

**Centro de Investigación Científica y de Educación
Superior de Ensenada, Baja California**



**Maestría en Ciencias
en Electrónica y Telecomunicaciones
con orientación en Telecomunicaciones**

**Clasificación del audio cardiaco mediante
representación escasa de señales y aprendizaje
automático**

Tesis

para cubrir parcialmente los requisitos necesarios para obtener el grado de
Maestro en Ciencias

Presenta:

Javier Iván Gutiérrez Uribe

Ensenada, Baja California, México

2019

Tesis defendida por

Javier Iván Gutiérrez Uribe

y aprobada por el siguiente Comité

Dr. Miguel Ángel Alonso Arévalo
Director de tesis

Dr. Gabriel Alejandro Galaviz Mosqueda

Dr. Marco Antonio Panduro Mendoza

Dr. Israel Marck Martínez Pérez



Dr. Daniel Saucedo Carvajal
Coordinador del Posgrado en Electrónica y Telecomunicaciones

Dra. Rufina Hernández Martínez
Directora de Estudios de Posgrado

Javier Iván Gutiérrez Uribe © 2019

Queda prohibida la reproducción parcial o total de esta obra sin el permiso formal y explícito del autor y director de la tesis

Resumen de la tesis que presenta Javier Iván Gutiérrez Uribe como requisito parcial para la obtención del grado de Maestro en Ciencias en Electrónica y Telecomunicaciones con orientación en Telecomunicaciones.

Clasificación del audio cardiaco mediante representación escasa de señales y aprendizaje automático

Resumen aprobado por:

Dr. Miguel Ángel Alonso Arévalo

Director de tesis

Las principales causas de muerte alrededor del mundo son las enfermedades cardiovasculares (CVDs). El fonocardiograma (PCG) es la herramienta más simple, de bajo costo y ampliamente utilizada para ayudar a los fisiólogos en el diagnóstico de CVDs. Investigaciones en el procesamiento de señales y *Machine Learning* han tenido resultados prometedores en el área de la clasificación de audio cardiaco, motivando a aún más investigaciones. En este trabajo, se propone un esquema de clasificación de audio cardiaco para la detección de la presencia/ausencia de un estado patológico utilizando características tiempo-frecuencia de tipo *sparse*, específicamente funciones de Gabor, y el análisis de su residual; predicción lineal, características en el dominio de tiempo, características en el dominio de la frecuencia, coeficientes cepstrales en la escala de frecuencias Mel y su velocidad. Cada uno de estos fue probado con 9 clasificadores, utilizando una base de datos que contiene 2,873 PCGs previamente segmentados, la cual fue utilizada para crear una base de datos balanceada de 740 PCGs (600 para entrenar al clasificador y 140 para prueba) utilizando submuestreo aleatorio. El clasificador de bosques aleatorios tuvo el mejor rendimiento, y fue utilizado para clasificar el vector completo de todas las 300 características. Posteriormente, se implementó reducción de la dimensionalidad para disminuir el vector de características a 60, con análisis de componentes principales (PCA-60) y evaluación de la importancia de las características (RF-60); resultando en que RF-60 tenía mejor desempeño. El algoritmo propuesto mostró un rendimiento de $SP=0.87$, $SE=0.798$ y $MAcc=0.834$ en términos de *specificity*, *sensitivity* y *modified accuracy*. Además, este algoritmo, fue probado utilizando la base de datos desbalanceada que cuenta con 2,873 PCGs mostrando un desempeño de $SP=0.974$, $SE=0.780$ y $MAcc=0.877$.

Palabras clave: Codificación escasa, aprendizaje máquina, Señales de audio cardiaco, procesamiento de señales, clasificación, Matching Pursuit

Abstract of the thesis presented by Javier Iván Gutiérrez Uribe as a partial requirement to obtain the Master of Science degree in Electronics and Telecommunications with orientation in Telecommunications.

Cardiac audio classification by sparse signal representation and automatic learning.

Abstract approved by:

Dr. Miguel Ángel Alonso Arévalo
Thesis Director

The main cause of death worldwide are the cardiovascular diseases (CVDs). Phonocardiogram (PCG) is the most simple, low cost and highly used tool to assist physicians in diagnosing CVDs. Researches in signal processing and machine learning have had promising results in the area of heart sound signal classification, and motivated even more researches. In this work, we propose a heart sound classification scheme to detect the presence/absence of a pathological state using sparse time-frequency features, specifically Gabor wavelets, and its residual analysis; linear predictive, time domain features, frequency domain features, Mel-frequency cepstral coefficients and its velocity. Each one of these were tested with 9 classifiers, using a database containing 2,873 PCGs previously segmented, which was used to create a balanced dataset of 740 PCGs (600 for train the classifier and 140 for testing) using random undersampling. The random forest classifier had the best performance, and was used to classify the complete vector of all the 300 features. Then principal components analysis and an evaluation of the feature importance were used to decrease the dimensionality of the features vector to 60; having this the best performance. The proposal showed performances of $SP=0.87$, $SE=0.798$ and $MAcc=83.4$ in terms of specificity, sensitivity and modified accuracy. Also, the whole unbalanced dataset containing 2,873 PCGs was tested showing a performance of $SP=0.974$, $SE=0.780$ and $MAcc=0.877$.

Keywords: Sparse coding, machine learning, heart sound signals, signal processing, classification, Matching Pursuit

Dedicatoria

***A mis padres y mis hermanas; a mis sobrinas
y a mis amigos.***

Agradecimientos

A los investigadores del Centro de Investigación Científica y de Educación Superior de Ensenada por brindarme de sus conocimientos en este proceso de formación.

Al Consejo Nacional de Ciencia y Tecnología (CONACyT) por brindarme el apoyo económico para realizar mis estudios de maestría. No. de becario: 634091

A mi director de tesis el Dr. Miguel Ángel Alonso, por compartir sus conocimientos y guiarme en este trabajo.

Tabla de contenido

	Página
Resumen en español	ii
Resumen en inglés	iii
Dedicatoria	iv
Agradecimientos	v
Lista de figuras	viii
Lista de tablas	xi
Capítulo 1. Introducción	
1.1. Objetivos	2
1.1.1. Objetivo general	2
1.1.2. Objetivos específicos	2
1.2. Justificación	2
1.3. Metodología de la investigación	3
1.4. Organización de la tesis	4
Capítulo 2. Antecedentes	
2.1. Los sonidos cardiacos	5
2.2. Estado del arte en clasificación del audio cardiaco	8
2.3. Trabajos relacionados	13
2.4. Propuesta de trabajo	13
Capítulo 3. Extracción de características	
3.1. Algoritmos de obtención de características	16
3.1.1. Promedio de características	16
3.1.2. Promedio de ciclos cardiacos	17
3.1.3. Promedio de ciclos cardiacos ajustados	18
3.2. Representación escasa	19
3.3. Codificación Predictiva Lineal	21
3.4. Coeficientes Cepstrales en la Escala de Frecuencias Mel	25
3.5. Características en el dominio del tiempo y de la energía	29
3.6. Características en el dominio de la frecuencia	31
3.7. Características en el residual de Matching Pursuit	33
3.8. Vector de características	34
3.9. Análisis de componentes principales	34
3.10. Evaluación de la importancia de las características	36
Capítulo 4. Técnicas de clasificación del audio cardiaco	
4.1. Regresión logística	37
4.2. Análisis discriminante lineal	39
4.3. K vecinos más cercanos	39

Tabla de contenido (continuación)

4.4.	Árbol de decisión	41
4.5.	Bayesiano Gaussiano ingenuo	42
4.6.	Máquinas de vectores de soporte	44
4.7.	Bosques aleatorios	48
Capítulo 5. Resultados		
5.1.	Matriz de entrenamiento y prueba	51
5.2.	Métricas para evaluación de los sistemas de clasificación	51
5.3.	Comparación de algoritmos de obtención de características	52
5.4.	Características y clasificadores	53
5.5.	Resultados con todas las características y reducción de la dimensio- nalidad.	56
5.6.	Resultados por base de datos	58
5.7.	Discusión sobre los resultados	58
Capítulo 6. Conclusiones		
6.1.	Trabajo futuro	62
Literatura citada		64

Lista de figuras

Figura	Página
1. Anatomía del corazón.	6
2. Representación del ciclo cardiaco (Arthur Guyton (2016), Cap. 9).	8
3. Densidad espectral de potencia de cada sonido cardiaco en un audio normal.	9
4. Umbral de audición y espectro del audio cardiaco (Arthur Guyton (2016), Cap. 23).	10
5. Diagrama a bloques del sistema de clasificación de audio cardiaco.	11
6. Par de ciclos de un sonido normal. Dentro del recuadro azul se puede observar como el S2 está mal segmentado, saliendose de los límites de la segmentación.	17
7. Promedio de ciclos de un AC normal. En el cuadro azul se puede observar como el S2 se sale de los límites de su segmentación, debido a que no todos los S2 de todos los ciclos del audio empiezan donde las longitudes promedio de los sonidos lo indica. La longitud total del ciclo promedio corresponde a la de el ciclo de mayor tamaño del AC.	18
8. Promedio de ciclos ajustados de un AC normal. La longitud total del ciclo corresponde al promedio de los ciclos del AC.	19
9. Algoritmo de Matching Pursuit.	21
10. Ejemplo de un átomo de Gabor con longitud de 1024 muestras.	22
11. a) Sistema de análisis de LPC. b) Sistema de síntesis de LPC (O'Shaughnessy, 1988).	23
12. Filtros triangulares utilizados en el cálculo de los MFCC.	26
13. Cruces por cero (Lartillot, 2018).	29
14. Roll-off. Se detecta la frecuencia por debajo de la cual se encuentra cierto porcentaje de la energía de la señal (Lartillot, 2018).	30
15. Brillo. Porcentaje de la energía de la señal que está por encima de una frecuencia dada (Lartillot, 2018).	30
16. En círculos rojos están marcados los eventos (Lartillot, 2018).	31
17. Residuo de la extracción de 8 átomos de Gabor.	34
18. Las flechas negras son los ejes principales de los datos. La longitud de estas es una medida de la varianza de los datos cuando son proyectados en el eje correspondiente.	35
19. En azul fuerte se encuentran los datos proyectados sobre un solo eje principal. Reduciéndose los datos de 2 dimensiones a una sola.	36

Lista de figuras (continuación)

Figura	Página
20. Función sigmoide que calcula la probabilidad de la clase. Los círculos representan diferentes casos de entrenamiento que le dieron forma al modelo, el eje horizontal representa la característica y el eje vertical la clasificación.	38
21. Se proyectan las mismas muestras en dos líneas diferentes marcadas como w . La figura en la derecha muestra mayor separación entre clases, cada clase de un color diferente (Duda <i>et al.</i> , 2000).	40
22. Se representa el algoritmo de decisión de K vecinos más cercanos de un clasificador binario con $K=1$. A la izquierda para un espacio bidimensional, a la derecha un espacio tridimensional (Duda <i>et al.</i> , 2000).	41
23. Árbol de decisión. Las preguntas se realizan empezando desde arriba, las hojas del árbol están representadas en rectángulos y es donde se toma la decisión final de la clasificación (Huang <i>et al.</i> , 2001).	42
24. Segmentación de los datos por medio de un árbol de decisiones (VanderPlas, 2016).	43
25. Probabilidad de las clases utilizando un modelo Bayesiano Gaussiano ingenuo (VanderPlas, 2016).	44
26. Dentro de los círculos negros se encuentran los vectores de soporte, las líneas entrecortadas son los márgenes (VanderPlas, 2016).	45
27. Se puede observar que agregar mayor cantidad de muestras en el entrenamiento de la máquina de vectores de soporte no modifica el modelo final, siempre que las nuevas muestras no sobrepasen su margen (VanderPlas, 2016).	46
28. Es imposible separar los datos con una función lineal, por lo que necesitan ser proyectados con un kernel a una dimensión superior (VanderPlas, 2016).	46
29. Una vez los datos son proyectados con un kernel pueden ser separados por una función lineal (VanderPlas, 2016).	47
30. Se puede observar que la segmentación que realizan los bosques aleatorios toman formas más complejas que se adecuan mejor a los datos que en el caso del árbol de decisión (VanderPlas, 2016).	49
31. Entre más árboles mayor es el desempeño. En cruces negras se presentan resultados de pruebas y en rojo la regresión logarítmica. Se consideró un desempeño de 1 al resultado obtenido con 10,000 árboles.	50

32. Análisis del residual de *Matching Pursuit* con MFCC. Se puede observar que la cantidad de átomos extraídos que da el mejor MAcc es 8. Los puntos probados están representados por puntos negros. 54

Lista de tablas

Tabla	Página
1.	Base de datos de Physionet. 13
2.	Audios cardiacos utilizados. 14
3.	Vector de características. 35
4.	Matriz de entrenamiento y de prueba. 51
5.	Matriz de confusión. Contiene la cantidad de resultados TP, FN, FP y TN. 52
6.	Resultados del promedio de características MFCC. En negritas se presentan los valores máximos. 53
7.	Resultados de características MFCC en promedio de ciclos cardiacos. En negritas se presentan los valores máximos. 53
8.	Resultados de características MFCC en promedio de ciclos cardiacos ajustados. En negritas se presentan los valores máximos. 53
9.	Resultados de delta MFCC. En negritas se presentan los valores máximos. 54
10.	Resultados de <i>Matching Pursuit</i> de 10 átomos. En negritas se presentan los valores máximos. 54
11.	Resultados del promedio de características LPC. En negritas se presentan los valores máximos. 55
12.	Resultados del promedio de características TIMEF. En negritas se presentan los valores máximos. 55
13.	Resultados de SPTR. En negritas se presentan los valores máximos. . 55
14.	Resultados de características MFCC en el residual de la extracción de 8 átomos de <i>Matching Pursuit</i> (MFCC8). En negritas se presentan los valores máximos. 55
15.	Resultados con todas las características y reducción de la dimensionalidad. (T.E.= tiempo de entrenamiento en segundos, T.P.=Tiempo de predicción en segundos). En negritas se presentan los valores máximos. 56
16.	Las 60 características más importantes. Entre más arriba y más a la izquierda tienen mayor importancia. 57
17.	Resultados de características con bosques aleatorios, evaluadas en cada base de datos. 58

Capítulo 1. Introducción

En México los gastos económicos provocados por epidemias de enfermedades cardiovasculares y de diabetes presentan una amenaza a futuro para la viabilidad del sistema público de salud. De acuerdo al Instituto Nacional de Estadística, Geografía e Informática (INEGI), en México las enfermedades cardiovasculares son la primera causa de muertes en el país provocando más de 140 mil decesos en 2017 (INEGI, 2018). La auscultación es una técnica no invasiva y de bajo costo que ayuda al médico a conocer de manera inmediata el estado del paciente y a decidir si son necesarias algunas otras pruebas más especializadas. De esta forma un cardiólogo experimentado puede distinguir patologías cardíacas y estimar su severidad utilizando solo un estetoscopio. El problema es que en México el número de especialistas con oído clínico entrenado es bastante reducido, lo cual vuelve difícil el acceso a un diagnóstico experto. Además los recursos y especialistas de la salud se encuentran concentrados en las áreas metropolitanas lo cual hace que el problema sea mayor en las zonas marginadas debido a la distancia entre el médico y el paciente.

Para solucionar el problema de la distancia se busca desarrollar servicios de telemedicina y telesalud, que faciliten el intercambio de recursos clínicos y consultas a distancia entre el paciente y el especialista. Como apoyo en la solución del problema de la falta de especialistas se han estado desarrollando sistemas de diagnóstico automático (Dwivedi *et al.*, 2019).

El propósito de un sistema como el desarrollado en esta tesis es servir como apoyo médico; el sistema no detecta patologías en concreto sino que detecta la presencia de audio "anormal".

El sistema ya implementado en el mundo real serviría como guía para, después de haber recibido una etiqueta de "anormal", dirigir al paciente a exámenes especializados en busca de alguna patología concreta.

1.1. Objetivos

1.1.1. Objetivo general

- Desarrollar un algoritmo de clasificación basado en un modelo de descomposición escasa de la señal de audio cardiaco.

1.1.2. Objetivos específicos

- Aprender los fundamentos de las técnicas de descomposición escasa de señales, en particular de los algoritmos de la familia *matching pursuit*.
- Revisar los diferentes métodos de clasificación de datos y seleccionar los algoritmos que resulten más adecuados para identificar el estado normal/anormal en la señal de audio cardiaco.
- Investigar el uso de técnicas de reducción de la dimensionalidad para disminuir la carga computacional de los algoritmos de clasificación.

1.2. Justificación

Una manera económica y sencilla de diagnosticar patologías cardiacas es mediante la auscultación, de la cual se obtienen señales de audio cardiaco que proporcionan información sobre la actividad del corazón y el estado de sus válvulas.

El audio cardiaco presenta ciertos sonidos que son considerados normales en un ciclo cardiaco sano (el sonido 1 y el sonido 2) pero debido a la presencia de patologías cardiacas se pueden encontrar sonidos adicionales, debido al flujo anormal de la sangre y/o al mal funcionamiento de las válvulas (Arthur Guyton (2016), Cap. 23). Estas

alteraciones en el audio cardíaco son las que dan el potencial de detección de patologías. En una situación normal, un médico con oído entrenado sería capaz de detectar estas patologías y brindar un diagnóstico; el problema es cuando no se dispone de un médico especialista que atienda a cada paciente.

Se pretende utilizar técnicas de clasificación automática de señales en conjunto con extractores de características que describan de manera eficiente a las señales normales y patológicas. Dentro de estas características se utilizará codificación escasa (*Matching Pursuit*), codificación predictiva lineal (LPC), coeficientes cepstrales en la escala de frecuencias Mel (MFCC), características en el dominio del tiempo y en el dominio de la frecuencia. Con el propósito de poder brindar un apoyo computacional a los médicos especialistas.

1.3. Metodología de la investigación

Para lograr los objetivos planteados en este trabajo se presentaron las siguientes etapas:

- Revisión de la bibliografía sobre el audio cardíaco y sistemas de clasificación automática.
- Investigación sobre representación escasa y modelado de sonidos cardíacos.
- Investigación sobre codificación predictiva lineal y otros tipos de extractores de características.
- Extracción de características en la base de datos y construcción de la matriz de características.
- Investigación de diferentes tipos de clasificadores.
- Evaluación de los diferentes tipos de características con los clasificadores.
- Evaluación de las características delta-MFCC.
- Evaluación del residual de *Matching Pursuit* para clasificación.

- Evaluación de técnicas de reducción de la dimensionalidad en el vector de características.
- Evaluación de la clasificación en las diferentes bases de datos.

1.4. Organización de la tesis

En el capítulo 2 se tratan algunos temas previos al trabajo presente como son el origen mecánico de los sonidos cardiacos, el origen histórico de la auscultación cardiaca, trabajos relacionados en clasificación del audio cardiaco y trabajos desarrollados en CICESE relacionados directamente con este. Además se describe la base de datos de audio cardiaco a utilizar en este trabajo y la propuesta de trabajo.

En el capítulo 3 se describen los diferentes algoritmos para la extracción de las características que describen los audios cardiacos a través de los ciclos, también se explica qué se extrae con cada característica, se retrata de qué características está compuesto el vector de características final y además las técnicas de reducción de la dimensionalidad.

En el capítulo 4 se presentan las diferentes técnicas de clasificación a utilizar con los diferentes tipos de características.

En el capítulo 5 se describen las métricas a utilizar para la evaluación del desempeño de los diferentes algoritmos de clasificación; además de describir cómo serán construidas las matrices de entrenamiento y de prueba. Se muestran los resultados de la combinación de cada tipo de característica expuesta en el capítulo 3 con cada tipo de clasificador presentado en el capítulo 4. Se presentan los resultados utilizando todos los tipos de características en una sola matriz e implementando el clasificador que mejor desempeño mostró, para posteriormente usar técnicas de reducción de la dimensionalidad. Se analizan los resultados con las diferentes bases de datos.

El presente trabajo finaliza en el capítulo 6 con conclusiones sobre el trabajo realizado, además se exponen algunas ideas que podrían servir para el desarrollo de trabajos futuros en el área de clasificación de audio cardiaco.

Capítulo 2. Antecedentes

2.1. Los sonidos cardiacos

La sangre es un elemento vital para la subsistencia de las células (Arthur Guyton (2016), Cap. 14). Es necesaria su circulación en el cuerpo a través de los vasos sanguíneos. El corazón sirve como sistema de bombeo de la sangre para su flujo a través de los vasos sanguíneos, y es debido a este flujo y el cerrado y apertura de válvulas en el corazón por lo que se generan los sonidos cardiacos (SC).

El corazón

El corazón, una representación esquemática se muestra en la Figura 1, está formado por dos bombas separadas, la bomba derecha que bombea sangre hacia los pulmones y la bomba izquierda que aporta flujo sanguíneo a los demás órganos (Arthur Guyton (2016), Cap. 9). Cada una de estas bombas está dividida en dos partes que hacen fluir la sangre por medio de contracciones musculares, una aurícula y un ventrículo. Cada una de las aurículas es una bomba débil que alimenta al ventrículo correspondiente. Los ventrículos después aportan la mayor fuerza de bombeo hacia los pulmones (ventrículo derecho) y hacia los demás órganos (ventrículo izquierdo).

La sangre, en el corazón, entra por las aurículas y a través de válvulas unidireccionales (Figura 1) fluye desde las aurículas hacia los ventrículos y desde los ventrículos hacia las arterias. La apertura y cierre de las válvulas se debe a diferencias de presión entre cada uno de sus lados.

Ciclo cardiaco y sus sonidos

Se denomina ciclo cardiaco (CC) a los fenómenos que se producen desde el inicio de un latido hasta el comienzo del siguiente (Arthur Guyton (2016), Cap. 9). El CC está dividido en dos periodos, un periodo de relajación denominado diástole y un periodo de contracción denominado sístole. En la Figura 2 se representan los acontecimientos

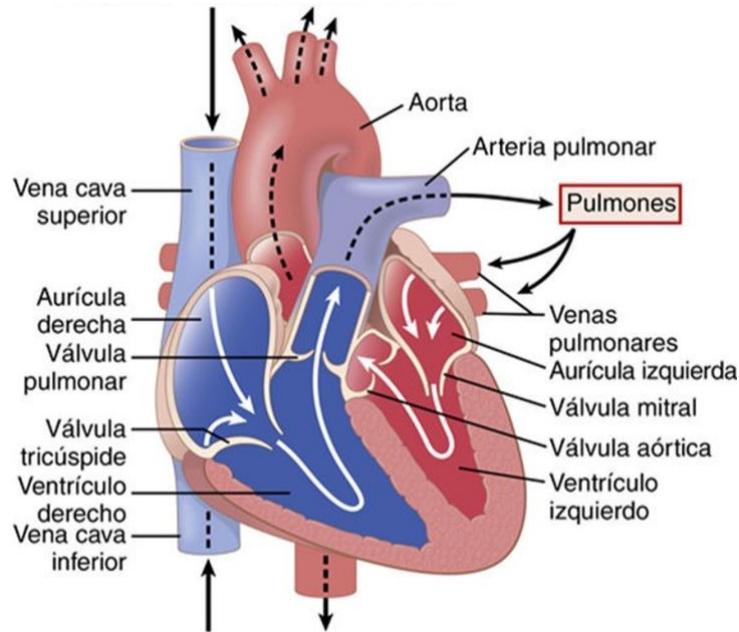


Figura 1. Anatomía del corazón.

que se producen durante un ciclo cardiaco en el lado izquierdo del corazón, las tres curvas superiores muestran los cambios de presión en la aorta, ventrículo izquierdo y aurícula izquierda, respectivamente. La cuarta curva representa los cambios de volumen ventricular izquierdo, la quinta curva es un electrocardiograma (ECG, representa el potencial eléctrico), y la sexta curva es un fonocardiograma (FCG, representa los sonidos que produce el corazón).

Al inicio de un CC los ventrículos están llenos casi en su totalidad, y la diferencia de presión entre los ventrículos y la arteria aórtica mantiene cerradas las válvulas pulmonar y aórtica; mientras que la diferencia de presión entre las aurículas y los ventrículos mantiene abiertas las válvulas tricúspide y mitral. La onda P marca el inicio del CC iniciando la contracción de las aurículas lo cual produce un aumento de volumen y presión de las ventrículas. Poco después se producen las ondas QRS que inician la contracción de las ventrículas, generando un aumento en la presión ventricular. La presión elevada en los ventrículos produce el cierre de las válvulas tricúspide y mitral, produciendo el primer sonido cardiaco S1. El cierre de estas válvulas produce un aumento en la presión ventricular que cuando acumula una presión suficiente para superar la presión de las arterias produce la apertura de las válvulas pulmonar y aórtica,

comenzando la expulsión de sangre hacia las arterias. A causa de esto las ventrículas comienzan a perder presión que ocasiona el cierre súbito de las válvulas pulmonar y aórtica produciendo el segundo sonido cardiaco S2. El CC termina cuando se tiene una mayor presión en las aurículas que en las ventrículas lo cual ocasiona la apertura de las válvulas tricúspide y mitral.

S1 marca el inicio de la sístole y se compone de dos sonidos principales: el de la válvula mitral y el de la válvula tricúspide. S2 marca el inicio de la diástole y se compone de dos sonidos principales: el de la válvula pulmonar y el de la válvula aórtica (Arthur Guyton (2016), Cap. 24).

Aunque los SC S1 y S2 son los sonidos principales que se esperan escuchar en cualquier audio cardiaco (AC), también se pueden encontrar otros sonidos que muchas veces se relacionan con cardiopatías (Arthur Guyton (2016), Cap. 23):

Tercer sonido: Un sonido débil y grave que ocurre después de S2, a veces inaudible. Se dice que se debe al flujo de sangre que entra a las ventrículas. Este sonido suele estar presente sin significar una patología en niños, adolescentes, y adultos jóvenes, aunque en general en adultos mayores indica insuficiencia cardiaca.

Cuarto sonido: Es muy raro que se escuche este sonido con un estetoscopio debido a su debilidad y baja frecuencia. Se produce cuando las aurículas se contraen, al igual que el tercer sonido se dice que se debe a flujo acelerado que entra a las ventrículas.

Soplos cardiacos: Son sonidos cardiacos producidos por flujo cardiaco turbulento debido a alteraciones de las válvulas.

Chasquido: Producido por el cierre en destiempo de alguna válvula o debido a la tensión de esta durante la contracción.

Fricciones: Producido por el rozamiento del corazón inflamado con los órganos que le rodean.

Los sonidos cardiacos se pueden diferenciar por su contenido espectral (Figura 3). Los componentes en frecuencia de los SC suelen estar por debajo de los 100 Hz, aunque S2 tiende a tener componentes más altos que S1, por lo que tiene un ancho de

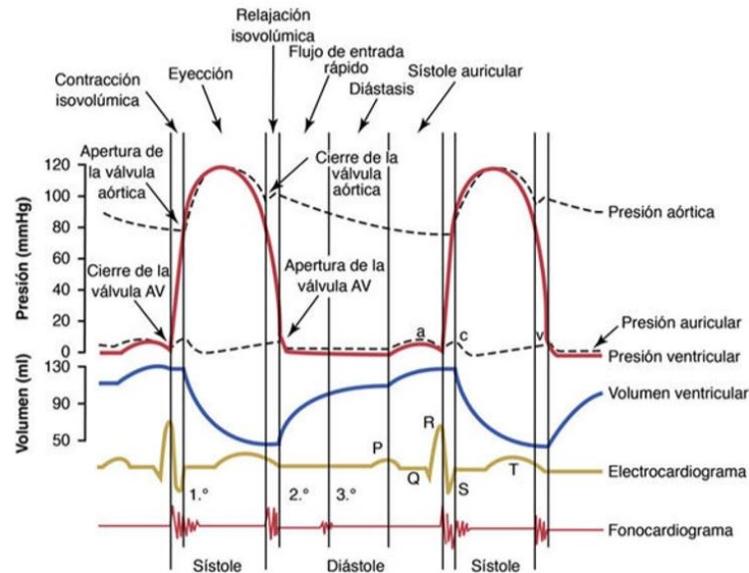


Figura 2. Representación del ciclo cardíaco (Arthur Guyton (2016), Cap. 9).

banda mayor aunque suele tener menor potencia que S1 que es donde se concentra la mayor cantidad de energía del AC.

2.2. Estado del arte en clasificación del audio cardíaco

Análisis clínico

Robert Hooke, en el siglo XVII, fue la primera persona en atribuir potencial de diagnóstico a la auscultación cardíaca debido a que fue capaz de escuchar claramente los sonidos provenientes del corazón y también planteó que podría ser posible indagar en el funcionamiento de otros órganos a partir de sus sonidos (Wood y Barry, 1995; Luisada, 1959).

El estetoscopio fue inventado por Lænnec en 1816, convirtiendo rápidamente la auscultación en un instrumento esencial en el diagnóstico clínico que se utiliza hasta la actualidad (Roguin, 2006). Sin embargo todavía no se tenía un buen entendimiento

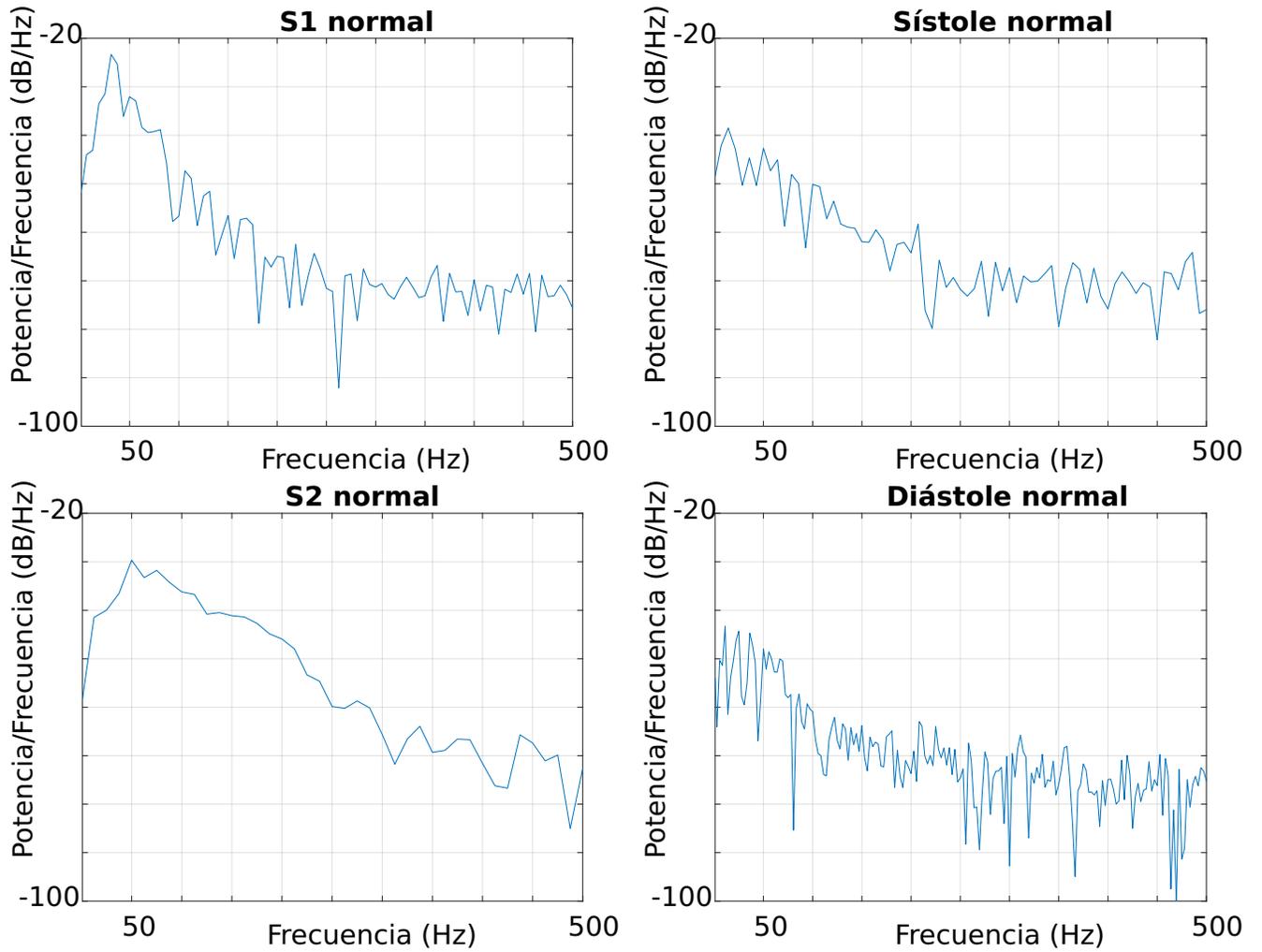


Figura 3. Densidad espectral de potencia de cada sonido cardiaco en un audio normal.

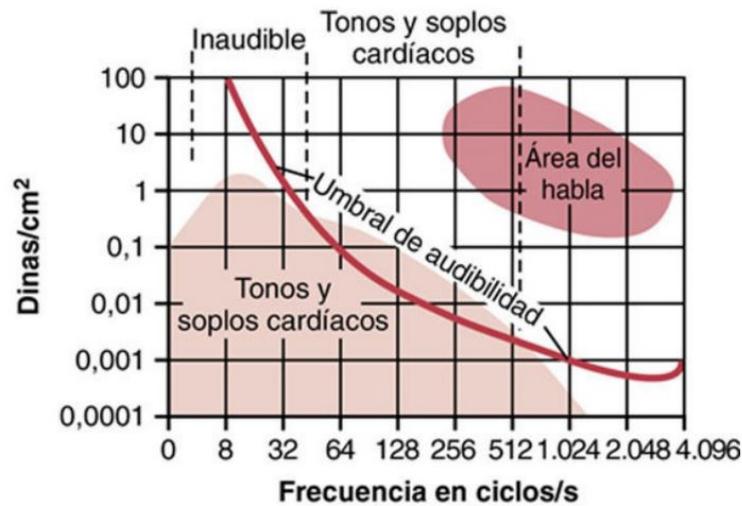


Figura 4. Umbral de audición y espectro del audio cardíaco (Arthur Guyton (2016), Cap. 23).

de los mecanismos de los sonidos cardíacos debido a la falta de estandarización en la nomenclatura, transductores y la localización de las mediciones (Wood y Barry, 1995). Læennec investigó los sonidos del corazón y de los pulmones con su invento y encontró que sus diagnósticos concordaban con observaciones hechas en autopsias. En 1819 publicó el primer trabajo en el uso de escuchar los sonidos del cuerpo, titulado "De l'auscultation médiate ou Traité du Diagnostic des Maladies des Poumon et du Coeur".

Sin embargo, algunos factores han demostrado que la auscultación es una herramienta limitada, debido en parte a que el oído humano solo es capaz de detectar una fracción de la energía del AC como se ve en la Figura 4. Para mejorar la detección de patologías se han empleado grabaciones de AC, con la ventaja de poder hacer una inspección visual de la señal. Actualmente con los avances del procesamiento digital de señales se han empezado a emplear herramientas computacionales para el análisis de las señales (Mahnke, 2009).

Diagnóstico asistido por computadora

Con el fin de desarrollar herramientas de apoyo para el diagnóstico de patologías cardíacas se han realizado una gran cantidad de trabajos utilizando procesamiento

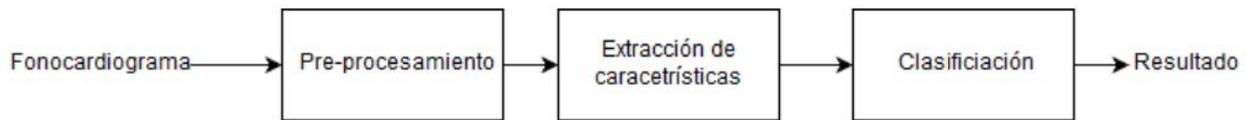


Figura 5. Diagrama a bloques del sistema de clasificación de audio cardíaco.

computacional; los sistemas de clasificación, como se describe en Duda *et al.* (2000), generalmente empiezan por un pre-procesamiento de la señal de AC, seguido por la extracción de características que describan de forma adecuada la señal y que luego puedan ser analizadas por un clasificador (Figura 5).

Uno de los pasos de pre-procesamiento más importantes es la segmentación del AC, debido a que contiene diferentes eventos que se pueden dividir en S1, S2 y fases de reposo (visto en la sección 2.1). Se busca identificar el inicio y fin de cada uno de estos eventos, para posteriormente extraer sus características, es común la utilización de algoritmos de segmentación como el de Springer *et al.* (2016). Aunque existen algunos algoritmos de clasificación que no utilizan la segmentación del AC (Zabihi *et al.*, 2016). También es común disminuir la frecuencia de muestreo, debido a que la mayor cantidad de energía del AC suele estar en las frecuencias bajas entre 20 y 400 Hz, por lo que algunos métodos sugieren incluso utilizar una frecuencia de muestreo de 1 KHz (Abdollahpur *et al.*, 2016). Algunos trabajos también realizan un análisis de la calidad de los audios cardíacos durante el pre-procesamiento (Zabihi *et al.*, 2016).

Después del pre-procesamiento se realiza la extracción de características. Se utilizan características de los siguientes tipos (se explican más a detalle en el capítulo 3):

En el dominio de la frecuencia: se refiere a características que describen el comportamiento en frecuencia del AC, ya sea el ancho de banda, energía en el dominio de la frecuencia, o características estadísticas en la frecuencia como oblicuidad o planitud (Homsí *et al.*, 2016).

En el dominio del tiempo: describen el comportamiento del PCG desde el dominio del tiempo. Características como el ritmo cardíaco, cruces por zero o duración en el

tiempo (Homsí *et al.*, 2016).

Perceptuales: son características que describen la señal basadas en la percepción auditiva humana, un ejemplo son los coeficientes cepstrales en la escala de frecuencias Mel (MFCC) (Kay y Agarwal, 2016).

Codificación escasa: busca representar una señal con una combinación lineal de unas cuantas formas de onda (Whitaker y Anderson, 2016).

Una vez que se tienen las características se analizan en algún esquema de clasificación. En este caso, se han utilizado redes neuronales (Zabihi *et al.*, 2016), máquinas de soporte vectoriales (Whitaker y Anderson, 2016), bosques aleatorios (Banerjee *et al.*, 2016), regresión logística o incluso esquemas de votación que utilizan diferentes tipos de clasificadores y la decisión final se realiza a partir del resultado de cada clasificador, ya sea el resultado con más votos o asignando algún valor de ponderación a cada clasificador (Nilanon *et al.*, 2016).

Base de datos de audio cardíaco

El análisis de señales de AC en las últimas décadas, especialmente para su segmentación y clasificación automática, ha reportado tener potencial para la detección automática de patologías en aplicaciones clínicas. Una gran ayuda para la comparación de los diversos algoritmos que han estado surgiendo es utilizar bases de datos de grabaciones de audio cardíaco abiertas al público. Actualmente, la base de datos de AC con mayor cantidad de grabaciones es la base de PhysioNet (Liu *et al.*, 2016). Existen otras bases de datos como las de PASCAL (P. Bentley y Getz., 2011), *The open Michigan heart sound & murmur library (OMHSML)* (Judge y Mangrulkar., 2015), y *The cardiac auscultation of heart murmurs database* (eGeneral Medical, s.f.).

2.3. Trabajos relacionados

En CICESE se han realizado trabajos sobre la representación y segmentación del audio cardiaco utilizando Matching Pursuit (MP) y átomos de Gabor. En estos trabajos se encontró que este tipo de átomos son de gran utilidad para modelar sonidos de audio cardiaco (Nieblas-Hernández, 2014).

En otro trabajo se implementó codificación predictiva lineal (LPC) en el residual de MP, de esta forma con MP se modela la parte determinística del audio cardiaco, y con LPC se modela la parte estocástica (Ibarra-Hernandez, 2014).

También se desarrolló un método de segmentación de audio cardiaco con análisis tiempo-frecuencia (Cruz Gutierrez, 2016).

2.4. Propuesta de trabajo

Base de datos de Physionet

En esta tesis de investigación se utilizará la base de datos de Physionet (Liu *et al.*, 2016), debido a que es el conjunto de sonidos cardiacos más grande disponible actualmente y además es de acceso libre, para realizar la evaluación de los diversos algoritmos de clasificación de AC. En la Tabla 1 se puede observar cómo está compuesta esta base de datos.

Tabla 1. Base de datos de Physionet.

Base de datos	Número de:		Porcentaje de audios:			Número de:
	pacientes	grabaciones	anormales	normales	inseguras	
Entrenamiento-a	121	409	67.5	28.4	4.2	14559
Entrenamiento-b	106	490	14.9	60.2	24.9	3353
Entrenamiento-c	31	31	64.5	22.6	12.9	1808
Entrenamiento-d	38	55	47.3	47.3	5.5	853
Entrenamiento-e	356	2054	7.1	86.7	6.2	59593
Entrenamiento-f	112	114	27.2	68.4	4.4	4259
Total	764	3153	18.1	73	8.8	84425

No forma parte del propósito de esta tesis implementar un algoritmo de segmen-

tación del ciclo cardíaco, por lo que se utilizaron solo los audios cardíacos de la base de datos Physionet que estaban ya segmentados por el proveedor, y solo aquellos ciclos cardíacos que están bien definidos por la segmentación. En la Tabla 2 se puede observar cuántos audios cardíacos sí fueron utilizados para la evaluación de cada algoritmo.

Tabla 2. Audios cardíacos utilizados.

Base de datos	Número de audios:			Porcentaje de audios:	
	Totales	Normales	Anormales	Normales	Anormales
Entrenamiento-a	392	116	276	29.59	70.41
Entrenamiento-b	368	295	73	80.16	19.84
Entrenamiento-c	27	7	20	25.93	74.07
Entrenamiento-d	52	26	26	50	50
Entrenamiento-e	1926	1780	146	92.42	7.58
Entrenamiento-f	108	77	31	71.3	28.7
Total	2873	2301	572	80.09	19.91

Etapas de los algoritmos

Las etapas de los algoritmos están descritas en el diagrama de la Figura 5. Se realizó un mínimo de pre-procesamiento a la señal (muestreo a 2 kHz y ajuste de *offset* al extraer características con *Matching Pursuit*).

Se pretenden evaluar varios extractores de características junto con nueve clasificadores diferentes, los extractores están descritos en el capítulo 3 y los clasificadores en el capítulo 4.

Primero se formará una matriz de características de los 2873 audios cardíacos que contienen segmentación. Cada línea de esta matriz representa las características de un audio cardíaco diferente, y contiene su respectiva etiqueta ("normal" o "anormal"). A partir de esta matriz se formarán las matrices de entrenamiento y de prueba.

Cada clasificador será entrenado con una matriz de características de entrenamiento generada de manera aleatoria y balanceada, con 50% de señales "anormales" y 50% de señales "normales". Y también se evaluará su rendimiento con una matriz de prueba balanceada.

Contribuciones

Se evaluará la utilización del residual de Matching Pursuit , con la hipótesis de que los átomos de Gabor (que fueron utilizados en trabajos anteriores, sección 2.3) tienen la cualidad de estar correlacionados con la parte normal de un audio cardíaco, dejando solo el audio anormal en el residual.

También se evaluarán los extractores Delta-MFCC, de los cuales no se ha encontrado algún trabajo en la literatura donde se utilicen para la clasificación de audio cardíaco.

Capítulo 3. Extracción de características

El aprendizaje automático adecua modelos de clasificación a datos para poder realizar predicciones. Estos modelos tienen que ser alimentados con características. Una característica es una representación numérica de algún aspecto de los datos crudos (Zheng y Casari, 2018).

La elección de las características para una función deseada es probablemente el problema más importante al que se puede enfrentar un diseñador de un algoritmo de aprendizaje. Además de influir en la naturaleza del algoritmo de aprendizaje, puede determinar si un problema se puede solucionar o no, o qué tan bien se está describiendo el problema para que el algoritmo de aprendizaje trabaje con él (Liu y Motoda, 1998).

3.1. Algoritmos de obtención de características

Para realizar la extracción de las características se tomaron en cuenta tres formas de extracción, descritas a continuación.

3.1.1. Promedio de características

Este método de extracción toma las características de cada ciclo en un audio cardiaco, posteriormente se obtiene el promedio de cada característica a lo largo de los ciclos. Este promedio de las características representa al audio cardiaco y es el que se utiliza para formar la matriz de características. Este método de extracción es el que requiere de mayor procesamiento computacional, además algunos ciclos cardiacos pueden presentar una segmentación errónea de los sonidos, como se puede observar en la Figura 6.

Algoritmo:

1. Se detectan los ciclos de acuerdo a la segmentación proporcionada.

2. Se extraen las características de cada ciclo en cada sonido cardiaco.
3. Se crea una matriz (número de ciclos, número de características) para cada sonido cardiaco.
4. Se obtiene el promedio de cada característica (promedio de cada columna en la matriz).

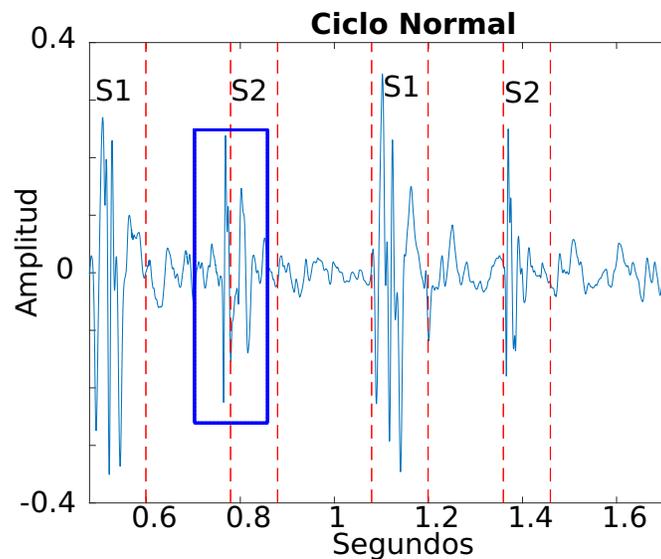


Figura 6. Par de ciclos de un sonido normal. Dentro del recuadro azul se puede observar como el S2 está mal segmentado, saliéndose de los límites de la segmentación.

3.1.2. Promedio de ciclos cardiacos

Primero se detectan los ciclos y posteriormente se calcula un "superciclo" que es el promedio de los ciclos del audio cardiaco. Se ve como una alternativa debido a que reduce en gran medida el coste computacional en comparación al método de promedio de características (sección 3.1.1). Aunque debido a que cada ciclo cardiaco tiene diferentes tamaños de sus sonidos, estos suelen deformarse en el "superciclo" resultante, como se ve en la Figura 7.

Algoritmo:

1. Se detectan los ciclos de acuerdo a la segmentación proporcionada.

2. Se crea una matriz (número de ciclos, longitud del ciclo + *zero padding*) para cada sonido cardiaco.
3. Se obtiene el promedio de los ciclos y se elimina el *zero padding* excedente.
4. Se extraen las características al ciclo resultante, tomando en cuenta la longitud promedio de cada sonido cardiaco para la segmentación.

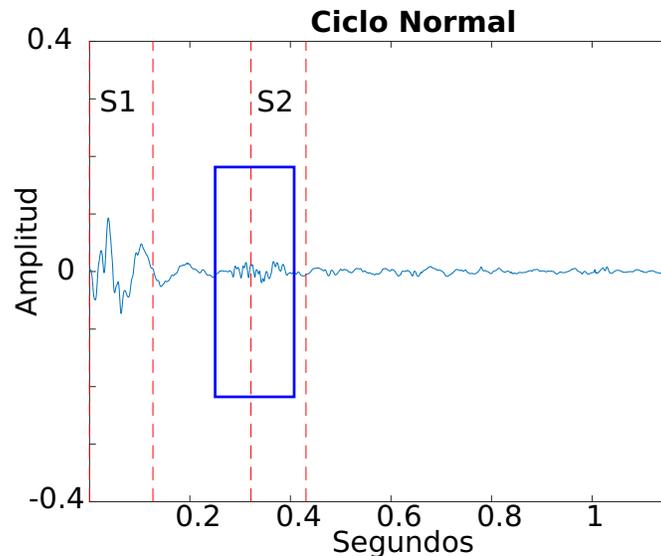


Figura 7. Promedio de ciclos de un AC normal. En el cuadro azul se puede observar como el S2 se sale de los límites de su segmentación, debido a que no todos los S2 de todos los ciclos del audio empiezan donde las longitudes promedio de los sonidos lo indica. La longitud total del ciclo promedio corresponde a la de el ciclo de mayor tamaño del AC.

3.1.3. Promedio de ciclos cardiacos ajustados

El algoritmo es el mismo que en el promedio de ciclos cardiacos (sección 3.1.2), pero se le agrega un ajuste antes de la obtención del promedio del ciclo. Se obtiene el promedio de la longitud de cada sonido cardiaco, y cada sonido cardiaco en cada ciclo es ajustado a su longitud respectiva utilizando un método de interpolación polinomial. De esta forma se corrige el desajuste de la segmentación (Figura 8).

Algoritmo:

1. Se detectan los ciclos de acuerdo a la segmentación proporcionada.

2. Se obtiene el promedio de la duración de cada sonido cardiaco en cada grabación, cada sonido cardiaco es ajustado para que su duración corresponda al promedio.
3. Se crea una matriz (número de ciclos, longitud promedio de los ciclos) para cada sonido cardiaco.
4. Se obtiene el promedio de los ciclos.
5. Se extraen las características al ciclo resultante, tomando en cuenta la longitud promedio de cada sonido cardiaco para la segmentación.

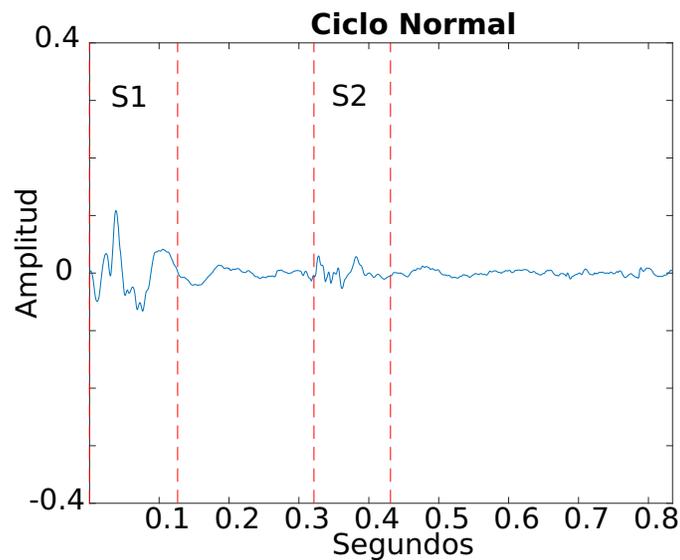


Figura 8. Promedio de ciclos ajustados de un AC normal. La longitud total del ciclo corresponde al promedio de los ciclos del AC.

3.2. Representación escasa

Matching Pursuit (MP) es parte de una clase de algoritmos de análisis de señales conocidas como descomposición atómica (Krstulovic y Gribonval, 2006; Mallat y Zhang, 1993). Estos algoritmos consideran una señal x como una combinación lineal de señales elementales w_m , conocidas como átomos, seleccionadas de un diccionario D :

$$x = \sum_{m=1}^M \alpha_m w_m, \quad \text{donde } w_m \in D. \quad (1)$$

Usualmente, el diccionario D es sobrecompleto en la dimensión N , esto significa que D tiene más de N elementos y abarca el espacio entero. En este caso, la descomposición no es única, puede haber un número infinito de soluciones. Sobre todas las posibles descomposiciones las preferidas son las más compactas (o escasas), lo que quiere decir que solo los primeros pocos átomos en la ecuación, acomodadas de mayor peso a menor peso α_m , son necesarias para obtener una buena aproximación de la señal. En general, entre más grande sea el diccionario mayor número de soluciones potenciales y mayor será la oportunidad de encontrar una aproximación compacta de la señal original.

Algoritmo de Matching Pursuit

En cada iteración m , el algoritmo de MP busca el átomo w_m con mayor correlación con la señal x , el cual es el que tenga el mayor producto interno absoluto con la señal original. Se siguen los siguientes pasos (Figura 9):

1. Inicialización de las variables: $m = 0$, $x_m = x_0 = x$.
2. Cálculo de las correlaciones entre la señal x_m y cada átomo en D , usando el producto interno:

$$\forall w \in D : \text{CORR}(x_m, w) = |\langle x_m, w \rangle|. \quad (2)$$

3. Se busca el átomo más correlacionado, buscando el producto interno máximo:

$$\widehat{w}_m = \underset{w \in D}{\operatorname{argmax}} \text{CORR}(x_m, w). \quad (3)$$

4. Sustracción del átomo ponderado $\alpha \widehat{w}_m$ de la señal x_m :

$$x_{m+1} = x_m - \alpha_m \widehat{w}_m, \quad \text{donde } \alpha_m = \langle x_m, \widehat{w}_m \rangle. \quad (4)$$

5. Si se alcanzó el nivel de precisión, ya sea en términos del número de átomos o en la proporción de energía entre la señal original y el residual x_{m+1} , el algoritmo se detiene aquí; de otra manera se realiza una nueva iteración sobre el residual: $m \leftarrow m + 1$ y se empieza en el paso 2.

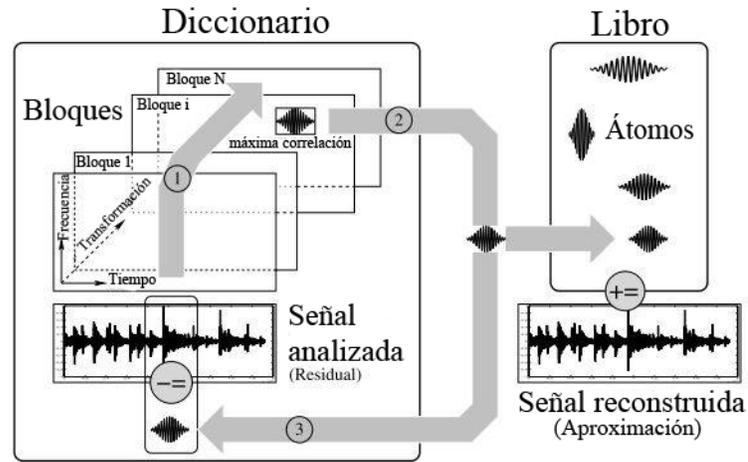


Figura 9. Algoritmo de Matching Pursuit.

Diccionario de Gabor

Para realizar el diccionario se implementaron átomos de Gabor (ver Figura 10), los cuales en un trabajo previo se demostró que tienen una buena correlación con el audio cardiaco (Nieblas-Hernández (2014)).

Cada átomo de Gabor es representado a partir de 3 características, las cuales son: posición u , longitud s y frecuencia ξ ; se pueden observar en la siguiente ecuación:

$$g_{\gamma}(t) = \frac{1}{\sqrt{s}} w\left(\frac{t-u}{s}\right) e^{i2\pi\xi(t-u)}, \quad (5)$$

donde w es una ventana Gaussiana $w(t) = \sqrt[4]{2}e^{-\pi t^2}$. Además, a este átomo se le hace un ajuste de amplitud, por lo que por cada átomo se están extrayendo 4 características. Al representar una señal con 10 átomos se estarían extrayendo un total de 40 características.

3.3. Codificación Predictiva Lineal

Considerando una señal $s(t)$ muestreada cada T segundos, de tal manera que $s(n) = s(nT)$ para $n \in \mathbb{Z}$. Esta señal se puede representar en el dominio z como $S(z)$

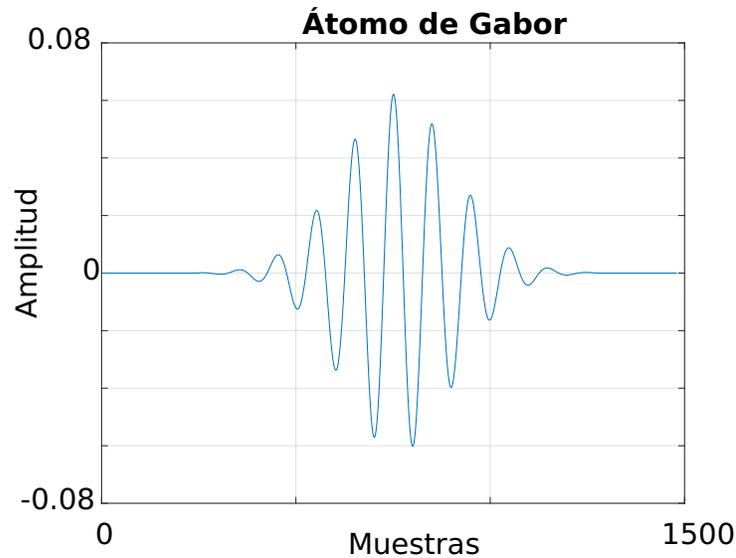


Figura 10. Ejemplo de un átomo de Gabor con longitud de 1024 muestras.

y $s(n)$ se puede modelar a partir de una fuente de excitación $U(z)$ como entrada a un filtro conformador espectral $H(z)$, para producir $S(z) = U(z)H(z)$. La codificación predictiva lineal (LPC) de $s(n)$ "deconvoluciona" la señal de excitación y el filtro $\hat{U}(z)$ y $\hat{H}(z)$ siguiendo la restricción de que su producto sea lo más cercano posible en un sentido de medio cuadrático a la señal original $s(n)$ (O'Shaughnessy, 1988).

Para simplificar el problema del modelado, normalmente a $\hat{U}(z)$ se le asigna una envolvente espectral plana de tal forma que solo se tiene que calcular $\hat{H}(z)$. Para simplificar el cálculo de $\hat{H}(z)$ se considera que $s(n)$ es estacionaria durante una ventana de tiempo de N muestras. El filtro $H(z)$ puede ser modelado con coeficientes constantes. En este caso, se asume que $H(z)$ tiene p polos y q zeros en un caso general. Esto quiere decir que $\hat{s}(n)$ es una combinación lineal de las p muestras previas de salida y las $q + 1$ muestras previas de entrada de un sintetizador LPC:

$$\hat{s}(n) = \sum_{k=1}^p a_k \hat{s}(n-k) + G \sum_{l=0}^q b_l \hat{u}(n-l), \quad (6)$$

donde G es un factor de ganancia para la excitación (asumiendo $b_0 = 1$). Equivalentemente:

$$\hat{H}(z) = \frac{\hat{S}(z)}{\hat{U}(z)} = G \frac{1 + \sum_{l=1}^q b_l z^{-l}}{1 - \sum_{k=1}^p a_k z^{-k}}. \quad (7)$$

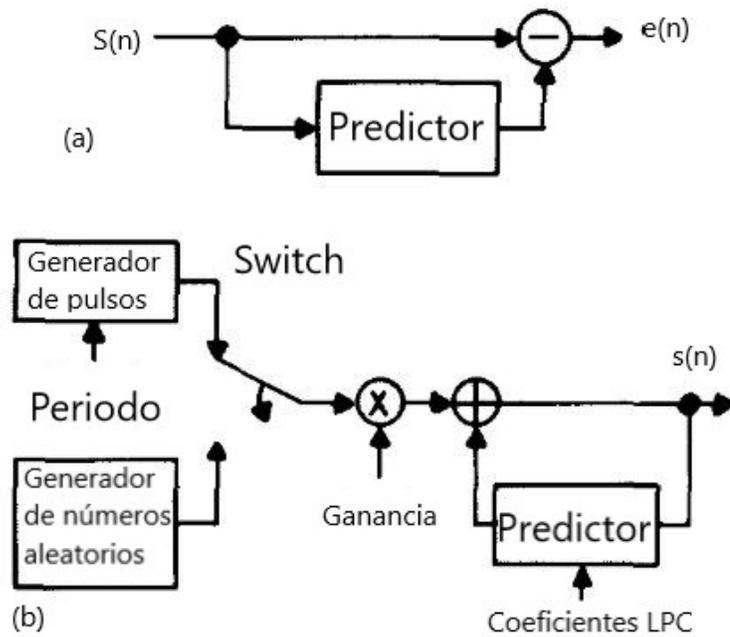


Figura 11. a) Sistema de análisis de LPC. b) Sistema de síntesis de LPC (O'Shaughnessy, 1988).

La mayoría de trabajos asumen un modelo de solo polos llamado modelo autoregresivo (AR), donde $q = 0$. A un modelo de solo ceros donde $p = 0$ se le denomina modelo de promedio móvil (MA) debido a que la salida es un promedio ponderado de las $q + 1$ entradas posteriores. El modelo LPC general con polos y ceros es denominado modelo de media móvil autoregresiva (ARMA). El modelo AR LPC es el más común debido a que las señales físicas en la naturaleza suelen ser caracterizadas por la presencia de energía a ciertas frecuencias más que por su ausencia. Por lo tanto las señales con picos espectrales son más comunes que aquellas con ceros.

Si $s(n)$ se filtra por la inversa de un filtro de síntesis $\hat{H}(z)$ (ecuación 8), la salida $e(n)$ es llamada la señal de error (ecuación 9):

$$A(z) = 1 - \sum_{k=1}^p a_k z^{-k}, \quad (8)$$

$$e(n) = s(n) - \sum_{k=1}^p a_k s(n-k). \quad (9)$$

Entre más parecidos sean $E(z)$ y $\widehat{H}(z)$ el modelo será considerado más adecuado. En audio no vocalizado la señal de error debe emular ruido blanco, y en la entrada del sintetizador LPC se utiliza un generador de números pseudo-aleatorios.

Existen dos enfoques para obtener los coeficientes a_k de LPC de un modelo de solo polos $\widehat{H}(z)$. El método clásico de mínimos cuadrados selecciona a_k para minimizar la energía media en $e(n)$ sobre una ventana de datos, mientras que el enfoque enrejado permite una actualización instantánea de los coeficientes. En la técnica antigua, tanto $s(n)$ como $e(n)$ son ventaneadas para limitar la extensión de la señal bajo análisis. El análisis de ventanas consecutivas permiten seguir los cambios espectrales en una señal variante en el tiempo.

La primera de dos técnicas de mínimos cuadrados es el método de autocorrelación, el cual multiplica la señal por una ventana de tiempo $w(n)$, ya sea de Hamming o similar, por lo que $x(n) = w(n)s(n)$ tiene una duración finita de N muestras. Así, $x(n) = 0$ fuera de el rango $0 \leq n \leq N - 1$. LPC modela todas las muestras de $x(n)$ en cada ventana; así cuando la señal es no estacionaria, los coeficientes LPC describen un promedio suavizado. Siendo E la energía del error:

$$E = \sum_{-\infty}^{\infty} e^2(n) = \sum_{-\infty}^{\infty} [x(n) - \sum_{k=1}^p a_k x(n-k)]^2, \quad (10)$$

donde $e(n)$ es el residual correspondiente a la señal ventaneada $x(n)$. Los valores de a_k que minimizan E se encuentran asignando a $\delta E / \delta a_k = 0$ los valores $k = 1, 2, 3, \dots, p$. Esto produce p ecuaciones lineales ($i = 1, 2, 3, \dots, p$).

$$\sum_{n=-\infty}^{\infty} x(n-i)x(n) = \sum_{k=1}^p a_k \sum_{n=-\infty}^{\infty} x(n-i)x(n-k), \quad (11)$$

en p no conocidos a_k . Siendo el primer término la autocorrelación $R(i)$ de $x(n)$ y $x(n)$ teniendo duración finita,

$$\sum_{k=1}^p a_k R(i-k) = R(i), \quad 1 \leq i \leq p, \quad (12)$$

donde:

$$R(i) = \sum_{n=i}^{N-1} x(n)x(n-i). \quad (13)$$

El método de covarianza de mínimos cuadrados ventanea la señal de error $e(n)$ en vez de $s(n)$,

$$E = \sum_{-\infty}^{\infty} e^2(n)w(n). \quad (14)$$

Usualmente el error es ponderado uniformemente en el tiempo con una ventana rectangular de N muestras, que en efecto reemplaza $R(i)$ con la función de covarianza:

$$\phi(i, k) = \sum_{n=0}^{N-1} s(n-k)s(n-i). \quad (15)$$

Las dos técnicas varían en los efectos de ventaneado, lo cual les da diferencias en el cálculo y en la estabilidad de los filtros de síntesis. El enfoque en autocorrelación genera distorsión en el procedimiento de estimación ya que el ventaneado en el tiempo maneja tanto el espectro de tiempo corto de la señal original como la respuesta en frecuencia de la ventana. La mayoría de las ventanas tienen una respuesta en frecuencia pasa-bajas; así, el espectro de la señal ventaneada es una versión suavizada de la versión original. La extensión y el tipo de suavizado depende de la forma y duración de la ventana.

En este trabajo se obtienen 10 coeficientes LPC para representar cada ventana de tiempo en cada uno de los cuatro sonidos cardiacos, se obtiene el promedio de los 10 coeficientes para cada sonido cardiaco dando como resultado un total de 40 coeficientes LPC por audio cardiaco.

3.4. Coeficientes Cepstrales en la Escala de Frecuencias Mel

Los coeficientes cepstrales en la escala de frecuencias Mel (MFCC) son una representación definida como el cepstrum real de una señal de tiempo corto ventaneada derivada de la transformada rápida de Fourier de esa señal (Huang *et al.*, 2001). La diferencia del cepstrum real es que se utiliza una escala de frecuencia no lineal, lo cual la aproxima al comportamiento del sistema auditivo. Dada la transformada discreta de

Fourier de una señal de entrada,

$$X_a[k] = \sum_{n=0}^{N-1} x[n] e^{-j2\pi nk/N}, \quad 0 \leq k < N. \quad (16)$$

Definimos un banco de filtros con M filtros ($m=1,2,\dots,M$), donde el filtro m es un filtro triangular dado por:

$$H_m[k] = \begin{cases} 0 & k < f[m-1] \\ \frac{2(k-f[m-1])}{(f[m+1]-f[m-1])(f[m]-f[m-1])} & f[m-1] \leq k \leq f[m] \\ \frac{2(f[m+1]-k)}{(f[m+1]-f[m-1])(f[m+1]-f[m])} & f[m] \leq k \leq f[m+1] \\ 0 & k > f[m+1]. \end{cases} \quad (17)$$

Tales filtros calculan el espectro promedio alrededor de cada frecuencia central con anchos de banda crecientes como se ve en la Figura 12. Alternativamente, los filtros

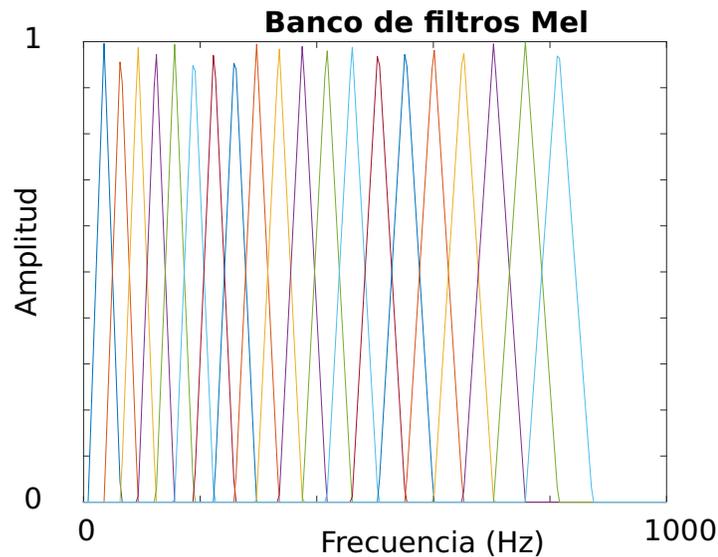


Figura 12. Filtros triangulares utilizados en el cálculo de los MFCC.

pueden ser escogidos como,

$$H'_m[k] = \begin{cases} 0 & k < f[m-1] \\ \frac{(k-f[m-1])}{(f[m]-f[m-1])} & f[m-1] \leq k \leq f[m] \\ \frac{(f[m+1]-k)}{(f[m+1]-f[m])} & f[m] \leq k \leq f[m+1] \\ 0 & k > f[m+1], \end{cases} \quad (18)$$

la cual satisface $\sum_{m=0}^{M-1} H'_m[k] = 1$. El Mel-cepstrum calculado con $H_m[k]$ o $H'_m[k]$ se diferenciará por un vector constante para todas las entradas, aunque la elección no es importante cuando se usa para un sistema que ha sido entrenado con los mismos filtros.

Definiendo f_1 y f_h como la menor y mayor frecuencia, respectivamente, del banco de filtros en Hertz, F_s la frecuencia de muestreo en Hertz, M el número de filtros, y N el tamaño de la transformada rápida de Fourier. Los puntos límites $f[m]$ están uniformemente espaciados en la escala Mel:

$$f[m] = \left(\frac{N}{F_s}\right) B^{-1}\left(B(f_1) + m \frac{B(f_h) - B(f_1)}{M+1}\right), \quad (19)$$

donde la escala Mel B está dada por:

$$B(f) = 1125 \ln(1 + f/700), \quad (20)$$

y la inversa B^{-1} es

$$B^{-1}(b) = 700(\exp(b/1125) - 1), \quad (21)$$

después se calcula el logaritmo de la energía de la salida de cada filtro,

$$S[m] = \ln\left[\sum_{k=0}^{N-1} |X_a[k]|^2 H_m[k]\right], \quad 0 \leq m < M. \quad (22)$$

El cepstrum en la escala de frecuencias Mel es la transformada de coseno discreta de las M salidas de los filtros:

$$c[n] = \sum_{m=0}^{M-1} S[m] \cos(\pi n(m+1/2)/M), \quad 0 \leq n < M. \quad (23)$$

En este trabajo se han estado utilizando 20 filtros triangulares, de los cuales se extraen 14 coeficientes. Estos 14 coeficientes son extraídos en ranuras de tiempo a lo largo de cada sonido cardiaco, después se obtiene el promedio de cada coeficiente de MFCC a lo largo de cada sonido cardiaco. Esto da como resultado 14 coeficientes por sonido lo cual da un total de 56 características para cada audio cardiaco.

Delta MFCC y Delta-Delta MFCC

Las características en el dominio del tiempo pueden darnos información necesaria, al utilizar MFCC estamos obteniendo información en el dominio de la frecuencia. Para poder extraer información del dominio del tiempo a partir de MFCC (a los que en esta sección les denominamos "coeficientes estáticos") se puede medir el cambio de MFCC a lo largo del tiempo, calculando la razón de cambio de MFCC a través de ranuras de tiempo. A estas características que miden el cambio de MFCC a lo largo del tiempo se le denomina delta MFCC (Huang *et al.*, 2001; Das *et al.*, 2016). Para obtener los delta MFCC se utiliza la siguiente ecuación,

$$d_t = \frac{\sum_{n=1}^N n(c_{t+n} - c_{t-n})}{2 \sum_{n=1}^N n^2}, \quad (24)$$

donde d_t es un coeficiente delta, de la ranura t calculada en términos de los coeficientes estáticos c_{t+N} y c_{t-N} , en este trabajo se utilizó $N = 2$. El número de características delta MFCC que se obtienen es la misma cantidad que el de características estáticas. Al obtener los delta MFCC se está midiendo la velocidad de cambio de MFCC, para medir la aceleración de MFCC se puede aplicar la misma ecuación 24 pero utilizando a los delta MFCC como características estáticas. La aceleración de MFCC es también llamada delta-delta MFCC.

Al agregar delta MFCC y delta-delta MFCC se obtiene un vector de características con la forma:

$$X_k = \begin{bmatrix} c_k \\ \Delta c_k \\ \Delta\Delta c_k \end{bmatrix}, \quad (25)$$

donde c_k son los MFCC, Δc_k son los delta MFCC y $\Delta\Delta c_k$ son los delta-delta MFCC. Δc_k

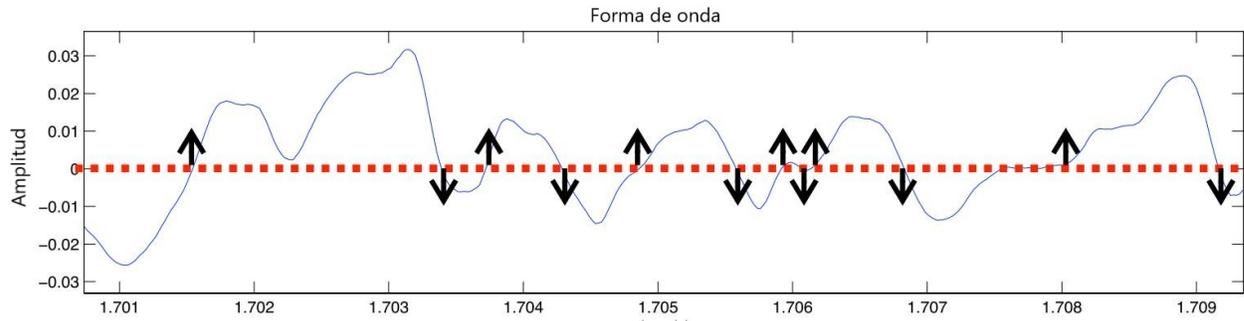


Figura 13. Cruces por cero (Lartillot, 2018).

aporta 56 características al vector utilizado en este trabajo de tesis.

3.5. Características en el dominio del tiempo y de la energía

Longitud: Se obtiene la longitud temporal de la secuencia dada en la entrada (Lartillot, 2018). Se pretende utilizar para medir la duración de cada sonido cardiaco.

Cruces por cero: Es un indicador simple del ruido, consiste en contar el número de veces que la señal cambia de signo (Figura 13, Lartillot (2018)).

Roll-Off: Es la frecuencia debajo de la cual se contiene un cierto porcentaje de la energía de la señal (Figura 14, Lartillot (2018)). En el caso de este trabajo se busca la frecuencia debajo de la cual se encuentra el 85 % de la energía.

Brillo: Medir la cantidad de energía que se encuentra por encima de una frecuencia dada (en este trabajo se utilizó 200 Hz). El resultado es un número dentro de 0 y 1 (Figura 15, Lartillot (2018)).

Baja energía: Se obtiene el porcentaje de ranuras de tiempo que tienen una energía RMS inferior a un umbral dado (Lartillot (2018)). El tamaño de las ranuras en este trabajo es de 50 ms, y el umbral es la media de la energía RMS en las ranuras de tiempo.

Densidad de eventos: Calcula la frecuencia promedio de los eventos. Se refiere

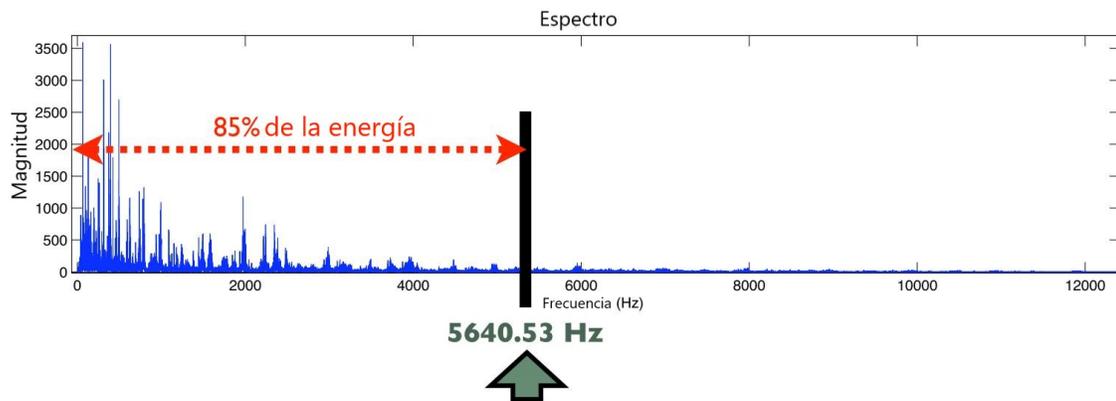


Figura 14. Roll-off. Se detecta la frecuencia por debajo de la cual se encuentra cierto porcentaje de la energía de la señal (Lartillot, 2018).

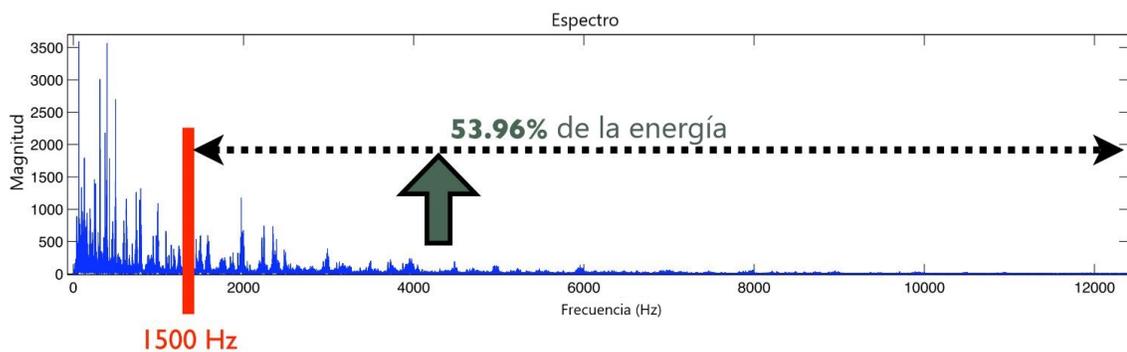


Figura 15. Brillo. Porcentaje de la energía de la señal que está por encima de una frecuencia dada (Lartillot, 2018).

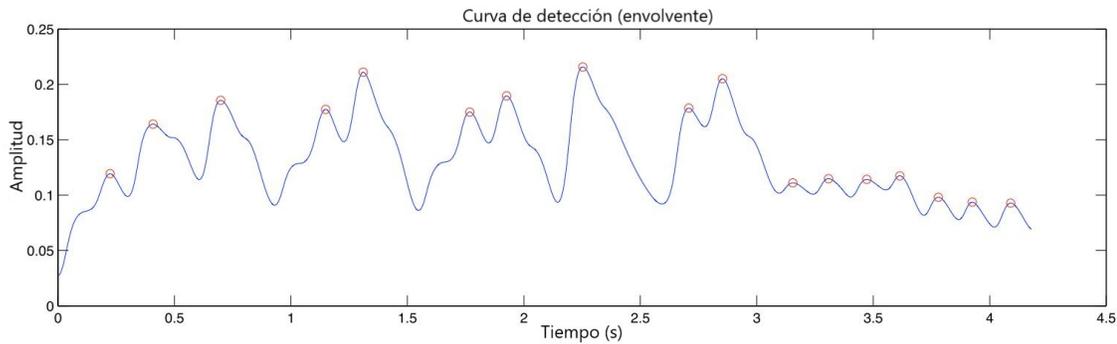


Figura 16. En círculos rojos están marcados los eventos (Lartillot, 2018).

al número de eventos detectados por segundo. Para calcular los eventos primero se calcula la curva de detección, que muestra las subidas de energía correspondiente a los pulsos sucesivos en la señal analizada. Después se detectan los picos en la curva de detección.(Figura 16, Lartillot (2018)).

Las características en el dominio del tiempo y de energía se calculan en cada sonido cardiaco, por lo que suman un total de 48 características.

3.6. Características en el dominio de la frecuencia

Parámetros de la forma del espectro

Para describir el comportamiento de los sonidos cardiacos en el dominio de la frecuencia se recurrió a cuatro diferentes métricas de probabilidad, a las cuales se les da el nombre de parámetros de la forma del espectro (Gillet y Richard (2004)). Las cuales se obtienen de los primeros cuatro momentos de la distribución en frecuencia de la señal de audio en cada sonido de cada ciclo.

El i -ésimo momento se obtiene aplicando la ecuación 26, donde $A(k)$ es la amplitud del k -ésimo componente de la transformada de Fourier de la señal de entrada. El centroide se obtiene directamente a partir del primer momento (ecuación 27). Los si-

güentes parámetros de forma espectrales son: la desviación estándar (ecuación 28), *skewness* (ecuación 29) y kurtosis (ecuación 30). La desviación estándar da una aproximación del ancho de banda de la señal de entrada, *skewness* y kurtosis miden la simetría y la oblicuidad respectivamente, del espectro de la señal. Las ecuaciones utilizadas para el cálculo de los parámetros de forma son:

$$\mu_i = \frac{\sum_{k=0}^{N-1} k^i A(k)}{\sum_{k=0}^{N-1} A(k)}, \quad (26)$$

$$S_c = \mu_1, \quad (27)$$

$$S_w = \sqrt{\mu_2 - \mu_1^2}, \quad (28)$$

$$S_a = \frac{2(\mu_1)^3 - 3\mu_1\mu_2 + \mu_3}{(S_w)^3}, \quad (29)$$

$$S_f = \frac{-3\mu_1^4 + 6\mu_1\mu_2 - 4\mu_1\mu_3 + \mu_4}{(S_w)^4} - 3. \quad (30)$$

Planitud espectral

La planitud espectral es una medida de lo "ruidosa" que es una señal (qué tan parecida es una señal al ruido blanco), obteniendo un valor de 1 si la señal es completamente ruidosa o un valor de 0 si la señal es tonal. Se puede definir como el cociente entre la media geométrica y la media aritmética (Peeters, 2004). Se calcula como:

$$SFM = \frac{(\prod_k A(k))^{(1/k)}}{\frac{1}{k} \sum_k A(k)}. \quad (31)$$

Irregularidad del espectro

La irregularidad del espectro es la medición de la variación de picos sucesivos del espectro (Lartillot, 2018). Primero se realiza una segmentación de la señal en ranuras de 50 milisegundos con la mitad de superposición, y se obtienen los picos de cada

ranura. Existen diferentes formas de calcular esta variación:

1. Jensen: Donde la irregularidad es la suma de el cuadrado de la diferencia en amplitud entre parciales adyacentes. Utiliza la ecuación:

$$\frac{\sum_{k=1}^N (a_k - a_{k+1})^2}{\sum_{k=1}^N a_k^2}. \quad (32)$$

2. Krimphoff: Donde la irregularidad es la suma de la amplitud menos el promedio de la amplitud anterior, la presente y la siguiente. Expresado de la siguiente manera:

$$\sum_{k=2}^{N-1} \left| a_k - \frac{a_{k-1} + a_k + a_{k+1}}{3} \right|. \quad (33)$$

Para este trabajo se calculó la variación utilizando la ecuación de Jensen (ecuación 32).

Entropía espectral

Se obtiene la entropía de Shannon relativa de la entrada (Lartillot, 2018; Shannon, 1948). La entropía de Shannon, utilizada en teoría de la información, está basada en la ecuación siguiente:

$$H(X) = - \sum_{i=1}^n p(x_i) \log_b p(x_i), \quad (34)$$

donde b es la base del logaritmo.

Las características en el dominio de la frecuencia se calculan en cada sonido cardiaco, por lo que suman un total de 28 características.

3.7. Características en el residual de Matching Pursuit

Primero se extraen átomos de Gabor a la señal de AC, después al residuo de este proceso (Figura 17) se le realiza la extracción de características con algunos de los métodos descritos en este capítulo.

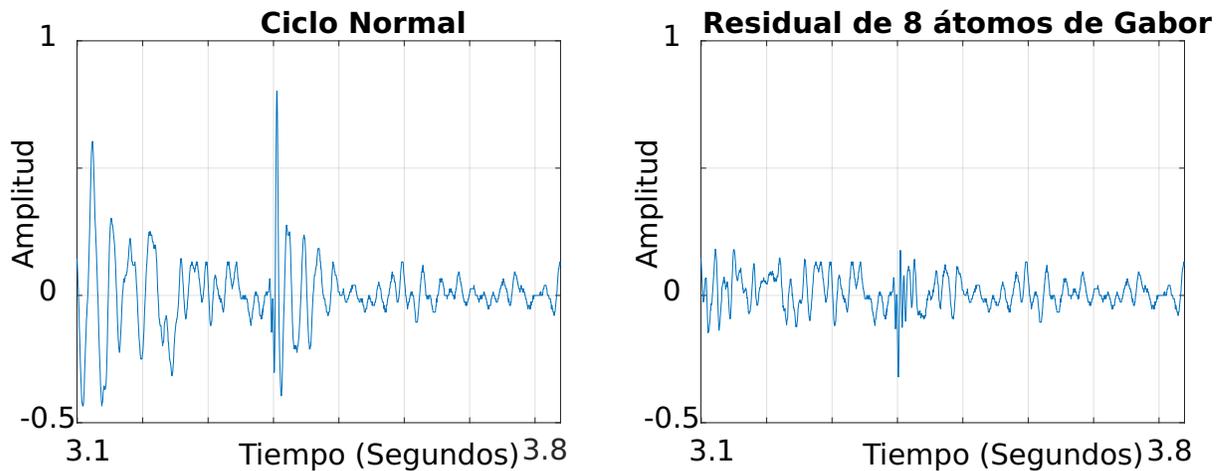


Figura 17. Residuo de la extracción de 8 átomos de Gabor.

En trabajos anteriores se ha demostrado que los átomos de Gabor son muy eficientes para representar el audio cardíaco "normal" (Nieblas-Hernández, 2014). Con este método de extracción se pretende primero eliminar la parte del audio cardíaco que está bien correlacionada con los átomos de Gabor. De esta manera se deja expuesto solo el audio que no está correlacionado y se trabaja sobre él.

Para el análisis del residual se utilizará MFCC, debido a que es una de las características que muestran mejor desempeño en el audio en general. Por lo que al final se obtendrán 56 características por audio cardíaco.

3.8. Vector de características

El vector de características sobre el que se estará trabajando contiene un total de 300 características, y se muestra en la Tabla 3.

3.9. Análisis de componentes principales

Es una manera de identificar patrones en los datos, y expresarlos de forma que realce sus similitudes y diferencias. Mientras que puede ser difícil encontrar patrones

Tabla 3. Vector de características.

Características	# valores
MP	40
LPC	40
MFCC	56
Delta MFCC	56
Tiempo y energía	24
Dominio de la frecuencia	28
Residual de MP	56
Total	300

**Figura 18.** Las flechas negras son los ejes principales de los datos. La longitud de estas es una medida de la varianza de los datos cuando son proyectados en el eje correspondiente.

en datos de alta dimensionalidad, donde no nos podemos dar el lujo de representarlos gráficamente, análisis de componentes principales (PCA) es una herramienta poderosa como algoritmo de reducción de la dimensionalidad y análisis de datos (VanderPlas, 2016; Smith, 2002).

En PCA, se busca una lista de los ejes principales de los datos, y utilizando esos ejes se describe el *dataset* (ver Figura 18). La proyección de cada dato en los ejes principales son los "componentes principales" de los datos.

Usar PCA para la reducción de la dimensionalidad involucra poner en cero uno o más de los componentes principales más pequeños, resultando en una proyección a

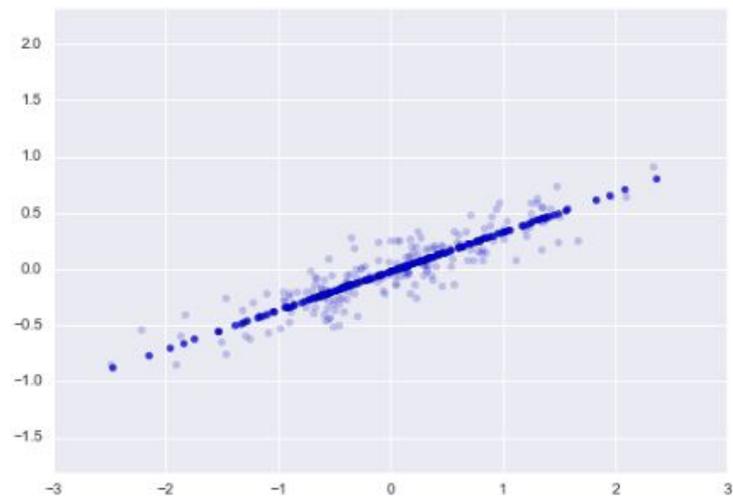


Figura 19. En azul fuerte se encuentran los datos proyectados sobre un solo eje principal. Reduciéndose los datos de 2 dimensiones a una sola.

una dimensión menor de los datos que tenga la máxima varianza (ver Figura 19).

3.10. Evaluación de la importancia de las características

En muchas tareas de *machine learning*, el objetivo no es solo encontrar el modelo más adecuado para obtener una respuesta sino también identificar cuáles de las variables de entrada son más importantes para la predicción. Esto con el propósito de tener un mayor entendimiento del problema en estudio.

Las características utilizadas en lo alto de un árbol de decisiones contribuyen a la predicción final de una gran fracción de las muestras de entrada. La fracción esperada de las muestras a las que contribuye una característica pueden ser utilizadas como un estimado de la importancia. La fracción de muestras a las que una característica contribuye es combinada con el descenso de impurezas al dividir las para crear un estimado normalizado del poder de predicción de esa característica (Loupe, 2014).

Promediando el poder de predicción a lo largo de muchos árboles aleatorios se puede reducir la varianza de tal estimador y utilizarlo para selección de características.

Capítulo 4. Técnicas de clasificación del audio cardiaco

El aprendizaje automático (AA) es un área de investigación interdisciplinaria que combina ideas de muchas ramas de la ciencia tales como: inteligencia artificial, estadística, teoría de la información, matemáticas, etc (Moorthy y Gandhi, 2017). El principal foco de las investigaciones en AA es el desarrollo de algoritmos de aprendizaje rápidos y eficientes que puedan hacer predicciones sobre datos. Cuando se utiliza con análisis de datos, el AA se utiliza para crear modelos de predicción. Las tareas del AA se pueden dividir en tres grupos: supervisado, no supervisado y reforzado. El AA supervisado requiere de datos de entrenamiento con etiquetas. Cada conjunto de datos de entrenamientos contiene valores de entrada y un valor de salida deseado. El AA supervisado analiza los datos de entrenamiento y crea una función de inferencia, la cual puede ser utilizada para analizar nuevos valores. En el AA no supervisado, ideas ocultas son dibujadas en sets de datos no etiquetados, por ejemplo, análisis de *cluster*. La tercera categoría, el AA reforzado permite a una máquina aprender su comportamiento por una retroalimentación recibida a través de las interacciones con un entorno externo. Desde un punto de vista de procesamiento de datos, tanto el AA supervisado como no supervisado son preferibles para análisis de datos y el AA reforzado es preferible para problemas de toma de decisiones (Moorthy y Gandhi, 2017).

En este trabajo se utilizan algoritmos de AA supervisados, debido a que se está utilizando una base de datos con etiquetas para realizar el entrenamiento y posterior prueba de cada clasificador. Los clasificadores que se evaluaron son descritos en este capítulo. Fueron elegidos estos clasificadores debido a la simpleza de su implementación y a que no requieren una carga computacional excesiva.

4.1. Regresión logística

Tomando en cuenta ciertos supuestos generales se puede ver la probabilidad de una clase C_1 descrita como una sigmoide logística actuando en una función lineal del vector de características ϕ (ver Figura 20) (Bishop, 2006). De tal manera que,

$$p(C_1|\phi) = y(\phi) = \sigma(w^T \phi), \quad (35)$$

donde w es el vector de parámetros ajustables y $p(C_2|\phi) = 1 - p(C_1|\phi)$. Además $\sigma(\cdot)$ es la función sigmoide logística definida como,

$$\sigma(a) = \frac{1}{1 + \exp(-a)}. \quad (36)$$

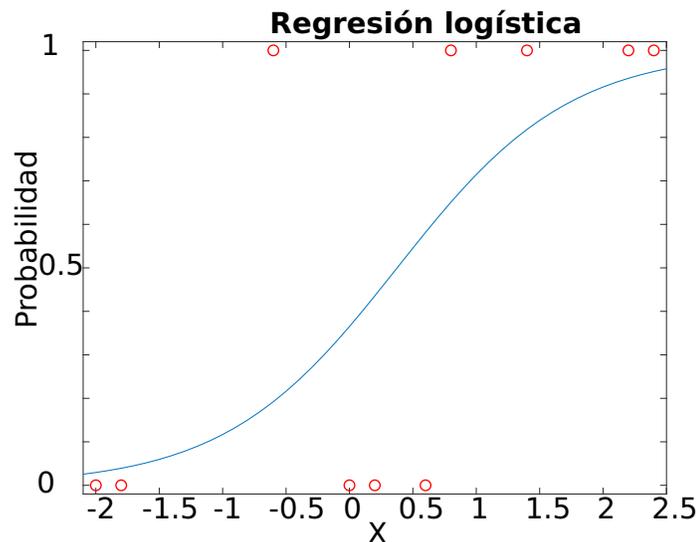


Figura 20. Función sigmoide que calcula la probabilidad de la clase. Los círculos representan diferentes casos de entrenamiento que le dieron forma al modelo, el eje horizontal representa la característica y el eje vertical la clasificación.

En terminología de estadística este modelo se conoce como regresión logística (LR), aunque se debe enfatizar en que es un modelo de clasificación y no uno de regresión.

Para un espacio de características M -dimensional ϕ , este modelo tiene M parámetros ajustables. Por lo que utilizando valores grandes de M hay una clara ventaja utilizando LR debido a que no requiere de mucha potencia computacional. Estos parámetros se ajustan para adecuarlos a los casos de entrenamiento para generar el modelo.

4.2. Análisis discriminante lineal

Análisis discriminante lineal (LDA) busca aquellos vectores en el espacio subyacente que mejor discrimina entre clases (Martinez y Kak, 2001). Más formalmente, dado un número de características independientes relativas a las cuales los datos son descritos, LDA crea una combinación lineal de estas que genere la mayor diferencia media entre las clases deseadas, como se muestra en la Figura 21. Matemáticamente hablando, para todas las muestras de todas las clases definimos dos medidas, la primera es llamada matriz de dispersión dentro de la clase (DDC), dada por:

$$S_w = \sum_{j=1}^C \sum_{i=1}^{N_j} (x_i^j - \mu_j)(x_i^j - \mu_j)^T, \quad (37)$$

donde x_i^j es la i -ésima muestra de la clase j , μ_j es la media de la clase j , c es el número de clases, y N_j el número de muestras en la clase j ; la segunda medida es llamada matriz de dispersión entre clases (DEC),

$$S_b = \sum_{j=1}^C (\mu_j - \mu)(\mu_j - \mu)^T, \quad (38)$$

donde μ representa la media de todas las clases.

El objetivo es maximizar la DEC mientras que se minimiza la DDC. Una forma de hacer esto es maximizando la proporción $\frac{\det[S_b]}{\det[S_w]}$.

4.3. K vecinos más cercanos

Los algoritmos de vecinos más cercanos se encuentran entre los algoritmos más simples de AA (Duda *et al.*, 2000; Shalev-Shwartz y Ben-David, 2014). La idea es calcular los puntos del *set* de entrenamiento que estén más cercanos a un nuevo caso, la clasificación se toma a partir de las etiquetas de estos puntos. El fundamento de este método se basa en la suposición de que las características que son usadas para localizar un caso en un punto del espacio de características son relevantes para su etiquetado, y entonces se asume que puntos cercanos suelen tener la misma etiqueta.

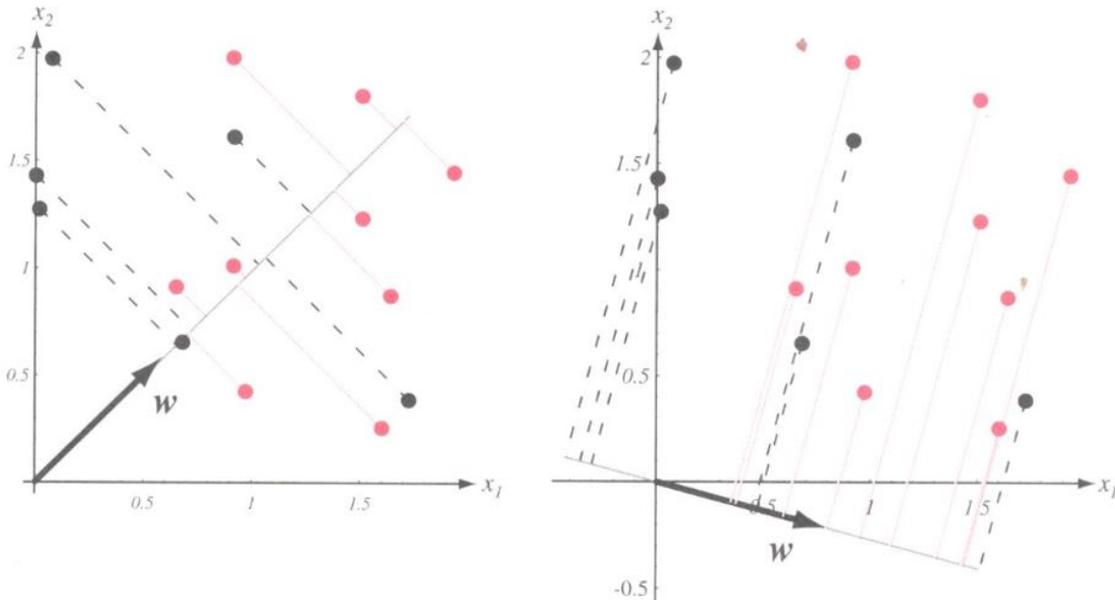


Figura 21. Se proyectan las mismas muestras en dos líneas diferentes marcadas como w . La figura en la derecha muestra mayor separación entre clases, cada clase de un color diferente (Duda *et al.*, 2000).

Además, en algunas situaciones, incluso cuando el set de entrenamiento es inmenso, encontrar los vecinos más cercanos puede hacerse extremadamente rápido.

Nuestro dominio de casos, X , está dotado por una función métrica ρ . Esto es, $\rho : X \times X \rightarrow \mathbb{R}$ es una función que regresa la distancia entre dos elementos de X . Por ejemplo, si $X = \mathbb{R}^d$ entonces ρ puede ser la distancia euclidiana, $\rho(x, x') = \|x - x'\| = \sqrt{\sum_{i=1}^d (x_i - x'_i)^2}$.

Siendo $S = (x_1, y_1), \dots, (x_m, y_m)$ una secuencia de ejemplos de entrenamiento. Por cada $x \in X$, $\pi_1(x), \dots, \pi_m(x)$ es un reordenamiento de $1, \dots, m$ de acuerdo a su distancia de x , $\rho(x, x_i)$. Esto es, para todo $i < m$,

$$\rho(x, x_{\pi_i(x)}) \leq \rho(x, x_{\pi_{i+1}(x)}). \quad (39)$$

Para un número k el algoritmo de K vecinos más cercanos para un clasificador binario está definido como:

1. Entrada: muestras de entrenamiento $S = (x_1, y_1), \dots, (x_m, y_m)$

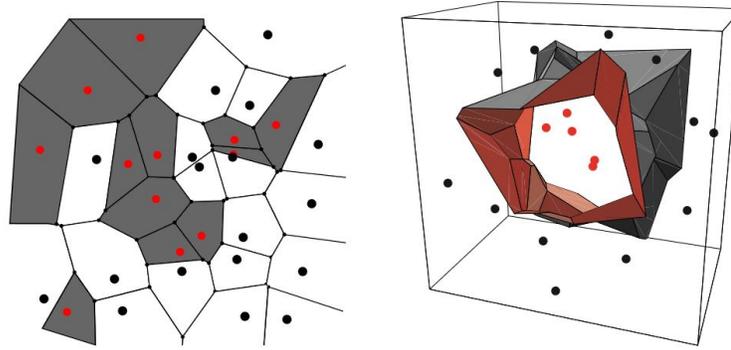


Figura 22. Se representa el algoritmo de decisión de K vecinos más cercanos de un clasificador binario con $K=1$. A la izquierda para un espacio bidimensional, a la derecha un espacio tridimensional (Duda *et al.*, 2000).

2. Salida: para cada punto $x \in X$, regresa la etiqueta mayoritaria en $\{y_{\pi_i(x)} : i \leq k\}$.

4.4. Árbol de decisión

Es natural e intuitivo clasificar un patrón a través de una serie de preguntas, en la cual la siguiente pregunta depende de la respuesta a la pregunta presente (Duda *et al.*, 2000; VanderPlas, 2016). Este enfoque de varias preguntas es particularmente útil para datos cualitativos, ya que todas las preguntas pueden ser respondidas por un estilo de "sí/no", "cierto/falso" o "valor(propiedad) \in set de valores" que no requiere ninguna noción de métrica.

Tal secuencia de preguntas se presentan en un árbol de decisión directo o simplemente en un árbol, que por convención el primer nodo o nodo raíz se presenta en la cima, conectado por líneas sucesivas a otros nodos. Estos están conectados de manera similar hasta que se alcanza la terminal o el nodo hoja (ver Figura 23).

Asumimos que tenemos un set D de datos de entrenamiento etiquetados y que hemos elegido un set de propiedades que pueden ser usadas para discriminar patrones, pero no se sabe como organizar las pruebas en un árbol. Claramente, cualquier árbol de decisión progresivamente dividirá el set de entrenamiento en subsets cada vez más pequeños, como se ve en la Figura 24. Sería ideal si todas las muestras en cada subset

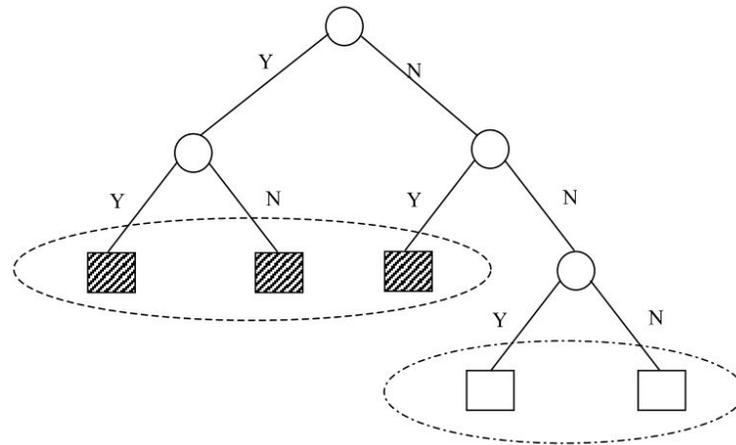


Figura 23. Árbol de decisión. Las preguntas se realizan empezando desde arriba, las hojas del árbol están representadas en rectángulos y es donde se toma la decisión final de la clasificación (Huang *et al.*, 2001).

tuvieran la misma etiqueta. En ese caso, diríamos que cada subset es puro, y podríamos terminar esa porción del árbol. Sin embargo, usualmente hay una combinación de etiquetas en cada subset debido a que por cada rama deberemos decidir si terminar de partir el subset y aceptar la imperfección, o seleccionar alguna otra propiedad y hacer crecer más el árbol.

Esto sugiere un proceso obvio de crecimiento del árbol: dados los datos representados en un nodo, se declara ese nodo como una hoja (y se indica que categoría asignar), o se busca alguna otra propiedad para usar en la partición de los datos en subsets más pequeños. Este es un ejemplo de CART (*classification and regression trees*).

4.5. Bayesiano Gaussiano ingenuo

Los clasificadores Bayesianos ingenuos son un grupo de algoritmos de clasificación extremadamente rápidos y simples que son utilizables para *datasets* de dimensionalidad muy alta (VanderPlas, 2016). Debido a que son demasiado rápidos y tienen pocos parámetros ajustables.

Los clasificadores Bayesianos ingenuos son construidos por un método de clasi-

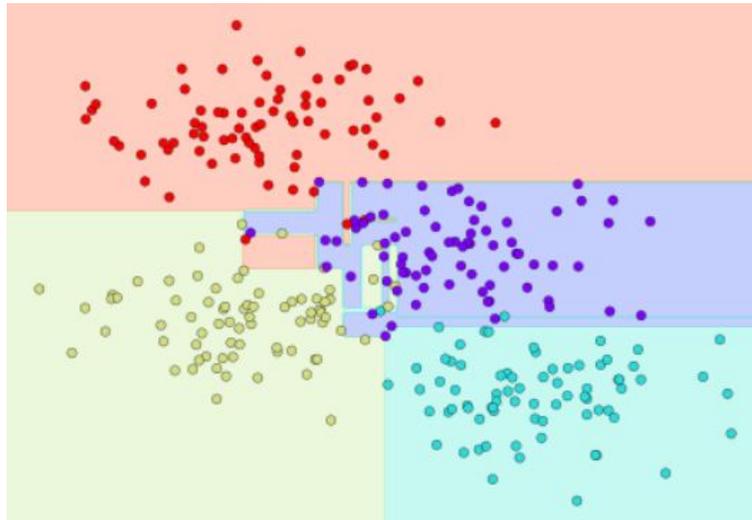


Figura 24. Segmentación de los datos por medio de un árbol de decisiones (VanderPlas, 2016).

ficación Bayesiano. Estos dependen del teorema de Bayes, el cual es una ecuación que describe la relación de probabilidades condicionales de cantidades estadísticas. En clasificación Bayesiana, estamos interesados en encontrar la probabilidad de una etiqueta dadas algunas características observadas, las cuales podemos escribir como $P(L|\text{características})$. El teorema de Bayes nos dice como expresar esto en términos de cantidades que podamos calcular más directamente:

$$P(L|\text{características}) = \frac{P(\text{características}|L)P(L)}{P(\text{características})}. \quad (40)$$

Si tratamos de decidir entre dos etiquetas L_1 o L_2 , entonces una manera de tomar esta decisión es calcular la proporción de las probabilidades para cada etiqueta:

$$\frac{P(L_1|\text{características})}{P(L_2|\text{características})} = \frac{P(\text{características}|L_1)P(L_1)}{P(\text{características}|L_2)P(L_2)}. \quad (41)$$

Lo que necesitamos ahora es algún modelo en el cual se pueda calcular $P(\text{características}|L_i)$ para cada etiqueta. Ese modelo es llamado generador porque especifica el hipotético proceso aleatorio que genera los datos. Especificar este modelo para cada etiqueta es la pieza principal de entrenar un clasificador Bayesiano. La versión general del paso de entrenamiento es una tarea muy difícil, pero podemos hacerlo

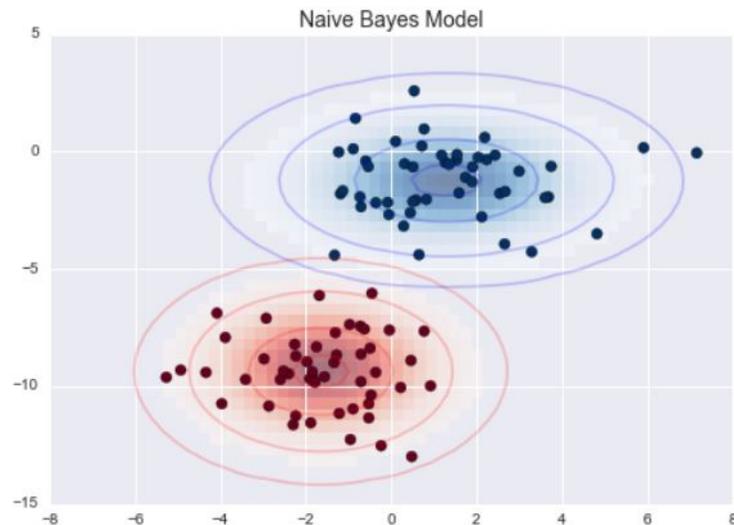


Figura 25. Probabilidad de las clases utilizando un modelo Bayesiano Gaussiano ingenuo (VanderPlas, 2016).

más simple por el uso de algunas suposiciones acerca de la forma de este modelo.

Aquí es donde la palabra "ingenuo" entra, si hacemos una suposición ingenua acerca del modelo generador para cada etiqueta, podemos encontrar una aproximación del modelo generador para cada clase, y entonces proceder con la clasificación Bayesiana. Diferentes tipos de clasificadores de Bayes ingenuos significan diferentes suposiciones ingenuas sobre los datos.

El clasificador Bayesiano ingenuo más fácil de entender es el Gaussiano. En este clasificador la suposición es que los datos de cada etiqueta están trazados como una distribución Gaussiana, sin covarianza entre dimensiones. Podemos ajustar este modelo simplemente encontrando una media y una desviación estándar de los puntos en cada etiqueta, lo cual es lo único que se necesita para definir dicha distribución. El resultado de la suposición ingenua Gaussiana se representa en la Figura 25.

4.6. Máquinas de vectores de soporte

En las máquinas de vectores de soporte (SVM) la idea básica es buscar una línea o curva (en dos dimensiones) o *manifold* (en múltiples dimensiones) que dividan las

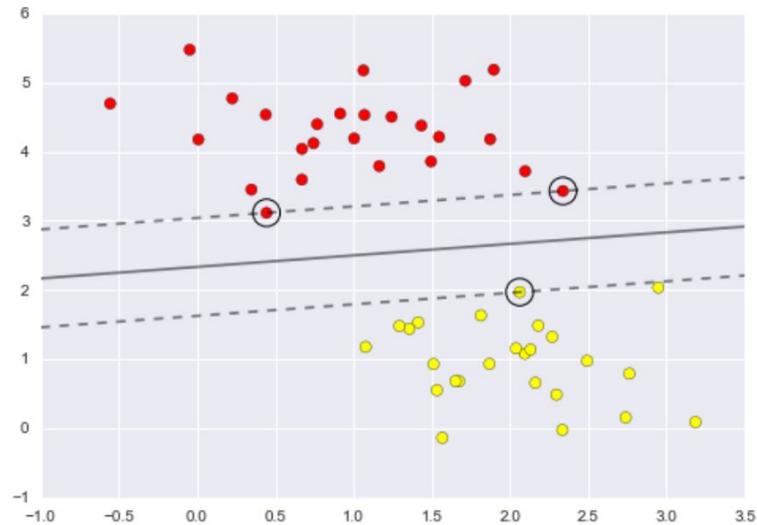


Figura 26. Dentro de los círculos negros se encuentran los vectores de soporte, las líneas entrecortadas son los márgenes (VanderPlas, 2016).

clases entre sí (VanderPlas, 2016).

En la Figura 26 se puede observar un caso simple en el que dos clases están claramente separadas. Se podría trazar una línea para separar estas dos clases, pero hay un número infinito de líneas que podrían realizar esta acción. Para elegir una de este infinito número de líneas se dibuja un margen de algún grosor en cada línea, hasta el punto más cercano. En máquinas de soporte vectoriales la línea que maximiza este margen es la elegida como el modelo óptimo. Los puntos que tocan los márgenes son conocidos como vectores de soporte, y le dan nombre al algoritmo. Una clave para el éxito de este clasificador es que solo la posición de los vectores de soporte importa; el modelo no se ve afectado con nuevas muestras de entrenamiento siempre que no sobrepasen el margen (Figura 27).

SVM se vuelve un algoritmo muy útil cuando es combinado con kernels. Se proyectan los datos en un espacio de dimensión mayor definido por polinomios o funciones Gaussianas, y de esta forma es posible ajustar el clasificador lineal para relaciones no lineales.

En la Figura 28 se observa que los datos no pueden ser separables por una función lineal. En estos casos se recurre a proyectarlos en una dimensión adicional utilizando

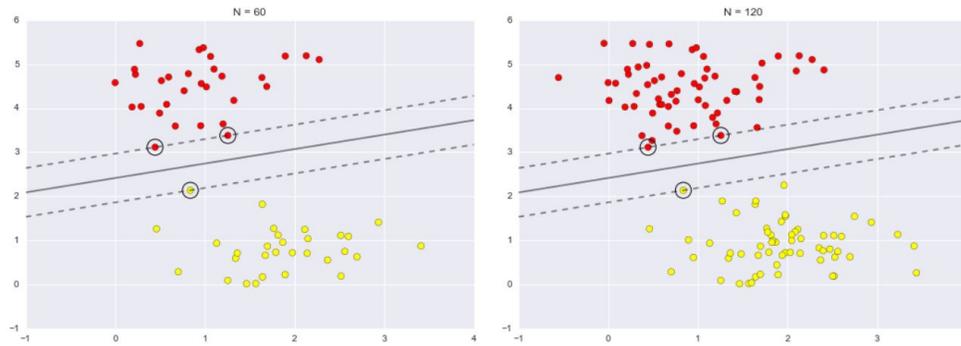


Figura 27. Se puede observar que agregar mayor cantidad de muestras en el entrenamiento de la máquina de vectores de soporte no modifica el modelo final, siempre que las nuevas muestras no sobrepasen su margen (VanderPlas, 2016).

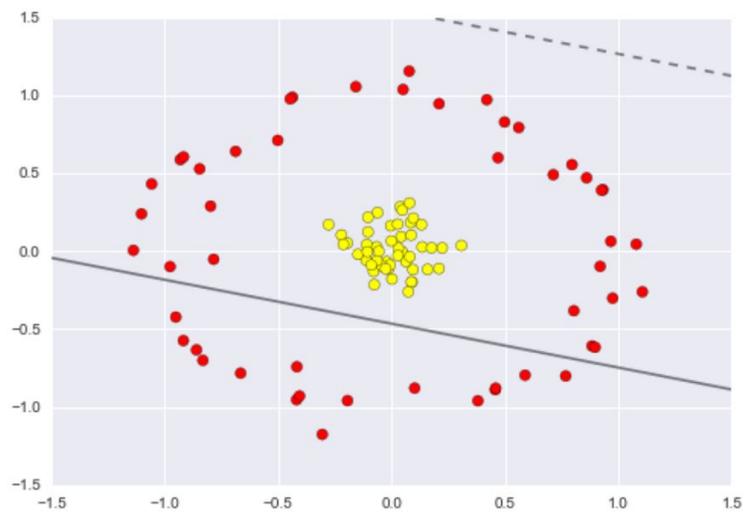


Figura 28. Es imposible separar los datos con una función lineal, por lo que necesitan ser proyectados con un kernel a una dimensión superior (VanderPlas, 2016).

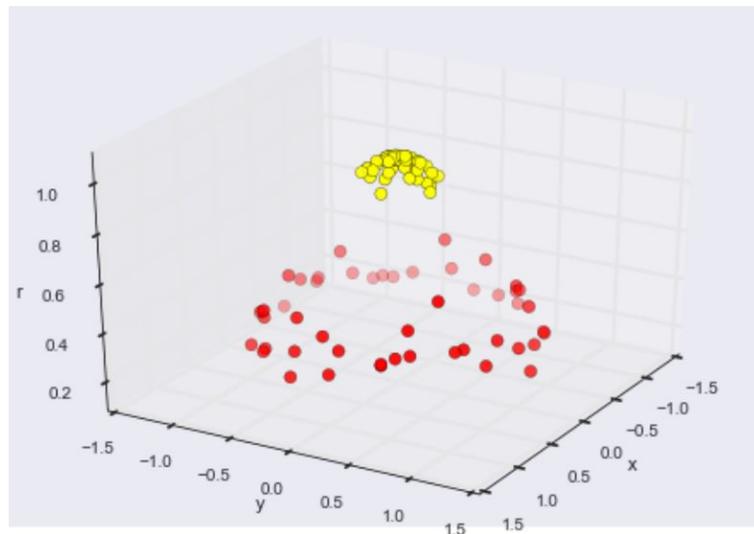


Figura 29. Una vez los datos son proyectados con un kernel pueden ser separados por una función lineal (VanderPlas, 2016).

algún kernel, de tal forma que las clases puedan separarse por medio de una función lineal. En el caso de la Figura 29 una sola proyección de los datos es suficiente. A esta transformación basada en una función se le conoce como transformación de kernel. Estas son las ecuaciones de los kernel utilizados:

Kernel lineal:

$$K(x, x') = \langle x, x' \rangle . \quad (42)$$

Kernel polinomial:

$$K_p(x, x') = [\gamma \langle x, x' \rangle + r]^d . \quad (43)$$

Kernel función de base radial:

$$K(x, x') = \exp(-\gamma \|x - x'\|^2), \gamma > 0 . \quad (44)$$

Donde x y x' son vectores de características extraídas de la matriz de entrena-

miento (ver sección 5.1). El parámetro γ define qué tanto influye una sola muestra de entrenamiento, en este trabajo se utilizó γ como el inverso del número de características; el parámetro r es igual a cero. Además, d es el grado del polinomio. En este trabajo se utilizó una función de tercer grado.

4.7. Bosques aleatorios

Es un clasificador que consiste en una colección de clasificadores de estructura de árbol $\{h(x, \theta_k), k = 1, \dots\}$ donde θ_k son vectores aleatorios independientes e idénticamente distribuidos y cada árbol da un único voto para la clase más popular de la entrada x (Breiman, 2001).

Cada árbol depende de valores de un vector aleatorio muestreado independientemente y con la misma distribución para todos los árboles en el bosque. El error de generalización converge a un límite entre más árboles hay en el bosque. El error de generalización de un bosque depende de la fuerza de cada árbol en el bosque de forma individual y la correlación entre ellos.

Múltiples modelos que realizan sobreajuste pueden ser combinados para reducir los efectos del sobreajuste, esto es lo que subyace un método de ensamble llamado *bagging* (VanderPlas, 2016). *Bagging* hace uso de un ensamble de modelos paralelos, cada uno el cual se sobreajusta a los datos, y promedia los resultados para buscar una mejor clasificación. Un ensamble de árboles de decisión aleatorios es llamado bosque aleatorio.

En la Figura 30 podemos ver que promediando cien modelos perturbados aleatoriamente terminamos con un modelo general que es más cercano a la intuición acerca de cómo el espacio de parámetros debería ser segmentado.

En la Figura 31 se graficaron los resultados de pruebas con diferente cantidad de árboles, se realizó regresión logarítmica para generar una función y se tomó el resultado MAcc dado con 10,000 árboles como 1 de desempeño.

En este trabajo se utilizan 1,000 árboles para la clasificación con bosques aleatorios debido a que, como se ve en la Figura 31, con esta cantidad de árboles se está ob-

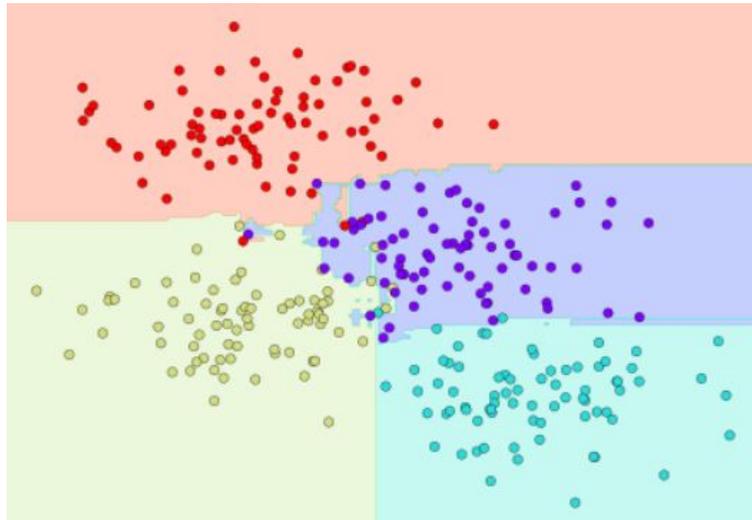


Figura 30. Se puede observar que la segmentación que realizan los bosques aleatorios toman formas más complejas que se adecuan mejor a los datos que en el caso del árbol de decisión (VanderPlas, 2016).

teniendo cerca de 0.99 del desempeño que se tendría con 10,000 árboles (los cuales sería impráctico utilizar debido al coste computacional alto). Esta cantidad de árboles que se está utilizando no ocasiona un gasto computacional excesivo ni exige mucho tiempo de procesamiento. El crecimiento de rendimiento del modelo, con la cantidad de árboles, corresponde a una función logarítmica; por lo que aumentar más la cantidad de árboles no significaría una mejora significativa.

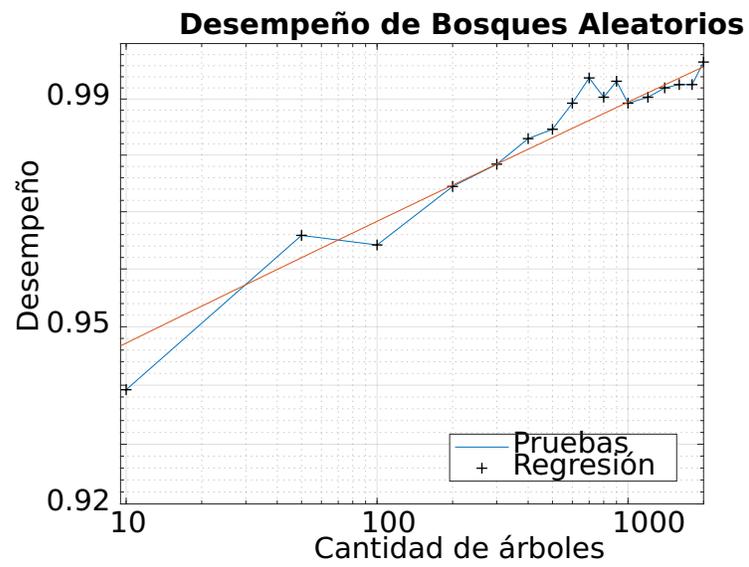


Figura 31. Entre más árboles mayor es el desempeño. En cruces negras se presentan resultados de pruebas y en rojo la regresión logarítmica. Se consideró un desempeño de 1 al resultado obtenido con 10,000 árboles.

Capítulo 5. Resultados

5.1. Matriz de entrenamiento y prueba

Para evaluar el rendimiento, de los vectores de características utilizados en este trabajo, se generaron tablas donde se muestran los resultados del promedio de 20 iteraciones en cada tipo de clasificador.

Tabla 4. Matriz de entrenamiento y de prueba.

Base de datos	A	B	C	D	E	F	Total
Audios de entrenamiento	174	110	10	40	218	48	600
Audios de prueba	42	26	2	8	52	10	140
% Audios en matriz de entrenamiento	29	18.33	1.67	6.67	36.33	8	100
% Audios en matriz de prueba	30	18.57	1.43	5.71	37.14	7.14	100
Porcentaje de audios de entrenamiento	80.56	80.88	83.33	83.33	80.74	82.76	81.08

Como se vio en la sección 2.4, en la Tabla 2, se están utilizando 572 audios "anormales" y 2301 "normales". De los cuales, en cada iteración, se seleccionan de manera aleatoria de cada base de datos (A, B, C, D, E y F) el número de audios de entrenamiento y de prueba que se muestran en la Tabla 4. Dando un total de 600 muestras en la matriz de entrenamiento, utilizada para generar el modelo de clasificación; y 140 muestras en la matriz de prueba, utilizada para evaluar el modelo de clasificación. Por cada base de datos se extraen 50 % de señales anormales y 50 % de señales normales.

5.2. Métricas para evaluación de los sistemas de clasificación

Para realizar la evaluación del modelo primero se realiza la clasificación sobre la matriz de prueba. Después se comparan los resultados de las predicciones con las etiquetas de los audios cardiacos y se genera una matriz de confusión como la que se puede ver en la Tabla 5. Para generarla se consideran cuatro tipos de resultados:

TP (*true positive*): Los resultados positivos (en nuestro caso es la etiqueta "anormal") que fueron correctamente clasificados.

FP (*false positive*): Resultados positivos que fueron incorrectamente clasificados.

FN (*false negative*): Resultados negativos (en nuestro caso la etiqueta "normal") que fueron incorrectamente clasificados.

TN (*true negative*): Resultados negativos que fueron correctamente clasificados.

Tabla 5. Matriz de confusión. Contiene la cantidad de resultados TP, FN, FP y TN.

	Predicho positivo	Predicho negativo
Realmente positivo	TP	FN
Realmente negativo	FP	TN

En esta tesis se utilizarán las métricas utilizadas en el desafío de 2016 de PhysioNet (Clifford *et al.*, 2016), las cuales son: *specificity*, *sensitivity* y MAcc, donde el desempeño de las características y clasificadores se medirán con el puntaje MAcc.

$$\text{specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}, \quad (45)$$

$$\text{sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (46)$$

$$\text{MAcc} = \frac{\text{specificity} + \text{sensitivity}}{2}, \quad (47)$$

donde *specificity* es la proporción de muestras "normales" que fueron clasificadas correctamente. *Sensitivity* es la proporción de muestras "anormales" que fueron clasificadas correctamente. Ambas manejan valores entre 0 y 1, significando 0 que ninguna muestra fue clasificada correctamente y 1 que todas fueron clasificadas correctamente.

5.3. Comparación de algoritmos de obtención de características

Se evalúan los tres algoritmos de extracción de características vistos en la sección 3.1. Se utilizaron las características MFCC para esta prueba. Se entrenaron y evaluaron clasificadores de regresión logística (LR), de análisis discriminante lineal (LDA), de K vecinos más cercanos (KNN), de árboles de decisión (CART), Bayesiano Gaussiano

ingenuo (NB), de máquinas de soporte vectorial con kernels de función de base radial (RBF), lineal (Linear) y polinomial (Poli); y clasificador de bosques aleatorios (RFC).

Tabla 6. Resultados del promedio de características MFCC. En negritas se presentan los valores máximos.

	LR	LDA	KNN	CART	NB	RBF	Linear	Poli	RFC
Specificity	0.740	0.727	0.753	0.745	0.639	0.701	0.717	0.806	0.835
Sensitivity	0.781	0.759	0.737	0.750	0.741	0.815	0.794	0.769	0.785
MAcc	0.761	0.743	0.745	0.748	0.690	0.758	0.755	0.788	0.810

Tabla 7. Resultados de características MFCC en promedio de ciclos cardiacos. En negritas se presentan los valores máximos.

	LR	LDA	KNN	CART	NB	RBF	Linear	Poli	RFC
Specificity	0.694	0.692	0.675	0.647	0.621	0.628	0.670	0.714	0.757
Sensitivity	0.678	0.699	0.729	0.655	0.709	0.796	0.720	0.740	0.761
MAcc	0.686	0.695	0.702	0.651	0.665	0.712	0.690	0.727	0.759

Tabla 8. Resultados de características MFCC en promedio de ciclos cardiacos ajustados. En negritas se presentan los valores máximos.

	LR	LDA	KNN	CART	NB	RBF	Linear	Poli	RFC
Specificity	0.693	0.698	0.653	0.652	0.603	0.613	0.660	0.690	0.731
Sensitivity	0.708	0.677	0.671	0.651	0.723	0.764	0.740	0.699	0.731
MAcc	0.700	0.688	0.662	0.652	0.663	0.689	0.700	0.695	0.731

De las Tablas 6, 7 y 8, se observa que el promedio de características es el algoritmo que da mejores resultados en el puntaje MAcc. Y, a pesar de ser el que requiere de mayor procesamiento computacional, se observará que su gasto computacional no es excesivo como para considerarlo impráctico; como se verá posteriormente, generar un vector con este algoritmo requiere de unos pocos segundos. A partir de este punto todas las pruebas se hicieron utilizando el algoritmo de promedio de características.

5.4. Características y clasificadores

Se generaron matrices de entrenamiento y prueba utilizando cada tipo de características vistas en el capítulo 3, y se utilizaron los clasificadores vistos en el capítulo 4 para su entrenamiento y predicción. Se presentan los resultados para MFCC (Tabla 6), delta-MFCC (Tabla 9), codificación escasa (Tabla 10), codificación predictiva lineal (Tabla 11), características en el dominio del tiempo (Tabla 12), características en el

dominio de la frecuencia (Tabla 13) y características MFCC en el residual de *Matching Pursuit* (Tabla 14).

Se hizo análisis del residual de *Matching Pursuit* con diferente número de átomos, y se observó que el análisis del residual de la extracción de 8 átomos es el que da como resultado el MAcc más alto, por lo que se consideró este el número adecuado de átomos a extraer (Figura 32). A la extracción de los MFCC del residual de la extracción de 8 átomos se le abrevia como MFCC8 en este trabajo.

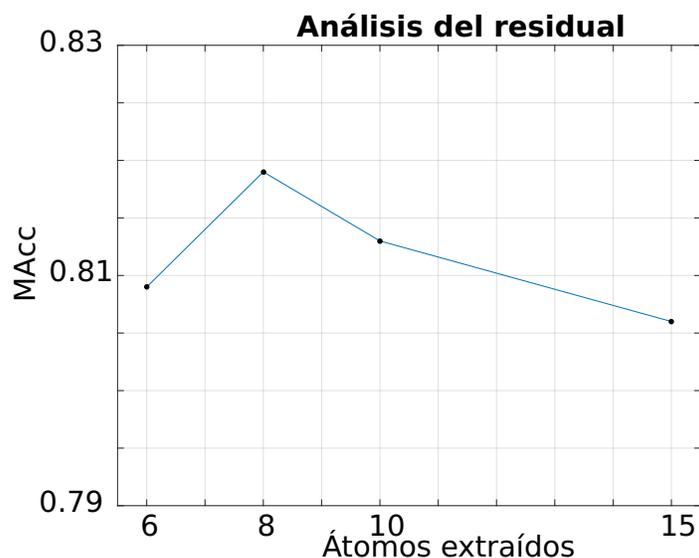


Figura 32. Análisis del residual de *Matching Pursuit* con MFCC. Se puede observar que la cantidad de átomos extraídos que da el mejor MAcc es 8. Los puntos probados están representados por puntos negros.

Tabla 9. Resultados de delta MFCC. En negritas se presentan los valores máximos.

	LR	LDA	KNN	CART	NB	RBF	Linear	Poli	RFC
Specificity	0.774	0.739	0.757	0.715	0.660	0.671	0.745	0.796	0.821
Sensitivity	0.772	0.747	0.738	0.719	0.709	0.827	0.756	0.804	0.801
MAcc	0.773	0.743	0.748	0.717	0.684	0.749	0.750	0.800	0.811

Tabla 10. Resultados de *Matching Pursuit* de 10 átomos. En negritas se presentan los valores máximos.

	LR	LDA	KNN	CART	NB	RBF	Linear	Poli	RFC
Specificity	0.664	0.629	0.581	0.627	0.670	0.994	0.595	0.552	0.660
Sensitivity	0.607	0.621	0.643	0.609	0.549	0.006	0.647	0.532	0.760
MAcc	0.635	0.625	0.612	0.618	0.609	0.500	0.621	0.542	0.710

En la Tabla 10, en RBF, se puede ver que la *specificity* es casi 1, lo que significa

que casi todas las muestras "normales" fueron clasificadas correctamente. También se puede observar que la *sensitivity* tiende a 0, debido a que casi ninguna muestra "anormal" fue clasificada correctamente. Esto pasa cuando un clasificador está clasificando la gran mayoría de las muestras con una misma etiqueta. Un caso parecido sucede con el clasificador Poli en la Tabla 11.

Tabla 11. Resultados del promedio de características LPC. En negritas se presentan los valores máximos.

	LR	LDA	KNN	CART	NB	RBF	Linear	Poli	RFC
Specificity	0.762	0.772	0.836	0.689	0.337	0.721	0.786	0.015	0.838
Sensitivity	0.739	0.761	0.733	0.701	0.856	0.730	0.739	1.000	0.769
MAcc	0.750	0.766	0.784	0.695	0.596	0.725	0.762	0.508	0.803

Tabla 12. Resultados del promedio de características TIMEF. En negritas se presentan los valores máximos.

	LR	LDA	KNN	CART	NB	RBF	Linear	Poli ¹	RFC
Specificity	0.730	0.771	0.780	0.676	0.838	0.457	0.795	0.671	0.801
Sensitivity	0.656	0.686	0.629	0.684	0.479	0.550	0.616	0.771	0.721
MAcc	0.693	0.729	0.705	0.680	0.659	0.504	0.706	0.721	0.761

Tabla 13. Resultados de SPTR. En negritas se presentan los valores máximos.

	LR	LDA	KNN	CART	NB	RBF	Linear	Poli	RFC
Specificity	0.753	0.748	0.764	0.651	0.515	0.434	0.799	0.743	0.764
Sensitivity	0.705	0.700	0.656	0.664	0.720	0.608	0.715	0.643	0.704
MAcc	0.729	0.724	0.710	0.658	0.618	0.521	0.757	0.693	0.734

Tabla 14. Resultados de características MFCC en el residual de la extracción de 8 átomos de *Matching Pursuit* (MFCC8). En negritas se presentan los valores máximos.

	LR	LDA	KNN	CART	NB	RBF	Linear	Poli	RFC
Specificity	0.762	0.743	0.793	0.739	0.659	0.795	0.734	0.808	0.848
Sensitivity	0.759	0.738	0.730	0.735	0.742	0.780	0.784	0.784	0.789
MAcc	0.761	0.740	0.761	0.737	0.700	0.788	0.759	0.796	0.819

Las Tablas 6, 9, 10, 11, 12, 13 y 14 nos muestran que el clasificador de bosques aleatorios (RFC) es el que nos entrega mejores puntajes de MAcc para los diferentes tipos de características. Con excepción de las características SPTR donde Linear da el mejor resultado MAcc, pero la diferencia con RFC es pequeña.

¹Con este clasificador el gasto computacional fue excesivo. Solo se realizó una iteración.

5.5. Resultados con todas las características y reducción de la dimensionalidad.

En casi todos los casos el clasificador de bosques aleatorios fue el que mostró mayores valores de MAcc. Por lo que fue escogido para evaluar el desempeño del vector de características completo (VCC), el cual se compone de 300 características. Además se utilizó PCA para reducir el número de características a 60 formando así el vector PCA-60; por otro lado también se hizo reducción del número de características a las 60 más importantes de bosques aleatorios formando el vector RF-60. Estos resultados se muestran en la Tabla 15, donde se ve que el vector RF-60 tiene el mejor puntaje MAcc, incluso más que el VCC.

Tabla 15. Resultados con todas las características y reducción de la dimensionalidad. (T.E.= tiempo de entrenamiento en segundos, T.P.=Tiempo de predicción en segundos). En negritas se presentan los valores máximos.

	VCC	PCA-60	RF-60
Specificity	0.851	0.792	0.870
Sensitivity	0.808	0.805	0.798
MAcc	0.830	0.799	0.834
T.E.	3.911	2.378	2.008
T.P.	0.061	0.062	0.058

Cada vez que se hacía una iteración para evaluar los vectores PCA-60 y RF-60 se formaba un modelo de reducción de la dimensionalidad nuevo, de acuerdo a las 600 muestras de entrenamiento, y se utilizaba para la reducción de la matriz de prueba y entrenamiento.

El vector PCA-60 demostró generar pérdidas en los resultados del puntaje MAcc de 0.031 con respecto a la utilización de VCC, por lo que fue desechado. A pesar de tener la misma cantidad de características que RF-60 demuestra tener menor desempeño, y por tanto ser inviable.

Se observaron 5 modelos para formar vectores RF-60 de diferentes iteraciones, con el fin de buscar las 60 características que suelen tener mayor importancia en la clasificación; se muestran en la Tabla 16 (cabe destacar que no había mucha variación entre los 5 vectores RF-60 observados).

Tabla 16. Las 60 características más importantes. Entre más arriba y más a la izquierda tienen mayor importancia.

1. Quinto MFCC8 de la diástole	2. Quinto MFCC de la diástole	3. Cruces por cero de la sístole	4. Quinto MFCC8 de S2	5. Cruces por cero de la diástole
6. Quinto MFCC de la sístole	7. Quinto MFCC8 de S1	8. Quinto MFCC de S2	9. Quinto LPC de la sístole	10. Quinto MFCC8 de la sístole
11. Cruces por cero de S2	12. Segundo MFCC de S2	13. Segundo MFCC8 de S2	14. Primer LPC de la diástole	15. Segundo MFCC8 de diástole
16. Quinto MFCC de S1	17. Tercer LPC de S2	18. Segundo MFCC de la diástole	19. Sexto MFCC de S1	20. Cuarto MFCC de la diástole
21. Quinto LPC de la diástole	22. Primer LPC de S2	23. Sexto MFCC8 de S2	24. Primer LPC de la sístole	25. Sexto MFCC de S2
26. Cuarto MFCC8 de la diástole	27. Segundo LPC de la diástole	28. Cuarto MFCC de la sístole	29. Segundo MFCC8 de la sístole	30. Frecuencia del sexto átomo
31. Tercer MFCC8 de la diástole	32. Cuarto MFCC8 de S2	33. Tercer MFCC de la diástole	34. Octavo MFCC8 de S1	35. Sexto MFCC8 de la diástole
36. Sexto LPC de la sístole	37. Sexto MFCC de la diástole	38. Segundo LPC de la sístole	39. Tercer LPC de la diástole	40. <i>Roll-off</i> de la sístole
41. Frecuencia del octavo átomo	42. Desviación estándar del espectro de la diástole	43. Primer MFCC8 de la diástole	44. Frecuencia del cuarto átomo	45. Décimo primer MFCC8 de la diástole
46. Frecuencia del décimo átomo	47. Posición del sexto átomo	48. Frecuencia del séptimo átomo	49. Segundo LPC de S2	50. Frecuencia del quinto átomo
51. Primer MFCC de la diástole	52. <i>Skewness</i> del espectro de la sístole	53. Quinto delta MFCC de la sístole	54. Centroide del espectro de la sístole	55. Séptimo LPC de la sístole
56. Noveno LPC de la diástole	57. Cuarto MFCC de S2	58. Frecuencia del segundo átomo	59. Tercer MFCC de S1	60. Primer MFCC8 de S1

Otra ventaja que tiene el vector RF-60 con los valores de la Tabla 16 es el tiempo, se tomó como ejemplo el segundo audio cardiaco de la base E y se necesitaron 80 segundos para generar su VCC; en cambio para generar su vector RF-60 se necesitan 11 segundos. Aunque estos tiempos están sujetos a los algoritmos que se han estado utilizando, los cuales podrían ser optimizados.

A la base de datos completa (Tabla 2) por medio de validación cruzada con fold-10 se comparó VCC con el vector RF-60, utilizando en este las características de la Tabla 16. Con VCC se obtuvo una cantidad de aciertos del 92.725%, 0.973 de *specificity*, 0.743 de *sensitivity* y una MAcc de 0.859; con RF-60 se obtuvieron el 93.526% de aciertos, 0.974 de *specificity*, 0.780 de *sensitivity* y una MAcc de 0.877. La base de datos E es la que mejores resultados muestra en la clasificación, y además la mayoría de muestras que contiene están etiquetadas como "normal"; es por esto que *specificity* tiene puntajes tan altos en esta prueba, *sensitivity* por su lado obtiene puntajes más bajos que en las pruebas anteriores debido al desbalance entre la cantidad de muestras con etiqueta "normal" y "anormal".

5.6. Resultados por base de datos

De los resultados anteriores se seleccionaron los conjuntos de características y RFC que dieron mayores puntajes M_{Acc}, con el fin de evaluarlos a lo largo de las diferentes bases de datos de manera individual, de esta forma no se están dando prioridad a las bases de datos con mayor número de elementos en la matriz de prueba para el resultado final. Estos resultados se muestran en la Tabla 17. También se obtuvieron los resultados promedio de las 6 bases.

De la Tabla 17 se puede ver que la base de datos E es la que presenta los mejores puntajes M_{Acc}, en cambio la base F presenta los peores puntajes. También se observa que RF-60 da el mejor resultado promedio.

Tabla 17. Resultados de características con bosques aleatorios, evaluadas en cada base de datos.

bases de datos \ características	A	B	C	D	E	F	Promedio
MFCC	0.731	0.723	0.775	0.775	0.993	0.620	0.770
delta MFCC	0.715	0.715	0.850	0.706	0.993	0.635	0.769
VCC	0.731	0.702	0.825	0.794	0.994	0.640	0.781
RF-60	0.718	0.723	0.850	0.775	0.994	0.635	0.783
MFCC8	0.698	0.758	0.725	0.736	0.988	0.545	0.742

5.7. Discusión sobre los resultados

El VCC consta de 300 características y la matriz de entrenamiento consta de 600 muestras, las cuales son un número muy reducido en comparación a la cantidad de características. Para poder entrenar de mejor manera el VCC se necesitaría utilizar un número de muestras mucho más grande para su entrenamiento. Este problema se resuelve al seleccionar características y generar el vector RF-60. Aunque a la vez se está perdiendo información que podría ser útil para tener un mayor desempeño.

El vector RF-60 dio el mayor puntaje M_{Acc}, incluso mayor que VCC, además de haber demostrado el mejor desempeño a lo largo de las diferentes bases de datos,

teniendo el puntaje MAcc promedio más alto de la Tabla 17. Por lo que se considera la mejor opción de todas las evaluadas.

Los delta MFCC, en la forma en que se están implementando, no representaron una mejoría en comparación a los MFCC, a pesar de que aportan 56 características adicionales. Además, como se ve en la Tabla 16, los delta MFCC solo aportan una característica al vector de las 60 características más importantes. Por estas razones, cuando menos de la forma en que los delta MFCC fueron implementados, se considera un esfuerzo computacional inútil calcularlos.

Los MFCC8 aportan 16 características al vector RF-60 mientras que los MFCC aportan 15 (Tabla 16), además la característica más importante es el quinto MFCC8 de la diástole. Los MFCC8, utilizando RFC, presentan ligeramente mayor puntaje MAcc que los MFCC (Tablas 6 y 14). Entonces se puede decir que los MFCC8 sí presentan una mejoría con respecto a los MFCC, tanto para el vector RF-60 como en cuestiones de puntaje MAcc individual. Aunque acorde a la Tabla 17, utilizar solo características de MFCC8 empeora por mucho los puntajes en algunas bases de datos con respecto a MFCC.

En la Tabla 16 es notable que el análisis del quinto coeficiente es el más importante tanto para MFCC8 como para MFCC. Además se puede observar que en las fases S1 y S2 el quinto MFCC8 es más importante que el quinto MFCC.

En la extracción de átomos de Gabor a la señal original para el análisis del residual se observa, según la gráfica de la Fig. 32, que después de los 10 átomos extraídos el puntaje MAcc comienza a decaer, esto debido a que la información de la patología se extrae mayormente después de los 10 átomos. Es importante no excederse en la cantidad de átomos extraídos para no dejar de lado esa parte patológica de la señal.

Las características extraídas con *Matching Pursuit* y átomos de Gabor, no presentaron resultados destacables con bosques aleatorios (Tabla 10), aunque demostraron ser parte importante del vector de características resultado de RF-60 (Tabla 16); aquí se puede observar además que las características extraídas con *Matching Pursuit* que más aportan son las frecuencias y posiciones de los átomos de Gabor.

Los cruces por cero en sístole y en la diástole resultaron ser de las características

más importantes en el vector obtenido por RF-60. Probablemente se deba al hecho de que estas son las etapas de silencio, y la presencia de sonido en estas etapas son debidas a patologías que afectan la cantidad de cruces por cero.

Como se ve en la Tabla 17 el desempeño varía mucho en función de la base de datos, y no solo tiene que ver con el número de muestras que tiene cada base, siendo la base C la de menor cantidad de muestras y en muchos casos es la segunda con mejores puntajes MACc. La base E por su parte, además de ser la que tiene mayor número de muestras (2054 grabaciones en la base original (Tabla 1), de las cuales se utilizaron 1780 para este trabajo (Tabla 2)), es en la que se facilita más la clasificación. Es muy importante considerar esto para no darle excesiva importancia a esta base de datos, ni en la matriz de entrenamiento ni en la de prueba.

Capítulo 6. Conclusiones

Durante el desarrollo de este trabajo se obtuvieron las siguientes conclusiones:

- El objetivo general de la tesis se considera cumplido, debido a que se ha desarrollado un algoritmo de clasificación que utiliza descomposición escasa para extracción de características y además se utiliza para analizar su residual.
- De la extracción de características de codificación escasa por medio del algoritmo de *Matching Pursuit* y la función de Gabor las características más importantes para realizar la clasificación, por medio de un RFC, son la frecuencia y la posición.
- Los delta MFCC no representan una mejoría para la clasificación de audio cardíaco, cuando menos no de la forma en que se utilizaron en este trabajo. Hace falta evaluar formas diferentes de como implementarlos.
- Los MFCC8 son un fuerte competidor para los MFCC, demostrando que en el residual de *Matching Pursuit* con átomos de Gabor se encuentra información muy importante que se puede utilizar para clasificación de sonidos cardíacos "anormales" y "normales".
- Al obtener el residual para los MFCC8 es necesario tener en cuenta hasta qué momento parar con la extracción de átomos, después de cierto punto lo que se está extrayendo es parte de la señal patológica y esto disminuye el rendimiento de la clasificación.
- La evaluación de características importantes por medio de bosques aleatorios, en sistemas que utilizan bosques aleatorios, es un método poderoso para la reducción de la dimensionalidad.
- Al momento de evaluar un modelo es importante considerar métricas que no den prioridad a una base de datos sobre las demás, cada base de datos fue grabada en condiciones diferentes y el modelo desarrollado debe funcionar bien en ellas.

6.1. Trabajo futuro

Se realizan las siguientes recomendaciones para trabajos futuros que puedan surgir en el área de clasificación de audio cardiaco:

- Implementar *Correlation Feature Selection* antes de la selección por características más importantes (RF-60); de esta forma disminuye el número de características correlacionadas entre sí.
- Aumentar el número de señales de muestra en la matriz de entrenamiento mediante SMOTE (Chawla, N.V., Bowyer, K.W., Hall, L.O., 2002), para mejorar el entrenamiento de los modelos. Este aumento se podría hacer en cada base de datos, para que todas tengan el mismo número de muestras en la matriz de entrenamiento.
- Utilizar moléculas (conformada por diferentes tipos de átomos) con el algoritmo de *Matching Pursuit*. De tal forma que un tipo de átomo esté correlacionado con la parte "normal" y los otros n tipos de átomos estén correlacionados con la parte patológica de la señal.
- Implementar delta MFCC y delta-delta MFCC en ranuras de tiempo pequeñas, con mayor resolución, y utilizar inteligencia artificial para su análisis (esto requeriría una matriz de entrenamiento mucho mayor a la que es posible con las bases de datos disponibles).
- Obtener grabaciones de audio cardiaco de diferentes personas (etiquetadas para cada persona), y utilizarlas para el entrenamiento y evaluación de un clasificador que realice identificación biométrica.
- Obtener el residual con *Matching Pursuit*, sin extraer los átomos que tienen su centroide dentro de la fase de sístole o diástole; ya que estos átomos podrían estar modelando alguna patología dentro de las fases de silencio.
- Obtener el residual de *Matching Pursuit* con átomos de Gabor pero utilizando como métrica para detener el algoritmo la energía extraída y no el número de átomos, y buscar el porcentaje de energía óptimo a extraer que deje el residual que sirva mejor para clasificación.

- Grabar audios cardiacos propios con diferentes condiciones ("normal", y diferentes tipos de condiciones "anormales") para enriquecer la base de datos.
- Implementar algoritmos de detección de los sonidos cardiacos para realizar segmentación antes de la extracción de las características.
- Realizar análisis de la calidad del audio de los ciclos y dejar de lado los que tengan menor calidad o tratarlos de forma diferente.
- Un análisis más profundo de la utilización de SVM con diferentes tipos de kernel, y tomando en cuenta funciones con diferentes grados para el kernel polinomial.

Literatura citada

- Abdollahpur, M., Ghiasi, S., Mollakazemi, M. J., y Ghaffari, A. (2016). Cycle selection and neuro-voting system for classifying heart sound recordings. En: *2016 Computing in Cardiology Conference (CinC)*. pp. 1–4.
- Arthur Guyton, J. E. H. (2016). *Tratado de Fisiología médica*. Elsevier, decimotercera edición.
- Banerjee, R., Biswas, S., Banerjee, S., Choudhury, A. D., Chattopadhyay, T., Pal, A., Deshpande, P., y Mandana, K. M. (2016). Time-frequency analysis of phonocardiogram for classifying heart disease. En: *2016 Computing in Cardiology Conference (CinC)*. pp. 573–576.
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag. Berlin, Heidelberg.
- Breiman, L. (2001). Random forests. *Machine Learning*, **45**(1): 5–32.
- Chawla, N.V., Bowyer, K.W., Hall, L.O., K. W. (2002). SMOTE: Synthetic Minority Over-Sampling Technique. *Journal of Artificial Intelligence Research*. *Journal of Artificial Intelligence Research*, **16**: 321–357.
- Clifford, G. D., Liu, C., Moody, B., Springer, D., Silva, I., Li, Q., y Mark, R. G. (2016). Classification of normal/abnormal heart sound recordings: The physionet/computing in cardiology challenge 2016. En: *2016 Computing in Cardiology Conference (CinC)*. pp. 609–612.
- Cormen, T. H., Leiserson, C. E., Rivest, R. L., y Stein, C. (2009). *Introduction to Algorithms*. The MIT Press.
- Cruz Gutierrez, A. (2016). *Segmentación robusta de audio cardíaco mediante análisis tiempo-frecuencia y métodos de optimización*. Tesis de maestría, Centro de Investigación Científica y de Educación Superior de Ensenada, Baja California.
- Das, P. P., Allayear, S. M., Amin, R., y Rahman, Z. (2016). Bangladeshi dialect recognition using mel frequency cepstral coefficient, delta, delta-delta and gaussian mixture model. En: *2016 Eighth International Conference on Advanced Computational Intelligence (ICACI)*. pp. 359–364.
- Duda, R. O., Hart, P. E., y Stork, D. G. (2000). *Pattern Classification*. Wiley-Interscience, segunda edición. New York, NY, USA.
- Dwivedi, A. K., Imtiaz, S. A., y Rodriguez-Villegas, E. (2019). Algorithms for automatic analysis and classification of heart sounds—a systematic review. *IEEE Access*, **7**: 8316–8345.
- eGeneral Medical (s.f.). Cardiac auscultation of heart murmurs database. Recuperado el 24/09/19 de: <http://www.egeneralmedical.com/listohearmur.html>.
- Gillet, O. y Richard, G. (2004). Automatic transcription of drum loops. En: *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*. Vol. 4, pp. iv–iv.

- Homsí, M. N., Medina, N., Hernández, M., Quintero, N., Perpiñan, G., Quintana, A., y Warrick, P. (2016). Automatic heart sound recording classification using a nested set of ensemble algorithms. En: *2016 Computing in Cardiology Conference (CinC)*. pp. 817–820.
- Huang, X., Acero, A., y Hon, H.-W. (2001). *Spoken Language Processing: A Guide to Theory, Algorithm, and System Development*. Prentice Hall PTR, primera edición.
- Ibarra-Hernández, R. F. (2014). *Desarrollo de un códec para la transmisión de audio cardiaco sobre redes de bajas tasas de datos*. Tesis de maestría, Centro de Investigación Científica y de Educación Superior de Ensenada, Baja California.
- INEGI (2018). Características de las defunciones registradas en México durante 2017. Recuperado el 24/08/19 de: <https://www.inegi.org.mx/contenidos/saladeprensa/boletines/2018/EstSociodemo/DEFUNCIONES2017.pdf>.
- Judge, R. D. y Mangrulkar, R. (2015). The open michigan heart sound murmur library (omhsm). university of michigan. Recuperado el 29/05/19 de: <http://www.med.umich.edu/lrc/psb/heartsounds/>.
- Kay, E. y Agarwal, A. (2016). Dropconnected neural network trained with diverse features for classifying heart sounds. En: *2016 Computing in Cardiology Conference (CinC)*. pp. 617–620.
- Krstulovic, S. y Gribonval, R. (2006). Mptk: Matching pursuit made tractable. En: *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*. Vol. 3, pp. III–III.
- Lartillot, O. (2018). *MIRtoolbox 1.7.1 User's Manual*.
- Liu, C., Springer, D., Li, Q., Moody, B., Juan, R., Chorro, F., Castells, F., Roig, J., Silva, I., Johnson, A., Syed, Z., Schmidt, S., Papadaniil, C., Hadjileontiadis, L., Naseri, H., Moukadem, A., Dieterlen, A., Brandt, C., Tang, H., y Clifford, G. (2016). An open access database for the evaluation of heart sound algorithms. *Physiological Measurement*, **37**: 2181–2213.
- Liu, H. y Motoda, H. (1998). *Feature Selection for Knowledge Discovery and Data Mining*. Kluwer Academic Publishers. Norwell, MA, USA.
- Louppe, G. (2014). *Understanding Random Forests: From Theory to Practice*. Tesis de doctorado.
- Luisada, A. A. (1959). Cardiovascular sound in health and disease. *American Journal of Cardiology*, **3**(6): 835–836.
- Mahnke, C. B. (2009). Automated heartsound analysis/computer-aided auscultation: A cardiologist's perspective and suggestions for future development. En: *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. pp. 3115–3118.
- Mallat, S. G. y Zhang, Z. (1993). Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, **41**(12): 3397–3415.
- Martínez, A. M. y Kak, A. C. (2001). Pca versus lda. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **23**(2): 228–233.

- Moorthy, U. y Gandhi, U. D. (2017). A Survey of Big Data Analytics Using Machine Learning Algorithms. *HCI Challenges and Privacy Preservation in Big Data Security*, pp. 95–123.
- Nieblas-Hernández, C. I. (2014). *Compresión del audio cardiaco mediante diccionarios redundantes y codificación escasa: aplicación a la transmisión en redes de sensores*. Tesis de maestría, Centro de Investigación Científica y de Educación Superior de Ensenada, Baja California.
- Nilanon, T., Yao, J., Hao, J., Purushotham, S., y Liu, Y. (2016). Normal / abnormal heart sound recordings classification using convolutional neural network. En: *2016 Computing in Cardiology Conference (CinC)*. pp. 585–588.
- O'Shaughnessy, D. (1988). Linear predictive coding. *IEEE Potentials*, **7**(1): 29–32.
- P. Bentley, G. Nordehn, M. C. S. M. y Getz., R. (2011). The pascal classifying heart sounds challenge 2011 (chsc2011). Recuperado el 29/05/19 de: <http://www.peterjbentley.com/heartchallenge/>.
- Peeters, G. (2004). A large set of audio features for sound description (similarity and classification) in the CUIDADO project. *CUIDADO IST Project Report*, **54**(0): 1–25.
- Roguin, A. (2006). Rene theophile hyacinthe laënnec (1781–1826): The man behind the stethoscope. *Clinical Medicine and Research*, **4**(3): 230–235.
- Shalev-Shwartz, S. y Ben-David, S. (2014). *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press. New York, NY, USA.
- Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell Systems Technical Journal*, **27**: 379–423.
- Smith, L. I. (2002). A tutorial on principal components analysis. Recuperado el 30/10/19 de: http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf.
- Springer, D. B., Tarassenko, L., y Clifford, G. D. (2016). Logistic regression-HSMM-based heart sound segmentation. *IEEE Transactions on Biomedical Engineering*, **63**(4): 822–832.
- VanderPlas, J. (2016). *Python Data Science Handbook: Essential Tools for Working with Data*. O'Reilly Media, Inc., primera edición.
- Whitaker, B. M. y Anderson, D. V. (2016). Heart sound classification via sparse coding. En: *2016 Computing in Cardiology Conference (CinC)*. pp. 805–808.
- Wood, J. C. y Barry, D. T. (1995). Time-frequency analysis of the first heart sound. *IEEE Engineering in Medicine and Biology Magazine*, **14**(2): 144–151.
- Zabihi, M., Rad, A. B., Kiranyaz, S., Gabbouj, M., y Katsaggelos, A. K. (2016). Heart sound anomaly and quality detection using ensemble of neural networks without segmentation. En: *2016 Computing in Cardiology Conference (CinC)*. pp. 613–616.
- Zheng, A. y Casari, A. (2018). *Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists*. O'Reilly Media, Inc, USA, primera edición.