \ldots, n & \text{es el conjunto de índices de las observaciones} \\ j = 1, \ldots, p & \text{es el conjunto de índices de las variables} \end{array}$$

además, la media de la variable escalar x_j es,

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^{n} x_{ij} \tag{4.9}$$

y su respectiva desviación,

$$s_j = \sqrt{\frac{\sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2}{n}} \tag{4.10}$$

así, el coeficiente de kurtosis se expresa como,

$$K_j = \frac{1}{n} \frac{\sum (x_{ij} - \bar{x}_j)^4}{s_j^4} \tag{4.11}$$

El coeficiente de kurtosis es siempre mayor o igual que uno y mide la relación entre la variabilidad de las desviaciones y la desviación media. Tomando en cuenta distribuciones gaussianas el valor de $K_j \to 3$ (Fig. 4.12). Por tal motivo, con el coeficiente de kurtosis es

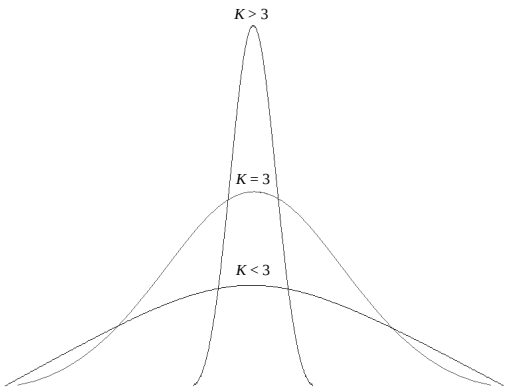


Figura 4.12. Distribuciones con kurtosis mayores y menores que 3 comparadas a la gaussiana

posible determinar qué variables presentan heterogeneidad, es decir, aquellos coeficientes alejados de $K_j = 3$, en particular y como regla empírica se pueden eliminar por falta de homogeneidad aquellas variables que tengan un coeficiente de kurtosis menor que 2 ($K_j < 2$), porque cuando el valor de kurtosis es bajo, tiende a uno o es menor que 2, es posible que la variable refleje la mezcla de dos o más poblaciones. Por otra parte, si hay unos pocos datos atípicos muy alejados del resto, la variabilidad de las desviaciones será grande, y por ende, los coeficientes de kurtosis arrojaran valores altos, usualmente mayores

que 7 u 8 ($K_j > 7$) [23], en este caso, no se deben eliminar directamente las variables, sino detectar las observaciones con dicho comportamiento atípico y eliminar estas observaciones. En general, no debe existir más de un 10 % de datos atípicos por cada variable [27], si esto no se cumple, como criterio empírico es posible eliminar la variable tratada.

Los datos atípicos (*outliers*) son las observaciones que parecen tener un comportamiento distinto, o haber sido generados de forma diferente al resto de los datos. Cuando existe más de un dato atípico en la muestra, es posible que se presenten efectos de enmascaramiento, en el cual observaciones atípicas similares se ocultan entre sí. Los datos atípicos puedes ser detectados bajo dos diferentes perspectivas, bien sea de manera univariada o multivariada. Una regla para la detección de valores atípicos de forma univariada está dada por señalar como datos atípicos aquellos que cumplan la siguiente condición,

$$\frac{|x_i - \operatorname{med}(x)|}{\operatorname{MEDA}(x)} > 4.5 \tag{4.12}$$

donde $\operatorname{med}(x)$ es la mediana de las observaciones, que es el valor que se encuentra en la posición central al ordenar las observaciones y $\operatorname{MEDA}(x)$ es la mediana de las desviaciones absolutas con respecto a la mediana, que es una medida robusta de dispersión.

Para la detección de datos atípicos de forma multivariada se propone en [24] el Algoritmo 4.2.

Algoritmo 4.2 – Detección multivariada de datos atípicos

1: Buscar p direcciones ortogonales de máxima kurtosis.
2: Buscar p direcciones ortogonales de mínima kurtosis.
3: Eliminar provisionalmente los datos extremos en estas direcciones.
4: Calcular la media y la matriz de covarianza con los datos no sospechosos.
5: Finalmente identificar los datos atípicos como aquellos que son extremos con la distancia de Mahalanobis (ec. (4.13)) calculada con las estimaciones no contaminadas, sobre toda la muestra.

$$d_i = \left[(\mathbf{x}_i - \bar{\mathbf{x}})^\top \mathbf{S}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}) \right]^{1/2} \tag{4.13}$$

4.5.3. Verificación de normalidad

El juicio sobre la estructura Gaussiana de los datos se puede realizar a partir del análisis de los histogramas y de la respectiva prueba de hipótesis.

La división de valores que toman las variables aleatorias en rangos de alguna vecindad genera diferentes clases de representación, de las cuales la más conocida corresponde al diagrama del número de observaciones en función de la localización de cada rango de la variable aleatoria, conocida como *histograma*. Sin embargo, la cantidad de intervalos k de agrupación dependen del volumen n de la muestra. Entre las recomendaciones para la selección del número de intervalos del histograma están las siguientes [25]:

1. $k = \lfloor 10 \lg n \rfloor$, $5 \leq k \leq 30$,
2. $k = \lfloor 5 \lg n \rfloor$, $6 \leq k \leq 20$,

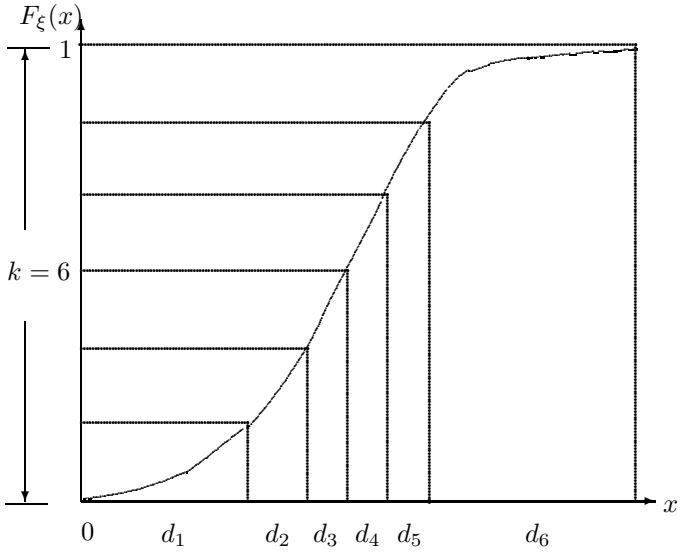


Figura 4.13. Intervalos de longitud equiprobable

3. $k = \begin{cases} \lfloor 3.26 \lg n + 0.25 \rfloor + 1, & n \leq 100 \\ \text{mín}\left(\lfloor 0.1n \rfloor, 0.25\right) + 1, & n > 100 \end{cases}$

4.

n	50	100	500	1000	10000
k	8	10	13	15	20

5. $k = \begin{cases} \lfloor 4 \left(0.75 \left(n-1\right)^2\right)^{1/5} \rfloor, & n > 200 \\ \lfloor 0.2n \rfloor, & n \leq 200 \end{cases}$

6. $k = \text{mín}\left(\lfloor \sqrt{n} \rfloor, 30\right)$

Asumiendo la igualdad en la longitud d de los intervalos, ésta se calcula de la relación:

$$d = 1.02\xi_{pp}/k$$

De tal manera, que los límites del intervalo k se definen como $x_{min} - (k-1)d - \Delta, x_{min} + (k-1)d - \Delta$, siendo $\Delta \simeq 0.01d$.

Cuando la distribución $F_\xi(x)$ es conocida o por lo menos, se puede asociar a una conocida, teniendo en cuenta la relación $P_i = F_\xi(x_i) - F_\xi(x_{i-1}) = i/k - (i-1)/p = 1/k = p$, la división es preferible realizarla sobre k intervalos equiprobables $d_i = (x_{i-1}, x_i)$, siendo $x_i = f^{-1}\left\{F_\xi(x)\right\}$ los puntos calculados de la función entre $[0, 1]$, como se muestra en la Figura 4.13.

Al superponer los histogramas o distribuciones obtenidos empíricamente con las respectivas curvas teóricas de densidad o distribución, se tener un elemento de juicio, por lo menos cualitativo, sobre su concordancia.

Prueba de Kolmogorov–Smirnov La estadística para este criterio es el máximo valor de desviación entre la distribución observada de (empírica) $F_n(x)$ de la distribución $\tilde{F}(x)$ pronosticada por la hipótesis H_0:

$$D_n = \max_x \left| F_n(x) - \tilde{F}(x) \right|$$

El cálculo de T se lleva a cabo de la siguiente manera:

- Construcción de la serie variacional $\{x_{(i)}\}$ a partir de la observación $\{x_i : i = 1, \ldots, n\}$.
- Cálculo de la función empírica de distribución
- Cálculo de D_n
- Los parámetros de la distribución normal, la media y la varianza se estiman de la observación. La estadística del criterio se asume $\tilde{T} = D_n$ para valores $n > 100$. Para valores $n < 100$ es conveniente realizar ajustes en la estadística [25].
- Cálculo del valor crítico $T_{1-\alpha}$ de acuerdo a la expresión:

$$T_{1-\alpha} \approx \sqrt{\frac{1}{2}\left(y - \frac{1}{18n}\left(2y^2 - 4y - 1\right)\right)} - \frac{1}{6\sqrt{n}}$$

  siendo $y = -\ln \alpha/2$
- Cálculo de p empleando la distribución de la estadística $\sqrt{n}T$.

4.5.4. Transformación de distribuciones

Si la prueba de verificación de la distribución da como resultado el rechazo de la hipótesis de normalidad, entonces se deben tomar las medidas para transformar la observación de tal manera que pueda cumplirse aceptarse la hipótesis sobre la normalidad de los datos. En este sentido el análisis de los histogramas puede ayudar a sugerir alguna forma de acomodación de los datos.

Inicialmente se puede partir de un conjunto de posibles transformaciones. En tareas asociadas al proceso de bioseñales, es común el empleo de la distribución logarítmica normal. Después, de realizar la transformación, se debe realizar de nuevo la prueba de verificación, y se toma aquella transformación que permita aceptar la hipótesis de normalidad, o bien, aquella que más se aproxime.

La familia de transformaciones utilizada para resolver los problemas relacionados a la no-normalidad es comúnmente conocida como *Box-Cox*, mediante la cual se transforma la variable x, cuyos valores muestrales se suponen positivos, en caso contrario se suma una cantidad fija k_x tal que $x + k_x > 0$, de la forma

$$y(\lambda) = \begin{cases} \dfrac{x^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \lg x, & \lambda = 0 \end{cases}$$

El parámetro de ajuste λ representa diferentes funciones de transformación, así por ejemplo, la siguiente malla de valores corresponde a las funciones de transformación:

λ	-1	-0.5	0	0.5	1
Transformación	$1/x$	$1/\sqrt{x}$	$\lg x$	$\sqrt{x}$	x

El valor λ se puede estimar por criterio de máxima verosimilitud, así, para diferentes valores de λ se realiza la transformación

$$U(\lambda) = \begin{cases} \dfrac{x^{\lambda} - 1}{\lambda \tilde{m}_{xg}^{(\lambda-1)}}, & \lambda \neq 0 \\ \tilde{m}_{xg} \lg x, & \lambda = 0 \end{cases}$$

siendo $\tilde{m}_{xg}$ la media geométrica de la variable x. Para cada λ, se obtiene el conjunto de valores $U_i(\lambda),\ i = 1, \ldots, n$. La función de verosimilitud es

$$\Lambda(\lambda) = \frac{n}{2} \ln \left(\sum_{i=1}^{n} \left(U_i(\lambda) - \tilde{U}(\lambda) \right) \right)$$

de la cual se obtiene el valor óptimo de λ_{op}.

4.6. Modelos desarrollados para reducir dimensiones

Después de extraer las medidas de representación, se obtienen tres ensambles de características representados por las siguientes matrices:

1. $\mathbf{X}_{ECG}$: Conjunto de características extraídas sobre la base de datos BD-UNAL-GCPDS-ECG.

2. $\mathbf{X}_{STT}$: Conjunto de características extraídas sobre la base de datos BD-MIT E-STT.

3. $\mathbf{X}_{FCG}$: Conjunto de características extraídas sobre la base de datos BD-UNAL-GCPDS-FCG.

Mientras no se especifique algo diferente, la clasificación se hace usando la regla de los vecinos más cercanos y la validación se obtiene mediante validación cruzada con 15 particiones balanceadas, usando el total de las observaciones para la etapa de procesamiento (entrenamiento y validación), aunque para algunos procedimientos se hizo necesario usar un 30 % adicional de observaciones (recientemente adquiridas y etiquetadas por los médicos especialistas) con el fin de verificar el desempeño del sistema simulando una prueba de concordancia, y de esta manera, analizar la capacidad de generalización.

4.6.1. Nivel 1

Para evaluar la reducción de dimensiones a este nivel de operación, se escogieron dos técnicas frecuentemente usadas en la literatura, debido a que de ellas se pueden derivar bajos ciertas restricciones las otras técnicas presentadas en la Sección 3.3.1. Las técnicas son: el análisis de componentes principales y el análisis de componentes independientes.

Análisis de componentes principales

Sobre los 3 conjuntos de características ($\mathbf{X}_{ECG}$, $\mathbf{X}_{STT}$ y $\mathbf{X}_{FCG}$) se obtienen resultados relacionados a la capacidad de esta técnica para reducir dimensiones siguiendo el procedimiento basado en las expresiones (3.6) y (3.12), mientras que también se evalúa el poder discriminante del espacio reducido para distinguir entre las clases.

Análisis de componentes independientes

El Análisis de Componentes Independientes (ICA) es una herramienta que busca separar fuentes estadísticamente independientes a partir de mezclas obtenidas mediante diferentes sensores [5]. En este sentido, se toman $\mathbf{X}_{ECG}$, $\mathbf{X}_{STT}$ y $\mathbf{X}_{FCG}$ como conjuntos de información que han sido resultado de mezclar diferentes fuentes independientes.

Procedimiento Existen diferentes versiones para desarrollar este algoritmo, sin embargo, para este trabajo se propone un procedimiento que funciona bajo las siguientes consideraciones [34]: para el conjunto inicial de entrenamiento $\mathbf{X}_{n\times p} = [\boldsymbol{l}_1 \ \boldsymbol{l}_2 \ \cdots \ \boldsymbol{l}_p]$, las fuentes son estadísticamente independientes, solo una de las fuentes puede tener una distribución gaussiana y las señales capturadas en los sensores $\{\boldsymbol{l}_j = l_j(\ell), \ \ j = 1, \ldots, p \ \ \wedge \ \ \ell = 1, \ldots, n\}$ son combinaciones lineales de las fuentes $s_j(\ell)$, de forma que para dos fuentes y dos sensores, el problema puede ser formulado de la siguiente manera:

$$l_1(\ell) = a_{11}s_1(\ell) + a_{12}s_2(\ell)$$

$$l_2(\ell) = a_{21}s_1(\ell) + a_{22}s_2(\ell)$$

con el objetivo de encontrar a_{11}, a_{12}, a_{21}, a_{22}, $s_1(\ell)$ y $s_2(\ell)$ para $l_1(\ell)$ y $l_2(\ell)$ conocidas. De forma general, para un número de sensores igual al de fuentes, se tiene que:

$$\mathbf{x} = \begin{bmatrix} l_1(\ell) \\ \vdots \\ l_p(\ell) \end{bmatrix}, \ \mathbf{A} = \begin{bmatrix} a_{11} & \cdots & a_{1p} \\ \vdots & \ddots & \vdots \\ a_{p1} & \cdots & a_{pp} \end{bmatrix} \quad \text{y} \quad \mathbf{s} = \begin{bmatrix} s_1(\ell) \\ \vdots \\ s_p(\ell) \end{bmatrix}$$

de tal manera, que las combinaciones lineales de las fuentes se pueden reescribir como $\mathbf{x} = \mathbf{A}\mathbf{s}$. Cuando $\mathbf{A}$ es encontrada, $\mathbf{s}$ puede ser calculada como $\mathbf{s} = \mathbf{A}^{-1}\mathbf{x}$. La solución encontrada cuenta con la limitación de que tanto los signos como el orden de las fuentes no pueden ser determinados [16].

Sea, $y_j(\ell)$, la estimación de un componente independiente ($\approx s_j(\ell)$) y $\mathbf{w}_j$ una fila de la matriz estimada de transformación inversa $\mathbf{W} \approx \mathbf{A}^{-1}$. Un componente independiente se estima de acuerdo a la expresión $y_j(\ell) = \mathbf{w}_j^{\top}\mathbf{x}$. Bajo ciertas condiciones, la distribución que resulta de la suma de variables aleatorias estadísticamente independientes, es mucho más cercana a la gaussiana que cualquiera de las distribuciones de las variables originales [17]. Este teorema resulta básico para los algoritmos de ICA. Como $l_j(\ell)$ es una suma ponderada de los componentes en $\mathbf{s}$ y los componentes son no gaussianos e independientes, el problema se reduce a encontrar $\mathbf{w}_j$ de tal manera que la no gaussividad de $y_j(\ell)$ sea maximizada [35]. Una manera de medir la no gaussividad, es por medio de la siguiente aproximación de

negentropía [15]:

$$J(y) \propto \left(\mathrm{E}\left\{G(y)\right\} - \mathrm{E}\left\{G(v)\right\}\right)^2$$

donde $\propto$ denota proporcionalidad; v es una variable gaussiana con media cero y varianza unitaria, de manera que el término $\mathrm{E}\left\{G(v)\right\}$ es una constante; $G(y)$ es una función no cuadrática, cuya elección depende del problema, y las comúnmente usadas son: $G_1(y) = \frac{y^4}{4}$, $G_2(y) = \log(\cosh(y))$ y $G_3(y) = -\exp\left(-\frac{1}{2}y^2\right)$ [26]. Finalmente, se obtiene $\mathbf{w}_j$ de forma que $J(\mathbf{w}_j^\top \mathbf{x})$ sea maximizada.

El procedimiento se puede extender a la técnica conocida como *kernel–ICA*, para la cual se propone el uso del kernel Gaussiano. El valor de γ se ajusta donde se obtienen los mejores resultados de clasificación, después de realizar pruebas con diferentes valores desde 0.1 hasta 9 usando los conjuntos de datos artificiales. El valor de κ se obtiene con un procedimiento experimental similar.

4.6.2. Nivel 2

La reducción de dimensiones a este nivel de operación es evaluada sobre los conjuntos $\mathbf{X}_{ECG}$, $\mathbf{X}_{STT}$ y $\mathbf{X}_{FCG}$, inicialmente con un método muy simple pero de alta complejidad computacional (i.e., búsqueda heurística wrapper), después de esto, son desarrollados 3 procedimientos que se acercan en su efectividad pero evidencian menos costo computacional.

Búsqueda heurística wrapper

Este procedimiento sigue el mismo planteamiento expuesto en el Algoritmo 3.2, donde la función de evaluación $f_{\hat{\boldsymbol{\xi}}_j}$ es el resultado de clasificación usando la estrategia de los vecinos más cercanos.

Búsqueda heurística con función de costo MANOVA

Para este procedimiento, de la misma manera que en el procedimiento anterior, se implementa el Algoritmo 3.2, con la diferencia que la función de evaluación $f_{\hat{\boldsymbol{\xi}}_j}$ se basa en la prueba de hipótesis relacionada al análisis de varianza multivariada (MANOVA), donde la hipótesis nula es descrita en la expresión (3.22) y se impone una confiabilidad del 95 %.

Modelo híbrido: Árboles de decisión, algoritmos genéticos y k-NN

Un sistema híbrido formado por árboles de decisión, algoritmos genéticos y la regla de los k vecinos más cercanos es propuesto para lograr una selección de características basada en pesos que son ajustados de acuerdo al nivel de relevancia de cada característica respecto al concepto de clasificación. La medida de relevancia propuesta para este modelo híbrido, está dada por la siguiente expresión:

$$\mu(\check{\boldsymbol{\zeta}}) = \left(\sum_{j=1}^{u} w_j \left(\xi_{(j,k_1)} - \xi_{(j,k_2)}\right)^2\right)^{1/2} \tag{4.14}$$

donde u es el cardinal del subconjunto de prueba $\check{\boldsymbol{\zeta}} \in \varphi(\boldsymbol{\xi})$, w_j es el peso asignado a la característica j, k_1 y k_2 son las clases 1 y 2 respectivamente. La Figura 4.14 expone

los parámetros de salida del modelo propuesto, donde q es el número más pequeño de características con que se maximiza la medida de relevancia μ, descrita en la expresión (4.14), tomando $u = q$. En la Figura 4.15, se presenta el diagrama de bloques del modelo

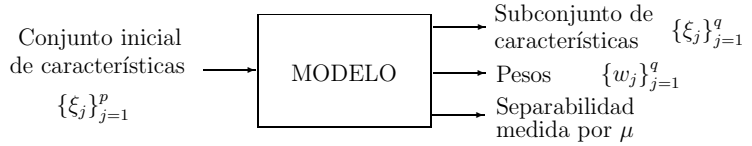


Figura 4.14. Reducción de dimensiones basada en el peso de las características

híbrido propuesto para reducir dimensiones al nivel 2 de operación, donde se puede observar el efecto de la relevancia débil (rechazar las características irrelevantes) y la relevancia fuerte (escoger el subconjunto minimal con mayor capacidad de representación de acuerdo a la medida de relevancia μ), en el sentido de la Observación 3.1, de tal manera que $q < g < p$. Los árboles de decisión aportan una primera reducción del espacio de representación

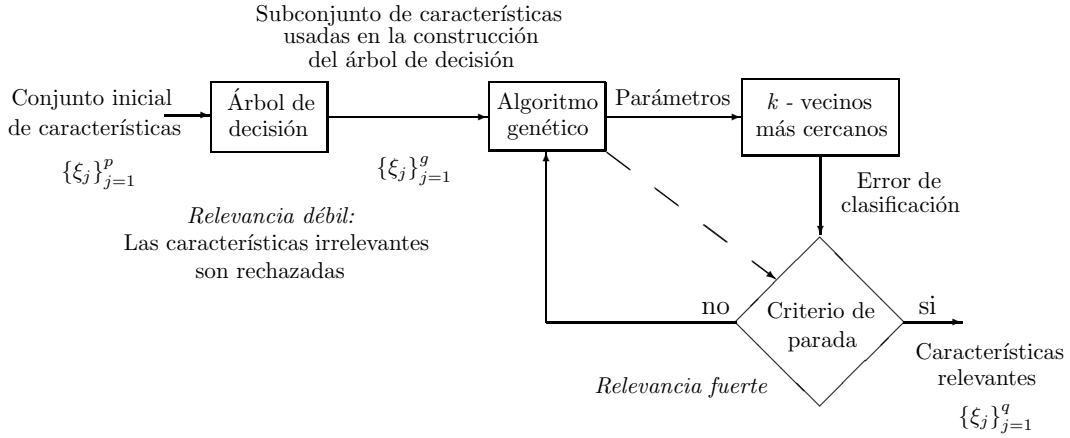


Figura 4.15. Modelo híbrido propuesto para la selección de características

mediante una medida μ basada en la ganancia de información, siguiendo el esquema de la Figura 3.6. Luego, los algoritmos genéticos encuentran un subconjunto de características de bajo orden y alto poder discriminante, mientras simultáneamente minimiza el error de clasificación de un conjunto de observaciones dado. Los parámetros del método son:

- $\{w_j\}_{j=1}^{p}$: Peso de las características (i.e., w_j es el peso de la características ξ_j).
- θ_w: Umbral de decisión. Si $w_j < \theta_w$ entonces ξ_j es rechazada.
- $\mathcal{P}_m$: Probabilidad de mutación.
- k: Número de vecinos más cercanos.

Asimismo, los parámetros de funcionamiento del algoritmo genético fueron tomados de la siguiente manera:

– Esquema de codificación:

$$\underbrace{0\ldots10}_{\substack{w_1\\20\,bits}}\underbrace{1\ldots00}_{\substack{w_2\\20\,bits}}\ldots\underbrace{1\ldots10}_{\substack{w_q\\20\,bits}}\underbrace{0\ldots10}_{\substack{\theta_w\\20\,bits}}\underbrace{0\ldots10}_{\substack{\mathcal{P}_m\\20\,bits}}\underbrace{0\ldots10}_{\substack{k\\4\,bits}}$$

– Tamaño de la población: 40

– Probabilidad de recombinación (χ): 0.9

– Generation gap: 0.95

– Número de generaciones: 300 con mutación auto-adaptativa.

El procedimiento relacionado al funcionamiento del algoritmo genético sigue la rutina presentada en el Algoritmo 3.3.

Variante del modelo híbrido usando la complejidad de Rademacher

Al modelo híbrido expuesto en la Figura 4.15, se le hace una variante relacionada con la medida de evaluación, de forma que el error de clasificación es perturbado por un conjunto de variables aleatorias del tipo Rademacher. A continuación se presenta la lista de consideraciones:

– $\mathbf{x}_i \in \mathbb{R}^g$: La i-ésima observación está asociada al conjunto de características $\{\xi_j\}_{j=1}^{g}$.

– y_i: Etiqueta de clase correspondiente a la observación $\mathbf{x}_i$.

– h: Clasificador entrenado usando el conjunto de observaciones $\{\mathbf{x}_i\}_{i=1}^{n}$ y etiquetas de clase y_i.

– $h(\mathbf{x}_i)$: Clase obtenida mediante el clasificador h, dada la observación $\mathbf{x}_i$.

– $\{\sigma_i\}_{i=1}^{n}$: Variables aleatorias Rademacher (i.i.d.) independientes de los datos. Una variable aleatoria Rademacher toma valores de +1 y -1, cada una con probabilidad de 1/2.

– Función indicadora: $I_{\{h(\mathbf{x}_i)\neq y_i\}} = \begin{cases} 1, & \text{cuando } h(\mathbf{x}_i) \neq y_i \\ 0, & \text{cuando } h(\mathbf{x}_i) = y_i \end{cases}$

– *Error de entrenamiento*: $\hat{e}(h) = \frac{1}{n}\sum_{i=1}^{n} I_{\{h(\mathbf{x}_i)\neq y_i\}}$

Para la estimación de la *complejidad de Rademacher*, se sigue el siguiente procedimiento:

1. Cálculo de la penalización de Rademacher
   – Nuevo conjunto de etiquetas de clase: $z_i = \sigma_i y_i$
   – El clasificador $h_{\mathcal{R}}$ es entrenado usando $\{\mathbf{x}_i, z_i\}_{i=1}^{n}$
   – La penalización Rademacher se obtiene mediante: $\mathcal{R}_m = \left| \frac{1}{n}\sum_{i=1}^{n} \sigma_i I_{\{h_{\mathcal{R}}(\mathbf{x}_i)\neq y_i\}} \right|$

2. La complejidad de Rademacher se estima como:

$$RC = \hat{e}(h) + \mathcal{R}_m$$

3. Estrategia de validación: validación cruzada – 15×2 *cv*.

De esta manera, la función de evaluación $f_{\check{\boldsymbol{\zeta}}}$ no se basa en el error de clasificación como en el caso anterior, sino en la estimación de la complejidad de Rademacher RC.

4.6.3. Nivel 3

Reducción de dimensiones mediante cargas agrupadas

Como el objetivo de este nivel de reducciones es visualizar estructuras embebidas en altas dimensiones con el fin de hallar interpretación fisiológica, se propone el análisis sobre la estructura multivariada $\mathbf{X}_{STT}$, debido a que ninguna característica de este ensamble tiene sentido fisiológico directo, y precisamente, esto es lo que se quiere resolver en este nivel de operación. El esquema presentado en la Figura 4.16 se diseñó siguiendo la restricción presentada en la Observación 3.2, de forma que tomando en cuenta la expresión (3.23), se puedan hallar los mapeos de reconstrucción para cada etapa en cascada de reducción de dimensiones. En el Algoritmo 4.3 se describe detalladamente este procedimiento, tomando

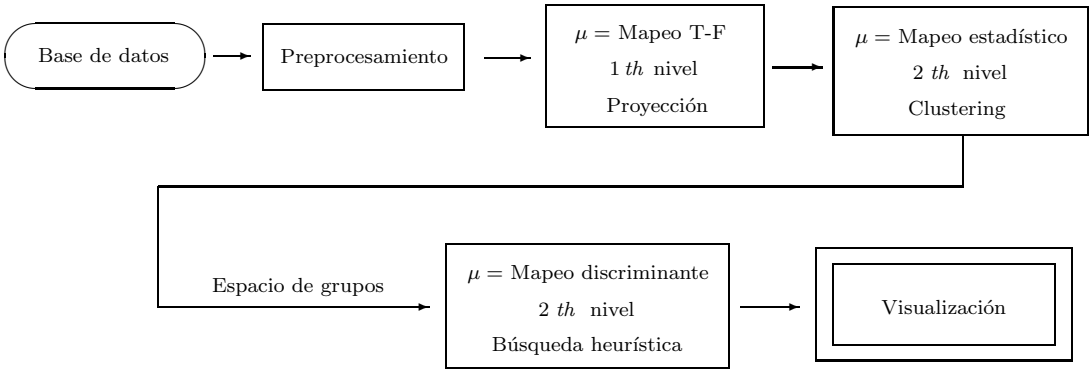


Figura 4.16. Esquema del procedimiento de cargas agrupadas

como espacio inicial de entrenamiento, la representación de n registros mediante la estimación de p características que pertenecen a $\boldsymbol{\xi}$ y que son ubicadas en una matriz $\mathbf{X}_{n\times p}$, donde las columnas contienen las características y las filas los registros u observaciones.

El procedimiento propuesto consiste de manera particular en las siguientes rutinas: para escoger un subconjunto de las variables originales, o características, del vector aleatorio $\boldsymbol{\xi}$, se parte de que

$$\boldsymbol{\Sigma}_{\boldsymbol{\xi}} = \mathbf{A}\boldsymbol{\Lambda}\mathbf{A}^T \tag{4.15}$$

donde $\boldsymbol{\Lambda}$ es una matriz diagonal que contiene los autovalores de $\boldsymbol{\Sigma}_{\boldsymbol{\xi}}$, ordenados en orden descendente (i.e., $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p$). Después de construir $\mathbf{A}_q$, ubicando por columnas los q autovectores asociados a los q mayores autovalores, se toma a $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p \in \mathbb{R}^q$ como las filas de la matriz $\mathbf{A}_q$. Cada vector $\mathbf{v}_i$ representa la proyección de la i-ésima

característica (variable) del espacio inicial $\boldsymbol{\xi}$ al espacio de menor dimensión $\boldsymbol{\zeta}$, esto es, los q elementos de $\mathbf{v}_i$ corresponden a los pesos (cargas) de la i-ésima característica en cada eje del subespacio. Una observación importante es que las características que son altamente correlacionadas, o tienen alta información mutua, tendrán vectores de cargas $\mathbf{v}_i$ con valores absolutos similares (el cambio de signo en las cargas no tiene significancia estadística [27]). Tomando los dos casos extremos, para dos variables independientes los vectores de cargas tienen máxima separación, mientras que dos variables totalmente correlacionadas tienen vectores de cargas idénticos (excepto por algún cambio de signo). Para encontrar la mejor representación reducida, se usa la estructura de las filas de $\mathbf{A}_q$ con el objetivo de primero encontrar las m agrupaciones de características que están altamente correlacionadas usando el Algoritmo 3.4 (los grupos se denotan por $\{\boldsymbol{\nu}_j : j = 1, \ldots, m\}$), y luego, con la media estimada a partir de los valores que corresponden a las características de cada grupo se construyen componentes. El conjunto de estos componentes constituyen el espacio $\hat{\boldsymbol{\xi}} \in \mathbb{R}^m$. Luego, usando el algoritmo de búsqueda heurística SFFS (ver Algoritmo 3.2), se hace una reducción supervisada para encontrar el número mínimo de componentes (≥ 3 para poder lograr la visualización), de forma que este espacio de representación $\boldsymbol{\varsigma}$ permita de manera suficiente la clasificación de patrones.

Algoritmo 4.3 Visualización mediante cargas agrupadas

Require: $\mathbf{X}_{n\times p}$.

1. Estimar la matriz de covarianza $\boldsymbol{\Sigma}_{\mathbf{X}}$.
2. Hallar los autovalores $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p$ de $\boldsymbol{\Sigma}_{\mathbf{X}}$.
3. Se calcula q $\star$\número de autovalores mediante algún criterio\$\star$
4. $\mathbf{A}_q \leftarrow q$ autovectores asociados a los q mayores autovalores por columnas.
5. Se toman las filas de $\mathbf{A}_q$ como $\mathbf{v}_i$, con $i = 1, \ldots, p$.
6. $\boldsymbol{\nu}_j = \text{ISODATA}\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ con $j = 1, \ldots, m$. $\star$\agrupación de características en m clusters\$\star$
7. $\hat{\boldsymbol{\xi}} = [mean\{\boldsymbol{\nu}_1\}, mean\{\boldsymbol{\nu}_2\}, \ldots, mean\{\boldsymbol{\nu}_m\}]$ $\star$\cada columna de esta matriz corresponde a la media de las agrupaciones\$\star$
8. $\boldsymbol{\varsigma} = \text{SFFS}\{\hat{\boldsymbol{\xi}}_1, \hat{\boldsymbol{\xi}}_2, \ldots, \hat{\boldsymbol{\xi}}_m\}$, usando como función de evaluación un clasificador k-nn.
9. Validación: clasificador k-nn (50 % para entrenamiento y 50 % para validación).

$\boldsymbol{\varsigma} = \{$ conjunto efectivo de características$\}$

Reducción de dimensiones para el reconocimiento de múltiples formas

Sea, $\mathbf{X} = [\boldsymbol{l}_1, \boldsymbol{l}_2, \ldots, \boldsymbol{l}_n]$, una matriz de dimensiones $m \times n$, donde cada columna de $\mathbf{X}$ se toma como un vector $\boldsymbol{l}_s$ de m elementos con media cero. La *descomposición en valores singulares* (SVD) de $\mathbf{X}$ da como resultado el producto de tres matrices:

$$\mathbf{X} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T = \sum_{j=1}^{p} \sigma_j \boldsymbol{u}_j \boldsymbol{v}_j^T \tag{4.16}$$

donde $p = \min(m, n)$, $\mathbf{U} = [\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_m]$ es una matriz ortonormal de dimensiones $m \times m$, $\mathbf{V} = [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n]$ es una matriz ortonormal de dimensiones $n \times n$ y $\boldsymbol{\Sigma}$ es una matriz de dimensiones $m \times n$ con valores diferentes de cero en la diagonal (notados como σ_j) e iguales a cero en las demás posiciones. Supóngase que se quieren agrupar vectores $\boldsymbol{l}_s$ usando un criterio geométrico basado en la distancia euclidiana. Con la finalidad de disminuir la dimensión de los vectores $\boldsymbol{l}_s$, conservando estrictamente la distancia entre cada uno, se considera la Propiedad 4.1 relacionada con SVD, expuesta y demostrada en [20].

Propiedad 4.1 (SVD - Distancia entre variables)

Si $\mathbf{X}$ es una matriz de dimensiones $m \times n$ y rango r, con una descomposición en valores singulares $\mathbf{X} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T$, entonces la distancia Euclidiana entre cualquier par de columnas de la matriz $\mathbf{X}$ es igual a la distancia Euclidiana ponderada entre las correspondientes columnas de $\mathbf{V}^T$, y esa ponderación se obtiene mediante los valores singulares σ_j. Es decir,

$$\|\boldsymbol{l}_s - \boldsymbol{l}_t\|^2 = \sum_{j=1}^{r} \sigma_j^2 \left(v_{sj} - v_{tj}\right)^2$$

El esquema presentado en la Figura 4.17 se diseñó siguiendo la restricción presentada en la Observación 3.2, de forma que tomando en cuenta la expresión (3.23), se puedan hallar los mapeos de reconstrucción para cada etapa en cascada de reducción de dimensiones.

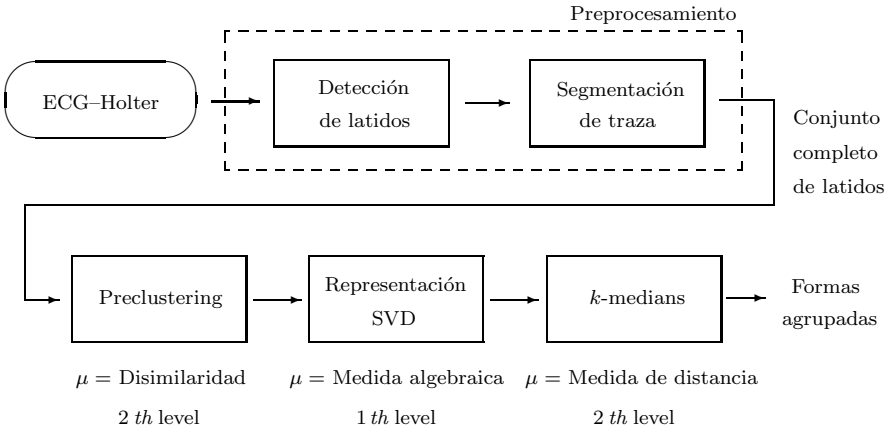


Figura 4.17. Esquema del procedimiento para formas agrupadas

En el Algoritmo 4.4 se expone el procedimiento diseñado para la representación reducida de registros de larga duración orientada a la visualización de un conjunto reducido de latidos cardíacos con alto contenido de información clínica, tomando como espacio inicial de entrenamiento, n latidos de un registro ECG-holter con m puntos cada uno, ubicados en las columnas de la matriz $\mathbf{X}_{m \times n}$, para obtener finalmente agrupaciones de latidos que morfológicamente se asemejan, o en general, grupos de variables con comportamiento similar. Es importante resaltar que el algoritmo trabaja sobre los latidos que han sido reducidos previamente usando SVD, y el procedimiento conserva a la salida el orden de los latidos del espacio de entrada, por lo cual, finalmente se pueden visualizar los latidos completos por agrupaciones, además de asociarse cada latido al índice correspondiente de la entrada

para poder realizar posteriormente la validación.

Algoritmo 4.4 Reducción de dimensiones para la visualización de formas agrupadas

Require: $\mathbf{X}_{m\times n}$.
$k = 0$ y $acumulado = 0$.

1. $\hat{\mathbf{X}} = \text{preclustering}\{\mathbf{X}\}$ $\star$\La matriz $\hat{\mathbf{X}}$ es de dimensiones $m \times p$\$\star$
2. $\text{SVD}\{\hat{\mathbf{X}}\} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T$ $\star$\descomposición en valores singulares\$\star$
3. $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_p$ $\star$\se ordenan los valores singulares\$\star$
4. $\delta = \alpha \left(\sum_{j=1}^{\text{rank}(\hat{\mathbf{X}})} \sigma_j^2 \right)$ $\star$\$\alpha$ se calcula de forma experimental\$\star$
5. **while** $acumulado \leq \delta$ **do**
   $k = k + 1$
   $acumulado = acumulado + \sigma_k^2$
6. **end while** $\star$\en k queda el número de agrupaciones que se requieren\$\star$
7. Se genera un espacio de representación reducida $\mathbf{Z}$ con columnas $[\sigma_h v_{jh}]^T$ con $h = 1, \ldots, k$ y $j = 1, \ldots, p$.
8. $\hat{\boldsymbol{\nu}}_i = \text{kmedians}\{\mathbf{Z}\}$ con $i = 1, \ldots, k$.

Output: $\boldsymbol{\nu}_i = \{$ variables agrupadas $\}$ con $i = 1, \ldots, k$.

Para reducir la complejidad computacional se propone una etapa de pre-clustering para disminuir los elementos del conjunto de análisis, de cientos de miles de latidos a solo un par de miles, sin perder información relevante (latidos poco repetidos) del registro [6]. Sea P el conjunto de l latidos del registro, el objetivo es encontrar un conjunto $R \in P$, con r latidos, donde $r << l$, de manera que todos los tipos de latidos presentes en P se hallen representados en R, con lo cual se eliminan aquellos redundantes. Para hacer la comparación entre los latidos se propone el uso de la técnica DTW (*Dynamic Time Warping*) como medida de disimilitud, la cual permite hallar una función de alineamiento óptimo entre dos secuencias de distinta longitud [6]. Inicialmente a R se le asigna el primer elemento de P y luego, se van tomando latidos de forma que cada uno se compara con un subconjunto de aquellos ya presentes en R, comenzando por el más cercano temporalmente, así, si la disimilitud es menor de un cierto umbral, se omite el latido al encontrarse ya uno muy similar en el conjunto final R y se continua al latido siguiente, repitiendo el proceso hasta completar todos los latidos [6]. En la etapa de clustering, se propone una modificación del algoritmo k-means, la cual consiste usar como medida la DTW en lugar de alguna medida de distancia estándar para la construcción de cada cluster $\boldsymbol{\nu}_i$, a partir del conjunto de secuencias con diferente longitud. La mediana es usada para re-calcular los centroides m_i, ya que el algoritmo k-means requiere estimar el conjunto de medias para obtener el nuevo centroide en cada iteración, lo cual no es posible en un espacio no-euclidiano. Así,

$$m_i = \underset{r_j \in \boldsymbol{\nu}_i}{\text{argmin}} \left(\sum_{r' \in \boldsymbol{\nu}_i} d(r_j, r') \right)$$

y la función criterio es ahora:

$$J = \sum_{i=1}^{k} \sum_{r_j \in \boldsymbol{\nu}_i} d(r_j, m_i)$$

Con estos cambios, el algoritmo de clustering k-means queda convertido en el algoritmo k-medians [6].

Para este caso, el algoritmo es ajustado para el reconocimiento de formas relacionadas a la isquemia sobre la derivación $V4$ de la base de datos BD-MIT E-STT. Así, la validación del algoritmo se realiza usando las anotaciones que ofrece la base de datos BD-MIT E-STT, para comparar tanto el número como la clase de grupos encontrado en la base de datos con el número y la clase de los grupos que encuentra el algoritmo propuesto en este trabajo (este procedimiento se realiza latido a latido). Las agrupaciones relacionadas a las formas de onda de la señal ECG con evidencias isquémicas se han etiquetado de la siguiente forma: $A =$ Normal, $B = ST$ disminuido, $C = ST - T$ disminuido, $D = ST - T$ aumentado, $E = T$ aumentada y $F = T$ invertida. El análisis se hizo sobre 59377 latidos correspondientes a 12 registros ECG-holter tomando la derivación $V4$.

BIBLIOGRAFÍA

[1] C. Ahlstrom, O. Liljefeldt, P. Hult, and P. Ask. Heart sound cancellation from lung sound recordings using recurrence time statistics and nonlinear prediction. *IEEE Signal Processing Letters*, 12(12):812–815, 2005.

[2] V. Babovic, M. Keijzer, and M. Stefansson. Chaos theory, optimal embedding and evolutionary algorithms. Technical Report 9800463, Danish Technical Research Council (STVF), 2001.

[3] J. B. Bassingthwaighte and G. M. Raymond. Evaluating rescaled range analysis for time series. *Annals of Biomedical Engineering*, 22(4):432–444, 1994.

[4] J. B. Bassingthwaighte and G. M. Raymond. Evaluation of the dispersional analysis method for fractal time series. *Annals of Biomedical Engineering*, 23(4):491–505, 1995.

[5] A. Cichocki. Blind signal processing methods for analyzing multichannel brain signals. *International Journal of Bioelectromagnetism*, 6(1), 2004.

[6] D. Cuesta-Frau, J. C. Pérez-Cortés, and G. A. García. Clustering of electrocariograph signals in computer-aided holter analysis. *Elsevier Computer Methods and Programs in Biomedicine*, (72):179–196, 2003.

[7] D. Cuesta-Frau, D. Novak, V. Eck, J. C. Pérez-Cortés, G. Andreu-García, C. Sastre, and M. Llorca. Reducción del ruido en señales electrocardiográficas mediante la transformada wavelet. In *Congreso Anual de la Sociedad Española de Ingeniería Biomédica*, pages 103–106, 2000.

[8] I. Daskalov. Developments in ECG acquisition, preprocessing, parameter measurement and recording. *EMB Magazine*, 17(2):50–58, 1998.

[9] K. A. DeJong. *An analysis of the Behavior of a class of Genetic Adaptative System*. PhD thesis, University of Michigan, 1975.

[10] P. S. R. Diniz. *Adaptive Filtering. Algorithms and Practical Implementation*. Kluwer Academic Publishers, 2 edition, 2002.

[11] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. John Wiley Sons, Inc., 2 edition, 2001.

[12] G. M. Friesen, T. C. Janett, M. A. Jadallah, S. L. Yates, S. R. Quint, and H. T. Nagle. A comparison of the noise sensitivity of nine QRS detection algorithms. *IEEE Transactions on Biomedical Engineering*, 37(1):85–98, 1990.

[13] J. García, G. Wagner, L. Sörnmo, P. Lander, and P. Laguna. Identification of the occluded artery in patients with myocardial ischemia induced by prolonged percutaneous transluminal coronary angioplasty using traditional vs transformed ECG-based indexes. *Computers and Biomedical Research*, 32(5):470–482, 1999.

[14] P. S. Hamilton and W. J. Tompkins. Quantitative investigation of QRS detection rules using the MIT/BIH arrhythmia database. *IEEE Transactions on Biomedical Engineering*, BME-33(12):1157–1165, 1986.

[15] P. Hubka, V. Rosik, J. Zdinak, M. Tysler, and I. Hulín. Independent component analysis of electrogastrographic signals. *Measurement Science Review*, 5(2):21–24, 2005.

[16] A. Hyvarinen and E. Oja. A fast fixed-point algorithm for independent component analysis. *Neural Computation*, 9(7):1483–1492, 1997.

[17] A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. Wiley series on adaptive and learning systems for signal processing, comunications and control. John Wiley & Sons, Inc, New York, 2001.

[18] J. F. Kaiser. On a simple algorithm to calculate the "energy" of a signal. In *International Conference on Acoustics, Speech, and Signal Processing*, pages 381–384, 1990.

[19] D. Kumar, P. Carvalho, M. Antunes, P. Gil, J. Henriques, and L. Eugénio. A new algorithm for detection of S1 and S2 heart sounds. *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2:1180–1183, 2006.

[20] S. Lee and M. H. Hayes. Properties of the singular value decomposition for efficient data clustering. *IEEE Signal processing letters*, 11(11):862–866, 2004.

[21] G. Melli. Data set generator. http://www. datasetgenerator.com/.

[22] R. Orozco, M. Pérez, J. Lorenzo, R. Grau, and R. Ramos. Evaluation of QRS morphological classifiers in the presence of noise. *Computers and Biomedical Research*, 30(3):200–210, 1997.

[23] D. Peña. *Análisis de datos multivariantes*. Mc Graw Hill, 2002.

[24] D. Peña and F. J. Prieto. Multivariate outlier detection and robust covariance matrix estimation. *Technometrics*, 43(3):286–310, 2001.

[25] M. L. Petrovich and M. I. Davidovich. *Estimación Estadística y Prueba de Hipótesis en el Computador (Rus)*. Finansy i Statistika, Moscú, 1989.

[26] P. A. Regalia and E. Kofidis. Monotonic convergence of fixed-point algorithms for ica. *IEEE Transactions on Neural Networks*, 14(4):943–949, 2003.

[27] A. C. Rencher. *Methods of Multivariate Analysis*. Wiley Series in Probability and Statistics, New York, 2 edition, 2002.

[28] M. Rosenstein, J. Collins, and C. D. Luca. A practical method for calculating largest lyapunov exponents from small data sets. *Physica D*, 65(1-2):117–134, 1993.

[29] J. S. Sahambi, S. Tandon, and R. Bhatt. Using wavelet transform for ECG characterization. *IEEE Engineering in Medicine and Biology*, 16(1):77–83, 1997.

[30] H. C. So. Adaptive algorithm for sinusoidal interference cancellation. *Electronics Letters*, 33(22):1910–1912, 1997.

[31] A. Taddei, A. Biagini, G. Distante, C. Marchesi, M. G. Mazzei, P. Pisani, P. Roggero, and C. Zeelenberg. An annotated database aimed at performance evaluation of algorithms for ST-T change analysis. In *Computers in Cardiology*, volume 16, pages 117–120, 1989.

[32] A. Taddei, G. Distante, M. Emdin, P. Pisani, G. B. Moody, C. Zeelenberg, and C. Marchesi. The european ST-T database: standard for evaluating systems for the analysis of ST-T changes in ambulatory electrocardiography. *European Heart Journal*, 13:1164–1172, 1992.

[33] B. T. Tan, M. Fu, A. Spray, and P. Dermody. The use of wavelet transforms in phoneme recognition. In *Fourth International Conference on Spoken Language*, volume 4, pages 2431–2434, 1996.

[34] R. Vigário, J. Särelä, V. Jousmäki, M. Hämäläinen, and E. Oja. Independent component approach to the analysis of eeg and meg recordings. *IEEE Transactions on Biomedical Engineering*, 47(5):589–593, 2000.

[35] Z. Wei, L. Jin, and Y. Jin. Independent component analysis. Technical report, Department of Statistics: Stanford University, 2005.

[36] B. Widrow, J. R. Glover, J. M. McCool, J. Kaunitz, C. S. Williams, R. H. Hearn, J. R. Zeidler, E. Dong, and R. C. Goodlin. Adaptive noise cancelling: Principles and applications. *Proceedings of the IEEE*, 63(12):1692–1716, 1975.

[37] Y. Zigel, A. Cohen, and A. Katz. The weighted diagnostic distortion (WDD) measure for ECG signal compression. *IEEE Transactions on Biomedical Engineering*, 47(11):1422–1430, 2000.

CAPÍTULO 5

Resultados y discusión

Bona pars bene dicendi est acite mentiri.

(Desiderio Erasmo de Rotterdam)

En este capítulo se presentan los resultados obtenidos con los procedimientos metodológicos propuestos en el capítulo anterior, donde se pueden apreciar las bondades de los diferentes niveles de reducción de dimensiones. Además, se valida el método propuesto mediante resultados de precisión de clasificación, capacidad de generalización y facilidad de interpretación fisiológica y clínica.

5.1. Etapa de preprocesamiento

5.1.1. Filtración

Señales ECG

A continuación se muestran las tablas con los resultados de las pruebas de distorsión realizadas a los dos filtros empleados para la reducción de la interferencia de la red: cancelador de interferencia sinusoidal adaptativo (ASIC – *adaptive sinusoidal interference canceller*) y de estimación incremental (IE – *incremental estimation*). Las pruebas se realizaron sobre el conjunto de registros seleccionados de las doce derivaciones de la base de datos BD-UNAL-GCPDS-ECG, para intervalos de señal con longitud de 300 s. Las pruebas se realizaron para tres valores distintos de SNR de entrada (-6 dB, -3 dB y 0 dB). En las Tablas 5.1, 5.2 y 5.3 se muestran la media y la desviación estándar de los valores obtenidos para el filtro IE, el cual mostró mejor desempeño contra los resultados del filtro ASIC, expuestos en las Tablas 5.4, 5.5 y 5.6. Sin embargo, debido a que el algoritmo IE permite la identificación específica de una frecuencia $\omega = 2\pi f_0$, que para este caso la frecuencia de la interferencia f_0 es la frecuencia de la red eléctrica ($f_0 = 60$ Hz), y únicamente estima la amplitud de la perturbación con una desviación de la frecuencia $\Delta\omega$, el algoritmo IE no soporta armónicos de la frecuencia fundamental, ni variaciones de fase de la interferencia. Con el filtro ASIMC, se pueden eliminar M interferencias, donde $M-1$ es el número de

Derivación	PRD		SNR		R_{xy}	
	μ	σ^2	μ	σ^2	μ	σ^2
DI	15.555	3.0748	37.578	3.7673	0.9876	0.0049
DII	14.546	4.4757	39.485	6.0819	0.9886	0.0067
DIII	16.999	5.3174	36.565	6.9462	0.9846	0.0083
aVR	14.059	3.0759	39.688	4.1905	0.9898	0.0045
aVL	16.536	6.9301	37.426	7.2510	0.9845	0.0139
aVF	16.204	4.8020	37.394	6.5235	0.9860	0.0072
V1	19.631	5.1705	33.223	5.1294	0.9801	0.0106
V2	15.799	2.7539	37.206	3.4689	0.9874	0.0043
V3	15.440	2.7895	37.689	3.6056	0.9879	0.0042
V4	19.631	5.1705	33.223	5.1294	0.9801	0.0106
V5	15.799	2.7539	37.206	3.4689	0.9874	0.0043
V6	15.440	2.7895	37.689	3.6056	0.9879	0.0042

Tabla 5.1. Distorsión filtro IE $SNR_i = -6dB$

Derivación	PRD		SNR		R_{xy}	
	μ	σ^2	μ	σ^2	μ	σ^2
DI	15.902	4.1685	37.429	5.0773	0.9868	0.0069
DII	15.813	5.1568	37.936	6.4686	0.9865	0.0083
DIII	23.034	17.3611	37.679	8.4247	0.9848	0.0097
aVR	14.974	3.7892	38.598	4.9764	0.9883	0.0059
aVL	17.164	7.8731	36.989	8.0208	0.9830	0.0163
aVF	20.360	19.1545	37.716	7.6804	0.9855	0.0090
V1	21.240	5.7314	31.687	5.3038	0.9768	0.0125
V2	24.658	21.8770	36.650	5.6235	0.9856	0.0064
V3	16.712	3.3688	36.183	3.9940	0.9857	0.0055
V4	21.240	5.7314	31.687	5.3038	0.9768	0.0125
V5	24.658	21.8770	36.650	5.6235	0.9856	0.0064
V6	16.712	3.3688	36.183	3.9940	0.9857	0.0055

Tabla 5.2. Distorsión filtro IE $SNR_i = -3dB$

Derivación	PRD		SNR		R_{xy}	
	μ	σ^2	μ	σ^2	μ	σ^2
DI	17.070	4.4336	35.994	4.9952	0.9848	0.0078
DII	17.355	5.2358	35.915	5.9335	0.9840	0.0092
DIII	25.621	20.2412	35.674	7.8589	0.9822	0.0108
aVR	16.431	3.9203	36.666	4.6528	0.9860	0.0066
aVL	18.733	8.1733	35.093	7.7088	0.9802	0.0180
aVF	25.824	21.7061	35.723	7.1255	0.9829	0.0100
V1	22.756	6.0038	30.271	5.1415	0.9736	0.0140
V2	18.007	4.4423	34.975	5.4204	0.9832	0.0073
V3	18.053	3.5550	34.618	3.8970	0.9835	0.0063
V4	22.756	6.0038	30.271	5.1415	0.9736	0.0140
V5	18.007	4.4423	34.775	5.6404	0.9832	0.0073
V6	18.053	3.5550	34.618	3.8970	0.9835	0.0063

Tabla 5.3. Distorsión filtro IE $SNR_i = 0dB$

Derivación	PRD		SNR		R_{xy}	
	μ	σ^2	μ	σ^2	μ	σ^2
DI	77.886	0.0292	4.998	0.0074	0.7925	0.0002
DII	77.888	0.0155	4.998	0.0060	0.9840	0.0001
DIII	77.897	0.0262	4.996	0.0066	0.9822	0.0002
aVR	77.910	0.0153	4.992	0.0040	0.9860	0.0001
aVL	77.886	0.0334	4.999	0.0083	0.9802	0.0002
aVF	77.892	0.0205	4.997	0.0064	0.9829	0.0001
V1	77.912	0.0704	4.991	0.0181	0.9736	0.0005
V2	77.903	0.0518	4.994	0.0135	0.9832	0.0003
V3	77.908	0.0468	4.993	0.0126	0.9835	0.0004
V4	77.912	0.0704	4.991	0.0181	0.9736	0.0005
V5	77.903	0.0518	4.994	0.0135	0.9832	0.0003
V6	77.908	0.0468	4.993	0.0126	0.9835	0.0004

Tabla 5.4. Distorsión filtro ASIC $SNR_i = -6dB$

Derivación	PRD		SNR		R_{xy}	
	μ	σ^2	μ	σ^2	μ	σ^2
DI	55.178	0.0296	11.891	0.0108	0.8780	0.0001
DII	55.179	0.0157	11.891	0.0044	0.8780	0.0001
DIII	55.186	0.0275	11.889	0.0106	0.8779	0.0001
aVR	55.201	0.0167	11.883	0.0076	0.8779	0.0001
aVL	55.175	0.0333	11.893	0.0132	0.8788	0.0001
aVF	55.182	0.0210	11.890	0.0087	0.8780	0.0001
V1	55.210	0.0683	11.881	0.0231	0.8778	0.0003
V2	55.194	0.0525	11.885	0.0193	0.8779	0.0002
V3	55.201	0.0489	11.884	0.0190	0.8779	0.0002
V4	55.210	0.0683	11.881	0.0231	0.8778	0.0003
V5	55.194	0.0525	11.885	0.0193	0.8779	0.0002
V6	55.201	0.0489	11.884	0.0190	0.8779	0.0002

Tabla 5.5. Distorsión filtro ASIC $SNR_i = -3dB$

Derivación	PRD		SNR		R_{xy}	
	μ	σ^2	μ	σ^2	μ	σ^2
DI	39.121	0.0306	18.769	0.0159	0.9328	0.0001
DII	39.119	0.0152	18.769	0.0082	0.9328	0.0001
DIII	39.126	0.0280	18.765	0.0152	0.9327	0.0001
aVR	39.142	0.0194	18.759	0.0099	0.9327	0.0001
aVL	39.115	0.0337	18.772	0.0174	0.9328	0.0001
aVF	39.122	0.0211	18.768	0.0110	0.9327	0.0001
V1	39.159	0.0661	18.751	0.0344	0.9326	0.0002
V2	39.136	0.0544	18.761	0.0270	0.9327	0.0001
V3	39.144	0.0517	18.758	0.0276	0.9327	0.0002
V4	39.159	0.0661	18.751	0.0344	0.9326	0.0002
V5	39.136	0.0544	18.761	0.0270	0.9327	0.0001
V6	39.144	0.0517	18.758	0.0276	0.9327	0.0002

Tabla 5.6. Distorsión filtro ASIC $SNR_i = 0dB$

armónicos, como se muestra en la expresión (4.1). Este algoritmo estima la amplitud y la fase, además de tener en cuenta los armónicos de la perturbación, de modo que el algoritmo de filtración ASIMC presenta mejores resultados en cuanto a rendimiento que el algoritmo IE, como se puede observar en la Tabla 5.7. Es importante decir que los valores mostrados en la Tabla 5.7, son el promedio sobre todos los registros analizados de ECG tomando en cuenta las 12 derivaciones. Además, los parámetros de eficiencia mostrados en las Tablas 5.1 a 5.7, son obtenidos usando las expresiones (4.2), (4.4), (4.5) y (4.6).

Algoritmo	máx	mín	μ	σ^2
	PRD			
ASIMC	25.48	7.65	11.26	3.42
IE	128.99	87.45	105	10.94
	SNR			
ASIMC	51.39	27.34	44.31	4.75
IE	2.68	-5.09	-0.87	2.05
	R_{xy}			
ASIMC	0.997	0.988	0.995	2.26E-03
IE	0.608	0.568	0.595	7.27E-03
	$\bar{\varepsilon}^2$			
ASIMC	1.46E-03	1.099E-04	2.85E-04	2.74E-04
IE	0.043	0.011	0.022	0.008

Tabla 5.7. Filtración de señales perturbadas con 2 armónicos y SNR=-6 dB

Con respecto a la reducción de las perturbaciones relacionadas a la línea base, los resultados obtenidos con el filtro basado en la transformada wavelet se muestran en la Tabla 5.8.

PRD	SNR	R_{xy}	PRD	SNR	R_{xy}	PRD	SNR	R_{xy}
-6 dB			-3 dB			0 dB		
0.826	1.657	0.563	0.742	2.593	0.670	0.647	3.788	0.763
0.837	1.543	0.548	0.760	2.382	0.651	0.674	3.422	0.741
0.874	1.168	0.491	0.820	1.723	0.581	0.762	2.357	0.660
0.835	1.566	0.550	0.753	2.462	0.658	0.661	3.598	0.751
0.926	0.668	0.410	0.899	0.922	0.501	0.872	1.192	0.595
0.847	1.446	0.534	0.776	2.205	0.634	0.698	3.120	0.721
0.824	1.687	0.570	0.735	2.680	0.684	0.633	3.967	0.782
0.860	1.307	0.524	0.794	1.999	0.633	0.722	2.828	0.730
0.833	1.588	0.557	0.752	2.472	0.671	0.662	3.584	0.774
0.829	1.631	0.560	0.746	2.544	0.667	0.653	3.700	0.758
0.824	1.677	0.567	0.739	2.628	0.676	0.643	3.842	0.770
0.824	1.685	0.571	0.737	2.644	0.686	0.639	3.888	0.785

Tabla 5.8. Filtración de línea base usando la transformada wavelet

Señales FCG

Todos los registros FCG fueron filtrados con un algoritmo de paso bajo a una frecuencia de corte de 2000 Hz. El nivel de descomposición hallado para reducir los efectos indeseables de las perturbaciones acústicas fue de 8, con la wavelet $Coiflet$ 4, usando como regla de

umbralización una variante heurística del principio de riesgo no sesgado de *Stein*, la cual es una umbralización suave, y con nivel de reescalamiento basado en la estimación del ruido. Para confirmar la hipótesis de que no se hubiera afectado la información acústica de los registros, fueron escuchados por 3 médicos especialistas en semiología, los cuales corroboraron la hipótesis.

5.1.2. Segmentación

Señales ECG

El algoritmo de segmentación para las señales ECG, fue desarrollado y ajustado para las derivaciones $V3$, $V4$ y $V5$ de la base de datos BD-MIT E-STT, ya que esta base de datos ofrece un banco de etiquetas para cada latido, entre las cuales están los puntos de segmentación, debido a esto, esta base de datos se presta para validar el nivel de precisión de los algoritmos de segmentación. Es importante tener en cuenta, que después de tener una derivación adecuadamente segmentada, estos puntos pueden hacerse corresponder al resto de las derivaciones para el caso de segmentar un ECG de 12 derivaciones. En la Tabla 5.9 se muestra con detalle el desempeño del algoritmo de detección de los complejos *QRS*, de acuerdo a las expresiones (4.7) y (4.8), sobre un total de 22 registros de 30 minutos cada uno, usando la transformada wavelet. Es importante anotar que el segmentador propuesto tiene un rendimiento adecuado para aplicaciones médicas, ya que el valor predictivo positivo (P) es de 100 %, lo cual indica que el sistema siempre va a operar sobre segmentos que coinciden correctamente con los que el médico usa para tomar decisiones clínicas.

Registro	Derivación	# latidos	P (%)	S (%)
E0103	$V4$	1802	100	100
E0108	$V4$	1594	100	99.88
E0115	$V5$	2722	100	99.56
E0118	$V4$	1865	100	98.13
E0123	$V4$	2310	100	98.22
E0124	$V4$	2402	100	99.64
E0125	$V4$	2214	100	98.09
E0126	$V4$	2115	100	98.43
E0127	$V4$	2222	100	99.96
E0151	$V3$	1715	100	99.71
E0170	$V4$	2061	100	99.95
E0202	$V5$	2580	100	98.18
E0212	$V5$	2790	100	99.82
E0404	$V5$	1712	100	99.77
E0405	$V5$	2556	100	99.92
E0417	$V5$	2411	100	98.03
E0418	$V5$	2742	100	98.63
E0605	$V5$	3302	100	99.21
E0606	$V5$	2014	100	100
E0615	$V5$	1814	100	98.07
E0704	$V5$	2241	100	98.47
E0818	$V5$	2271	100	99.96
Total		49455	100	99.17

Tabla 5.9. Detección *QRS* para la base de datos BD-MIT E-STT

Así, el algoritmo de detección del complejo QRS propuesto en esta tesis por su baja complejidad computacional se propone como una alternativa para aplicaciones en tiempo real, además de presentar una eficiencia promedio del 99.03 % sobre toda la base de datos BD-MIT E-STT, el cual es un resultado comparable con los que son frecuentemente citados en la literatura por su alto desempeño, como se muestra en la Tabla 5.10.

Algoritmos para la detección del QRS	Error (%)	P (%)	S (%)
Martínez et al. [12]	0.34	99.86	99.80
Hamilton et al. [6]	0.54	99.69	99.77
Pan et al. [13]	0.71	99.54	99.75
Madeiro et al. [2]	2.57	98.96	98.47
Propuesto en este trabajo	0.87	100	99.17

Tabla 5.10. Comparación con otros algoritmos de segmentación del complejo QRS reportados en la literatura usando la base de datos BD-MIT E-STT

Señales FCG

La evaluación del desempeño correspondiente a los algoritmos desarrollados para la segmentación de señales FCG, se hizo usando la base de datos BD-UNAL-GCPDS-FCG. La segmentación entre-latidos basada en la información sincrónica de la señal ECG dio una precisión del 100 %. La segmentación intra-latido para detectar las 4 componentes del latido FCG ($S1$, $S2$, sístole y diástole) se hizo sobre 360 latidos (180 normales y 180 con soplo), de forma que se consideraron 1440 segmentos. A partir de la Tabla 5.11 y las expresiones (4.7)

	Falsos positivos (FP)	Falsos negativos (FN)
Latidos normales	0/720	0/720
Latidos con evidencia de soplo	32/720	0/720

Tabla 5.11. Resultados de la segmentación intra-latido en señales FCG

y (4.8) se obtienen los valores de valor predictivo positivo (P) y sensibilidad (S), teniendo en cuenta que el valor total de Verdaderos Positivos (TP) es de 1408/1440. En la Tabla 5.12 se muestran los resultados obtenidos en este trabajo y se comparan con los reportados en trabajos que son frecuentemente citados en la literatura por el alto desempeño de sus algoritmos, notando que el algoritmo propuesto en esta tesis basado en la estadística de tiempos de recurrencia y la transformada wavelet supera los mejores reportados en la literatura en cuanto a eficiencia en la detección de los eventos intra-latido.

Algoritmos de segmentación para FCG	P (%)	S (%)	Eficiencia (%)
Carvalho et al. [3, 9]	95.5	89.1	96.2
Kumar et al. [8]	96.1	94.7	95.4
Propuesto en este trabajo	97.7	100	98.6

Tabla 5.12. Comparación con otros algoritmos de segmentación FCG intra-latido reportados en la literatura

5.2. Generación del espacio de características

Usando los modelos de representación presentados en el Capítulo 2, se estimaron las medidas de representación siguiendo los procedimientos expuestos en la Sección 4.4.

Particularmente, para las señales ECG correspondientes a la base de datos BD-UNAL-GCPDS-ECG, como es de 12 derivaciones, se obtuvo la caracterización sobre 27 señales por derivación para cada una de las clases (normal y con evidencia de isquemia), de forma que la matriz de características quedó conformada por 648 filas (324 por clase) y 856 columnas, de las cuales 13 son características de diagnóstico, 840 son características wavelet y 3 son características derivadas del análisis de complejidad. Asimismo, con el fin de hacer un análisis tiempo-frecuencia sobre la dinámica cardíaca perturbada por eventos isquémicos, se usó la base de datos BD-MIT E-STT, para la cual se conformó una matriz de características con dimensiones 1800×840, donde las filas corresponden a 900 latidos por clase y las columnas a 840 características wavelet. De la base de datos BD-UNAL-GCPDS-FCG, los médicos especialistas escogieron un conjunto representativo de 360 latidos FCG adquiridos en los diferentes focos de auscultación (180 latidos normales y 180 latidos con evidencia de soplo cardíaco), sobre los cuales se estimó un espacio de características representado por una matriz de dimensiones 360×150, donde las filas corresponden a los latidos FCG y las columnas a la estimación de 3 grupos de características (62 características de diagnóstico, 85 características espectro-temporales y 3 características derivadas del análisis de complejidad).

En general, se conformaron los siguientes ensambles de características, compuestos a su vez por subconjuntos relacionados a los modelos de representación, para los cuales se obtuvieron resultados de clasificación sin aplicar procedimientos de reducción de dimensiones, usando un clasificador 5-NN:

– $\mathbf{X}_{ECG} = [\mathbf{X}_{1-ECG}\ \mathbf{X}_{2-ECG}\ \mathbf{X}_{3-ECG}]$ es el conjunto de características extraídas de la base de datos BD-UNAL-GCPDS-ECG, donde $\mathbf{X}_{1-ECG}$ son las características derivadas de las medidas de diagnóstico, $\mathbf{X}_{2-ECG}$ son el conjunto de características wavelet y $\mathbf{X}_{3-ECG}$ son las características derivadas del análisis de complejidad. En la Tabla 5.13 se muestran los resultados de *precisión de clasificación* (PC) para estos subconjuntos de características. Los resultados de clasificación para las características derivadas del análisis de complejidad ofrecieron los mejores resultados, lo cual era de esperarse debido a que estas características intentan capturar la *dinámica intrínseca* del sistema, y entre las afecciones que más generan cambio dinámico en el *sistema cardíaco* se encuentra la isquemia (objetivo de reconocimiento para este estudio). Además, de acuerdo al procedimiento propuesto en este trabajo, para cada sujeto se extrajeron los parámetros de embebimiento sujeto a la capacidad de predic-

Ensambles	Dimensiones	PC (%)
$\mathbf{X}_{ECG}$	648×856	56.6±3.9
$\mathbf{X}_{1-ECG}$	648×13	89.3±2.5
$\mathbf{X}_{2-ECG}$	648×840	63.4±4.3
$\mathbf{X}_{3-ECG}$	648×3	95.6±0.4

Tabla 5.13. Resultados de clasificación para el ensamble $\mathbf{X}_{ECG}$

ción, lo cual asegura la reconstrucción del espacio de embebimiento difeomórfico al original, por lo tanto, la información no lineal proporcionada por los coeficientes de complejidad no se encuentra perturbada por la variabilidad fisiológica que inducen las características propias de cada paciente (edad, contextura, sexo, etc), y de esta manera, estos parámetros ofrecen una mayor capacidad de representación y separabilidad entre las clases de análisis. De otro modo, las características de diagnóstico, a pesar de que son las que contienen información de interpretación clínica y son las que usan los especialistas para tomar decisiones diagnósticas, en los sistemas automatizados tienen la desventaja de depender fuertemente de la efectividad del segmentador, el cual puede inducir altas desviaciones debidas a errores que se suscitan en señales fuertemente alteradas ante la presencia de enfermedades coronarias agudas. Los problemas de segmentación, no permiten que el entrenamiento se realice de forma adecuada, provocando errores frecuentes en la clasificación. En cuanto a las características wavelet, era de esperarse que los resultados de clasificación no fueran prometedores, debido a que la extracción se hizo usando una gran cantidad de familias wavelet ante diversas condiciones de operación, esto, de antemano provoca que la representación conjunta contenga una gran cantidad de información redundante, correlacionada e irrelevante, y que la información discriminante este embebida en todo ese espacio. Esta estructura se generó con el propósito de probar más adelante, si los procedimientos de reducción de dimensiones son capaces de rescatar la información relevante de toda esa cantidad no útil y repetida de estimaciones.

– $\mathbf{X}_{STT}$ es el conjunto de características wavelet extraídas de la base de datos BD-MIT E-STT. En la Tabla 5.14 se muestran los resultados de precisión de clasificación (PC) para este conjunto de características. El análisis tiempo-frecuencia sobre la dinámica

Ensambles	Dimensiones	PC (%)
$\mathbf{X}_{STT}$	1800×840	61.2±1.3

Tabla 5.14. Resultados de clasificación para el ensamble $\mathbf{X}_{STT}$

cardíaca perturbada por eventos isquémicos, seguramente debe ofrecer información discriminante para el reconocimiento de este tipo de enfermedades a partir de las señales ECG, como ya se mencionó, esta estructura formada por 840 características, contiene una alta cantidad de información correlacionada, ya que muchos de los coeficientes WT para las diferentes familias, generan una estructura muy similar de representación. De acuerdo a esto, con el resultado de clasificación mostrado en la Tabla 5.14 se demuestra que el alto nivel de redundancia e irrelevancia, sin importar que también halla algún contenido de información relevante dentro del espacio de representación, el nivel de precisión en la clasificación se afecta negativamente dificultando el reconocimiento de patrones.

– $\mathbf{X}_{FCG} = [\mathbf{X}_{1-FCG}\ \mathbf{X}_{2-FCG}\ \mathbf{X}_{3-FCG}]$ es el conjunto de características extraídas de la base de datos BD-UNAL-GCPDS-FCG, donde $\mathbf{X}_{1-FCG}$ son las características derivadas de las medidas de diagnóstico, $\mathbf{X}_{2-FCG}$ son el conjunto de características espectro-temporales y $\mathbf{X}_{3-FCG}$ son las características derivadas del análisis de

complejidad. En la Tabla 5.15 se muestran los resultados de precisión de clasificación (PC) para estos subconjuntos de características. La contribución de las características

Ensambles	Dimensiones	PC (%)
$\mathbf{X}_{FCG}$	360×153	90.5±2.5
$\mathbf{X}_{1-FCG}$	360×65	88.3±2.1
$\mathbf{X}_{2-FCG}$	360×85	95.6±1.3
$\mathbf{X}_{3-FCG}$	360×3	94.9±1.5

Tabla 5.15. Resultados de clasificación para el ensamble $\mathbf{X}_{FCG}$

extraídas muestran que las características espectro-temporales ofrecen el mejor resultado de precisión en la clasificación, seguido cercanamente por el resultado de clasificación de las características derivadas del análisis de complejidad. La metodología de representación usando características espectro-temporales dio resultados equivalentes con los mejores que han sido reportados en la literatura [10]. Las características derivadas del análisis de complejidad tienen la ventaja con respecto a las espectro-temporales de que no requieren de la segmentación intra-latido, lo cual las convierte en la mejor opción de representación, ya que además toma en cuenta la variabilidad inducida por los desórdenes valvulares contenidos en la señales FCG, al ser reconstruida la estructura dinámica mediante las correlaciones de alto rango resaltando las heterogeneidades que se presentan para los diferentes estados funcionales, es decir, la caracterización fractal explica de cierto modo las interacciones no lineales producidas por el flujo turbulento de la sangre cuando pasa a través de las válvulas cardíacas. Por otro lado, las características de diagnóstico, igualmente que para las señales ECG, dependen fuertemente de la segmentación, y además, la presencia de perturbaciones acústicas incluidas en la etapa de adquisición, dificultan la estimación correcta de estos coeficientes, ya que las perturbaciones se traslapan con la banda de frecuencias en las que se presentan los sonidos y los soplos cardíacos. Sin embargo, el resultado de clasificación es aceptable aunque se requiere de grandes esfuerzos en la detección de los eventos intra-latido.

5.3. Reducción de dimensiones: Nivel 1

Inicialmente, se obtuvieron los resultados aplicando PCA (considerando el 95 % de la varianza explicada) e ICA a los conjuntos de datos artificiales: Monk-1, Monk-3 y SynthData. En la Tabla 5.16 se muestra el porcentaje de reducción (PR) y la precisión de clasificación (PC) usando un clasificador 5-NN.

	PCA		ICA	
	PR (%)	PC (%)	PR (%)	PC (%)
Monk-1	50	94.1±1.2	50	73.3±3.2
Monk-3	50	97.3±0.6	50	85.6±2.1
SynthData	42.8	95.7±1.4	42.8	80.2±2.8

Tabla 5.16. Resultados obtenidos con PCA e ICA para los datos artificiales

En la Tabla 5.17, se muestran los resultados obtenidos después de aplicar PCA a los di-

ferentes ensambles de características, para una explicación de varianza del 95 %. Asimismo, se presentan los resultados después de aplicar ICA.

	PCA		ICA	
	PR (%)	PC (%)	PR (%)	PC (%)
$\mathbf{X}_{ECG}$	19.4	69.2±1.7	9.3	63.1±3.4
$\mathbf{X}_{STT}$	47.6	70.1±1.3	7.8	58.6±4.2
$\mathbf{X}_{FCG}$	33.8	85.7±2.1	13.4	53.9±2.9

Tabla 5.17. Resultados obtenidos con PCA e ICA para los ensambles de características

Después de aplicar variantes de PCA e ICA, con el fin de encontrar mejores resultados de clasificación, se obtienen los resultados mostrados en la Tabla 5.18. Los mejores resultados se obtuvieron para kernel–ICA, usando un kernel Gaussiano con valores de $\gamma = 1$ y $\kappa = 0.02$.

	PPCA		kernel–ICA	
	PR (%)	PC (%)	PR (%)	PC (%)
$\mathbf{X}_{ECG}$	15.8	70.2±1.2	13.1	90.1±0.8
$\mathbf{X}_{STT}$	42.5	74.7±1.5	9.6	92.6±1.7
$\mathbf{X}_{FCG}$	30.1	87.6±1.3	10.8	94.9±0.9

Tabla 5.18. Resultados obtenidos con PPCA y kernel–ICA

Comparando las Tablas 5.17 y 5.18, se puede observar que aunque PPCA toma ventaja de su enfoque probabilístico para direccionar la reducción de dimensiones hacia la diferencia que se presentan entre las distribuciones de las características extraídas para diferentes estados funcionales, la capacidad discriminante de sus componentes no se incrementa. De otro modo, el porcentaje de reducción de PPCA es menor con respecto al obtenido con PCA, se entiende que las componentes obtenidas retienen la mayor varianza posible, sin dejar de lado la distribución de probabilidad de las características para las diferentes clases. Finalmente, kernel–ICA presenta los mejores resultados para estos tres espacios de representación, ya que los datos presentan heterogeneidad en la naturaleza de sus características, y kernel–ICA toma en consideración las estructuras ocultas no lineales embebidas en el espacio de características para aumentar la separabilidad entre clases.

5.4. Reducción de dimensiones: Nivel 2

En las Tablas 5.19 y 5.20 se presenta una comparación entre los resultados de error de clasificación y porcentaje de reducción obtenidos con los procedimientos propuestos en este nivel de reducción de dimensiones. Las notaciones $\mathrm{MH}_{\mathcal{R}}$ y $\mathrm{MH}_{\Re}$ corresponden al modelo híbrido (i.e., árboles de decisión, algoritmos genéticos y k-NN) cuando es usada en la medida de evaluación el error de clasificación y la complejidad de Rademacher, respectivamente. Para este caso, se incluyó al estudio un conjunto de nuevos ejemplos correspondiente al 30 % del total de ejemplos usados en los procedimientos anteriores, con el fin de evaluar la capacidad de generalización del método mediante los resultados de verificación. En la Tabla 5.21 se presentan los resultados de validación del clasificador obtenidos mediante validación cruzada ($15 \times 2cv$) y el error de verificación (ε_v) estimado para el modelo híbrido simple y

Database	5-NN	ID3	SFFS/5-NN	SFFS/MANOVA	$MH_{\mathcal{R}}$	$MH_{\mathcal{R}}$
Monk1	29.4±2.7	27.8±2.2	4.7±3.2	6.1±2.7	9.1±3.6	4.5±2.2
Monk3	31.7±0.6	26.1±1.7	5.6±3.8	10.2±1.6	4.1±2.1	3.3±1.2
SynthData	15.0±1.7	24.4±2.1	6.1±2.2	7.6±1.5	4.5±1.8	3.7±1.5
$\mathbf{X}_{ECG}$	43.4±3.9	30.2±0.9	6.4±1.7	7.3±1.8	3.4±2.1	**2.6±0.5**
$\mathbf{X}_{STT}$	38.8±1.3	28.6±1.2	4.9±1.2	6.5±0.9	2.0±0.6	**1.3±0.4**
$\mathbf{X}_{FCG}$	19.3±2.5	16.6±0.3	8.2±0.9	8.9±2.2	6.2±3.8	**1.9±1.1**

Tabla 5.19. Porcentaje de error de clasificación

Database	5-NN	ID3	SFFS/5-NN	SFFS/MANOVA	$MH_{\mathcal{R}}$	$MH_{\mathcal{R}}$
Monk1	0.0	50.0	61.7	31.7	59.2	53.3
Monk3	0.0	33.3	75.0	30.0	68.5	65.0
SynthData	0.0	14.3	55.3	48.0	35.7	29.3
$\mathbf{X}_{ECG}$	0.0	44.4	79.3	75.2	89.9	**91.5**
$\mathbf{X}_{STT}$	0.0	52.9	87.1	89.9	95.2	**97.2**
$\mathbf{X}_{FCG}$	0.0	35.7	74.7	77.4	92.8	**95.4**

Tabla 5.20. Porcentaje de reducción de dimensiones

con la complejidad de Rademacher sobre la base de datos $\mathbf{X}_{STT}$ para evaluar la detección de isquemia y la base de datos $\mathbf{X}_{FCG}$ para evaluar la detección de soplos cardíacos.

		$15 \times 2cv$ (%)	ε_v (%)
$MH_{\mathcal{R}}$	Detección de isquemia	1.3±0.4	2.1
	Detección de soplo cardíaco	1.9±1.1	3.9
$MH_{\mathcal{R}}$	Detección de isquemia	2.0±0.6	10.8
	Detección de soplo cardíaco	6.2±3.8	18.3

Tabla 5.21. Porcentaje de error de verificación

De la Tabla 5.21 se observa que la penalización del error usando el modelo de *complejidad de Rademacher* demuestra ser de gran utilidad para mejorar el proceso de detección de desórdenes cardíacos. Adicionalmente, los resultados muestran que la penalización de Rademacher incrementa la capacidad de generalización del modelo, lo cual es muy deseable, debido a la alta variabilidad intra-clase de las señales cardíacas. El aumento en la capacidad de generalización es debido a que la inclusión de componentes de incertidumbre en el entrenamiento del clasificador permite que la cota del error teórico se base en el error empírico para generar un intervalo de confianza donde la frontera de decisión no se restringe a tomar en cuenta los centroides de las clases, sino que considera que algunos ejemplos de una clase pueden asemejarse mucho a los de la otra, y viceversa. De esta forma, la incertidumbre incluida en la selección de características genera una efectiva reducción de dimensiones con adecuada precisión de clasificación y alta capacidad de generalización. Finalmente, una de las virtudes del modelo es que cuando converge a alguna solución, en una de las salidas se obtiene el peso de relevancia para cada una de las características seleccionadas, lo cual es de gran ayuda para efectos de interpretar el espacio inicial y final de características. Después de obtener los resultados de clasificación y porcentajes de reducción, tal y como se presentaron en las Tablas 5.19 y 5.20, se realizó sobre el subconjunto de características

seleccionado por el modelo, el análisis de los pesos correspondiente a cada característica, con el fin de conocer del espacio inicial cuáles son los parámetros que más aportan a la discriminación de los estados funcionales.

Particularmente, del ensamble $\mathbf{X}_{ECG}$, se seleccionaron por cada familia wavelet las 4 características más relevantes al concepto de clasificación. Las que ofrecieron mejores resultados se muestran en la Tabla 5.22. De las características del ensamble $\mathbf{X}_{STT}$, también

Familia wavelet	# de coeficientes	Error de clasificación
Sym9	4	4.7±1.1
db3	4	3.2±0.8
bior44	4	3.8±1.5
rbio55	4	4.2±0.9

Tabla 5.22. Resultado de clasificación por familia wavelet del ensamble $\mathbf{X}_{ECG}$

se escogieron por cada familia wavelet las 4 características más relevantes al concepto de clasificación y las que ofrecieron mejores resultados se muestran en la Tabla 5.23. Final-

Familia wavelet	# de coeficientes	Error de clasificación
Sym3	4	3.7±0.6
db3	4	4.6±0.9
bior55	4	2.6±1.1
rbio22	4	3.2±0.8

Tabla 5.23. Resultado de clasificación por familia wavelet del ensamble $\mathbf{X}_{STT}$

mente de las características correspondientes al ensamble $\mathbf{X}_{FCG}$, el algoritmo de selección dio como resultado que con sólo 6 características se obtenía la precisión de clasificación de 96.1±1.1. Las características que determinan efectivamente el estado funcional normal y patológico en las señales FCG, son:

1. La sumatoria de la energía en el espectrograma STFT concentrada entre la sístole y la diástole.
2. La energía sumada en el espectrograma de Gabor de todo el latido FCG.
3. El primer componente principal del espectrograma de Gabor.
4. El MFCC relativo.
5. La dimensión de correlación.
6. El máximo exponente de Lyapunov.

Estas 6 características capturan la dinámica discriminante de las señales FCG relacionada a la presencia de eventos valvulo–patológicos, ya que los componentes energéticos de la señal destacan los episodios patológicos en sístole y diástole, además de que el primer componente principal del espectrograma obtenido de la señal resume la información espectral que resulta fuertemente perturbada por los soplos, por otro lado el MFCC relativo de la cuarta banda

pasante considera el ancho de banda espectral en que caen la mayoría de los soplos e intenta emular el comportamiento auditivo en la adquisición de estructuras acústicas, y las medidas de complejidad correspondientes a la dimensión de correlación y el máximo exponente de Lyapunov son sensibles a la inclusión de componentes de turbulencia en la estructura no lineal de la señal normal.

En cuanto a la detección de soplos cardíacos, la metodología presentada en este trabajo permitió el desarrollo de un algoritmo que superó en cuanto a precisión de clasificación (PC) a los que son considerados de alto desempeño en la literatura, como se muestra en la Tabla 5.24.

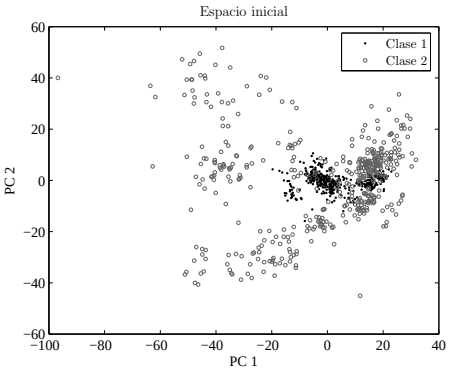
Algoritmos para la detección de soplos	PC (%)
Wang et al. [17]	97.2
El-Segaier et al. [4]	97.0
Leung et al. [10]	96.4
Wang et al. [18]	95.3
Propuesto en este trabajo	98.1

Tabla 5.24. Comparación con otros algoritmos de detección de soplos cardíacos reportados en la literatura

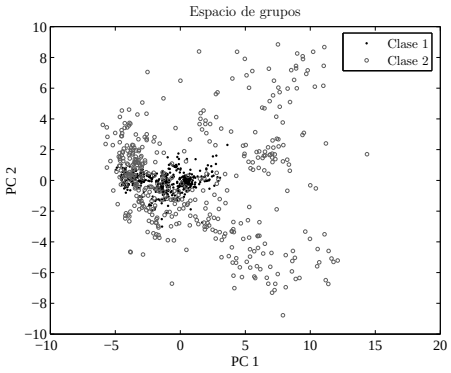
5.5. Reducción de dimensiones: Nivel 3

5.5.1. Reducción de dimensiones mediante cargas agrupadas

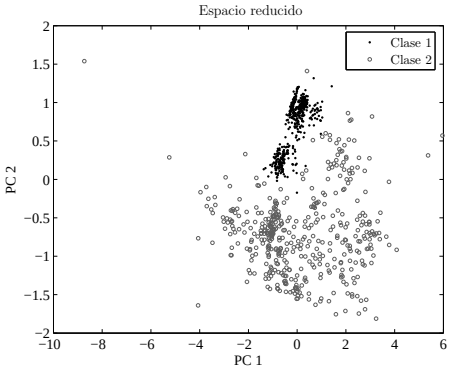
Los resultados de este procedimiento son obtenidos a partir de la estructura multivariada $\mathbf{X}_{STT}$, ya que es el conjunto de características que permiten evaluar el nivel de interpretación que ofrece la metodología propuesta, debido a que ninguna característica está relacionada directamente con algún sentido clínico. Así, siguiendo el procedimiento descrito en el Algoritmo 4.3, se obtuvo una matriz de proyección $\mathbf{A}_q$ con $q = 15$ componentes principales para el 99 % de la variabilidad de los datos (840 características representadas por 15 dimensiones). En la Fig. 5.1(a) se muestra la representación del espacio inicial de entrenamiento mediante la proyección en 2 componentes principales. Se puede observar que las observaciones para cada una de las clases tienen una disposición geométrica que dificulta la separabilidad de los patrones. El algoritmo *Isodata* agrupó el conjunto de características en 34 clusters $\boldsymbol{\upsilon}_j$, con los cuales se conformó una nueva matriz $\hat{\boldsymbol{\zeta}}$, que tiene el mismo número de observaciones del conjunto de entrenamiento pero con 34 columnas, siendo cada una de ellas el promedio de las características agrupadas. La representación por componentes principales del nuevo espacio $\hat{\boldsymbol{\zeta}}$ se muestra en la Fig. 5.1(b). Después de aplicar el algoritmo SFFS sobre $\hat{\boldsymbol{\zeta}}$, se encuentra que con un conjunto de 3 variables es suficiente para separar las clases (el clasificador usado como función de evaluación fue un 5-NN). El conjunto reducido de características obtenido fue $\boldsymbol{\varsigma} = \{\bar{\boldsymbol{\upsilon}}_7, \bar{\boldsymbol{\upsilon}}_8, \bar{\boldsymbol{\upsilon}}_{32}\}$, donde estos índices corresponden a las columnas de la matriz $\hat{\boldsymbol{\zeta}}$. El espacio de características reducido $\boldsymbol{\varsigma}$, se muestra mediante la proyección de los datos en el plano compuesto por los 2 componentes principales en la Fig. 5.1(c). Al comparar las Figs. 5.1(b) y 5.1(c), se puede observar el efecto de la selección de características al obtener un espacio reducido donde es visible la separabilidad de las observaciones.



(a) Componentes principales del espacio inicial



(b) Componentes principales del espacio de grupos $\hat{\boldsymbol{\zeta}}$



(c) Representación del espacio reducido $\boldsymbol{\varsigma}$

Figura 5.1. Visualización en las distintas etapas del procedimiento

Para los grupos de entrenamiento y validación, se crearon las respectivas matrices $\boldsymbol{\varsigma}_t$ y $\boldsymbol{\varsigma}_v$, de dimensiones 900×3, con lo cual se obtuvo la matriz de confusión para el clasificador 5-NN que se expone en la Tabla 5.25. La precisión de clasificación obtenida fue del 99.5 % para el conjunto de validación (ninguna observación de validación fue usada en el procedimiento de entrenamiento).

	Latido normal	Latido isquémico
Latido normal	449	3
Latido isquémico	1	447

Tabla 5.25. Matriz de confusión

Después de obtener los resultados de clasificación usando el espacio dimensional reducido $\boldsymbol{\varsigma} = \{\bar{\boldsymbol{\upsilon}}_7, \bar{\boldsymbol{\upsilon}}_8, \bar{\boldsymbol{\upsilon}}_{32}\}$, se buscó la conexión entre las características wavelet correspondientes a las agrupaciones 7, 8 y 32, y la representación en el tiempo de la señal ECG, con el fin de hallar la interpretación fisiológica de los resultados. En la Figura 5.2, pueden verse las regiones en las que está concentrada la representación temporal de las agrupaciones de

características halladas por el método.

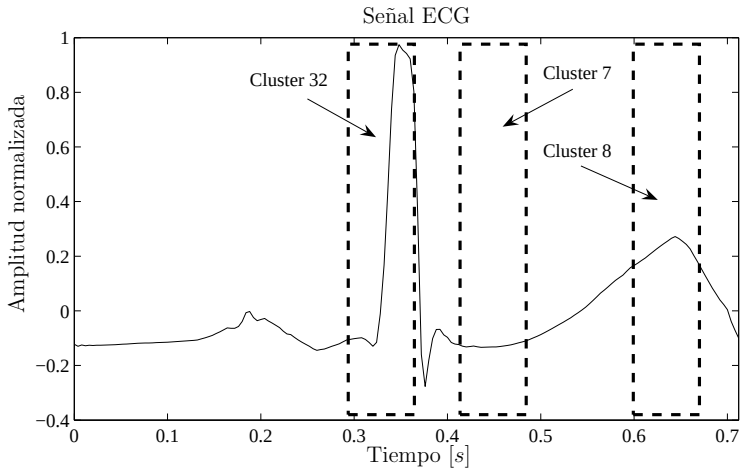


Figura 5.2. Interpretación visual de los grupos discriminates de características

Del análisis sobre las características agrupadas, se encuentra que la capacidad de representación de las diferentes wavelets utilizadas es muy similar, ya que subconjuntos de coeficientes de una wavelet quedaron agrupados en el cluster donde quedaron los mismos subconjuntos de coeficientes correspondientes a las otras wavelets. Esto quiere decir, que la construcción del espacio de representación usando varios tipos de wavelets conlleva en sí mismo variables correlacionadas, y en términos de separabilidad, las diferentes wavelets ofrecen un aporte similar. Por otro lado, los clusters asociados a los componentes resultantes de $\boldsymbol{\varsigma} = \{\bar{\boldsymbol{v}}_7, \bar{\boldsymbol{v}}_8, \bar{\boldsymbol{v}}_{32}\}$, revelaron que no solo las frecuencias del complejo $ST-T$ (1 - 5 Hz) son las que contienen información representativa de disfuncionalidades isquémicas, sino que en las bandas correspondientes al complejo QRS (15 - 40 Hz) también existe de manera influyente información discriminante para la detección automática. En general, los resultados obtenidos permiten asegurar que el análisis puede realizarse usando un solo tipo de wavelet, siempre y cuando, la wavelet ofrezca representación adecuada en la franja de frecuencias correspondiente a los complejos $ST-T$ y QRS, debido a que en $\boldsymbol{\varsigma}$ quedó contenida la información de estas franjas espectrales.

En general, el procedimiento presentado en este estudio ofrece los mejores resultados en cuanto a capacidad para reducir dimensiones y precisión de clasificación (PC = 99.5 %), respecto a otros modelos presentados en la literatura para la detección de isquemia sobre señales ECG. Particularmente, en la Tabla 5.26 se presenta una comparación del algoritmo propuesto en cuanto a sensibilidad (S = 99.7 %), obtenida mediante la expresión 4.8, y especificidad (Esp = 99.3 %), obtenida mediante la expresión 5.1, con varios métodos que son frecuentemente citados en la literatura, y es notorio que el algoritmo propuesto los supera considerablemente, además de facilitar la interpretación de los resultados mediante la conexión del espacio reducido con el espacio inicial. Adicionalmente, los mapeos de reconstrucción permiten obtener la abstracción de la información fisiológica embebida en el espacio inicial de representación. La Tabla 5.26 toma en consideración el nivel de *especificidad* de los clasificadores, el cual indica la capacidad que tiene el sistema para detectar

Algoritmos para la detección de isquemia	S (%)	Esp (%)
GA & MCDA [5]	91	91
ANN & PCA [16]	90	90
ANN & Modelos paramétricos [15]	81	84
ANN Backpropagation [11]	89	78
HMM [1]	86	85
Karhunen–Loève & Mahalanobis [7]	87	88
Sistema basado en reglas [14]	70	63
Cargas agrupadas (*propuesto en este trabajo*)	99	99

Tabla 5.26. Comparación con otros algoritmos de detección de isquemia reportados en la literatura

normalidad en los registros de entrenamiento. Este valor fue obtenido usando la siguiente relación:

$$\text{Esp} = \frac{TN}{TN + FP} \times 100\,\% \tag{5.1}$$

donde TN es el número de verdaderos negativos y FP el número de falsos positivos.

5.5.2. Reducción de dimensiones para el reconocimiento de múltiples formas

En la primera forma de onda de la Figura 5.3 se muestra el latido promedio del grupo normal correspondiente a la derivación $V4$ de un registro ECG-holter. La metodología propuesta en este trabajo permite visualizar en forma resumida el promedio de las formas de onda predominantes en registros ECG-holter (derivación $V4$) de la base de datos BD-MIT E-STT, los cuales evidencian disfuncionalidades isquémicas, usando el Algoritmo 4.4. Como caso particular, el resultado obtenido para la agrupación de 59377 latidos correspondientes a uno de los registros ECG-holter de la base de datos BD-MIT E-STT (derivación V4), se muestra a partir de la matriz de confusión expuesta en la Tabla 5.27, donde las filas representan las clases reales mientras que las columnas representan las 6 clases asignadas por el algoritmo no supervisado k-medians. Las agrupaciones se etiquetaron de la siguiente forma: $A =$ Normal, $B = ST$ disminuido, $C = ST-T$ disminuido, $D = ST-T$ elevado, $E =$ elevación de T y $F =$ inversión de T. Las formas de onda promedio para cada uno de estos grupos son como se muestran en la Figura 5.3. El método presenta una eficiencia global del

Caso	Asignación del algoritmo							Sensibilidad	Especificidad
real	A	B	C	D	E	F	Total	(%)	(%)
A	25341	4236	1203	1314	3428	34	355556	71.27	99.45
B	87	2974	32	21	903	10	4027	73.85	87.70
C	3	581	1735	10	0	2	2331	74.43	97.53
D	4	181	32	1293	349	0	1859	69.55	92.40
E	36	1806	57	3021	9098	0	14018	64.90	89.68
F	1	2	86	4	0	1493	1586	94.14	99.92
Total	25472	9780	3145	5663	13778	1539	59377		

Tabla 5.27. Matriz de confusión relacionada con la agrupación de latidos

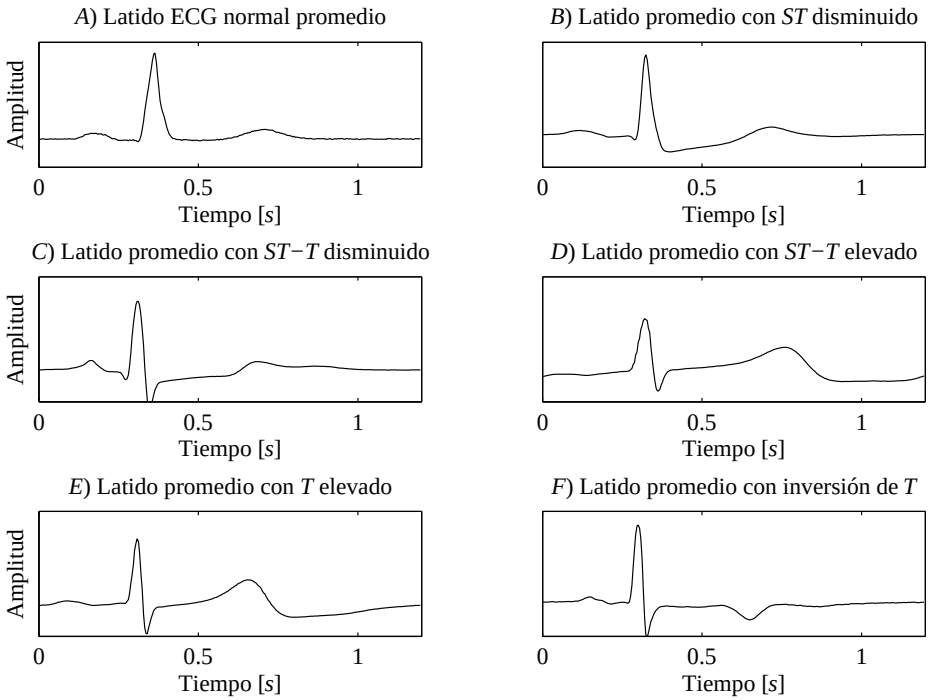


Figura 5.3. Visualización de las diferentes formas de onda relacionadas a episodios isquémicos presentes en registros ECG-holter (derivación $V4$)

70.6 % y considerando latidos patológicos frente a normales, una sensibilidad del 99.45 % y una especificidad del 71.27 %. Esto implica que pocos latidos anómalos serán considerados como normales, lo cual es deseable. El hecho de que latidos normales sean considerados como posibles patológicos implica que el cardiólogo deberá comprobarlos de forma visual, pero esto no es considerado como un problema grave. Por otro lado, cada latido del espacio de entrada (59377 latidos) tiene una longitud de 325 puntos y la representación reducida mediante SVD permite trabajar cada latido con una longitud de 6 puntos, lo cual permite agilizar el proceso y consumir menos recursos computacionales. Al intentar obtener la agrupación no supervisada con los latidos de 325 puntos, el algoritmo no converge por problemas de memoria dinámica del equipo de cómputo.

5.6. Discusión académica de los resultados

Los resultados obtenidos durante el desarrollo de la tesis fueron presentados en 10 eventos de difusión científica internacional, y en un número importante de congresos y simposios nacionales. Asimismo, se publicaron 2 artículos en revistas internacionales indexadas, 1 capítulo de libro de carácter internacional y 5 artículos en revistas nacionales indexadas por Colciencias. También se realizó el registro de software de 2 aplicativos para la detección de isquemia y deficiencias valvulares en la Dirección Nacional de Derechos de Autor. La tesis estuvo involucrada en el marco de 5 proyectos de investigación ejecutados por el Grupo de

Control y Procesamiento Digital de Señales. Finalmente, se lograron vínculos académicos con 5 grupos de investigación de cobertura internacional.

Artículos publicados en memorias de eventos internacionales

- **E. Delgado**, L. F. Giraldo y G. Castellanos. *Feature Selection Using a Hybrid Approach Based on the Rademacher Complexity Model Selection*. In: Computers in Cardiology, Vol. 34. pp. 257-260. 2007.

- **E. Delgado**, J. Jaramillo, A. F. Quiceno y G. Castellanos. *Parameter Tuning Associated with Nonlinear Dynamics Techniques for the Detection of Cardiac Murmurs by Using Genetic Algorithms*. In: Computers in Cardiology, Vol. 34. pp. 403-406. 2007.

- **E. Delgado**, J. L. Rodríguez, F. Jiménez, D. Cuesta y G. Castellanos. *Recognition of Cardiac Arrhythmias by Means of Beat Clustering on ECG-Holter Records*. In: Computers in Cardiology, Vol. 34. pp. 161-164. 2007.

- **E. Delgado**, A. F. Quiceno, C. Acosta y G. Castellanos. *Caracterización de Espectrogramas Usando Análisis de Componentes Principales y Medidas de Energía para Detección de Soplos Cardíacos*. In: IFMBE proceedings CLAIB 2007. Berlín: Springer-Verlag, Vol. 18. pp. 162-166. 2007.

- **E. Delgado**, M. Castaño, J. I. Godino y G. Castellanos. *Detección de Soplos Cardíacos usando Medidas Derivadas del Análisis Acústico en Señales Fonocardiográficas*. In: IFMBE proceedings CLAIB 2007. Berlín: Springer-Verlag, Vol. 18. pp. 202-206. 2007.

- F. Jiménez López, **E. Delgado Trejos**, D. Cuesta Frau y G. Castellanos Domínguez. *Agrupación de latidos en registros ECG-holter de pacientes isquémicos usando SVD y k-means*. In XXIV Congreso Anual de la Sociedad Española de Ingeniería Biomédica (CASEIB'06). Nov. 6 al 8 de 2006. Pamplona España. ISBN 84-9769-160-1.

- L. F. Giraldo, **E. Delgado**, J. C. Riaño y G. Castellanos. *Feature Selection using Hybrid Evaluation Approaches based on Genetics Algorithms*. In IEEE Computer Society Press Proceedings - Electronics, Robotics and Automotive Mechanics Conference. Vol. II. p.p. 242-247. México. Septiembre de 2006. ISBN 0-7695-2569-5.

- G. Castellanos, **E. Delgado**, G. Daza, L. G. Sánchez y J. F. Suárez. *Feature Selection in Pathology Detection using Hybrid Multidimensional Analysis*. In Proceedings 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC06), p.p 5503-5506, New York. Septiembre 2006.

- **E. Delgado**, G. Daza, J. F. Suárez y G. Castellanos. *Entrenamiento de sistemas para la detección de patologías empleando análisis de dimensión múltiple*. XXIII Congreso Anual de la Sociedad Española de Ingeniería Biomédica (CASEIB'05). Universidad Politécnica de Madrid. 10-12 nov de 2005. España. ISBN 84-7402-325-4.

– J.C. García, **E. Delgado**, V. E. Montes y G. Guarín. *Extracción de características de señales ECG con cardiopatía isquémica en tiempo real sobre DSP*. In IFMBE Proceedings 3rd Latin-American Congress on Biomedical Engineering - III CLAEB'04, Vol. 5, p. 1231-1234, 2004, Joao Pessoa, Brasil. Septiembre 22-25 de 2004.

Artículos publicados en revistas indexadas internacionales

– **E. Delgado**, A. Perera, M. Vallverdú, P. Caminal and G. Castellanos. *Dimensionality reduction oriented towards the feature interpretation for ischemia detection*. IEEE Transaction on Information Technology in Biomedicine. Special Issue. 2008.

– **E. Delgado**, J. I. Godino and G. Castellanos. *Digital Auscultation Analysis for Heart Murmur Detection*. Annals of Biomedical Engineering. 2008.

Capítulo de libro

– **E. Delgado**, G. Castellanos, L. G. Sánchez y J. F. Suárez. *Feature Selection in Pathology Detection using Hybrid Multidimensional Analysis*. Encyclopedia of healthcare information systems. 2008.

Artículos publicados en revistas indexadas nacionales

– A. M. Castaño, **E. Delgado**, J. I. Godino y G. Castellanos. *Acoustic analysis oriented to the cardiac murmur detection over digital auscultation signals*. Revista Avances en Sistemas e Informática. Vol. 4, No. 3, 2007.

– G. Castellanos, D. Marín, A. Cerquera y **E. Delgado**. *Análisis computarizado de registros fonocardiográficos para detección de soplos cardíacos*. Revista Colombiana de Cardiología. Vol. 13, No. 3, Noviembre-Diciembre 2006. pp 171-179.

– L. F. Giraldo, **E. Delgado**, J. C. Riaño y G. Castellanos. *Selección de características usando modelo híbrido basado en algoritmos genéticos*. Revista Ingeniería e Investigación. Vol. 26, No. 3, Diciembre 2006. pp 113-119.

– L. F. Giraldo, **E. Delgado** y G. Castellanos. *Cinemática inversa de un brazo de robot utilizando algoritmos genéticos*. Revista Avances en Sistemas e Informática. Vol. 3, Num. 1. 2006. pp 29-34.

– A. J. Hernández, **E. Delgado**, J. Rivera y G. Castellanos. *Reducción de dimensiones para clasificación de datos multidimensionales usando medidas de información*. Revista Scientia et Technica. Año XII, No. 32, Diciembre 2006. pp 181-186.

Registros de software

– Registro del software *Análisis automatizado de Señales Cardíacas* en la Unidad Administrativa Especial Dirección Nacional de Derechos de Autor. Autores: J. C. García, G. Castellanos, **E. Delgado** y J. L. Rodríguez. Libro 13 - Tomo 14 - Partida 236. Junio 09 de 2005.

– Registro del software *Visor y soporte de diagnóstico automatizado de soplos y arritmias cardíacas en señales fonocardiográficas (GCPDS-FONOSOP-V1.0)* en la Unidad Administrativa Especial Dirección Nacional de Derechos de Autor. Autores: G. Castellanos, **E. Delgado** y A. M. Matijasevic. Libro 13 - Tomo 17 - Partida 88. Octubre 19 de 2006.

Proyectos de investigación involucrados

– *Análisis automatizado de señales cardíacas*. Financiado por la Dirección de Investigaciones de Manizales (DIMA) con código 20201002545. Universidad Nacional de Colombia.

– *Auscultación y registro electrocardiográfico sobre la Web para servicios de teleconsulta médica*. Financiado por COLCIENCIAS con código 11271414907. Universidad Nacional de Colombia y Universidad de Caldas.

– *Métodos de generación y selección efectiva de características en la identificación automatizada de bioseñales de banda angosta*. Financiado por la Dirección de Investigaciones de Manizales (DIMA) a través del programa permanente de apoyo económico a la investigación en las Maestrías y los Doctorados de la Universidad Nacional de Colombia Sede Manizales.

– *Desarrollo de aplicativos de proceso de bio-señales para servicios de telemedicina en el departamento de Caldas*. Financiado por la Vicerrectoría de investigaciones de la Universidad de Caldas.

– *Técnicas de computación de alto rendimiento en la interpretación automatizada de imágenes médicas y bioseñales*. Financiado por la Dirección de Investigaciones de Manizales (DIMA) y la Vicerrectoría de investigaciones de la Universidad Nacional de Colombia, con código 20201004224.

Cooperación con otros grupos de investigación

– *Centre de Recerca en Enginyeria Biomèdica (CREB)*. Universidad Politécnica de Cataluña. España.

– *Grupo de Informática Industrial, Comunicaciones y Automática (GIICA)*. Universidad Politécnica de Valencia. España.

– *Instituto de Investigación en Ingeniería de Aragón*. Universidad de Zaragoza. España.

– *Grupo de Investigación en Algoritmos y Arquitecturas para Procesamiento de Señales Avanzado*. Universidad de Granada. España.

– *Grupo de Procesado de Señales Biológicas (GPSB)*. Universidad de Las Palmas de Gran Canaria. España.

BIBLIOGRAFÍA

[1] R. V. A. ao, S. M. T. Muller, J. Boudy, B. Dorizzi, T. F. Bastos-Filho, and M. Sarcinelli-Filho. Incremental HMM training applied to ECG signal analysis. *Computers in Biology and Medicine*, 38:659–667, 2008.

[2] J. ao P. V. Madeiro, P. C. Cortez, F. I. Oliveira, and R. S. Siqueira. A new approach to QRS segmentation based on wavelet bases and adaptive threshold technique. *Medical Engineering & Physics*, 29:26–37, 2007.

[3] P. Carvalho, P. Gil, J. Henriques, L. Eugénio, and M. Antunes. Low complexity algorithm for heart sound segmentation using the variance fractal dimension. In *IEEE International Workshop on Intelligent Signal Processing*, pages 194–199, 2005.

[4] M. El-Segaier, O. Lilja, S. Lukkarinen, L. Sörnmo, R. Sepponen, and E. Pesonen. Computer-based detection and analysis of heart sound and murmur. *Annals of Biomedical Engineering*, 33(7):937–942, 2005.

[5] Y. Goletsis, C. Papaloukas, D. I. Fotiadis, A. Likas, and L. K. Michalis. Automated ischemic beat classification using genetic algorithms and multicriteria decision analysis. *IEEE Transactions on Biomedical Engineering*, 51(10):1717–1725, 2004.

[6] P. S. Hamilton and W. J. Tompkins. Quantitative investigation of QRS detection rules using the MIT/BIH arrhythmia database. *IEEE Transactions on Biomedical Engineering*, BME-33(12):1157–1165, 1986.

[7] F. Jager, G. B. Moody, and R. G. Mark. Detection of transient *st* segment episodes during ambulatory ECG monitoring. *Computers and Biomedical Research*, 31(5):305–322, 1998.

[8] D. Kumar, P. Carvalho, M. Antunes, P. Gil, J. Henriques, and L. Eugénio. A new algorithm for detection of S1 and S2 heart sounds. *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2:1180–1183, 2006.

[9] D. Kumar, P. Carvalho, M. Antunes, J. Henriques, M. Maldonado, R. Schmidt, and J. Habetha. Wavelet transform and simplicity based heart murmur segmentation. In *Computers in Cardiology*, volume 33, pages 173–176, 2006.

[10] T. S. Leung, E. Brown, P. R. White, A. P. Salmon, and W. B. Collins. Classification of heart sounds using time-frequency method and artificial neural networks. In *Proceedings of 22nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 988–991, 2000.

[11] N. Maglaveras, T. Stamkopoulos, C. Pappas, and M. G. Strintzis. An adaptive backpropagation neural network for real-time ischemia episodes detection: development and performance analysis using the european ST–T database. *IEEE Transactions on Biomedical Engineering*, 45(7):805–813, 1998.

[12] J. P. Martínez, R. Almeida, S. Olmos, A. Rocha, and P. Laguna. A wavelet-based ecg delineator: evaluation on standard databases. *IEEE Transactions on Biomedical Engineering*, 51:570–581, 2004.

[13] J. Pan and W. J. Tompkins. A real-time QRS detection algorithm. *IEEE Transactions on Biomedical Engineering*, BME-32:230–236, 1985.

[14] C. Papaloukas, D. I. Fotiadis, A. Likas, A. P. Liavas, and L. K. Michalis. A knowledge-based technique for automated detection of ischemic episodes in long duration electrocardiograms. *Medical and Biological Engineering and Computing*, 39:105–112, 2001.

[15] C. Papaloukas, D. I. Fotiadis, A. Likas, and L. K. Michalis. An expert system for ischemia detection based on parametric modeling and artificial neural networks. In *Proceedings of the European Medical & Biological Engineering Conference*, page 742–743, 2002.

[16] C. Papaloukas, D. I. Fotiadis, A. Likas, and L. K. Michalis. An ischemia detection method based on artificial neural networks. *Artificial Intelligence in Medicine*, 24(2):167–178, 2002.

[17] P. Wang, Y. Kim, and C. B. Soh. Feature extraction based on mel-scaled wavelet transform for heart sound analysis. In *27th Annual International Conference of the Engineering in Medicine and Biology Society (IEEE-EMBS)*, pages 7572–7575, 2005.

[18] P. Wang, C. S. Lim, S. Chauhan, J. Y. A. Foo, and V. Anantharaman. Phonocardiographic signal analysis method using a modified hidden markov model. *Annals of Biomedical Engineering*, 35(3):367–374, 2007.

CAPÍTULO 6

Conclusiones

> Assiduus usus uni rei deditus et ingenium et artem saepe vincit.
>
> *(M. Tullius Cicero)*

A continuación se presentan conclusiones del trabajo con relación a: *i)* el análisis y comparación de las diferentes técnicas de generación de características indicando niveles de eficiencia, viabilidad y capacidad de generalización de los espacios de representación obtenidos, *ii)* el esquema propuesto basado en medidas de relevancia con el cual fue diseñado un método para reducir dimensiones señalando las bondades y limitaciones de los diferentes niveles de aplicación, además del método de selección de características fuertemente restringido para obtener interpretación visual y *iii)* la evaluación de los modelos desarrollados mediante clasificadores basados en algún principio discriminante. Adicionalmente, se presenta el trabajo futuro, las contribuciones científicas y el significado práctico de la tesis.

6.1. Conclusiones generales de la tesis

1. Se desarrolló un método de entrenamiento para el análisis de señales cardíacas mediante la generación de espacios de representación (estructura multivariada) compuestos por conjuntos características y un procedimiento basado en un esquema de reducción de dimensiones (extracción/selección de características), el cual permite considerar el problema de *información relevante* desde diferentes niveles de operación. En general, se obtuvo un método que de acuerdo a una medida de relevancia, permite interpretar, visualizar y clasificar patrones relacionados a estados funcionales cardíacos (isquemia y deficiencias valvulares), usando mapeos de reducción y reconstrucción de espacios, y adicionalmente, preservando la mayor cantidad de información relevante para lo cual se optimiza un criterio de evaluación.

2. Con respecto a la generación y estimación de características, se obtuvo que las derivadas del *análisis de complejidad* ofrecieron máxima capacidad de representación de los estados funcionales en la dinámica cardíaca. Como cada persona tiene un

sistema cardíaco diferente debido a la contextura del cuerpo, costumbres alimenticias, hábitos deportivos, entre otros, en este trabajo se comprobó que para la reconstrucción del espacio de estados correspondiente a diferentes sujetos, los parámetros de embebimiento cambian para cada caso aunque con baja variabilidad. Los parámetros asociados con la reconstrucción del espacio de estados (dimensión de embebimiento y el retardo de reconstrucción) fueron estimados haciendo uso de algoritmos genéticos. Aunque es conocido que este tipo de métodos agregan *alta complejidad de cómputo*, el uso de meta-algoritmos basados en la regresión por vectores de soporte, resultó ser una solución adecuada a este problema de *optimización*. De esta manera, fue posible reconstruir para cada registro el espacio de estados difeomórfico al que se reconstruiría si se tuvieran todas las variables dinámicas del sistema cardíaco particular del sujeto, permitiendo que las medidas estimadas sobre el atractor reconstruido fueran consistentes y altamente representativas con respecto a los diferentes estados de funcionalidad del corazón, además de que en este caso se toma como referencia de representación el sujeto mismo y se consideran especialmente las variaciones que se presenten con respecto a su propia dinámica. Es importante resaltar que la mejor forma encontrada para evaluar la calidad del espacio de estados reconstruido fue mediante la *capacidad de predicción* que tenía el atractor para reconstruir la señal en el tiempo. Así, se determinó que un atractor con alta capacidad para predecir la señal en el tiempo implica buenas propiedades de embebimiento. De otra parte, la generación de características basada en medidas de diagnóstico, permitieron la detección de anormalidades cardíacas con alta capacidad de separación, sin embargo, el desempeño de estas técnicas es altamente dependiente de etapas preliminares como la segmentación. Debido a eso, se requeriría un segmentador de señales ECG y FCG que asegure alto rendimiento a pesar de que se presenten condiciones patológicas que pueden llegar a destruir completamente el trazo del registro. Las características wavelet mostraron alta capacidad de representación de la dinámica cardíaca debido a que las señales ECG y FCG tienen componentes de naturaleza no estacionaria, de esta forma, la caracterización wavelet captura gran parte de la dinámica que determinan las disfunciones del aparato cardíaco. Además, se demostró que no es necesario el uso de diferentes tipos de wavelet porque cada una ofrece representaciones similares. Lo que se recomienda es que se elija la mejor base wavelet de acuerdo a los requerimientos del proceso y la señal de análisis.

3. Se desarrolló un esquema de representación efectiva basado en una definición generalizada de la relevancia que se expuso en este trabajo, el cual permitió formalizar la tarea de reducción de espacios usando diferentes métricas de representación.

4. Se desarrolló un método para reducir dimensiones de acuerdo a un esquema propuesto en este trabajo, el cual facilitó la interpretación de datos que residen en espacios de altas dimensiones. Los diferentes niveles del esquema presentado en este trabajo para reducir dimensiones de manera efectiva respecto a una medida de relevancia tienen bondades y limitaciones que deben ser consideradas para determinar el alcance que conllevan las diferentes soluciones experimentales buscando bajo costo operativo. El *nivel 1* permite transformar las variables originales, en general correlacionadas, en

variables independientes, facilitando la concepción abstracta de la estructura original multivariada mediante la identificación de las posibles variables *latentes*, ó no observadas, que se encuentran inmersas en los datos; sin embargo, la limitación de este primer nivel es que no siempre puede existir una explicación consistente, ó un canal de fácil interpretación, entre los datos originales y el nuevo espacio generado por el conjunto reducido de variables latentes. El *nivel 2* asegura la interpretación del nuevo espacio reducido, debido a que se crea un puente entre los resultados que se obtienen en el nuevo espacio y la influencia que tienen las variables de entrada de acuerdo a su naturaleza fisiológica y su explicación clínica. La limitación de este segundo nivel consiste en el costo de cómputo, ya que los algoritmos se basan en rutinas iterativas que dependen fuertemente de la población inicial escogida. El *nivel 3* asegura interpretación visual y las componentes resultantes están asociadas a elementos clínicos que a la vez sirven de soporte al médico. La limitación más sobresaliente en este tipo de procedimientos es el tiempo de cómputo, aunque el método que se propone en este trabajo involucra rutinas de baja complejidad computacional. En general, el esquema propuesto en este estudio promueve un avance relacionado a la interpretación clínica a partir de espacios de características de representación abstracta, lo cual es benéfico a la comunidad médica porque así no solo se involucran las medidas de diagnóstico convencionales para la toma de decisiones, sino un conjunto amplio de medidas de representación abstracta que llegan a enriquecer el análisis médico siempre y cuando el *sistema automático* ofrezca fácil interpretación fisiológica y clínica.

5. Se desarrolló un método de selección efectiva de características, en el cual quedaron relacionados los procedimientos de proyección, interpretación y visualización, de acuerdo a la profundidad de las capas del esquema propuesto, permitiendo visualizar los patrones inmersos en el espacio multidimensional, dando lugar al diseño de algoritmos que resuelven adecuadamente el problema de reconocimiento. Además, el esquema permite que el análisis de datos multidimensionales esté basado en medidas de relevancia, de forma que la elección de esta medida orienta la dirección y el contexto del análisis. En este sentido, el método facilita la creación de procedimientos para el hallazgo de subconjuntos de variables representativas respecto a una medida de evaluación, que adicionalmente mejora la precisión de clasificación para el reconocimiento de patrones. Es importante resaltar que la inclusión de un estimador de peso para cada característica ayuda a encontrar diferentes relaciones que pueden existir entre el conjunto de características y la interpretación fisiológica. Así, el método conecta el subespacio solución y el espacio inicial con el fin de hallar patrones relacionados a la naturaleza de las señales caracterizadas, en varios sentidos: patrones de funcionalidad fisiológica, patrones sobre contornos multivariados y patrones con información consistente y discriminante para la clasificación de ejemplos. La interpretación visual fue más allá de interpretar las estructuras residentes en los espacios de altas dimensiones, ya que ofreció la conexión con el fenómeno fisiológico subyacente.

6. Se desarrolló un modelo de evaluación de las características seleccionadas incluyendo componentes de incertidumbre en la validación de los clasificadores para incrementar la *capacidad de generalización* mediante el modelo de *complejidad de Rademacher*.

Este modelo crea un intervalo de confianza dado entre la cota del error teórico y el error empírico, para definir una frontera de decisión que toma en cuenta consideraciones relacionadas a la semejanza entre las clases. Así las características seleccionadas, no solo dotaron el espacio de representación de alta capacidad discriminante sino de alta capacidad de generalización para favorables niveles de concordancia en la verificación con nuevos ejemplos.

6.2. Trabajo Futuro

La dirección de trabajos futuros en esta área del conocimiento, consiste en tomar en cuenta los contornos que pueden suscitarse a lo largo del tiempo en una misma característica de representación, de esta manera, no sólo realizar el análisis sobre representaciones numéricas estáticas sino sobre representaciones sensibles a la variación en el tiempo de las características. De acuerdo al modelo presentado en las ecuaciones (3.1) y (3.2), esta tesis concentró su estudio en el caso estático de representación, igual a como se presenta en la expresión (3.4). Sin embargo, se pueden considerar en los modelos de representación cada cambio dinámico de las variables usando estrategias que permitan involucrar el cambio en el tiempo, de forma que ya no es útil la expresión (3.4), puesto que para estos casos $\mathbf{G} \neq \mathbf{0}$.

De otro modo, con los resultados obtenidos en esta tesis doctoral relacionados a las características que mejor representaron la dinámica cardíaca, se pueden proponer estudios que exploten otros mapeos lineales o no lineales de los espacios reducidos que permitan mejorar la comprensión fisiológica desde el punto de vista tiempo-frecuencia o de complejidad. Además de proponerse nuevas características de diagnóstico derivadas del análisis de representación abstracta.

6.3. Contribuciones científicas

1. Se desarrolló un método de *entrenamiento* y dos aplicaciones computacionales con alta tasa de acierto y baja complejidad computacional para la detección de isquemia y deficiencias valvulares sobre registros de la actividad cardíaca que sobrepasa los reportados en la literatura.

2. Se desarrolló en esquema nuevo de segmentación para señales FCG usando técnicas de dinámica no lineal, estadística de tiempos de recurrencia y un procedimiento alternativo basado en envolventes energéticos de alta frecuencia mediante la transformada wavelet.

3. Se presenta un análisis comparativo de los diferentes modelos de representación aplicados a la dinámica cardíaca.

4. Se desarrolló una nueva metodología para reconstruir el espacio de estados difeomórfico de la dinámica cardíaca correspondiente a cada sujeto analizado usando algoritmos genéticos, meta-algoritmos, superficies de regresión y predicción de series de tiempo caóticas.

5. Se desarrolló un esquema en 3 niveles de representación efectiva basado en la relevancia generalizada, lo cual no había sido reportado en la literatura.

6. Se desarrollo una método robusto con alto nivel de generalización para la detección de isquemia y deficiencias valvulares usando el modelo de complejidad de Rademacher.

7. Se desarrolló un sistema de reducción de dimensiones que facilita la interpretación de espacios de representación abstracta mediante la visualización de la dinámica discriminante.

6.4. Significado práctico de la tesis

1. Desarrollo de aplicaciones computacionales protegidas por la dirección nacional de derechos de autor para el soporte de diagnóstico relacionadas a la detección de isquemia y deficiencias valvulares sobre registros de la actividad cardíaca, validadas en diferentes centros clínicos del Departamento de Caldas.

2. Desarrollo y validación de un método de entrenamiento para sistemas automáticos de reconocimiento usando procesamiento digital de señales.

3. Los procedimientos propuestos en esta tesis doctoral presentan la facilidad de ser implementados en otro tipo de bioseñales diferentes a los de la actividad cardíaca.

APÉNDICE A

Algoritmos genéticos

> All truths are easy to understand once they are discovered; the point is to discover them.
>
> *(Galileo Galilei)*

Los *Algoritmos Genéticos* (GA) son métodos computacionales basados en los procesos de selección y evolución natural, que han sido satisfactoriamente aplicados sobre una gran cantidad de problemas en diferentes áreas de investigación. Su utilización puede ser dividida en dos enfoques principales: el desarrollo de modelos computacionales que ayuden en el estudio de *sistemas biológicos* [8, 16, 17], y como herramientas para la resolución de problemas de optimización [5, 10, 20]. En este capítulo se exponen los principios básicos que rigen los GA y algunas de las mejoras que se han aplicado sobre el modelo inicial, para convertirlo en una de las herramientas de búsqueda y optimización computacional más robustas que existen en el momento. Las secciones aquí expuestas son tomadas de [18].

A.1. El algoritmo genético simple

Los algoritmos genéticos fueron desarrollados por John Holland y sus estudiantes y colegas de la Universidad de Michigan en los 60's y 70's. El algoritmo genético de Holland, llamado algunas veces *Algoritmo Genético Simple* (SGA), es un método para moverse de una población de *cromosomas* a una nueva población, usando una especie de *selección natural* y los operadores de *cruce* y *mutación*, inspirados en la genética.

La ejecución de un SGA para resolver un problema de optimización, se puede sintetizar en las siguientes etapas:

Definición de la función de evaluación: La *codificación* y la *función de evaluación* son los dos únicos componentes que dependen de la naturaleza del problema a resolver. En optimización, la función de evaluación corresponde simplemente a la función que se quiere optimizar. De ahora en adelante, se considerará únicamente el problema de maximización, teniendo en cuenta que un problema de minimización puede ser expresado como uno de

maximización, de la forma:

$$\min_{\lambda_i} g(\lambda_1, \lambda_2, ..., \lambda_M) = \max_{\lambda_i}\{1 - g(\lambda_1, \lambda_2, ..., \lambda_M)\},\ i = 1, 2, ..., M \tag{A.1}$$

siendo $g(\lambda_1, ..., \lambda_M) \in [0, 1]$ la función a optimizar normalizada, y $\lambda_1, ..., \lambda_M$ los parámetros sobre los cuales se quiere hacer la optimización, que pueden ser expresados como un vector $\boldsymbol{\lambda} = [\lambda_1, ..., \lambda_M]$ de $1 \times M$.

Codificación de los parámetros de entrada y generación de la población inicial: Hace alusión a la tarea de cómo codificar el vector de parámetros de entrada $\boldsymbol{\lambda} = [\lambda_1, ..., \lambda_M]$, el cual representa las diferentes combinaciones de parámetros que pueden existir y cada una de las cuales es una solución potencial al problema de optimización. Cada posible solución debe ser codificada en una cadena de longitud finita, denominada *cromosoma*, conformada por la concatenación de las codificaciones de todos los parámetros $\lambda_i,\ i = 1, ..., M$

$$\mathbf{c} = [\mathcal{C}(\lambda_1), \mathcal{C}(\lambda_2), ..., \mathcal{C}(\lambda_M)] \tag{A.2}$$

donde $\mathbf{c}$ será el cromosoma resultante y $\mathcal{C}$ representa el operador de codificación. Estas cadenas están divididas en unidades denominadas *genes*, cada una de las cuales puede tomar diferentes valores llamados *alelos*. Holland [15] demostró a través de la teoría de esquemas que es preferible tener cromosomas de gran longitud con pocos posibles alelos por cada gen, lo que conlleva a que el mejor caso son las cadenas de genes binarios. El operador $\mathcal{C}$ queda entonces definido como una conversión a sistema binario

$$\mathcal{C} : \mathbb{R} \to \{0, 1\}^{l_i},\ i = 1, 2, ..., M \tag{A.3}$$

donde l_i es la longitud de la palabra usada para representar al parámetro λ_i y depende del planteamiento del problema. Así, cuando se ha definido la forma de codificación, es necesario generar una *población inicial* P_0 de r cromosomas

$$P_0 = \{\mathbf{c}_1, \mathbf{c}_2, ..., \mathbf{c}_r\} \tag{A.4}$$

En el SGA, esta población es generada de una forma completamente aleatoria. El parámetro r, denominado *longitud de la población inicial*, es el primer parámetro de control que es necesario sintonizar para el buen desempeño del GA.

Selección de los mejores individuos: Con el fin de emular el proceso de selección natural de los sujetos más aptos, es necesario asignar a cada individuo un valor de *calidad* (*fitness*), a partir del cual se hará dicha selección. Cuando los individuos más aptos han sido seleccionados, se forma la llamada *población intermedia*, que es un paso medio entre la población actual y la siguiente. Este proceso se ilustra en la Figura A.1 Aunque algunas veces se utilizan los términos *evaluación* y *calidad* sin ninguna distinción, es útil distinguir los dos términos al trabajar con algoritmos genéticos. La función de evaluación provee información sobre el desempeño de un individuo (cromosoma) en la solución del problema planteado. La calidad transforma esta evaluación en una asignación de oportunidades de reproducción. Aún más, la evaluación corresponde a un individuo y es calculada indepen-

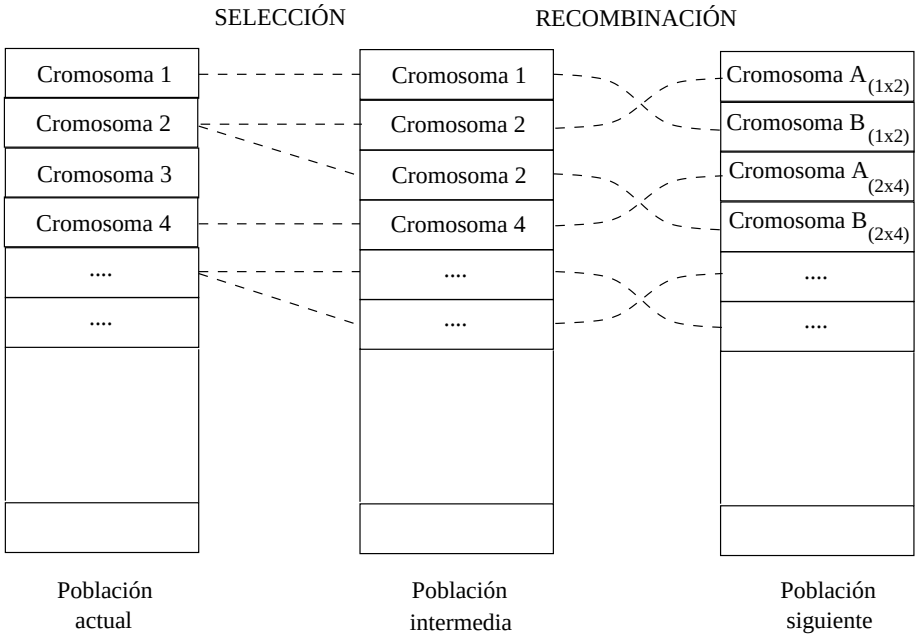


Figura A.1. Proceso de selección y recombinación en una generación del GA

dientemente del resto de la población, mientras la asignación de calidad depende de las evaluaciones de toda la población [30]. En el SGA, la calidad está definida por:

$$f(\mathbf{c}_i) = \frac{g(\mathbf{c}_i)}{\overline{g}},\ i = 1, 2, ..., r \tag{A.5}$$

donde $g(\mathbf{c}_i)$ es la evaluación del $i - simo$ individuo y $\overline{g}$ es el promedio de las evaluaciones de toda la población:

$$\overline{g} = \frac{1}{r}\sum_{i=1}^{r} g(\mathbf{c}_i) \tag{A.6}$$

Una vez se ha asignado un valor de calidad a cada uno de los individuos de la población, se escogen los más aptos a través de un muestreo estocástico, lo que le da oportunidad de reproducirse a todos los individuos, en proporción a su calidad. El primer método planteado es el de *muestreo estocástico con reemplazo*, que puede imaginarse como si se tuviera a todos los individuos en una ruleta, cada uno representado por un espacio proporcional al valor de su calidad. Girando repetidamente la ruleta, se van escogiendo individuos hasta completar la población intermedia. Otro método que puede concordar más con lo esperado a partir de los valores de calidad, es el de *muestreo estocástico sobre el restante* (*Remainder Stochastic Sampling*). Para cada cadena $\mathbf{c}_i$, $i = 1, 2, ..., r$, cuando $g(\mathbf{c}_i)$ sea mayor a 1.0, la parte entera de este número indica cuántas copias de esa cadena pasarán directamente a la población intermedia. Luego, todas las cadenas, incluyendo las que ya fueron copiadas directamente, tendrán una probabilidad de ser copiadas a la población intermedia,

correspondiente a la parte decimal de su valor de calidad. El *muestreo estocástico sobre el restante* es implementado de forma más eficiente si, para la parte de la selección no directa de copias, se utiliza el método conocido como *muestreo estocástico universal* (*Universal Stochastic Sampling*). Supóngase que se tienen todos los individuos ubicados al azar en una gráfica tipo *torta*, cada uno con un espacio proporcional a la parte decimal de su valor de calidad. Ahora, se superpone una ruleta de k lugares de igual tamaño, siendo k la cantidad de lugares disponibles en la población intermedia después de pasar las copias directas; entonces, con un solo giro de la ruleta, se escogen los k individuos que hacen falta.

Recombinación y mutación: La recombinación es el proceso por medio del cual se crea la población siguiente a partir de la población intermedia. La recombinación es aplicada a cromosomas apareados al azar con una probabilidad denotada por χ (se considera que la población ya ha sido suficientemente mezclada con el proceso de selección al azar). Lo que se hace es tomar un par de cromosomas y con dicha probabilidad χ decidir si los cromosomas se cruzarán para formar nuevos individuos o pasarán tal cual a la siguiente generación. El SGA utiliza la recombinación de un solo punto, que consiste en partir las dos cadenas por un punto escogido al azar e intercambiar los trozos de ambas, como se muestra en la Figura A.2. Después de la recombinación, se aplica el operador de mutación.

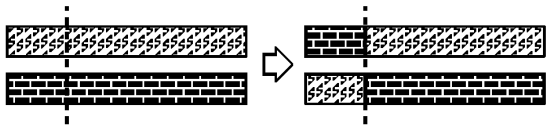


Figura A.2. Recombinación de un solo punto

Para cada bit de la población, se escoge si se aplica mutación o no, con una probabilidad $\mathcal{P}_m$. La mutación consiste en la *negación* del bit, es decir, se cambia la condición del bit de cero a uno o de uno a cero según sea el caso. Las probabilidades de recombinación (χ) y mutación ($\mathcal{P}_m$), constituyen otros dos parámetros de control fundamentales para el buen desempeño del GA.

Las etapas de selección, recombinación y mutación, son repetidas iterativamente hasta que la población haya convergido, es decir, hasta que el 95 % de los individuos compartan los mismos genes [5].

A.2. Fundamentos teóricos de los algoritmos genéticos

A pesar de que la descripción e implementación los GA es simple, los fundamentos teóricos acerca de por qué realmente funcionan, es aún un campo abierto. En esta sección se presentan dos aproximaciones que intentan describir el funcionamiento macroscópico de los GA: la *teoría de esquemas* utilizada por John Holland para describirlos originalmente [15] y los *caminos reales (Royal Roads)*, desarrollados más recientemente por Melanie Mitchell, Stephanie Forrest y John Holland como una alternativa para investigar más a fondo el funcionamiento de los GA [21].

A.2.1. Teoría de esquemas

Esta es la teoría tradicional de los GA, donde se asume su funcionamiento a través del descubrimiento y recombinación de *bloques constructores* de soluciones, en una forma de proceso paralelo. La noción de *bloques constructores* puede ser formalizada por medio de la introducción del concepto de *esquemas*, que pueden ser vistos como plantillas de unos, ceros y posiciones de *no importa* denotadas por asteriscos. Por ejemplo, el esquema $H = 1****1$ representa el conjunto de todas las cadenas de seis bits de longitud, que empiezan y terminan en 1. Las cadenas que cumplen con la plantilla H, son llamadas *ejemplos* de H. El número de bits que están dados en un esquema (que no son asteriscos) determina el *orden* del esquema y la distancia entre sus bits definidos más externos, es denominada *distancia definitiva*. Una cadena dada de longitud l, será entonces un ejemplo de 2^l esquemas diferentes. Por lo tanto, cuando un GA procesa r cadenas de una población, está procesando un número de esquemas entre 2^l (si toda la población es idéntica) y $2^l r$. Esto significa que, en una generación dada, mientras el GA evalúa explícitamente la calidad de r cadenas, está estimando implícitamente la calidad promedio de un número de esquemas mucho mayor, siendo la calidad promedio de un esquema, el promedio de la calidad de todos los ejemplos que cumplen con ese esquema. Esta idea es denominada *paralelismo implícito*. Desde este punto de vista, la dinámica aproximada de incremento y decremento de los ejemplos de un esquema determinado, puede calcularse como se explica a continuación.

Supóngase que H es un esquema con al menos un ejemplo presente en la población P_t de la generación t. Sea $m(H,t)$ el número de ejemplos del esquema H, en la generación t y sea $\hat{u}(H,t)$ la calidad promedio observada de H en la generación t (es decir, el promedio de la calidad de los ejemplos de H presentes en P_t). Se busca calcular el valor esperado de ejemplos de H, $E(m(H,t+1))$ en la generación $t+1$. Utilizando las ecuaciones (A.5) y (A.6), el número esperado de copias de todos los ejemplos $\{\mathbf{c}|\mathbf{c} \in H\}$ después de pasar por el proceso de selección, será:

$$\begin{aligned} E(m(H,t+1)) &= \textstyle\sum_{\mathbf{c}\in H} \frac{f(\mathbf{c})}{f} \\ &= \frac{\hat{u}(H,t)}{f} m(H,t) \end{aligned} \tag{A.7}$$

donde se puede observar que el incremento o decremento de los ejemplos de determinado esquema H en la generación t, depende de $\hat{u}(H,t)$ aunque este valor no sea calculado explícitamente.

La recombinación y la mutación pueden destruir o crear ejemplos de H. A continuación se considerarán los efectos destructivos de ambas. Sea χ la probabilidad de recombinación y supóngase que un ejemplo de H es escogido para ser uno de los padres. Se dice que el esquema H sobrevivió a la recombinación, si al menos una de las 2 cadenas resultantes pertenece a H. Una cota superior de la probabilidad $S_c(H)$ de que H sobreviva a la recombinación de un solo punto, estará dada por:

$$S_c(H) \geq 1 - \chi \frac{d(H)}{l-1} \tag{A.8}$$

donde $d(H)$ es la distancia definitiva de H y l es la longitud de las cadenas en el espacio de búsqueda. Esto significa que las recombinaciones que ocurran dentro de la distancia defini-

tiva de H, pueden destruir el esquema. Esta probabilidad, multiplicada por la probabilidad de recombinación χ, entrega la probabilidad de destrucción, por lo tanto, al restarla a uno, se tiene la probabilidad de supervivencia. Observando la ecuación (A.8), se puede deducir entonces que la probabilidad de supervivencia es mayor para los esquemas de menor longitud. A continuación, sea $\mathcal{P}_m$ la probabilidad de que algún bit sea mutado. La probabilidad $S_m(H)$ de que el esquema H sobreviva después de la etapa de mutación, será:

$$S_m(H) = (1 - \mathcal{P}_m)^{o(H)} \tag{A.9}$$

donde $o(H)$ es el orden del esquema. Esto es, para cada bit definido del esquema, la probabilidad de que no ocurra mutación es $1 - \mathcal{P}_m$, así que la probabilidad de que ninguno de los bits definidos del esquema sea mutado, será esta cantidad multiplicada $o(H)$ veces. Examinando la ecuación (A.9), se puede deducir que los esquemas de menor orden tendrán más probabilidad de sobrevivir. Estos efectos destructivos, pueden ser usados para completar la ecuación (A.7) de la forma:

$$E(m(H, t+1)) \geq \frac{\hat{u}(H,t)}{\bar{f}} m(H,t) \left(1 - \chi \frac{d(H)}{l-1}\right) (1 - \mathcal{P}_m)^{o(H)} \tag{A.10}$$

Esta ecuación es conocida como el *Teorema de Esquemas* [15] y describe el crecimiento de un esquema dado, de una generación a otra. Este teorema es un límite inferior, dado que sólo considera los efectos destructivos de la recombinación y la mutación, más no sus efectos constructivos. Comúnmente es interpretado como una implicación de que los esquemas cortos, de orden bajo, cuya calidad promedio se encuentre por encima de la media, recibirán un número exponencialmente creciente de ejemplos a través de la ejecución del GA, mientras el número de ejemplos de aquellos esquemas que no sean destruidos y permanezcan sobre el promedio de calidad, será incrementado por un factor de $\hat{u}(H,t)/\bar{f}$ en cada generación. Algunas críticas a esta interpretación, han surgido durante años, pero aún no hay un consenso definitivo sobre su validez o no [21, 30].

A.2.2. Caminos reales (*royal roads*)

El teorema de esquemas tiene en cuenta los efectos destructivos de la recombinación y la mutación, pero no sus efectos constructivos, aunque para muchos esta es una de las principales fuentes del poder de búsqueda de los GA. Los caminos reales, son un conjunto de funciones de calidad desarrolladas por Melanie Mitchell, Stephanie Forrest y John Holland para tratar de capturar la esencia de los bloques constructores, de una forma idealizada [22, 9, 23]. Como ejemplo, la función R_1, mostrada en la figura A.3, se encuentra definida como:

$$R_1(x) = \sum_i \delta_i(x), \; donde \; \delta(x) = \begin{cases} 1 \; si \; x \in s_i \\ 0 \; otro \; caso \end{cases} \tag{A.11}$$

La calidad de la cadena x será entonces la sumatoria de los coeficientes c_i asignados a todos los esquemas s_i a los cuales pertenece (ver figura A.3). Por ejemplo, si x pertenece a 2 de los esquemas, su calidad será 16, mientras para la cadena óptima mostrada en A.3, $R_1(s_{opt}) = 64$. La comparación del desempeño de los GA en la solución de estas funciones,

s_1 = 11111111**********; c_1 = 8
s_2 = ********11111111**********; c_2 = 8
s_3 = ****************11111111**********; c_3 = 8
s_4 = ************************11111111**********; c_4 = 8
s_5 = ********************************11111111**********; c_5 = 8
s_6 = **11111111**********; c_6 = 8
s_7 = **11111111********; c_7 = 8
s_8 = **11111111; c_8 = 8
s_{opt} = 1111111111111111111111111111111111111111111111111111111111111111

Figura A.3. Cadena óptima, dividida en ocho bloques constructores (esquemas)

contra otros métodos de optimización bien conocidos, permitió plantear lo que sus autores llaman el *Algoritmo genético ideal* (IGA) de la siguiente forma:

- En cada iteración, escójase una nueva cadena aleatoriamente, con probabilidad uniforme por cada bit.
- La primera vez que sea encontrada una cadena que contenga uno o más esquemas deseados, consérvese esa cadena.
- Cuando se encuentre una cadena que contenga uno o más esquemas no descubiertos aún, recombínese la nueva cadena con la almacenada de forma que la cadena resultante conserve todos los esquemas que han sido descubiertos hasta el momento.

Nótese que el IGA no maneja una población, sino que trabaja en una sola cadena cada vez. La selección completamente independiente de las cadenas, permite muestrear independientemente los esquemas. Además, la forma de recombinación que utiliza, implementa perfectamente la idea del paralelismo implícito. Este GA es, por supuesto, no implementable en un problema práctico dado que en la realidad no se tiene información sobre cuáles son los esquemas deseados o no deseados, pero su análisis puede dar ideas sobre cómo hacer más eficiente un GA. Para hacer un GA lo más parecido posible al IGA, se pueden tomar en cuenta las siguientes ideas:

Muestras Independientes. La población debe ser lo suficientemente grande, el proceso de selección lo suficientemente lento y la tasa de mutación lo suficientemente alta para asegurar que ningún gen particular tenga el mismo valor en toda la población o incluso en la gran mayoría.

Almacenamiento de los esquemas deseados. La selección debe ser lo suficientemente fuerte para preservar los esquemas deseados que han sido descubiertos.

Recombinación instantánea. La tasa de recombinación debe ser de tal manera que el tiempo en que se recombinan los esquemas deseados sea menor al tiempo con que se descubren nuevos esquemas.

Estos mecanismos no son todos mutuamente compatibles (por ejemplo la alta mutación no permite el almacenamiento de los esquemas deseados) y por lo tanto deben ser cuidadosamente balanceados.

A.3. Mejoramientos al modelo básico

Además de los operadores clásicos de selección, recombinación y mutación descritos en el SGA, existen otra serie de operadores inspirados en fenómenos naturales, que buscan aumentar la robustez del SGA. A continuación serán descritos algunos de esos operadores, que serán utilizados en el presente trabajo.

A.3.1. Métodos de escalamiento de la calidad

Estos métodos son utilizados para evitar dos problemas comunes en los GA. El primero es la *convergencia prematura*, que consiste en que en las primeras generaciones del algoritmo, algunos individuos con una calidad muy alta en comparación con el resto de la población, pueden dominar el proceso de selección e impedir el proceso de exploración que debe realizar el GA. El segundo, es el *terminado lento*, que se presenta en las generaciones donde la población tiene una alta convergencia y por lo tanto todos los individuos tienen valores de calidad similares. En este caso, la simulación tiende a moverse sin que predomine ninguno de los individuos, y no converge a alguno de ellos. En el primer caso, la función objetivo debe escalarse para evitar grandes diferencias entre la calidad de los individuos, mientras en el segundo caso se debe buscar una acentuación de estas diferencias para promover la convergencia [10]. Algunos de los métodos de escalamiento más usados son:

Escalamiento lineal. La función de calidad es obtenida a partir de la función objetivo, usando una ecuación lineal de la forma:

$$f(\mathbf{c}_i) = ag(\mathbf{c}_i) + b,\ i = 1, 2, ..., r \tag{A.12}$$

En esta ecuación, los valores a y b son escogidos para hacer dos cosas: igualar los valores promedio de la función escalada y la función original, y hacer que el máximo valor de calidad asignado sea un múltiplo específico del valor promedio (usualmente es de 2, pero en general puede ser tratado como otro parámetro de control). Estas dos condiciones aseguran que los miembros promedio reciban un valor esperado de copias igual a uno, mientras los mejores reciben el número de copias que se haya asignado.

Escalamiento sigma. En este método de escalamiento, se utiliza la desviación estándar de la población como una medida de la variación en la información contenida en esta. Un ejemplo de escalamiento sigma es el siguiente [21]:

$$f(\mathbf{c}_i) = \begin{cases} 1 + \frac{g(\mathbf{c}_i) - \overline{g}}{2\sigma} & \text{si } \sigma \neq 0 \\ 1 & \text{si } \sigma = 0 \end{cases} \tag{A.13}$$

donde $\overline{g}$ es la media de la evaluación de la población y σ es la desviación estándar de la evaluación. Esta función le da a un individuo la posibilidad de tener calidad con una desviación estándar por encima de la media, 1.5 de descendencia esperada. Al principio de las iteraciones, cuando la desviación estándar de la calidad es típicamente

alta, los individuos mejor calificados no tendrán una calidad que esté muchas desviaciones estándar por encima de la media y por lo tanto no se llevarán la parte más grande de la descendencia. Igualmente, cuando el número de iteraciones sea elevado, la convergencia será típicamente alta y la desviación estándar será pequeña, por lo tanto los individuos mejor calificados sobresaldrán más, permitiendo que la evolución continúe.

Escalamiento exponencial. En este método, la calidad es dada por una potencia específica de la función objetivo, de la forma:

$$f(\mathbf{c}_i) = g(\mathbf{c}_i)^k \tag{A.14}$$

donde el exponente k se convierte en un parámetro de control que debe ser sintonizado.

A.3.2. Métodos de ordenamiento (*Ranking*)

La naturaleza de los procesos de escalamiento condujo a la proposición de procesos de selección no paramétricos en los que los individuos son ordenados de acuerdo a su evaluación y el valor esperado de cada individuo depende de este rango y no del valor de calidad [3]. Eliminando la importancia de las diferencias grandes o pequeñas en la calidad, el algoritmo permite más oportunidades de reproducción a los individuos no muy altamente calificados, disminuyendo el riesgo de convergencia prematura.

Ordenamiento lineal. El método lineal propuesto en [3] consiste en lo siguiente: Todos los individuos son organizados desde 1 hasta r según su valor de calidad, siendo 1 el peor calificado y r el mejor. A continuación, el diseñador escoge el valor esperado de copias $1 \leq s \leq 2$ para el individuo mejor calificado, y calcula el resto aplicando la fórmula:

$$f(\mathbf{c}_i) = 2 - s + \frac{2(i-1)(s-1)}{r-1},\ i = 1, 2, ..., r \tag{A.15}$$

donde i corresponderá al lugar que ocupa el cromosoma $\mathbf{c}_i$, de forma que para el individuo peor calificado reciba un valor esperado de $2 - s$. Aunque en [3] se recomienda $s = 1.1$, en general este valor es un parámetro de control adicional.

Ordenamiento Exponencial. En este método, después de ordenar los individuos desde 1 hasta r según su valor de calidad, siendo 1 el peor calificado y r el mejor, el valor de calidad es asignado según:

$$f(\mathbf{c}_i) = s^{r-i} \tag{A.16}$$

donde s es un valor cercano a 1 que debe ser ajustado y se convierte en un parámetro de control. En este caso, la presión de selección es proporcional a $1 - s$ por lo tanto $s = 0.994$ proporciona dos veces la tasa de convergencia que proporciona $s = 0.998$. Con $s = 0.968$, la tasa de convergencia es idéntica a la obtenida con ordenamiento lineal y $s = 1.8$ [14].

A.3.3. Nichos y especies

Los GA con nichos buscan encontrar el mayor número de puntos óptimos que sea posible, por medio de la conservación de la diversidad en la población. La mayoría de estos modelos están basados en la idea de preservar la diversidad por medio de alteraciones en los operadores genéticos, para evitar la convergencia a un solo óptimo.

El método más común y mejor estudiado es el de *Calidad Compartida* (*Fitness Sharing*) [12]. Este método añade una penalización a la calidad $f(\mathbf{x}_i)$, $i = 1, 2, ..., r$ de todos los individuos de la población, en la medida que sean similares entre sí de acuerdo a una métrica $\| \|$. Para ello, se define una distancia σ_{share} por debajo de la cual los individuos compartirán su calidad, definiendo una nueva calidad compartida f':

$$f'(\mathbf{c}_i) = \frac{f(\mathbf{c}_i)}{\sum_{j=0}^{r} s(\mathbf{c}_i, \mathbf{c}_j)} \tag{A.17}$$

donde $s(\mathbf{c}_i, \mathbf{c}_j)$ es una función que define qué tanto deben compartir su calidad las muestras $\mathbf{x}_i$ y $\mathbf{c}_j$, dada por:

$$s(\mathbf{x}_i, \mathbf{x}_j) = \begin{cases} 1 - \frac{\|\mathbf{c}_i, \mathbf{c}_j\|}{\sigma_{share}} & \text{si } \|\mathbf{c}_i, \mathbf{c}_j\| \leq \sigma_{share} \\ 0 & \text{en otro caso} \end{cases} \tag{A.18}$$

Esta penalización pretende conseguir una distribución uniforme de la población, alrededor de áreas altamente calificadas y evitar la convergencia de toda la población a un solo óptimo. Desafortunadamente, tiene una serie de inconveniente, como se menciona en [28]:

- Se define un radio fijo σ_{share} para todos los individuos en la población. Por lo tanto, todos los óptimos deben estar aproximadamente equidistantes.
- La selección del radio σ_{share} requiere de conocimiento *a priori* acerca del espacio de búsqueda.
- El uso de funciones de escalamiento puede evitar que el algoritmo converja al máximo real, por lo cual se hace necesario un algoritmo de búsqueda local para mejorar el desempeño [4].

En [31] se propone un *Algoritmo Adaptativo de Clustering* para evitar la estimación *a priori* de σ_{share}. En lugar de las ecuaciones (A.17) y (A.18), se proponen las ecuaciones:

$$f'(\mathbf{c}_i) = \frac{f(\mathbf{c}_i)}{s(\mathbf{c}_i)} \tag{A.19}$$

$$s(\mathbf{c}_i) = n_{z_j} - n_{z_j} \left(\frac{\|\mathbf{c}_i, \mathbf{c}_j\|}{2d_{max}} \right)^{\alpha} \tag{A.20}$$

donde α es una constante, $\|\mathbf{c}_i, \mathbf{z}_j\|$ es la distancia entre la muestra $\mathbf{c}_i$ y el centroide $\mathbf{z}_j$ del cluster al cual pertenece $\mathbf{c}_i$, y n_{z_j} es el número de individuos asociados a dicho cluster. Aunque este algoritmo elimina la necesidad de definir un parámetro σ_{share}, introduce dos nuevas variables d_{min} y d_{max} que representan los radios mínimo y máximo permitidos a un cluster. Sin embargo, este trabajo ha sido el origen de muchas otras investigaciones que

asocian algoritmos de clustering a algoritmos genéticos con nichos y especies, como es el caso de [28] y [19].

A.4. Sintonización de los parámetros de control

Como se ha visto en las secciones anteriores, existe un número de parámetros que es necesario sintonizar en el GA para su correcto funcionamiento. Estos parámetros son comúnmente denominados *parámetros de control* e incluyen valores como el tamaño de la población inicial (r), la probabilidad de recombinación (χ), la probabilidad de mutación ($\mathcal{P}_m$) y los parámetros resultantes de los métodos de escalamiento y ordenamiento como el exponente k del escalamiento exponencial o el número esperado de copias s para el ordenamiento exponencial, entre otros.

Actualmente existen dos tendencias acerca del establecimiento de los parámetros de control. La primera apunta hacia la *sintonización* de los parámetros, es decir, el establecimiento de unos valores estáticos que se mantendrán a lo largo de toda la ejecución del algoritmo; mientras la segunda sugiere realizar un *control* de los parámetros, es decir, tener parámetros dinámicos que cambien a través de la ejecución del algoritmo, ya sea de forma determinista o por medio de una retroalimentación que entregue información del desempeño del algoritmo [6]. Otro conjunto de parámetros ampliamente usado, es el propuesto en [13], donde se usó por primera vez un *meta-algoritmo* genético. Recientemente, en [26] se tomó el modelo de meta-algoritmo para optimizar los parámetros de control de un GA aplicado a la solución del problema de la mayor sub-gráfica común. En dicho trabajo se introdujo una red neuronal para la evaluación de la función de adaptabilidad, con el fin de simplificar la ejecución del meta-algoritmo, entrenando la red para predecir el desempeño del algoritmo genético. La introducción de la red resultó ser una buena estrategia para disminuir el tiempo de ejecución pero no presenta una sustentación acerca de la función utilizada para medir el desempeño del GA de bajo nivel, ni el criterio usado para escoger el modelo de la red neuronal (cantidad de neuronas, número de capas, etc.) además de que no permite explicar la interacción entre los parámetros, debido a la misma naturaleza de las redes neuronales. En [25] se realizó un estudio estadístico usando análisis de varianza (ANOVA) para encontrar los parámetros que tienen una mayor influencia estadística sobre el comportamiento y el desempeño del GA, teniendo en cuenta las interacciones que se presentan entre ellos. Este estudio es un análisis exhaustivo pero no sugiere una manera de sintonizar los parámetros para otras aplicaciones.

Finalmente, cabe señalar de que a pesar que se han realizado una considerable cantidad de estudios teóricos y experimentales sobre el efecto de cada parámetro de control por separado, i.e., la influencia del tamaño de la población [11, 1], la tasa mutación [7, 24, 2] y la tasa de recombinación [27, 29], debido a las relaciones no lineales que se originan entre los parámetros de control, no es recomendable sintonizarlos independientemente.

BIBLIOGRAFÍA

[1] J. Arabas, Z. Michalewicz, and J. Mulawka. Gavaps - a genetic algorithm with varying population size. In *IEEE Interational Conference on Evolutionary Computation*, pages 7–12. IEEE Press, 1997.

[2] T. Bäck. Optimal mutation rates in genetic search. In *Proceedings of the third International Conference on Genetic Algorithms*, pages 2–8, 1993.

[3] J. E. Baker. Adaptive selection methods for genetic algorithms. In *Proceedings of the 1st International Conference on Genetic Algorithms*, pages 101–111. Lawrence Erlbaum Associates, Inc., 1985.

[4] P. Darwen and X. Yao. How good is fitness sharing with a scaling function. Technical Report CS 8/95, University of New South Wales, 1995.

[5] K. A. DeJong. *An analysis of the Behavior of a class of Genetic Adaptative System*. PhD thesis, University of Michigan, 1975.

[6] A. E. Eiben, R. Hinterding, and Z. Michalewicz. Parameter control in evolutionary algorithms. *IEEE Transactions on Evolutive Computation*, 3(2):124–141, 1999.

[7] T. C. Fogarty. Varying the probability of mutation in the genetic algorithm. In *Proceedings of the third International Conference on Genetic Algorithms*, pages 104–109, 1989.

[8] S. Forrest and T. Jones. *Complex Systems: Mechanisms of Adaptation*, chapter Modeling Complex Adaptive Systems with Echo. IOS Press, 1994.

[9] S. Forrest and M. Mitchell. *Foundations of Genetic Algorithms 2*, chapter Relative Building-Block Fitness and the Building-Block Hypothesis, pages 109–126. Morgan Kaufmann, 1993.

[10] D. E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley, 1989.

[11] D. E. Goldberg, K. Deb, and J. H. Clark. Genetic algorithms, noise and the sizing of populations. *Complex Systems*, 6:333–362, 1992.

[12] D. E. Goldberg and J. Richardson. Genetic algorithms with sharing for multimodal function optimization. In J. J. Grefenstette, editor, *Genetic algorithms and their applications: Proceedings of the Second International Conference on Genetic Algorithms*, pages 41–49. Lawrence Erlbaum Assoc., 1987.

[13] J. J. Grefenstette. Optimization of control parameters for genetic algorithms. *IEEE transactions on Systems, Man, and Cybernetics*, 16(1):122–128, 1986.

[14] P. J. B. Hancock. An empirical comparison of selection methods in evolutionary algorithms. In *Proceedings of the ASIB workshop on Evolutionary Computation*, 1994.

[15] J. H. Holland. *Adaptation in Natural and Artifical Systems*. University of Michigan Press, 1975.

[16] J. H. Holland. *Complexity: metaphors, models, and reality*, chapter Echoing emergence: objectives, rough definitions, and speculations for ECHO-class models, pages 309–342. Perseus Books, 1994.

[17] P. Hraber, T. Jones, and S. Forrest. The ecology of echo. *Artificial Life*, 3(3):165–190, 1997.

[18] J. Jaramillo-Garzón. Metodología de optimización de los parámetros de control de un algoritmo genético. Master's thesis, Universidad Nacional de Colombia, Manizales, Colombia, 2007.

[19] C.-Y. Lin and W.-H. Wu. Niche identification techniques in multimodal genetic search with sharing scheme. *Advances in Engineering Software*, 33(11-12):779–791, 2002.

[20] M. Mitchell. *Genetic Algorithms in Problem Solving*, chapter 2, pages 35–81. Massachusetts Institute of Technology, 1996.

[21] M. Mitchell. *An Introduction to Genetic Algorithms*. Massachusetts Institute of Technology, 1996.

[22] M. Mitchell, S. Forrest, and J. H. Holland. The royal road for genetic algorithms: Fitness landscapes and GA performance. In F. J. Varela and P. Bourgine, editors, *Towards a Practice of Autonomous Systems: Proceedings of the First European Conference on Artificial Life*, pages 245–254. The MIT Press, 1992.

[23] M. Mitchell, J. H. Holland, and S. Forrest. *Advances in Neural Information Processing Systems*, chapter When will a Genetic Algorithm Outperform Hill Climbing, pages 51–58. Morgan Kaufmann Publishers, Inc., 1994.

[24] H. Mühlenbein. How genetic algorithms really work: Mutation and hillclimbing. In *Parallel Problem Solving from Nature 2*, pages 15–26, 1992.

[25] I. Rojas, J. González, H. Pomares, J. J. Merelo, P. A. Castillo, and G. Romero. Statistical analysis of the main parameters involved in the design of a genetic algorithm. *IEEE transactions on Systems, Man, and Cybernetics*, 32(1):31–37, 2002.

[26] S. F. Smith and V. A. Cicirello. Modeling GA performance for control parameter optimization. In D. Whitley, D. Goldberg, E. Cantu-Paz, L. Spector, I. Parmee, and H.-G. Beyer, editors, *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2000)*, pages 235–242. Morgan Kaufmann, 2000.

[27] W. M. Spears. Adapting crossover in evolutionary algorithms. In J. R. McDonnell, R. G. Reynolds, and D. B. Fogel, editors, *Proceedings of the Fourth Annual Conference on Evolutionary Programming*, pages 367–384. MIT Press, 1995.

[28] Streichert, Stein, Ulmer, and Zell. A clustering based niching EA for multimodal search spaces. In *International Conference on Artificial Evolution, Evolution Artificielle, LNCS*, pages 293–304, 2003.

[29] T. White and F. Oppacher. Adaptive crossover using automata. In *Proceedings of the International Conference on Evolutionary Computation. The Third Conference on Parallel Problem Solving from Nature: Parallel Problem Solving from Nature*, pages 229–238, 1994.

[30] D. Whitley. An overview of evolutionary algorithms: practical issues and common pitfalls. *Information and Software Technology*, 43(14):817–831, 2001.

[31] X. Yin and N. Germay. Improving genetic algorithms with sharing through cluster analysis. In S. Forrest, editor, *Proceedings of the Fifth International Conference on Genetic Algorithms*. Morgan Kaufmann, 1993.

APÉNDICE B

Predicción en series de tiempo caóticas

> You see things; and you say, "why?" But I dream things that never were; and I say, "why not?".
>
> *(George Shaw)*

En la década de los 80, la reconstrucción de atractores caóticos a partir de series de tiempo llegó a ser una técnica muy común. La estructura geométrica de atractores llegó a ser estudiada de esta forma, y en particular, aproximaciones sobre la estimación de la dimensión de correlación fue tratada usando el método de Grassberger y Procaccia [3]. El siguiente mejor resultado fue obtenido en 1985, cuando fue propuesta la reconstrucción no solo de la geometría del atractor, sino la dinámica presente en el atractor. En el estudio de Eckmann y Ruelle [1], se introdujo la técnica de reconstruir la dinámica lineal local de un atractor determinista a partir de una serie de tiempo experimental. Este estudio, junto a otra ideas reportadas en la literatura, llegaron a ser muy influyentes en la extensión del concepto de *caos* en la ciencia experimental. A partir de estos conceptos, la *predicción de series de tiempo caóticas* fueron introducidas con el fin de estimar con anterioridad el comportamiento del error en varios experimentos relacionados a la dinámica biológica.

Cuando los datos que van a ser analizados corresponden a una serie de tiempo unidimensional, inicialmente se utiliza el teorema de embebimiento (ver Teorema 2.1) para representar la señal en un espacio de estados. Si el atractor es de dimensión D, un mínimo requerimiento es que $m \geq D$. El siguiente paso es asumir dentro del atractor reconstruido una relación funcional entre el estado actual $\boldsymbol{x}(t)$ y el estado futuro $\boldsymbol{x}(t+T) = f_T\left(\boldsymbol{x}(t)\right)$. Para predecir $x(t+T)$, se impone una métrica en el espacio de estados denotada por $\|\,\|$, con la cual se busca encontrar los k vecinos más cercanos de $\boldsymbol{x}(t)$, i.e., los k estados $\boldsymbol{x}(t')$ con $t' < t$ que minimice $\|\boldsymbol{x}(t) - \boldsymbol{x}(t')\|$. Se construye un predictor local tomando cada vecino $\boldsymbol{x}(t')$ como un punto en el dominio y $x(t'+T)$ como el correspondiente punto

en el rango. Por conveniencia, el rango es tratado como un escalar, mapeando estados m-dimensionales en valores uni-dimensionales aunque para algunos propósitos es deseable que el rango sea m-dimensional [2]. El ajuste puede ser realizado de varias maneras; para este trabajo, se proponen dos métodos: cuando el problema de regresión es *bien-condicionado*, se usa el método basado en los mínimos cuadrados mediante la descomposición de valores singulares, de otra parte, cuando los k vecinos más cercanos convierten el procedimiento en un problema *mal-condicionado* (el número de condiciones de la matriz con los k vecinos más cercanos es demasiado grande), se usa el método de *regularización de Tikhonov* para estimar el valor de regresión [4]. Así, se obtiene la predicción de señales caóticas, la cual es usada en este trabajo para determinar la función de costo que evalúa el nivel de difeomorfismo de un atractor reconstruido con respecto al original construido con todas las variables de estado, si se tuvieran.

BIBLIOGRAFÍA

[1] J. P. Eckmann and D. Ruelle. Ergodic theory of chaos and strange attractors. *Rev. Mod. Phys.*, 57, 1985.

[2] J. D. Farmer and J. J. Sindorowich. *Coping with chaos: Analysis of chaotic data and the exploitation of chaotic systems*, chapter Predicting chaotic time series, pages 208–211. John Wiley & Sons, 1994.

[3] P. Grassberger and I. Procaccia. Measuring the strangeness of strange attractors. *Physica D*, 9:189–208, 1983.

[4] A. N. Tikhonov and V. Y. Arsenin. *Solution of Ill-Posed Problems*. Winston, Washington, 1977.

ÍNDICE ALFABÉTICO

árboles de decisión, 109

AAMI – *Association for the Advancement of Medical Instrumentation*, 147
AHA – *American Heart Association*, 133
algoritmos genéticos, 110, 203
  codificación, 204
  estructura ideal, 209
  estructura simple, 110, 203
  mutación, 112, 206
  operadores genéticos, 112
  parámetros de control, 112
  recombinación, 112, 206
  selección, 112, 204
  sintonización, 213
análisis
  acústico, 149
  de componentes independientes – ICA, 98, 121, 162, 183
  de componentes principales – PCA, 96, 121, 162, 183
    no lineal, 122
    probabilístico – PPCA, 96, 184
  de dependencia lineal, 112
  de varianza, 113
  discriminante lineal – LDA, 105
  espectro-temporal, 149
  factorial – FA, 97, 121
  multiresolución, 49
  temporal, 148
angina
  de pecho
    estable, 17
    inestable, 17
  variante ó de Prinzmetal, 17
ANSI – *American National Standards Institute*, 133
aprendizaje, 35
atractor, 54
  difeomorfo, 54
  dimensión del, 54, 55
  extraño, 55
  reconstrucción del, 54

búsqueda de la proyección – PP, 100, 121
búsqueda heurística, 108
  con función de costo MANOVA, 163
bioseñal(es), 9
  análisis de, 35, 37
  procesamiento digital de, 2
  registro de las, 31
Box-Cox, 160

calidad
  compartida, 212
  de la percepción, 28
  de un individuo, 204
  de un sonido, 3
  del entrenamiento, 32
  del veredicto clínico, 29
  en la representación, 1, 33
  métodos de escalamiento de la, 210
caminos reales, 208
capacidad de generalización, 93
características
  estimación de, 34, 35, 148, 181
  extracción de, 94
  irrelevantes, 89
  redundantes, 89
  relevantes, 89, 101
    débilmente, 90
    fuertemente, 90
  selección de, 101
caracterización, 73
  basada en componentes principales, 75
    no lineal, 75
  basada en dinámica no lineal, 76
  basada en distribuciones tiempo-frecuencia, 74
  basada en la expansión de Hermite, 75
  basada en medidas de diagnóstico, 73
  mediante análisis espectral, 77
ciclo cardíaco, 21
  fase del, 23
clasificación, 35, 37
clasificador bayesiano, 105
clustering, 115

coeficiente de correlación, 138
complejidad
    análisis de, 50, 142, 154
    computacional, 93, 104
    de Rademacher, 165, 185
    medidas de, 55
    y alto costo de adquisición, 31

datos
    dinámicos, 88
    estáticos, 88
defecto del tabique
    interauricular, 25
    interventricular, 25
deficiencias valvulares, 1
densidad espectral de potencia, 46, 47
descomposición en valores singulares – SVD, 167
    distancia entre variables en la, 168
diagnóstico clínico, 29
    medidas de, 148
    naturaleza cualitativa del, 29
    naturaleza local del, 30
    naturaleza subjetiva del, 29
    orientabilidad del, 31
    trascendencia heurística del, 30
DICOM, 32
difeomorfismo, 53
dimensión
    box-counting, 56
    de capacidad, 56
    de correlación, 56, 156, 186
    fractal, 55
        generalizada, 56
    intrínseca, 118, 119
distribución poblacional
    asimetría en la, 30

Eckmann y Ruelle
    técnica de, 217
energía
    cuadrática, 43
    de Shannon, 43
    por valor absoluto, 43
entrenamiento
    conjunto ó muestra de, 32
    de sistemas, 1, 5, 88, 93
    error de, 165
    espacio inicial de, 6
    médico, 28, 44
entropía de Shannon, 43
error cuadrático medio, 138
escalamiento
    exponencial, 211
    lineal, 210
    sigma, 210
escalamiento multidimensional – MDS, 97, 121
espacio de estados
    reconstrucción del, 52
espectrogramas, 151
    caracterización de los, 153
estadística de tiempos de recurrencia, 142
estado(s)
    matriz de transición de, 86
    variables de, 85
estenosis, 25

filtración, 60, 175
    adaptativa, 65
    ASIC – *adaptive sinusoidal interference canceller*, 134, 175
    basada en el promedio de latidos, 62
    basada en filtros globales, 63
    basada en la transformada wavelet, 63
    denoising, 138
    IE – *incremental estimation*, 135, 175
    por aproximación mediante funciones, 62
    por media móvil, 62
    técnicas clásicas de, 60
    variante en el tiempo, 67
focos de auscultación, 22
    aórtico, 22
    accesorio o de Erb, 22
    de la base, 22
    de la punta, 22
    mesocárdico, 23
    mitral, 22
    pulmonar, 22
    tricúspide, 22
Fourier
    espectro de, 47
    STFT – transformada de tiempo corto de, 48, 186
    transformada de, 46
    transformada en tiempo discreto de, 47
    transformada inversa de, 47
Frémito, 23
función
    de adaptabilidad, 112
    de evaluación, 93
    de Heaviside, 56

Gabor
    transformada de, 49, 186
gauss-markoviano
    proceso aleatorio, 86
gaussiano(a)
    distribución, 87
    modelo, 85, 87
Grassberger y Procaccia
    método de, 217

Heisenberg
    caja, 49
Hurst
    exponente de, 59, 156

infarto agudo de miocardio
    fase estabilizada crónica, 18
    fase evolutiva aguda, 18
    localización del, 18
insuficiencia, 25
ISODATA, 115, 187
ISOMAP, 124
isquemia, 1

Jitter, 45, 149

k – vecinos más cercanos, 107

Kolmogorov–Smirnov
prueba de, 160

LLE – *Local Linear Embedding*, 124
Lyapunov
máximo exponente de, 57, 156, 186

máquinas de soporte vectorial – SVM, 106
matriz de dispersión
entre-clases, 105
intra-clase, 105
meta-algoritmo, 155
MFCC, 44, 149, 186
miocardiopatía hipertrófica obstructiva, 26
muestreo estocástico
con reemplazo, 205
sobre el restante, 205
universal, 206

nichos y especies, 212
normalidad
verificación de, 158
normalización
estadística, 156
geométrica, 156

ordenamiento
exponencial, 211
lineal, 211
métodos de, 211

parámetros de embebimiento, 53
paralelismo implícito, 207
período diastólico, 22
período sistólico, 22
población intermedia, 204
potenciales eléctricos, 9
PRD – *percentage root difference*, 137
precisión, 93
de clasificación, 37, 92
del sistema, 120
en el diagnóstico, 28, 29
predicción de series de tiempo caóticas, 217
prolapso de la válvula mitral, 26
proyecciones aleatorias, 122

reconstrucción
error de, 104
reducción de dimensiones, 94, 120
mediante cargas agrupadas, 166, 187
mediante formas agrupadas, 167, 190
nivel 1 – proyección, 94, 161, 183
nivel 2 – interpretación, 101, 163, 184
nivel 3 – visualización, 117, 166, 187
regresión por vectores de soporte – SVR, 155
relevancia
algebraica, 91
análisis de, 34, 36, 85
como medida de representación, 89
débil, 90
definición generalizada, 89
estadística, 90
estratificación de la, 89
fuerte, 90
informativa, 91
medida de, 90, 94
por clasificación, 92
por consistencia, 92
rendimiento
de proceso, 104
representación
modelos irregulares de, 30
resonancia, 45

salud
normatividad en el área de la, 31
SBS – *sequential backward selection*, 109
señal(es)
biomédica, *véase* bioseñal(es)
electrocardiográficas, 9
área de la onda T, 42
área negativa del QRS, 42
área positiva del QRS, 42
amplitud de la onda T, 43
amplitud negativa del QRS, 42
amplitud positiva del QRS, 42
ancho del QRS, 42
complejo QRS, 15
de alta resolución, 16
derivaciones aumentadas, 12
derivaciones bipolares, 12
derivaciones precordiales, 13
en ejercicio, 16
en reposo, 16
holter, 16
intervalo PP, 15
intervalo PR, 15, 43
intervalo QT, 15, 43
intervalo RR, 15
morfología, 10
nivel del segmento ST, 43
onda P, 14
onda T, 15
onda T respecto al pico R, 43
onda U, 15
pendiente del segmento ST, 43
repolarización ventricular invertida, 42
segmento PR, 15
segmento ST, 15
fonocardiográficas, 19
detección de *S1*, 142
detección de *S2*, 144
origen fisiológico, 20
primer sonido *S1*, 21
segundo sonido *S2*, 21
tercer y cuarto sonido *S3* y *S4*, 21
segmentación, 68, 179
basada en la representación, 70
basada en la transformada wavelet, 68
basada en modelos ocultos, 69
basada en transformaciones no lineales, 71
sensibilidad, 148
SFFS – *sequential forward floating selection*, 109, 167, 187
SFS – *sequential forward selection*, 109
sistema
auditivo, 44

caótico, 58
cardíaco, 12, 45, 154, 181, 198
celular, 9
complejo, 30
corporal, 28
de detección, 2, 36
de entrenamiento, 91
de reconocimiento, 2, 5, 36, 120
de representación, 5, 35
dinámica intrínseca del, 181
eléctrico del corazón, 6, 9, 10, 32
estabilidad del, 135
fisiológico, 90
lineal, 87
mecánico del corazón, 6, 19, 32
QuickCog, 118
sistemas
automáticos, 6, 9, 28, 30, 31, 44, 199
biológicos, 9, 31, 45, 203
centralizados, 31
de cómputo, 1, 29
de clasificación, 97
de coordenadas, 53, 94
digitales, 20, 29
dinámicos, 51, 85
expertos, 28
gaussianos, 7, 85
identificación de, 87, 88
inteligentes, 5, 117
no lineales deterministas, 51, 52, 55, 57
orgánicos, 3
SNR – *signal to noise ratio*, 138
soplos cardíacos, 23
área de registro de los, 23
clasificación de los, 23
continuos, 26
diastólicos
mesodiastólicos, 26
precoces, 26
telediastólicos ó presistólicos, 26
inocentes, 25
intensidad de los, 23
irradiación de los, 23
sistólicos
de eyección, 24
pansistólicos, 25
tardíos, 25
subconjuntos
evaluación de, 103
tipo filtro, 104
tipo *wrapper*, 103, 163
generación de, 102
aleatoria, 103
completa, 102
heurística, 103
suma de correlación, 56
generalizada, 56

teoría de esquemas, 207
bloques constructores, 207
Thrill, *véase* Frémito
Tikhonov
regularización de, 218
transformación de distribuciones, 160

válvula(s)
aórtica, 20
auriculoventriculares – AV, 20
mitral, 20
pulmonar, 20
sigmoideas, 20
tricúspide, 20
validación, 104
simple, 37
valor predictivo positivo, 148
valores atípicos
remoción de, 156
variedad diferenciable, 53

wavelet
transformada, 49
WW – Weight Watcher, 118

Todos los derechos reservados.
©Universidad Nacional de Colombia – Manizales 2008

Esta tesis doctoral fue desarrollada en el Grupo de Control y Procesamiento Digital de Señales adscrito al Departamento de Ingeniería Eléctrica, Electrónica y Computación de la Universidad Nacional de Colombia Sede Manizales. El desarrollo experimental y todo el marco operativo de la tesis doctoral tuvo el apoyo financiero del Instituto Colombiano para el Desarrollo de la Ciencia y la Tecnología Francisco José de Caldas – COLCIENCIAS, la Dirección de Investigaciones de Manizales (DIMA) y la Universidad Nacional de Colombia.

UNIVERSIDAD NACIONAL DE COLOMBIA
SEDE MANIZALES

Biblioteca Universidad Nacional de Colombia Sede Manizales – Campus La Nubia